DISSECTING THE DRIVING FORCES OF MEMBRANE PROTEIN FOLDING UNDER NATIVE CONDITIONS

By

Kristen Ann Gaffney

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Biochemistry and Molecular Biology−Doctor of Philosophy

2018

ABSTRACT

DISSECTING THE DRIVING FORCES OF
MEMBRANE PROTEIN FOLDING UNDER NATIVE CONDITIONS

By

Kristen Ann Gaffney

Membrane proteins are a unique class of proteins which reside within cellular membranes. They comprise 20~30% of all proteins in most organisms. Membrane proteins are involved in a variety of important cellular processes including ATP synthesis, photosynthesis, catalysis, molecular transport and cell signaling. Missense mutations in the genes encoding membrane proteins cause several life-threatening diseases including cystic fibrosis, Alzheimer's disease, and Charcot-Marie Tooth's disease. These mutations are known to cause disease majorly by impacting protein stability, rather than function, *via* two mechanisms: 1) protein destabilization which leads to excessive degradation and low accumulation of functional protein, 2) stabilization of non-functional misfolded forms of a protein which overwhelm cellular degradation machinery. To fundamentally understand disease mechanisms, it is necessary to understand the molecular forces and mechanisms in the folding of membrane proteins.

Although the study of protein folding has been one of the major quests in molecular biology over the last ~60 years, the understanding of membrane protein folding lags far behind that of soluble proteins. This is primarily due to the lack of available methods to control the reversible folding of membrane proteins under native conditions. Recently, steric trapping, which couples the unfolding of a doubly-biotinylated protein to monovalent streptavidin binding, has emerged as a promising technique to study membrane protein folding directly under native conditions without the use of chemical denaturants, heat, or pulling force. This work presents generalized steric trapping techniques utilizing novel tripartite chemical probes to dissect the folding energy

landscape of the intramembrane protease GlpG from *Escherichia coli*. The new steric trap tools were employed to examine the thermodynamic stability of GlpG and the physical dimension of its unfolded state. Upon the discovery of subglobal unfolding events of GlpG in the region encompassing the active site, an intricate cooperativity network important for maintaining the stability of GlpG was identified using cooperativity profiling at side chain resolution. Finally, double-mutant cycle analysis coupled with stability measurement by steric trapping revealed the weakly coupled hydrogen bond network in the catalytic active site of GlpG.

ACKNOWLEDGEMENTS

Foremost, I would like to thank my mentor Dr. Heedeok Hong for his unwavering support of my scientific endeavors throughout my time at Michigan State University. Dr. Hong has taught me to become a thoughtful and creative scientist by training but most importantly by example. His lab is one of encouragement and teamwork.

I would also like to thank Dr. Ruiqiong Guo, Dr. Miyeon Kim, and Yiqing Yang, as well as other members of the Hong Lab, for fruitful scientific discussions as well as therapeutic coffee dates. I am grateful to all my committee members who have provided me with wonderful guidance in my professional development. Thank you to all the students, staff and faculty of the Biochemistry & Molecular Biology Department at MSU. The friendly and collaborative environment promotes great friendship and support both professionally and personally.

I thank my parents, Michael T. and Barbara A. Gaffney, for providing me with endless encouragement. My siblings Ericka, Michael, Anna, and Joeleen, who always know how to make me laugh. Most importantly, I would like to thank my amazing and supportive partner, Dean Ouellette, for standing by my side for the last decade.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO ABBREVIATIONS

β-OG	octyl-β-glucoside
bR	bacteriorhodopsin
CHAPS	3-[(3-Cholamidopropyl) dimethylammonio]-1-propanesulfonate
Da	Dalton
DDM	n-dodecyl-β-D- maltoside
DEER-EPR	double electron-electron resonance-electron paramagnetic resonance
DMPC	1,2-dimyristoyl-sn-glycero-3-phosphocholine
DMPG	1,2-dimyristoyl-sn-glycero-3-phospho-(1'-rac-glycerol)
DSE	denatured state ensemble
DTT	dithiothreitol
E. coli	*Escherichia coli*
ER	endoplasmic reticulum
FRET	Förster resonance energy transfer
IPTG	isopropyl β-thiogalactopyranoside
mSA	monovalent streptavidin
MW	molecular weight
NA2HPO4	sodium phosphate
NaCl	sodium chloride
NMR	nuclear magnetic resonance
PAGE	polyacrylamide gel electrophoresis
PEG	polyethylene glycol

SDS	sodium dodecyl sulfate
SN	Staphylococcal nuclease
SRP	signal recognition particle
TCEP	tris(2-carboxyethyl) phosphine
TEV	Tobacco etch virus
TM	transmembrane
WT	wild-type

CHAPTER 1

Introduction

1.1. Importance of membrane protein folding in maintaining cellular health

Eukaryotic cells are divided into subcellular compartments, each of which carry out unique functions important for cellular homeostasis. These compartments are enclosed by a membrane composed of a unique set of lipids, proteins, and other macromolecules (1). The membranes not only package the organellar components, but also carry out diverse cellular processes. The proteins in the membranes often function as large homo- or hetero-oligomeric protein complexes, or function as a monomer, and chemical gradients across cell membranes are harnessed to carry out a diverse set of processes. They are involved in the generation of metabolic energy, transduction extracellular stimuli into intracellular signals, maintenance of ionic balance, transport of metabolites and mediating catalysis (2). Membrane proteins comprise roughly 25% of the total proteins in the cell (3). Their proper folding and trafficking are essential to perform their cellular function.

Missense mutations in general affect various folding properties of proteins and are implicated in numerous diseases (4). These mutations can impact the folding and protein homeostasis by two mechanisms. First, the mutation can lower the protein stability or induce misfolding, leading to excessive degradation, i.e., a low steady-state level of the active protein. Second, the mutation can increase the stability of a misfolded variant of the protein, overwhelming the quality control mechanisms, which can lead to excessive accumulation of the protein within the cell. Inherently, proteins need to maintain function which may require flexibility to undergo conformational change. They may also be required temporarily, so they must be able to be cleared from the cell. This leaves many cellular proteins with only marginal thermodynamic stability. Proteins are known to fold to their native structure through a funnel-shaped folding energy landscape (5) (**Fig. 1.1**). The unfolded polypeptide chain will have many high-energy

conformations. As the protein makes native-like contacts the possible conformations drop until the protein reaches its native, free energy minimum state (6). Because this folding funnel is not completely smooth, a protein will encounter local free energy wells which may represent a folding intermediate or a misfolded state. The detailed energy landscape can be altered by environmental variables such as temperature, pH, oxidative stress, as well as mutations and ligands. To maintain proper cellular homeostasis, cells require a diverse set of molecular chaperones and degradation machinery to maintain cell health (7). Unfortunately, this delicate system is often challenged by proteins exhibiting aberrant behavior.

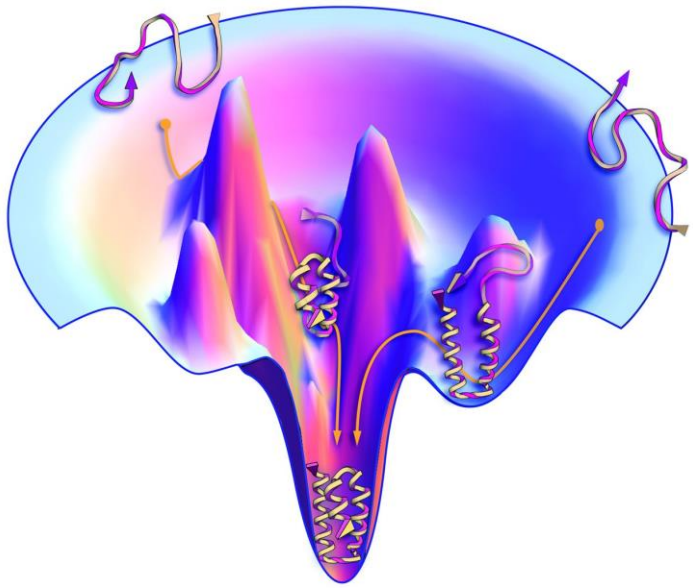


Figure 1.1 Protein folding energy landscape. Proteins fold through a funnel-like energy landscape (6). This image was reproduced from [Dill *et al.* (2012) The protein-folding problem, 50 years on. *Science*. **338**, 1042–6] with the permission of the American Association for the Advancement of Science via the Copyright Clearance Center.

1.1.1. CFTR misfolding and disease

The knowledge of driving forces, mechanisms, energy landscape and disease mechanisms associated with protein folding in cells has largely been obtained from studies of water-soluble proteins (6, 8, 9). To the same extent, the folding and stability of membrane proteins are under the risk of misfolding and implicated in several severe human diseases. Cystic fibrosis, an autosomal recessive disease can be caused by ~200 different point mutations in the gene encoding the cystic fibrosis transmembrane receptor (CFTR) protein (10). CFTR is an anion transporter, which is responsible for the balance of chloride ions across the plasma membrane. The disease mutations cause a general loss of hydration of membrane surfaces, specifically affecting mucus clearance within the lungs. Long-term complications in lung function can eventually lead to mortality (10). CFTR is a moderately stable protein such that ~50% of its wild type copies synthesized in the endoplasmic reticulum (ER) membranes reach the plasma membrane (11). The most common disease-causing mutation is at least one gene copy with the deletion of phenylalanine 508, ΔF508, which affects ~60% of cystic fibrosis patients. Although ΔF508 retains partial activity, the mutation facilitates the formation of an off pathway misfolding product which is a preferred target for degradation, leading to nearly ~0% accumulation of the active protein on the plasma membrane (12).

1.1.2. PMP22 misfolding and disease

Charcot-Marie-Tooth's (CMT) disease is characterized by demyelination of the myelin sheath in the peripheral nervous system. The disease is caused by missense mutations in the gene encoding peripheral myelin protein 22 (PMP22). Similar to the mechanism of cystic fibrosis, CMT occurs by the stability reduction and misfolding induced by mutation and the resulting impaired

trafficking to the plasma membrane (13). Only ~20% of wild-type PMP22 is successfully targeted to the plasma membrane. To fully understand disease mechanisms that are caused by misfolding and aberrant trafficking of integral membrane proteins, it is necessary to understand their driving forces and mechanisms of folding, i.e., the folding energy landscape of membrane proteins. Membrane proteins are known to follow the funneled energy landscape (14), but detailed information on the driving forces of membrane protein folding remains elusive. With a more complete knowledge of the folding energy landscape, it may be possible to design potential drugs which can increase the stability of the membrane proteins carrying disease mutations.

1.2. Membrane protein biogenesis

In eukaryotic cells, most membrane proteins are synthesized at the ER membrane and are inserted into the membrane mediated by the protein conduction channel, the Sec translocon. Specifically for Sec-dependent membrane protein biogenesis, a stretch of largely hydrophobic 8-12 amino acid residues that emerge from a translating ribosome is recognized as a signal sequence by the signal recognition particle (SRP) (15). SRP is a ribonucleoprotein which contains a GTPase domain. When SRP binds the ribosome-nascent chain complex in its GTP-bound state, translation is halted, and the translation complex is guided to the endoplasmic reticulum. SRP then binds the GTPase signal receptor (SR) protein and the nascent chain enters the protein conducting channel, the translocon Sec61 complex. At this stage, SRP and SR act as GTPase-activators for each other, and they hydrolyze GTP to GDP, which induces the release of the ribosome-nascent chain complex (15).

The translocon, Sec61, is a heterotrimeric complex consisting of Sec61αβγ along with translocation-associated membrane protein (TRAM) (16). Sec61α is involved in formation of the

pore-forming channel. As translation resumes, the transmembrane helices are thought to form their secondary structure within Sec61α, due to the large free energy cost of free hydrogen bond pairs within the hydrocarbon core of the bilayer (17). The transmembrane helices will then be inserted through the Sec61 lateral gate into the membrane. Any peripheral soluble domains and hydrophilic loops will be passed through the Sec61 pore. Several other proteins are associated with the Sec61 complex, many involved in post-translational modifications. For example, signal peptidase (SP) is located adjacent to Sec61, in order to cleave off the nascent chain's signal sequence (16). Oligosaccharyltransferase enzymes are also located near the Sec61 complex to carry out *N*-glycosylation at the end of the nascent chain. BiP has also emerged as an important protein that is associated with the Sec61 complex. BiP is a luminal protein which has been shown to be involved in a number of processes, mainly functioning as a luminal plug of the Sec61α pore when it is ribosome-free (18), as well as facilitating post-translational translocation, possibly by acting as a motor to drive protein translocation across the membrane (19).

Folding chaperones, such as calnexin, have also been shown to cross-link to nascent chains of membrane proteins co-translationally (20), although the chaperone doesn't appear to be part of the core translocon complex. Co-translational and post-translational tertiary folding as well as oligomerization, possibly facilitated by folding chaperones, will be carried out until the protein reaches its final, native structure. Once the membrane protein has reached its final form, it will then be shuttled through the Golgi complexes and on to its final cellular destination.

For prokaryotes, membrane proteins are co-translationally inserted into the plasma membraneby the SecYEG translocon complex. SecY, the pore forming unit, is composed of ten transmembrane segments and possesses a lateral gate where the transmembrane helices can make a contact with the lipid bilayer (21). The processes that govern the insertion *in vivo* and tertiary

folding of membrane proteins *in vitro* have been the focus of many biophysical studies in the recent decade.

1.3. Two-stage model for membrane protein folding

Membrane proteins are classified into β-barrel and α-helical types depending on the secondary structure of the membrane-embedded region. β-barrel membrane proteins exist in the outer membranes of Gram-negative bacteria, mitochondria and chloroplasts (22). α-helical membrane proteins are distributed in all membranes except for the outer membranes of Gram-negative bacteria (23). This work is focused on the folding of α-helical membranes which are more widely distributed in the kingdoms of life. Membrane proteins have evolved to maintain their structure and function in the hydrophobic lipid bilayer, which is a physically anisotropic and chemically heterogeneous environment.

Membrane protein folding can be divided into two thermodynamic stages (24) (**Fig. 1.2**). First, the hydrophobic segments of membrane proteins are inserted into the membrane, forming stable transmembrane helices. Second, the inserted transmembrane helices fold into a compact native structure through lateral helix-helix interactions. The insertion stage is known to be largely driven by the hydrophobic effect. Individual amino acids have been characterized by their hydrophobicity, *i.e.*, their propensity to partition from the aqueous phase into a non-polar phase. Over the past few decades, the hydrophobicity scales have been derived from measuring the partition of amino acid mimics between water and an organic solvent (*e.g.*, octanol in the Wimley-White scale (25)), measuring the partition of model transmembrane helices between the Sec61 translocon and the ER-derived membranes (*e.g.,* the Hessa-von Heijne scale (26)), and measuring

the unfolding free energy of a β-barrel scaffold protein which folds in the membrane from the aqueous phase (*e.g.,* the Moon-Fleming scale (27)).

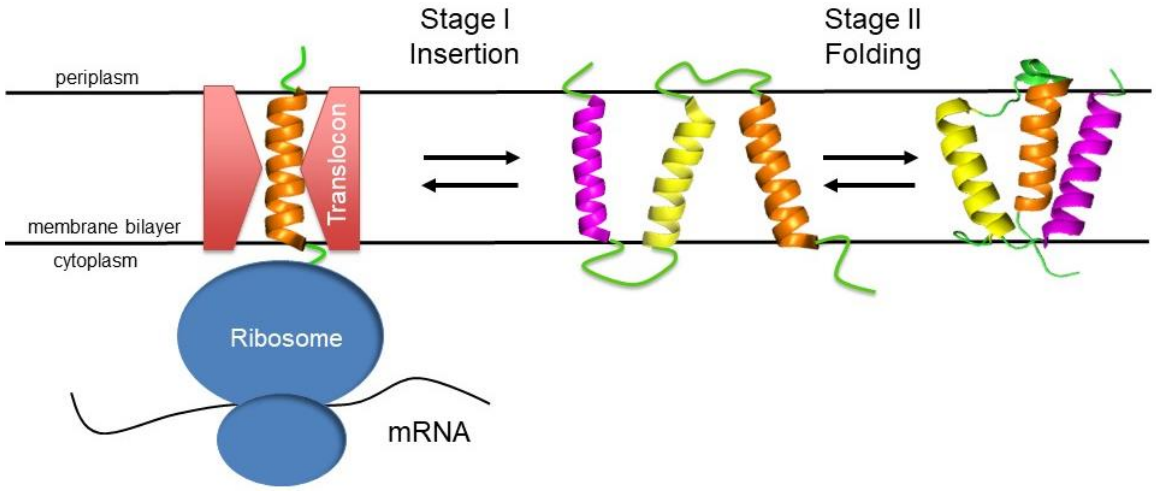


Figure 1.2 Two-stage model of membrane protein folding. Membrane proteins are co-translationally inserted into the lipid bilayer *via* the protein-conducting channel, the translocon. Stage I: insertion of individually stable transmembrane helices. Largely driven by the hydrophobic effect. Stage II: tertiary folding of transmembrane helices into the protein's native structure. Driving forces of this stage are largely unknown.

These scales commonly adapt a host-guest system, i.e., a "guest" residue forms a part of the host system (peptide or protein) and its partition free energy is measured relative to a reference residue (typically, Ala or Gly) in the host system. In recent years, depth-dependent hydrophobicity scales have been extensively developed by altering the position of the target amino acid along the membrane normal based on the statistical abundance from known membrane protein structures (28, 29).

Using these hydrophobicity scales, one can construct hydropathy plots which employ a sliding window (a continuous stretch of 18~20 residues) to identify possible transmembrane segments in an amino acid sequence (30). This hydropathy analysis has been proven to be very robust and allows for accurate predictions of transmembrane segments based solely on the amino acid sequence. As a transmembrane segment is inserted into the lipid bilayer *via* the translocon, it

must also decide on the topology or orientation within the bilayer. A common rule, called a "positive-inside rule" has been established that the peripheral region of the transmembrane segment which is more positively charged has a strong tendency to face the cytoplasmic side of the membrane (31–33). This could be due to charge gradients across the biological membrane or the local charge composition of the lipid headgroup.

The second stage of membrane protein folding, the formation of tertiary contacts, has remained more elusive. This is mainly due to the lack of appropriate methods to control the reversible folding of membrane proteins in their native lipid environment (9). Unlike their soluble counterparts, the hydrophobic effect is largely used up in the first insertion stage of membrane protein folding. Therefore, it is possible that other molecular forces will rise as important driving forces in the second stage of membrane protein folding, such as hydrogen bonding, van der Waals packing interactions, and salt bridge formation. In the recent years, several methods have been developed to study membrane protein folding. Chemical denaturation has evolved as a successful technique to control the reversible folding of α-helical membrane proteins. When this technique is used, membrane proteins of interest are isolated in non-denaturing detergent micelles, and then the denaturing detergent, sodium dodecyl sulfate (SDS) can be added for reversible unfolding (**Fig. 1.3**).

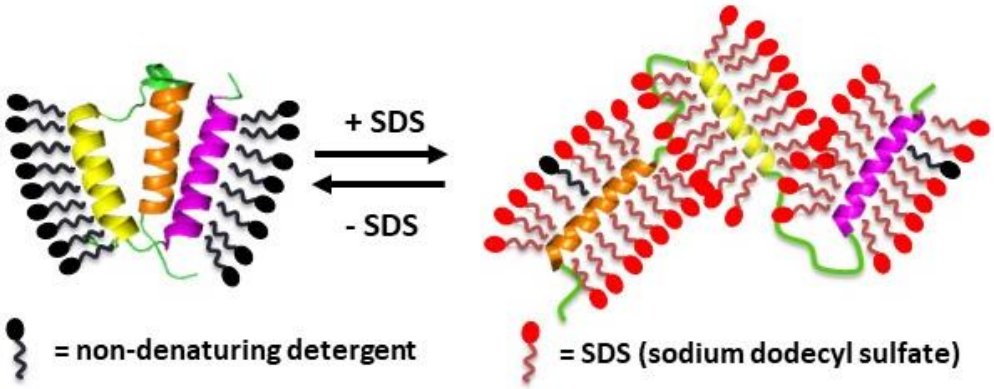


Figure 1.3 SDS-denaturation technique. The target membrane protein is purified in non-denaturing detergent micelles. The harsh detergent, SDS, can be titrated to unfold the protein into individual α-helices. SDS can be diluted to induce protein refolding.

The SDS denaturation has been applied to the seven-helical bundle light-driven proton pump bacteriorhodopsin (bR). bR displayed the ability to be denatured and renatured by the addition and dilution of SDS, respectively, in a variety of lipid-detergent mixed-micelles as well as detergent micelles (34, 35). Analysis of the secondary structure using circular dichroism indicated that bR retained ~70% of its helical content, relative to the native conformation in denaturing conditions. This demonstrated that the unfolded state likely retains its transmembrane helical structure. Amazingly, individual transmembrane fragments of bR were able to find their native tertiary structure without any information from the connecting loop regions (36). This indicates that the interactions between the transmembrane helices are major contributors to the stability of the protein. Despite its successes, the mixed micellar environment formed by SDS shares little physical features with the native lipid membrane.

1.4. Rhomboid proteases as a model system to study the driving forces of membrane protein folding

In this study, the rhomboid protease GlpG from *E. coli* was used as a model system to study the folding energy landscape of an α-helical membrane protein. Proteases, or peptidases, are enzymes that catalyze the hydrolysis of a peptide bond within a protein substrate. They are the largest class of proteins in all organisms, contributing to roughly 2% of the proteins in any given genome (37). Proteases can function in a variety of ways, such as protein processing, degradation, and activation. Intramembrane proteases are unique because they are integrated into cell membranes and mainly function to regulate transcription factors or signaling peptides by proteolysis. Five major classes of integral membrane proteases have been characterized (38): 1) site-2 metalloproteases (S2P), 2) γ-secretase aspartyl proteases, 3) signal peptide peptidase aspartyl proteases (SPP), 4) rhomboid serine proteases, and 5) the most recently described glutamyl proteases (39). Although the rhomboid protease family was recently discovered, it is one of the most structurally understood. Despite the vast knowledge of rhomboid structure, how they choose and accesses substrates is not fully realized (2).

1.4.1. Rhomboid proteases, a unique family of serine proteases

The rhomboid family proteins are conserved intramembrane serine proteases that site-specifically cleave single-pass membrane protein substrates (40). They were first discovered in 1984 as being involved in an early embryonic developmental stage of *Drosophila melanogaster* by regulating the epidermal growth factor receptor (EGFR) pathway through release of the membrane-anchored EGF Spitz by the proteolysis mediated by Rhomboid-1 (41). Genomic analysis revealed that Rhomboid-1 would be the first in a large family of rhomboid proteases that

possess multiple transmembrane segments and appear to perform hydrolysis within the membrane (42). The proposed proteolytic function in the membrane was surprising because water molecules that are necessary for proteolysis are scarce in the membrane.

Although rhomboid proteases show no sequence similarity to known proteases, *in vitro* mutagenic studies indicated that Rhomboid-1 was functioning as the first known intramembrane serine protease (43). By monitoring Spitz cleavage in mammalian cell lines with alanine scanning mutants of Rhomboid-1 at 18 conserved residues, researchers identified 6 residues which were required for function, of which included a GSAGG conserved motif around an essential serine (underlined) (43). A similar motif was found near the active site serine of other serine proteases. Reconstitution of Rhomboid-1 activity *in vitro* with detergent purified enzyme and substrate showed that rhomboid proteases were most likely functioning through a rare serine-histidine catalytic dyad instead of the more common catalytic triad (44). The ability of a rhomboid from a certain organism to cleave a variety of known biological substrates from various organisms suggested that they should function by a similar mechanism. Since their discovery, rhomboids have also been shown to be involved a variety of cellular processes, such as host cell invasion in apicomplexan parasites, quorum sensing in *Providencia stuartii*, and mitochondrial reorganization in *Saccharomyces cerevisiae* (45–47). Although, for many identified rhomboids, their biological function is still unknown.

1.4.2. Rhomboid protease substrate specificity

How do rhomboid proteases select their substrates? Answering this question would allow for more efficient identification of rhomboid substrates based on amino acid sequence as well as determination of their biological function. Rhomboid proteases are known to cleave single-pass

membrane proteins at a site-specific peptide bond that is located either buried within the membrane or near the bilayer interface (48). Their substrates typically exist as a single α-helix, which should be difficult to unravel in the bilayer environment due to the large energetic penalty for exposing unsatisfied hydrogen bonds in a non-polar environment (17). Yet, for proteolysis to occur it is necessary that the backbone hydrogen bonds near the scissile bond become unfolded (49). Similar to the other classes of intramembrane proteases, rhomboid substrates also display helix-destabilizing residues as a key sequence component to allow their cleavage (50). Strisovsky and researchers determined that rhomboid proteases from various bacterial organisms cleaved 4 different substrates at the same exact peptide bond (48). Although the detailed substrate recognition mechanism by rhomboids is not clear, they seem to function in a similar manner across the protein family.

1.4.3. Catalytic mechanism of rhomboid proteases

The feasibility of active rhomboid proteases for their overexpression and purification led to quick structural characterization of the rhomboid protease GlpG from *Escherichia coli* and *Haemophilus influenza* by X-ray crystallography (51–53). Interestingly, the active site was located ~10 angstroms deep into the membrane bilayer. Transmembrane segments 4 and 6 were tightly packed, which is mediated by two conserved glycines that allow the two helices to interact in close proximity. This allows for the active site Ser201 and His254 to be close in the structure to form the hydrogen bond necessary for catalysis. The crystal structures also identified coordinated water molecules that form a water-retention site contacting the active site, illustrating the ability of water to access His254 to be activated for hydrolysis of the substrate (51). This was further confirmed by examining the solvent-accessibility of the region near the active site (54). Single cysteine

mutants near the active site of GlpG were shown to be able to be modified by a membrane-impermeable thiol-alkylating reagent, proving the region near the active site is accessible to water.

By comparison with water-soluble serine proteases, as well as biochemical mechanistic studies and enzyme-inhibitor crystallographic studies, a catalytic mechanism for rhomboid proteases has been suggested (**Fig. 1.4**) (55, 56). For GlpG, the active site His254 activates Ser201 for nucleophilic attack on the carbonyl carbon of the peptide bond, followed by formation of the first tetrahedral intermediate. This tetrahedral intermediate is stabilized by a triad oxyanion hole containing the side chain of H150, the backbone amide of Ser201, and the side chain of N154. Next, the tetrahedral intermediate will collapse, allowing for the formal cleavage of the peptide bond. His254 will then activate a water molecule for nucleophilic attack and regeneration of the active site. With the knowledge of the catalytic mechanism and three-dimensional structures of rhomboid proteases, questions about its conformational dynamics and stability arise.

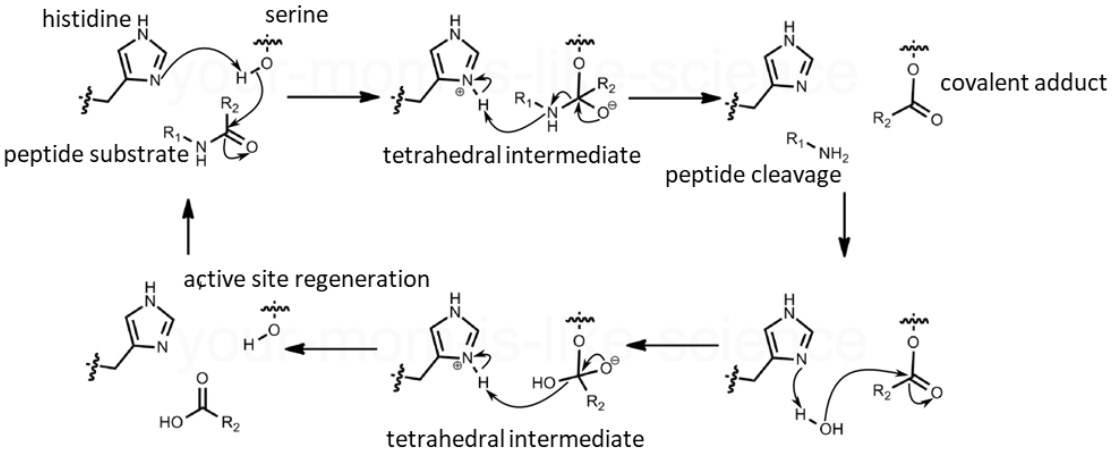


Figure 1.4 Proposed catalytic mechanism of rhomboid proteases. The His254 activates Ser201 for nucleophilic attack on the peptide bond, which forms the first tetrahedral intermediate. Intermediate collapse leads to the covalent adduct formation. His254 then activates a water molecule to attack the covalent adduct and form the second tetrahedral intermediate. Intermediate collapse leads to active site regeneration.

1.4.4. *In vitro* folding studies of rhomboid protease GlpG

In recent years, the stability and folding of GlpG have been studied using a variety of techniques. Thermal and SDS denaturation studies in dodecylmaltoside (DDM) micelles on ~150 mutants identified key packing regions in GlpG (57). GlpG was shown to cooperatively fold in a two-state manner (i.e., the folded and unfolded states) such that it did not possess any major folding intermediates. Four key regions that contributed to maintaining the structural stability of GlpG were identified. Helix-packing interactions between TM4 and TM6 form a glycine-zipper packing motif. TM4 displays a GXXXAXXG which tightly packs against the GXXXGXXXA on the interfacial region of TM6. Hydrogen bonding residues in the L1 loop, which are on the extracellular face, contribute > 2 kcal mol^{-1} to GlpG stability. Also, the key hydrogen bonding residue Glu166 on the cytoplasmic end of TM2 contributed ~2 kcal mol^{-1} to GlpG stability. Weak packing interactions were found to have a compounding effect, as triple and quadruple mutations could destabilize GlpG > 2 kcal mol^{-1}.

ϕ-value analysis was performed on 69 mutants of GlpG to study the folding mechanism (58). GlpG was unfolded in SDS to mimic the initial state for tertiary folding of membrane proteins, and then non-denaturing detergent DDM was added to refold the protein. By monitoring the folding and unfolding kinetics, as well as determining the thermodynamic stability, ϕ-value analysis allows for investigating the contacts formed during the transition state in the folding reaction through mutational perturbation. This provides information about which proteins are making native-like contacts during the folding process. The V-shaped chevron plot displayed by the folding and unfolding kinetics in this study indeed confirmed that GlpG folds through a two-state process, that it does not form a stable folding intermediate (58). The formation of a folding nucleus between TM1 and TM2 indicates that the folding starts by the formation of the packing

core in the N-terminal region. TM4 through TM6 seemed to lack structure during the folding transition, indicating that this region may be the last structural elements to fold and the first to unfold. This is reasonable, as TM4 and TM6 contain the active site residues, and TM5 is thought to have some conformation flexibility to allow substrate access into the active site. Frustrated regions of GlpG folding were identified, which were later determined to be "back-tracking" regions, in which native contacts were first formed during the folding transition, then had to be pulled apart to facilitate downstream folding events (59).

Single molecule pulling studies with magnetic tweezers showed that GlpG possesses a high kinetic barrier to unfolding in DMPC/CHAPSO bicelles and the force-induced unfolding is highly cooperative at a wide range of the applied forces (60). The unfolding is thought to involve complete unraveling of the secondary and tertiary structure of GlpG. Min and coworkers suggest that this high cooperativity may allow for the efficient formation of the native structure, avoiding any off-pathway intermediate structures that may be populated during folding in cells. Using force-jump experiments rather than slowly increasing the force they applied one large constant force, they were able to identify regions of intermediates. By calculating the degree of extension as well as destabilizing the local conformational stability by mutation, they were able to determine that the force-induced unfolding occurs from the C-terminus. This result agrees well with the ϕ -value analysis described above.

1.5. Lipid bilayer environment shapes the structure and function of membrane proteins

Membrane proteins have evolved to maintain their structure and function in the heterogeneous environment of the phospholipid bilayer. This bilayer environment is chemically and physically heterogeneous and anisotropic (i.e., the lipids are packed into a quasi-two-

dimensional space with a defined thickness). The bilayer consists of two monolayers of lipid that can be chemically diverse: a non-polar hydrocarbon tail which can vary in chain length and saturation and a polar head group that can have diverse chemical moieties and charged states at the interface with the aqueous solution (61). As determined from X-ray diffraction of the hydrated lipid bilayers, the total membrane thickness is ~55-60 Å, with a hydrocarbon core that occupies a total of 30 Å and each interfacial region account for 10-15 Å (62). The physical properties of the lipid bilayers are also sensitive to the environment, such as temperature, protein concentration, stress conditions (61).

Lipids have been known to affect the activity of membrane proteins, through specific binding or non-specific effects (63). Many membrane proteins resolved by X-ray crystallography have lipids bound to pockets on their surface, which should be tightly bound to the protein (64). The specific binding of lipids to membrane proteins has also been shown to affect their function (63, 65). One non-specific effect is through hydrophobic mismatch, that is the protein and lipid hydrocarbon thickness does not match well with each other. Because of the large energetic penalty of exposing hydrophobic surface area to aqueous phase, it is expected that the membrane will undergo local distortion to accommodate the protein. If the disparity between the protein and membrane is too large, either the protein will be released from the membrane (66), or the membrane will form non-bilayer structures (67).

Interestingly, GlpG has an unusually small hydrophobic thickness relative to the membrane bilayer, which is suggested to be compressed by 30% by analysis of the crystal structure (51). Molecular dynamics simulations of GlpG with mutations in loop L1, which have been shown to affect activity although it is distant from the active site, displayed alternate hydrogen bonding of L1 residues with lipid headgroups, which causes changes in the hydrophobic thickness of the

bilayer and shifts the orientation of GlpG relative to the bilayer normal (68). This also suggests that the ability of the lipid headgroups to hydrogen bond with the protein could have dramatic effects on membrane protein structure and function.

1.6. Conclusion

Membrane proteins are an important class of proteins, whose mutational defects in maintaining their stability are implicated in serious hereditary diseases (11). Understanding the driving forces of membrane protein folding within cell membranes is essential to gain insight to possible disease mechanisms as well as possible treatments. Membrane proteins are known to be co-translationally inserted into either the ER (eukaryotes) or plasma (prokaryotes) membranes *via* the protein conducting channel, the translocon. In the first stage of membrane protein folding, α-helical membrane proteins inserted are individually stable transmembrane helices and their insertion is largely driven by the hydrophobic effect. In the second stage of membrane protein folding, the transmembrane segments will form tertiary contacts. Although the first insertion stage has been thoroughly studied, the second folding stage is not well understood.

GlpG, the intramembrane rhomboid protease from *E. coli*, has emerged as a model membrane protein system for *in vitro* folding studies. It is a monomeric, 6 TM helical bundle, which has a convenient unfolding readout. Its folding properties have been studied in diverse lipid environments, including DDM micelles and DMPC/CHAPSO bicelles. Through a variety of methods, GlpG has been shown to cooperatively fold by a two-state process. Thorough mutational analysis has revealed key regions for maintaining GlpG stability, and an N-terminal folding nucleus in the transition state has been identified. Although the previous folding studies of GlpG have been thorough, they rely on irreversible thermal denaturation or chemically induced protein

unfolding, which does not have the ability to be applied to the native bilayer system. Force-induced unfolding studies displayed difficulty in obtaining thermodynamically relevant protein refolding. Therefore, there remains a need for the development of new techniques to quantitatively study the thermodynamics of membrane protein folding without disrupting the native bilayer environment, which must have a large effect on the structure and function of membrane proteins.

CHAPTER 2

Steric Trapping Reveals a Cooperativity Network in the Intramembrane Protease GlpG

Ruiqiong Guo[1,5], Kristen Gaffney[2,5], Zhongyu Yang[3,4], Miyeon Kim[1], Suttipun Sungsuwan[1], Xuefei Huang[1], Wayne L. Hubbell[3] and Heedeok Hong[1,2]

[1]Department of Chemistry & [2]Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, MI 48824, USA

[3]Jules Stein Eye Institute and Department of Chemistry and Biochemistry, University of California, Los Angeles, CA 90095, USA

[4]Present Address: Department of Chemistry and Biochemistry, North Dakota State University, Fargo, ND 58102, USA

[5]These authors equally contributed to this work.

Published in Nature Chemical Biology, May 2016, Volume 12, pg. 353-360.

This publication was completed as a close collaboration between Ruiqiong Guo and I, which is indicated by our listing as co-first authors. Dr. Yang performed the DEER-EPR experiments and data analysis in the lab of Dr. Hubbell. Dr. Kim prepared the EPR samples as well as carry out the CW-EPR experiments. Dr. Sungsuwan assisted Dr. Guo in the design and synthesis of the chemical probes, in the lab of Dr. Huang. I was responsible for the cloning and expression optimization for the GlpG and substrate protein constructs. I designed and optimized the fluorescence activity assay in Fig. 2.3. I determined all $K_{d,biotin}$ for mSA mutants at various [SDS], Table 2.2. I performed all proteolytic digestion of GlpG with chymotrypsin, Fig. 2.6. I carried out all SDS-denaturation curves, Fig. 2.7(b) and Fig. 2.10. Guo and I together screened numerous steric trapping pairs to find ideal sites, Fig. 2.4. Guo and I together performed the binding isotherms for stability analysis and cooperativity profiling of GlpG, Fig. 2.7(c) and Fig. 2.9.

2.1. Summary

Membrane proteins are assembled through balanced interactions among protein, lipids and water. Studying their folding while maintaining the native lipid environment is necessary but challenging. Here we present methods for analyzing key elements in membrane protein folding including thermodynamic stability, compactness of the unfolded state and folding cooperativity under native conditions. The methods are based on steric trapping which couples unfolding of a doubly-biotinylated protein to binding of monovalent streptavidin (mSA). We further advanced this technology for general application by developing versatile biotin probes possessing spectroscopic reporters that are sensitized by mSA binding or protein unfolding. By applying these methods to an intramembrane protease GlpG of *E. coli*, we elucidated a widely unraveled unfolded state, subglobal unfolding of the region encompassing the active site, and a network of cooperative and localized interactions to maintain the stability. These findings provide crucial insights into the folding energy landscape of membrane proteins.

2.2. Introduction

Understanding the free energy landscape of protein folding requires determination of the free energy levels of states populated during folding as well as analysis of energy barriers to reaching the native conformation (5). Experimentally, this task has been carried out by equilibrium and kinetic folding studies using denaturants that can readily shift population distribution between folded and unfolded states (69). However, in the presence of denaturants, the overall shape of the folding energy landscape substantially changes and certain short-lived higher-energy states may not be detected (70, 71). Thus, studying protein folding under native conditions is necessary for a full survey of the folding energy landscape. For water-soluble proteins, methods such as hydrogen-

deuterium exchange, NMR and proteolysis have revealed the dynamic and multistate nature of the native conformational ensemble (71–75), which is critical to protein function (76–78). For membrane proteins, however, such features remain largely unexplored because the poor accessibility of solvent water to the interior of micelles and bilayers, and the large sizes of protein-micellar and protein-liposomal complexes, have made it difficult to apply similar methods to characterize the native ensemble of membrane proteins (79, 80).

Steric trapping is a promising tool for investigating the thermodynamic stability and folding of membrane proteins directly under native conditions. The method couples the unfolding of a target protein labeled with two biotin tags to the competitive binding of bulky monovalent streptavidin molecules (mSA, MW = 52 kDa) (81–85) (**Fig. 2.1**). After conjugation of biotin tags to two specific residues that are spatially close in the folded state but distant in the amino acid sequence, the first mSA binds unhindered to either biotin label with intrinsic binding affinity (ΔG^{o}_{Bind}). Due to the steric hindrance with pre-bound mSA, the second mSA binds only when the native tertiary contacts between biotinylated sites are unraveled by transient unfolding. Coupling of mSA binding to unfolding leads to attenuation of the apparent binding affinity of the second mSA relative to that of the first mSA, whose degree is correlated with the protein stability. Thus, thermodynamic stability (ΔG^{o}_{U}) of the target protein can be determined by fitting of the second binding phase (see **equations (1)–(3)**). Overall, protein unfolding is driven by the affinity and concentration of mSA without perturbing the native solvent condition. Folding reversibility is tested upon addition of excess free biotin by which bound mSA molecules are released by competition.

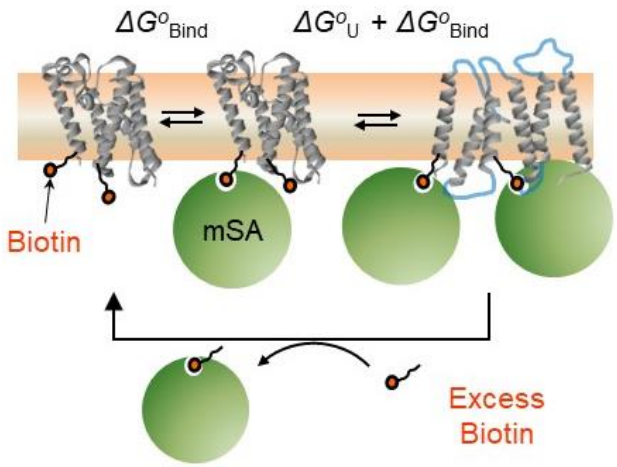


Figure 2.1 Steric trapping scheme. Bulky mSA binds and traps a target protein in its unfolded state without the use of chemical or physical denaturants. Free energy of mSA binding is coupled to protein unfolding.

Although promising, it is yet difficult to apply steric trapping to various types of membrane proteins. The method requires two features: two site-specifically conjugated biotin labels on a target protein and a probe to monitor mSA binding or protein unfolding. Site-specific biotinylation has been achieved by labeling of engineered cysteine residues with thiol-reactive biotin derivatives (81, 82, 85). For detection of unfolding, widely used tools such as tryptophan fluorescence and circular dichroism cannot be used because of the large signal interferences from mSA molecules. A method for direct detection of mSA binding has not yet been developed. Thus, its application has been limited to proteins possessing convenient unfolding readouts such as absorbance of a conformation-sensitive intrinsic chromophore (e.g., retinal in bacteriorhodopsin (82) and enzymatic activities (e.g., dihydrofolate reductase (81) and diacylglycerol kinase (85)). In this study, we developed a generalized steric trapping strategy that utilizes novel thiol-reactive biotin probes containing spectroscopic reporter groups for sensitive detection of mSA binding and protein unfolding. We used this strategy to analyze the thermodynamic stability, compactness of

the unfolded state and folding cooperativity of the six-helical-bundle intramembrane protease GlpG of *E. coli*.

GlpG is a member of the rhomboid protease family, which is widely conserved in all kingdoms of life. Rhomboid proteases act in diverse biological processes by activating membrane-bound signaling proteins or enzymes via the cleavage of a specific peptide bond near the membrane (43, 47, 86, 87). Because of the functional importance of rhomboid proteases and the detailed structural information available (>30 entries in the Protein Data Bank (PDB), http://www.rcsb.org/), GlpG has emerged as an important model for studying the folding of helical membrane proteins. Regions critical for its stability have been identified using heat and SDS denaturation tests of 151 variants (57). A kinetic folding study using SDS as a denaturant has suggested the existence of a compact folding nucleus in the folding transition state (58). A single-molecule magnetic tweezers study has shown that GlpG largely unfolds cooperatively at constant tension (60). Here, using steric trapping, we provide new insights into the folding energy landscape of GlpG in the absence of heat, chemical denaturants or pulling force. We elucidated an expanded heterogeneous conformational ensemble of the unfolded state, a structural region that undergoes subglobal unfolding, and an intricate network of cooperative and localized interactions to maintain the stability of GlpG.

2.3. Materials and Methods

2.3.1. Synthesis of BtnPyr-IA and BtnRG-TP

For design and synthesis of steric trapping probes please see ref. (88).

2.3.2. Preparation of glpg DNA constructs

The *glpG* gene was amplified from chromosomal DNA of *E. coli* strain MG1655 (Coli Genetic Stock Center at Yale University) using primers containing NdeI and BamHI restriction sites. The amplified gene was ligated into the pET15b vector with an N-terminal His_6-tag. Site-directed mutagenesis for introducing amino acid substitutions was per- formed using the QuikChange Site-Directed Mutagenesis Kit (Agilent).

2.3.3. Expression and purification of GlpG

GlpG was expressed in the *E. coli* BL21(DE3) RP strain. Cells were grown at 37 °C until OD_{600} = 0.6 was reached. Protein expression was induced with 0.5 mM isopropyl β-thiogalactopyranoside (IPTG, GoldBio), followed by additional cultivation at 15 °C for 16 h. GlpG was purified from the total membrane fraction obtained by ultracentrifugation (Beckman Coulter, Type 45 Ti rotor, 50,000g, for 2 h) using Ni^{2+}-NTA affinity chromatography (Qiagen) after solubilization with 2% n-dodecyl-β-D- maltoside (DDM, Anatrace).

2.3.4. Labeling of GlpG and determination of labeling efficiency using SDS-PAGE gel shift assay

For labeling, purified cysteine variants (0.2% DDM, 50 mM Tris-(hydroxymethyl) aminomethane hydrochloride (TrisHCl), 200 mM NaCl and pH 8.0) were diluted to less than 100 μM and incubated with a ten-fold molar excess tris(2-carboxyethyl) phosphine hydrochloride (TCEP-HCl, Pierce) for 1 h at room temperature. A 40 times molar excess of BtnPyr-IA or BtnRG-TP dissolved in dimethyl sulfoxide (~20 mg/ml) was added to the mixture while vortexing.

Labeling reaction was allowed to proceed at room temperature overnight in the dark. Excess free labels were removed by extensive washing of the proteins bound to Ni^{2+}-NTA affinity resin using 0.2% DDM, 50 mM TrisHCl, 200 mM NaCl and pH 8.0 solution. Labeled GlpG was dialyzed against 0.02% DDM, 50 mM TrisHCl, 200 mM NaCl, pH 8.0 buffer to remove imidazole. Typically, the labeling efficiency of BtnPyr-IA and BtnRG-TP ranged from 1.5 to 2.2 as estimated from SDS-PAGE gel shift assay or comparison of the concentration of BtnPyr determined by pyrene absorbance (ε_{346nm} = 43,000 $M^{-1}cm^{-1}$) and the concentration of GlpG determined by DC protein assay (Bio-Rad).

SDS-PAGE was employed using the facts that mSA maintains its tetrameric structure and the biotin–mSA complex is resistant to dissociation in the presence of SDS. SDS-PAGE gel shift assay was carried out as follows: 10 μL of 5 μM of labeled GlpG was incubated with 10 μL of 2% SDS sample-loading buffer with 10% (v/v) β-mercaptoethanol for 30 min. Then, wild-type monovalent streptavidin (mSA-WT) was added to labeled GlpG (GlpG:mSA-WT molar ratio of 1:3) and the mixture was incubated at room temperature for 30 min before SDS-PAGE without sample heating. The gel box was incubated in ice during electrophoresis to prevent heat-induced dissociation of mSA-WT bound to biotin label on GlpG. Labeling efficiency was determined by comparing the intensities that correspond to single-mSA bound GlpG and double-mSA bound GlpG after accounting for the molecular mass of GlpG and mSA (AlphaImager, ProteinSimple). GlpG with no label was not considered because this species does not bind mSA, thus not participating in steric trapping. mSA was prepared as described previously.

2.3.5. Expression and purification of GlpG substrate SN-LacYTM2

As a folding indicator for GlpG, we used its proteolytic activity mediating specific cleavage of a transmembrane (TM) substrate, the second TM domain of the lactose permease of *E. coli* fused to staphylococcal nuclease (SN-LacYTM2). The DNA construct for LacYTM2 was amplified from a DNA template containing full length lactose permease using primers containing XmaI and XhoI restriction sites, which was then ligated into a pET30a vector containing SN domain, TEV protease recognition site, and C-terminal His_6-tag (SN-TEV-LacYTM2-His_6). In the LacYTM2 region, the position which was five residues upstream from the scissile bond (P5 position) was substituted with cysteine for labeling with thiol-reactive, environment-sensitive fluorophore iodoacetyl-7-nitrobenz-2-oxa-1,3-diazol (IA-NBD amide, Setareh Biotech). SN-TEV-LacYTM2-His_6 containing the substituted cysteine was expressed in BL21(DE3) RP *E. coli* strain. The protein was expressed, purified and labeled using the protocol for SN-GpATM-His_6 described previously (83).

2.3.6. Fluorescence-based high-throughput activity assay for GlpG

Activity assay was initiated by addition of 10 times molar excess of NBD-labeled SN-LacYTM2 to purified GlpG. Time-dependent changes of NBD fluorescence were monitored in 96-well plate using SpectraMax M5e plate reader (Molecular Devices) with excitation and emission wavelengths of 485 nm and 535 nm, respectively. Fluorescence change was normalized to a control sample containing NBD-SN-LacYTM2 alone. The effect of mSA binding on the activity of single- and double-biotin variants of GlpG were tested by addition of excess mSA-WT (20 μM). The GlpG-mSA mixture was incubated overnight (single-biotin variants), for 2 days ($172/267_C$-$BtnPyr_2$) or for 5 days ($95/172_N$-$BtnPyr_2$). Folding reversibility was tested using the following steps: Each double-biotin GlpG variant was first inactivated with mSA-S27A variant (20 μM)

possessing a weaker biotin binding affinity ($K_{d,biotin}$ = 1.4 nM) for the same incubation time as with mSA-WT. Next, excess free biotin (2 mM) was added to induce competitive dissociation of bound mSA. The activity of refolded GlpG was measured after incubation overnight. The statistical significance of the activity changes upon unfolding and refolding were evaluated using Student's t-test (n = 3–5).

2.3.7. Double electron-electron resonance EPR spectroscopy (DEER-EPR)

DEER- EPR measurements were performed on a Bruker Elexsys 580 spectrometer with Super Q-FTu Bridge, Bruker ER 5107DQ resonator and 10 W Q-band amplifier at 80 K. The spin-labeled samples ranging from 80 to 160 μM GlpG were flash-frozen in quartz capillaries using a liquid nitrogen bath immediately before data collection. For data collection, a 36-ns π-pump pulse was applied to the low field peak of the nitroxide absorption spectrum, and the observer π/2 (16 ns) and π (32 ns) pulses were positioned 17.8 G (50 MHz) upfield, which corresponded to the nitroxide center resonance. A two-step phase cycling (+x, −x) was carried out on the first (π/2) pulse from the observer frequency. The time domain signal collected for each sample varied from 2.3 to 2.5 μs. Based on the collection time, the reliable interspin distance range was ~15−~60 Å. DEER data were analyzed using the program LongDistances, which was writ- ten in LabVIEW by Christian Altenbach (http://www.biochemistry.ucla.edu/ biochem/Faculty/Hubbell/).

2.3.8. Time-dependent proteolysis of GlpG using chymotrypsin

GlpG variants (5 μM) were incubated for 7 days ($95/172_N$-$BtnPyr_2$) or 4 days ($172/267_C$ - $BtnPyr_2$) in the presence and absence of 25 μM mSA-WT in 20 mM Na_2HPO_4 (pH 7.5), 20 mM DDM, 200 mM NaCl, 1 mM TCEP buffer. In the presence of mSA-WT, the residual activities were

~40% for 95/172$_N$-BtnPyr$_2$ and ~20% for 95/172$_N$-BtnPyr$_2$ relative to those without mSA before proteolysis. Proteolysis of 5 μM casein (from bovine milk, Sigma) was performed in the same buffer condition but without prolonged incubation. For all protein samples, proteolysis was initiated by the addition of 2.5 μM chymotrypsin (bovine chymotrypsin-α: sequencing grade, Sigma) to 10 μL aliquots, and quenched at specified time by the addition of 10 mM permethylsulfoxide. Time-dependent proteolysis was monitored by SDS-PAGE.

2.3.9. Construction of binding isotherms to determine thermodynamic stability of GlpG by steric trapping using FRET

1 μM of GlpG labeled with BtnPyr was titrated with mSA specifically labeled with DABCYL-plus-maleimide (AnaSpec) at Y83C-position of the active subunit (mSA$_{DAB}$) in 5 mM DDM, 0.25 mM TCEP, 20 mM Na_2HPO_4 and 200 mM NaCl (pH 7.5). The titrated samples were transferred to a 96-well UV-compatible microplate, sealed with a polyolefin tape, and incubated for 5 days (for 95/172$_N$-BtnPyr$_2$) or 2 days (for 172/267$_C$-BtnPyr$_2$) at room temperature. Binding was monitored by the decrease of pyrene-monomer fluorescence at 390 nm with an excitation wavelength of 345 nm using SpectraMax M5e plate reader. Data were averaged from four readings. Nonspecific FRET was obtained by measuring the fluorescence intensity of double-biotin GlpG variants which were pre-saturated with 10 μM of the high-affinity mSA-WT (without DABCYL-label) at increasing concentrations of the lower-affinity variant mSA$_{DAB}$-S45A. In this condition, mSA$_{DAB}$-S45A cannot compete for biotin label and only diffuses around in the solution.

2.3.10. Fitting of binding isotherm to obtain thermodynamic stability of GlpG

Fitting equation to obtain the thermodynamic stability of GlpG using steric trapping was based on the following reaction scheme (81):

$$\mathrm{F \cdot mSA} \xrightleftharpoons{K_U} \mathrm{U \cdot mSA}\text{, where} \quad K_U = \frac{[\mathrm{U \cdot mSA}]}{[\mathrm{F \cdot mSA}]} \qquad \textbf{(1)}$$

$$\mathrm{U \cdot mSA + mSA} \xrightleftharpoons[K_{d,biotin}]{} \mathrm{U \cdot 2mSA}\text{, where} \quad K_{d,biotin} = \frac{[\mathrm{U \cdot mSA}][\mathrm{mSA}]}{[\mathrm{U \cdot 2mSA}]} \qquad \textbf{(2)}$$

Fitting equation for the second mSA binding phase was:

$$F = \frac{1}{[1 + \left(K_{d,biotin} + \frac{K_{d,biotin}}{K_U} \right) \frac{1}{[\mathrm{mSA}]}]} \left(F_\infty - F_o \right) + F_o \qquad \textbf{(3)}$$

,where F is the measured fluorescence intensity, and F_0 and F_∞ are the fluorescence intensities from GlpG labeled with BtnPyr at [mSA] = 0 and at the saturated bound level, respectively. [mSA] is the total mSA concentration, $K_{d,biotin}$ is the dissociation constant for unhindered biotin binding affinity of mSA, and K_U is the equilibrium constant for unfolding of GlpG. After obtaining the fitted K_U, the thermodynamic stability was calculated using the equation $\Delta G^o_{U,ST} = -RT\ln K_U$.

2.3.11. Determination of biotin affinity ($K_{d,biotin}$) of mSA variants by FRET

Biotin binding affinity of a weaker binding mSA variant mSA-W79M was measured by titration of 50 nM GlpG single cysteine variants labeled with BtnPyr (FRET donor) with mSADAB-W79M (FRET acceptor) in 5 mM DDM, 0.25 mM TCEP, 20 mM Na_2HPO_4 and 200 mM NaCl (pH 7.5). The titrated samples were transferred to a 96-well UV-compatible microplate, sealed with a polyolefin tape, and incubated for 24 h at room temperature. Binding was monitored by the decrease of pyrene-monomer fluorescence at 390 nm with an excitation wavelength of 345 nm

using SpectraMax M5e plate reader. Data were averaged from four readings. For fitting of binding data to obtain $K_{d,biotin}$ of mSA_{DAB}-W79M, the following equation was used:

$$F = \mathrm{A1} \times \frac{(P_T + [\mathrm{mSA}] + K_{d,biotin}) - \sqrt{(P_T + [\mathrm{mSA}] + K_{d,biotin})^2 - 4P_T[\mathrm{mSA}]}}{2P_T} + \mathrm{A2} \qquad \textbf{(4)}$$

,where F is the measured fluorescence intensity, P_T is the total GlpG concentration, [mSA] is the total mSA concentration (variable), $K_{d,biotin}$ is the dissociation constant for biotin binding affinity of mSA_{DAB}, A1 is the net fluorescence change, and A2 is the fluorescence level without mSA_{DAB}. Fitted values include $K_{d,biotin}$, A1 and A2; other known values were fixed.

To determine $K_{d,biotin}$ of tight-binding mSA variants and mSA variants lacking DABCYL quencher label, a FRET-based competition assay was employed. 1 μM GlpG was pre-equilibrated with a 2–5 times excess of mSADAB variant for 1 h at room temperature. In this condition, pyrene fluorescence was suppressed. Next, weaker-affinity unlabeled mSA variant was titrated into the sample. Here either DABCYL-labeled or unlabeled mSA variant had known $K_{d,biotin}$. The titrated samples were transferred to a 96-well UV-compatible microplate, sealed with a polyolefin tape, and incubated for 24 h at room temperature. Resultant dissociation of mSADAB by competition was monitored by the increase of pyrene-monomer fluorescence at 390 nm with an excitation wavelength of 345 nm using SpectraMax M5e plate reader. Data were averaged from four readings. For fitting of competition data to obtain unknown $K_{d,biotin}$, the following equation was used:

$$F = \mathrm{A1} \times \frac{-\left[P_T + [\mathrm{mSA}] + \frac{K_{unlabel}}{K_{dabcyl}}(C_T - P_T)\right] + \sqrt{\left(P_T + [\mathrm{mSA}] + \frac{K_{unlabel}}{K_{dabcyl}}(C_T - P_T)\right)^2 + 4P_T[\mathrm{mSA}]\frac{K_{unlabel}}{K_{dabcyl}}}}{2P_T\frac{K_{unlabel}}{K_{dabcyl}}} + \mathrm{A2}$$

(5)

, where F is the measured fluorescence intensity, P_T is the total GlpG concentration, C_T is the total mSA_{DAB} concentration, [mSA] is the total unlabeled mSA concentration (variable), $K_{unlabel}$ is the $K_{d,biotin}$ for mSA without a DABCYL label, K_{dabcyl} is the $K_{d,biotin}$ for mSA_{DAB}, A1 is the amplitude of binding, and A2 is the initial fluorescence level. Fitted values include unknown $K_{unlabel}$ or K_{dabcyl}, A1 and A2; all other values are fixed.

2.3.12. SDS denaturation of GlpG variants labeled with BtnPyr

0.4 μM GlpG doubly labeled with BtnPyr was titrated with SDS in 5 mM DDM, 20 mM Na_2HPO_4, 200 mM NaCl, pH 7.5, ranging from 0 to 0.9 SDS mole-fraction (X_{SDS} = [SDS]/([SDS]+[DDM]). Samples were incubated overnight at room temperature. The detailed scheme for fluorescence spectroscopy to monitor SDS-induced equilibrium unfolding is described ref. (89). Unfolding curves were constructed using the average of three measurements. To determine thermodynamic stability of GlpG from SDS denaturation ($\Delta G^o_{U,SDS}$), the unfolding curves were fitted to the following two-state Santoro-Bolen equation (90):

$$F = \frac{\left[(Fl_F + m_F X_{SDS}) + (Fl_U + m_U X_{SDS}) \times \exp\left(\frac{m_{SDS}(X_{SDS} - X_{SDS,1/2})}{RT}\right)\right]}{1 + \exp\left(\frac{m_{SDS}(X_{SDS} - X_{SDS,1/2})}{RT}\right)} \qquad \textbf{(6)}$$

$$\Delta G^o_{U,SDS} = X_{SDS,1/2} \times m_{SDS} \qquad \textbf{(7)}$$

F is the net fluorescence change. Baselines for the pre- and post-transition regions were determined by the fitted parameters: Fl_F, the fluorescence value for fully folded GlpG; Fl_U, the fluorescence value for fully unfolded GlpG; m_F, the slope of the fully folded baseline; and m_U, the slope of the fully unfolded baseline. m_{SDS} is the slope of the transition region against X_{SDS} and $X_{SDS,1/2}$ is the

transition midpoint. Fitted values include Fl_F, Fl_U, m_F, m_U, m_{SDS}, and $X_{SDS,1/2}$. $\Delta G^o_{U,SDS}$ represents thermodynamic stability in the absence of denaturant.

2.4. Results and Discussion

2.4.1. Design and synthesis of new steric trapping probes

Our steric trapping probes are characterized by three features that are integrated into one molecular tag (**Fig. 2.2**): (i) a biotin group for binding mSA, (ii) a thiol-reactive group for conjugation to engineered cysteine residues on a target protein and (iii) a fluorescent or paramagnetic reporter group whose spectroscopic signal is sensitized by mSA binding or protein unfolding. Each probe was synthesized by stepwise substitutions of building blocks possessing characteristic features into a lysine or cysteine template. BtnPyr-IA (1) is a pyrene-based fluorescent sensor to detect mSA binding. When pyrenes are used to doubly label a target protein, their fluorescence is remarkably sensitive to the binding of quencher-labeled mSA by Förster resonance energy transfer (FRET). BtnRG-TP (2) is a paramagnetic sensor possessing a 1-oxyl-2,2,5,5-tetramethylpyrroline spin label to detect protein unfolding. The spin labels allow distance measurements to be obtained in the native and sterically trapped unfolded state using double electron-electron resonance spectroscopy (DEER).

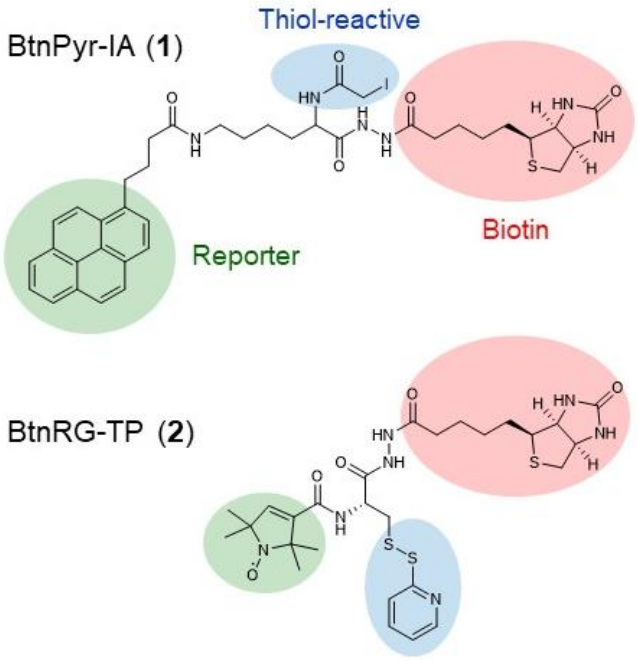


Figure 2.2 Steric trapping probes developed in this study. Thiol-reactive biotin derivatives possessing a spectroscopic reporter group developed in this study. BtnPyr-IA (1): biotin (red shaded)–pyrene (green shaded)–iodoacetamide (blue shaded) conjugated to a lysine template; BtnRG-TP (2): biotin (red shaded)–1-oxyl-2,2,5,5-tetramethylpyrroline spin label (green shaded)–thiopyridine (blue shaded) conjugated to a cysteine template.

2.4.2. Steric trapping controls reversible folding of GlpG

To prove the principle of our steric trapping strategy using the new probes, we used GlpG as a model and its proteolytic activity as a folding indicator. Here all studies were performed in DDM micelles, in which a majority of functional and folding studies of GlpG have been carried out (57, 91–94), and with the isolated transmembrane (TM) domain (residues 87–276) for which all structures of GlpG have been solved. For precise and efficient measurement of GlpG activity with membrane-bound substrates, we developed a fluorescence-based assay that can be transformed into a high-throughput format (**Fig. 2.3**). Second transmembrane segment of the lactose permease of *E. coli* fused to staphylococcal nuclease domain (SN-LacYTM2) (50). IA-

NBD, a thiol-reactive environment-sensitive fluorophore was conjugated to an engineered cysteine in the P5 position from the scissile bond. Cleavage of LacYTM2 led to a large decrease in the fluorescence intensity as NBD was transferred from the nonpolar micellar phase into the bulk aqueous phase. The rate of decrease in NBD fluorescence upon the addition of GlpG correlated with the cleavage of SN-LacYTM2 monitored by SDS-PAGE. Therefore, the fluorescence intensity monitored over time is a quantitative measure of GlpG activity.

Michaelis-Menten kinetic analysis of all GlpG variants was carried out in DDM micelles to ensure we obtained proteins with native-like activity. The activity of the TM domain of GlpG was indistinguishable from that of the full-length protein (**Table 2.1**, top row).

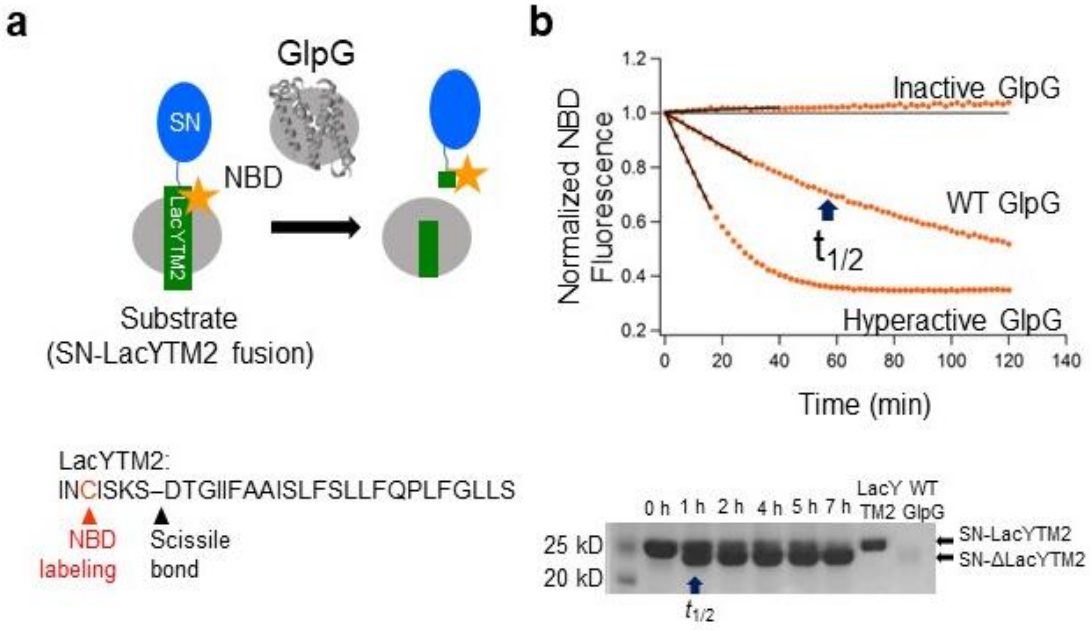


Figure 2.3 New high-throughput assay for measuring the proteolytic activity of GlpG. (a) Labeling of SN-LacYTM2 with IANBD in the P5 position relative to the scissile bond. After cleavage by GlpG, NBD will be released into bulk water with the SN domain. **(b)** (**Top**) Changes in the NBD fluorescence over time due to the proteolytic activity of GlpG. Addition of wild-type (WT) GlpG decreased NBD fluorescence. In contrast, addition of inactive GlpG variant (S201T) displayed negligible change in NBD fluorescence, and hyperactive GlpG mutant (W236A) increased the rate of NBD fluorescence change relative to WT. (**Bottom**) In the conventional SDS-PAGE assay for GlpG activity, a lower molecular weight band appeared, which corresponded to cleaved SN-LacYTM2 (SN-ΔLacYTM2).

	K_M (mM)	k_{cat} (min^{-1})	k_{cat}/K_M ($min^{-1}\,mM^{-1}$)
WT full length	30±2	0.20±0.01	0.0074±0.0007
WT TM	33±3	0.24±0.01	0.0067±0.0005
P95C-BtnPyr	33±7	0.24±0.03	0.0072±0.0017
P95C-BtnPyr+mSA	34±6	0.23±0.02	0.0070±0.0015
G172C-BtnPyr	29±5	0.31±0.02	0.011±0.0020
G172C-BtnPyr+mSA	27±4	0.28±0.02	0.010±0.0015
V267C-BtnPyr	34±2	0.31±0.01	0.0092±0.0007
V267C-BtnPyr+mSA	35±3	0.30±0.01	0.0082±0.0008

Table 2.1 Michaelis-Menten analysis of the proteolytic activity of GlpG. Summary of fitted parameters from Michaelis-Menten analysis of the kinetic activity data. All measurements were done in pH 7.5 Na_2HPO_4, 200 mM NaCl, 0.5 mM TCEP , 5 mM DDM detergent solution ([GlpG] = 1 μM and [SN-LacYTM2] = 20 μM). The errors denote mean ± s. d. from fitting.

For steric trapping, we first identified optimal residue pairs for cysteine substitution to conjugate thiol-reactive biotin labels. After testing multiple single- and double-cysteine variants, we selected two double-cysteine variants, P95C/G172C and G172C/V267C (**Fig. 2.4(a)**). The biotin pair conjugated to P95C/G172C is located in the approximate N-terminal half of GlpG (hereafter $95/172_N$), while the biotin pair conjugated to G172C/V267C is located in the C-terminal half ($172/267_C$). The individual single-cysteine variants P95C, G172C and V267C labeled with fluorescent BtnPyr-IA maintained the wild-type activity level (**Fig. 2.4(b)**, top), and this activity level was not significantly altered after binding of wild-type mSA (mSA-WT). Michaelis-Menten analysis of the proteolytic activity showed that K_m, k_{cat} and k_{cat}/K_m of the single-biotin variants bound with mSA were indistinguishable from those of unbound forms, demonstrating that binding of one mSA molecule to each biotin site did not perturb the structure and function of GlpG (**Table 2.1**, bottom 3 rows). The wild-type activity level was also maintained after labeling of double-

cysteine variants. In marked contrast, saturated binding of mSA to two biotin labels on each variant induced a substantial loss of activity, implying that GlpG was trapped in the unfolded state (**Fig. 2.4(b)**, bottom).

Next, we tested whether the sterically trapped unfolded state could refold after dissociation of bound mSA. Wild-type mSA binds biotin with an enormously high affinity ($K_{d,biotin} \approx 10{-}14$ M) and slow dissociation rate ($k_{off} \approx$ days) (95). Thus, we used the mSA-S27A variant, with a weaker biotin affinity, to facilitate dissociation of bound mSA by addition of excess free biotin (Fig. 2.1a) (96). Both double-biotin variants, when inactivated with mSA-S27A, regained the activity to a significant degree upon addition of free biotin (**Fig. 2.4(b)**, bottom): for $95/172_N$-$BtnPyr_2$, 50−70% of lost activity was regenerated, and for $172/267_C$-$BtnPyr_2$, >90% was regenerated. Thus, we achieved reversible folding of GlpG by steric trapping without using denaturants.

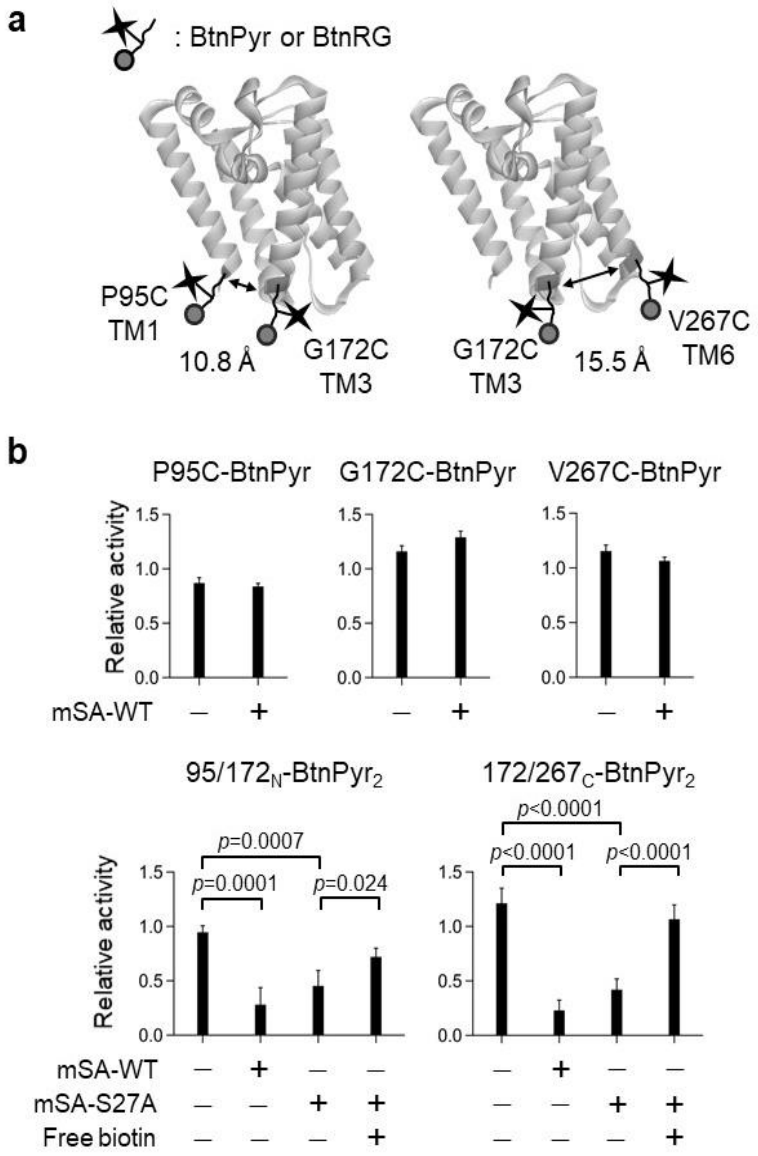


Figure 2.4 GlpG reversibly unfolds by double-binding of mSA. (**a**) Locations of two different biotin pairs for steric trapping in the structure of GlpG (PDB code: 3B45 (51)) and their C_α-C_α distances. (**b**) Reversible control of GlpG folding tested by the proteolytic activity as a folding indicator. All activity levels were normalized relative to the activity of wild-type GlpG. Error bars denote mean ± s. d. (n=5 for the data without mSA and n=3 for the data with mSA). **Top panels:** binding of wild-type mSA (mSA-WT) to individual single-cysteine variant labeled with BtnPyr did not affect the activity. **Bottom panels:** saturated binding of mSA-WT to each double-cysteine variant labeled with BtnPyr led to an inactivation of GlpG (the second bar from the left in each panel). To test folding reversibility, double-biotin GlpG variants were first inactivated with mSA-S27A possessing a weaker biotin binding affinity ($K_{d,biotin}$=1.4×10^{-9} M) for 2–5 days (the third bar). Next, excess free biotin was added to induce competitive dissociation of bound mSA (the fourth bar). All *p*-values obtained from Student's *t*-test were smaller than the threshold significance level (p=0.05), indicating that the activity changes for the unfolding and refolding reactions were significant.

2.4.3. Sterically trapped unfolded state is widely unraveled

So far, protein unfolding by steric trapping has been tested by the loss of enzymatic activity (81, 85), decrease of retinal absorbance (82) or increased susceptibility to proteolysis (81, 82). Although those features indicate unfolding, the possibility remains that the protein conformation trapped with mSA molecules is only locally distorted or still compact with residual tertiary interactions. Therefore, to elucidate the conformation of the sterically trapped unfolded state as well as to gain insights into the unfolded state ensemble of membrane proteins under non-denaturing conditions, we used a thiol-reactive biotin derivative possessing a spin label (BtnRG-TP) (**Fig. 2.2**). Labeling double-cysteine variants of GlpG with this probe provides the benefits of both trapping the unfolded state and measuring the distances between spin labels using DEER. DEER allows for measurements of long-range (15–60 Å) interspin distances (97) and provides not only the most probable distance but also the distance distribution, which is of great interest in characterization of the unfolded state (98, 99). Here, we obtained interspin distances for $95/172_N$-$BtnRG_2$ and $172/267_C$-$BtnRG_2$ in their native, SDS-induced unfolded and sterically trapped unfolded states (**Fig. 2.5**). In the native states, the distance distributions between BtnRG labels were overall similar to those between well-characterized R1 spin labels (100), demonstrating that our BtnRG label is capable of distance mapping of protein conformation.

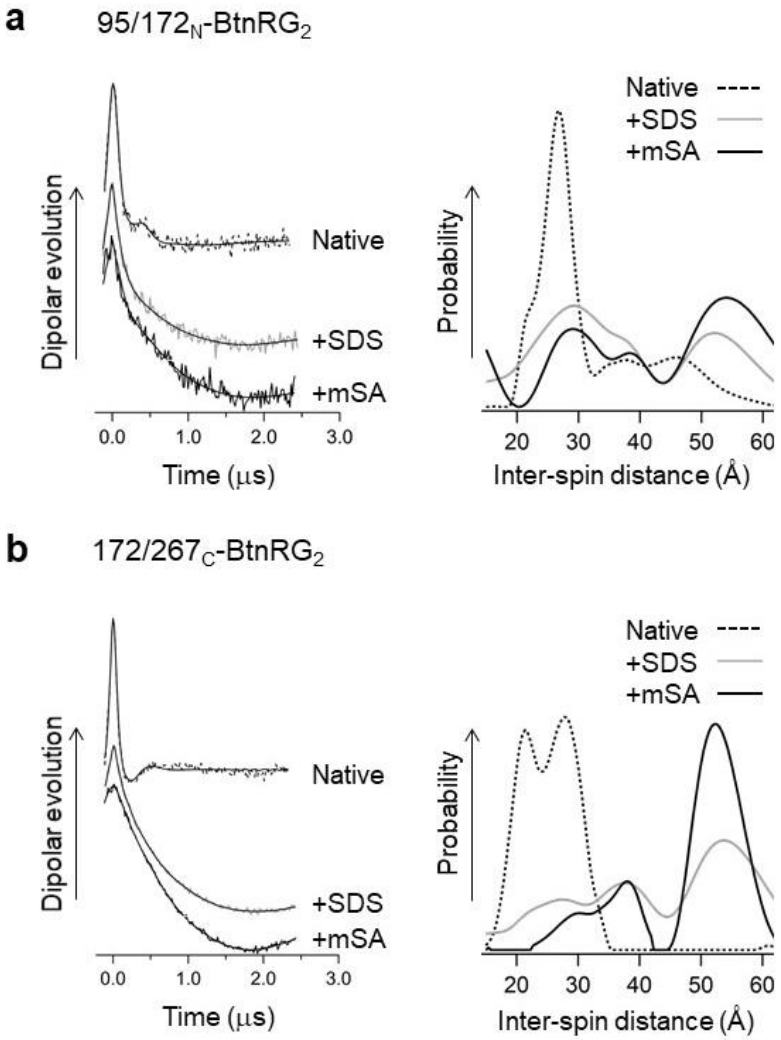


Figure 2.5 DEER suggests steric trapping induce wide separation of two biotinylated sites. Background-subtracted dipolar evolution data and their fits (left) and inter-spin distances (right) for the native (dashed lines), SDS-induced unfolded (gray solid lines, SDS mole fraction =[SDS]/([DDM]+[SDS])>0.8, in which the unfolded fraction exceeded 0.9), and steric-trapped (black solid lines) unfolded states for (**a**) 95/172$_N$-BtnRG$_2$ GlpG and (**b**) 172/267$_C$-BtnRG$_2$ GlpG. The approximate upper limit of the reliable mean distance was ~53 Å (97).

For both variants, SDS induced substantial broadening of the interspin distance distribution over the range from the native-like distances (15–35 Å) up to ~60 Å (**Fig. 2.5**, right panels), which indicates the existence of a heterogeneous conformational ensemble of the unfolded state in SDS. Interestingly, in non-denaturing DDM micelles, the sterically trapped unfolded states also exhibited similarly broad interspin distance distributions. The increase of the most probable distance from ~25 Å in the native state to ~55 Å in the sterically trapped unfolded state corresponds to an ~30 Å expansion of each half of the polypeptide chain covered by the respective biotin pair. This increased dimension is comparable to the whole diameter of native GlpG. Thus, our DEER data for GlpG rule out the existence of a compact unfolded state under non-denaturing conditions, such as has been observed for several water-soluble proteins (101).We note that, because of the detection limit of DEER, even longer-distance components (>60 Å) may have existed but not been detected. Addition of dithiothreitol to break the disulfide bond between GlpG and the biotin label bound with mSA led to regeneration of >70% of lost activity, indicating that a majority of the unfolded conformations were able to refold.

Steric repulsion between bound mSA molecules may have biased the conformational ensemble of the unfolded state. However, during the selection of optimal biotin pairs, we found that saturated binding of mSA to the biotin pairs conjugated to G94C/G172C and G172C/N271C, whose C_α-C_α distances were similar to those of $95/172_N$ and $172/267_C$, completely maintained the activity and therefore did not induce unfolding (88). This result implies that bound mSA molecules are allowed to coexist within close distances, probably also in the sterically trapped unfolded state. Therefore, steric repulsion may not fully explain the expanded unfolded state.

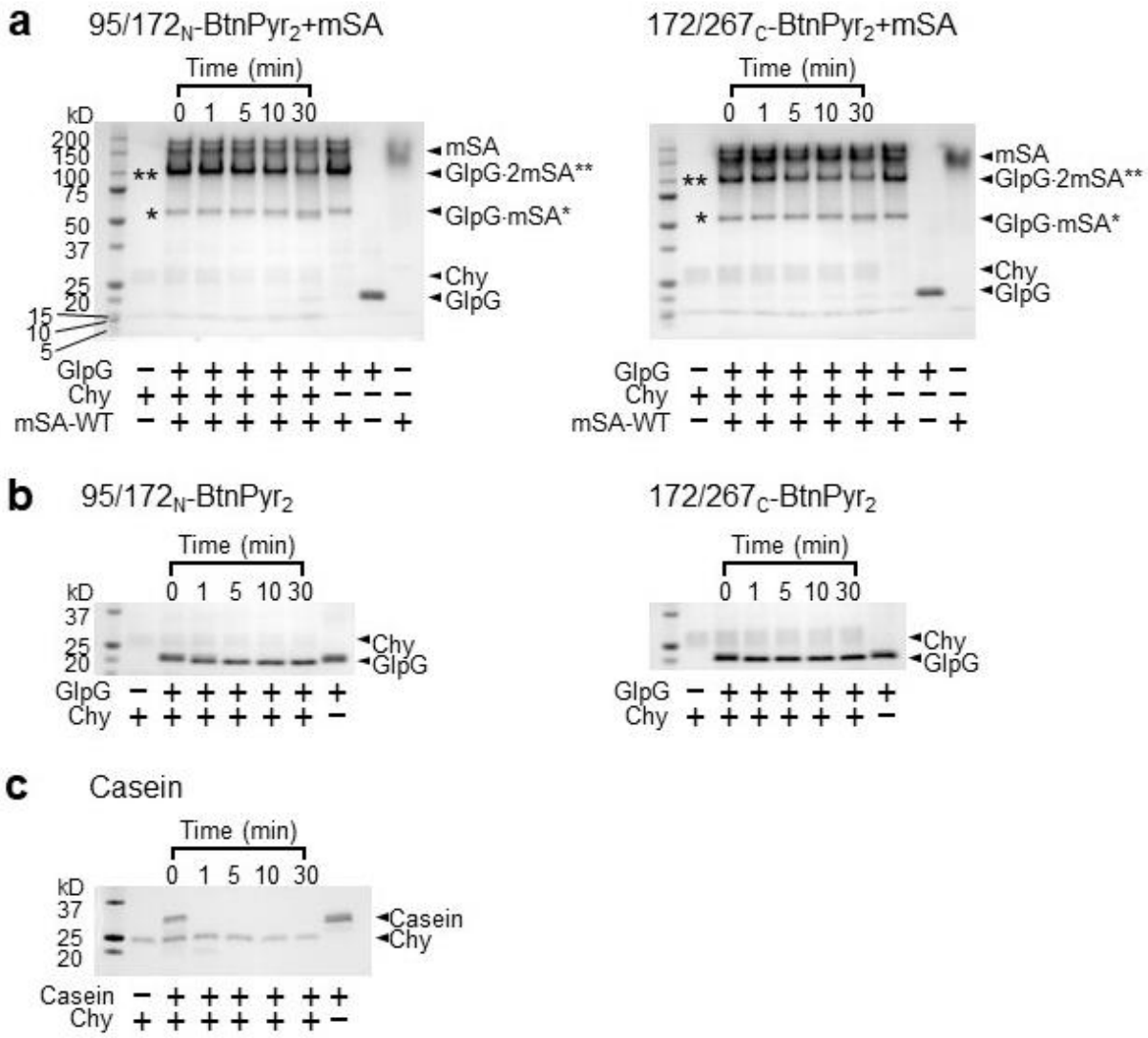


Figure 2.6 Probing the flexibility of the steric-trapped unfolded state of GlpG using proteolysis and SDS-PAGE. (a) Time-dependent proteolysis of the steric-trapped unfolded states (GlpG·2mSA, marked with **) of 95/172$_N$-BtnPyr$_2$ (left) and 172/267$_C$-BtnPyr$_2$ (right) by chymotrypsin (Chy). **(b)** Proteolysis of native double-biotin variants in the absence of mSA. Those samples exhibited only partial digestion of the terminal flexible regions (Wang *et al.* **2006** *Nature* 444, 179-180). **(c)** Proteolysis of casein as a control of protein substrate lacking significant ordered secondary structures. Detailed procedures are described in Materials and Methods.

We further characterized the conformational features of the sterically trapped unfolded state using proteolysis by chymotrypsin, which primarily targets aromatic residues prevalent throughout GlpG (**Fig. 2.6**). Whereas the unfolded state bound with two mSA molecules was gradually proteolyzed over ~30 min, either native GlpG or GlpG bound with one mSA molecule was not proteolyzed. As a control, we tested proteolysis of casein, which exists predominantly in random-coil conformation in aqueous solution (102). Casein was proteolyzed rapidly, within 1 min. Thus, we speculate that the sterically trapped unfolded state was mainly protected by secondary structures and micelles but possessed more dynamic features than the native state.

DEER and proteolysis results demonstrate that steric trapping induced a true unfolded state, which was an ensemble of expanded dynamic and heterogeneous conformations. This work also represents the first measurement of the physical dimension of a helical membrane protein in its unfolded state under non-denaturing conditions.

2.4.4. Stability of GlpG determined by steric trapping

To develop a general steric trapping strategy that does not depend on specific characteristics of a target protein, ideally the spectroscopic signal from the reporter group in our probe should sensitively change upon either mSA binding or protein unfolding. Here we achieved highly sensitive detection of mSA binding by employing FRET between the pyrene of the BtnPyr label and the non-fluorescent quencher DABCYL attached to mSA (mSA_{DAB}) (**Fig. 2.7(a)**).

SDS denaturation and linear extrapolation of the denaturation data to the zero-SDS mole fraction yielded the same stability ($\Delta G^{o}_{U,SDS}$) for $95/172_{N}$-$BtnPyr_2$ (8.4 ± 1.5 kcal/mol) and $172/267_{C}$-$BtnPyr_2$ (8.7 ± 1.2 kcal/mol) (**Fig. 2.7(b)**) which also was similar to that of the full-length wild-type GlpG (8.2 ± 1.4 kcal/mol) (58). This result indicates that the two double-biotin

variants possess the same global stability as wild-type GlpG. By design, steric trapping specifically captures transient unfolding of native interactions between a biotin pair. Thus, probing the stability with two biotin pairs located in different regions (**Fig. 2.4(a)**) provides a novel opportunity to test the folding cooperativity of GlpG.

An essential element of steric trapping to determine protein stability is choosing an mSA variant whose binding to a biotin label (ΔG^{o}_{bind}) optimally competes with folding (ΔG^{o}_{U}) to yield attenuated second binding in a desirable mSA concentration range (**Fig. 2.7(a)**). Among the mSA variants tested, mSA_{DAB}-S27A yielded an optimal separation of the first tight and second weaker binding phases (**Fig. 2.7(c)**). Parallel activity measurements showed that, for each GlpG variant, the second binding coincided with the activity loss (i.e., unfolding), which validated the unfolding-binding coupling.

Fitting of the second binding phases yielded a thermodynamic stability ($\Delta G^{o}_{U,ST}$, where ST signifies steric trapping) of 5.8 ± 0.2 kcal/mol for $95/172_{N}$-$BtnPyr_2$ and 4.7 ± 0.1 kcal/mol for $172/267_{C}$-$BtnPyr_2$ (equation (3) in Materials and Methods) in non-denaturing DDM micelles (**Fig. 2.7(c)**). Both $\Delta G^{o}_{U,ST}$ values were substantially lower than the extrapolated stability from SDS denaturation (8.4–8.7 kcal/mol) but higher than the stability in a bicelle (6.5 kBT, equivalent to ~4 kcal/mol) extrapolated to the zero force condition from a single-molecule magnetic tweezers study (60). If GlpG unfolded cooperatively, the same $\Delta G^{o}_{U,ST}$ would be expected regardless of the position of the biotin pair. However, while SDS denaturation yielded the same global stability for the two double-biotin variants, their stabilities obtained by steric trapping were comparable but significantly different by 1.1 ± 0.2 kcal/mol.

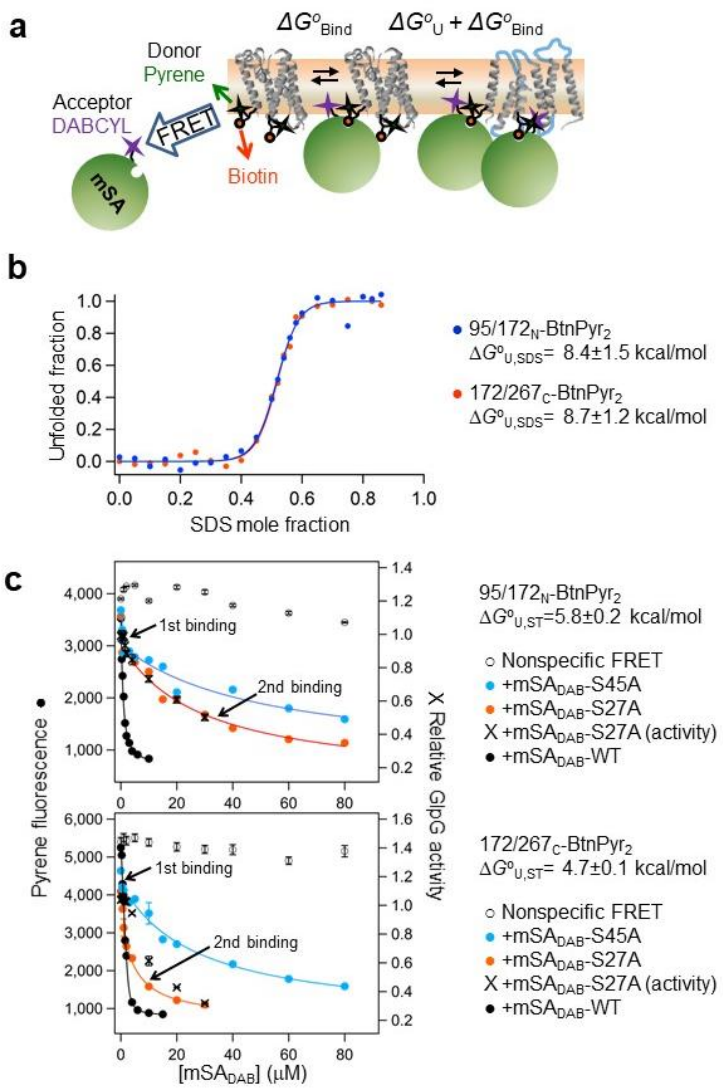


Figure 2.7 Thermodynamic stability of GlpG using steric trapping and SDS denaturation. (**a**) Steric-trapping strategy using FRET between fluorescent pyrene (donor) in BtnPyr labeled on GlpG and non-fluorescent quencher DABCYL (acceptor) thiol-specifically labeled near the biotin binding pocket (Y83C) of the active subunit of mSA (mSA_{DAB}). (**b**) Equilibrium unfolding of GlpG variants $95/172_N$-$BtnPyr_2$ and $172/267_C$-$BtnPyr_2$ as a function of SDS mole fraction measured by FRET between Trp residues (donor) of GlpG and pyrene (acceptor) on BtnPyr labels. Errors in $\Delta G^o_{U,SDS}$ values denote mean ± s. d. from fitting. (**c**) Binding isotherms of $95/172_N$-$BtnPyr_2$ and $172/267_C$-$BtnPyr_2$ with three mSA variants mSA_{DAB}-WT (black circles, $K_{d,biotin}$=~10^{-14} M), mSA_{DAB}-S27A (red circles, $K_{d,biotin}=1.4\times10^{-9}$ M) and mSA_{DAB}-S45A (blue circles, $K_{d,biotin}=9.0\times10^{-9}$ M). The activity change for each double-biotin variant (crosses, right y-axis) was measured at an increasing concentration of mSA_{DAB}-S27A. The thermodynamic stability ($\Delta G^o_{U,ST}$) of each variant was obtained by fitting of the second mSA-binding phase to **equation (3)** in Materials and Methods. Procedures to obtain nonspecific FRET (open circles) are described in Online Methods. Errors in fluorescence denote mean ± s. d. (n=4). Errors in activity denote ± s. d. from fitting. Errors in $\Delta G^o_{U,ST}$ values denote mean ± s. d. (n=3).

2.4.5. Subglobal unfolding of GlpG near the active site

To track down the origin of the discrepancy between the stability obtained by steric trapping ($\Delta G^{o}_{U,ST}$) under non-denaturing conditions and the extrapolated stability obtained by SDS denaturation ($\Delta G^{o}_{U,SDS}$), we directly measured the stability of the two double-biotin variants using steric trapping in the range of SDS mole fraction (X_{SDS} = 0–0.4) in which a major fraction of GlpG existed in the folded state (folded fraction >0.9). A plot of $\Delta G^{o}_{U,ST}$ versus X_{SDS} (**Fig. 2.8**) revealed two major features that clearly deviated from the behavior predicted from linear extrapolation of the SDS denaturation data. First, rather than following a linearly decreasing trend, the $\Delta G^{o}_{U,ST}$ plots for both variants exhibited an upward curvature as X_{SDS} increased. Second, while $\Delta G^{o}_{U,ST}$ values for of $95/172_{N}$-$BtnPyr_2$ were larger overall than those for $172/267_{C}$-$BtnPyr_2$, they converged remarkably at $X_{SDS} \approx 0.4$, where the main unfolding transition by SDS began, and this convergence was maintained up to $X_{SDS} = 0.5$. This result confirms that the two variants possess the same global stability. The effect of SDS on the $K_{d,biotin}$ of mSA accounted for (**Table 2.2**).

The overall nonlinearity of $\Delta G^{o}_{U,ST}$ against X_{SDS} implies a complex interaction between GlpG and DDM/SDS micelles. A similar disagreement between steric trapping and SDS denaturation has been reported for bacteriorhodopsin in DMPC/CHAPSO/SDS bicelles (82). In the case of GlpG, $\Delta G^{o}_{U,ST}$ of both variants reached a maximum at $X_{SDS} \approx 0.2$ but decreased linearly at higher X_{SDS} (**Fig. 2.8**). Notably, in the range of X_{SDS} = 0.2−0.4, the m value of $95/172_{N}$-$BtnPyr_2$ (14 ± 2 kcal/mol/X_{SDS}), which represents the slope of $\Delta G^{o}_{U,ST}$ against X_{SDS}, was substantially larger than that of $172/267_{C}$-$BtnPyr_2$ (8 ± 1 kcal/mol/X_{SDS}) but similar to those obtained by SDS denaturation (16–17 kcal/mol/X_{SDS}).

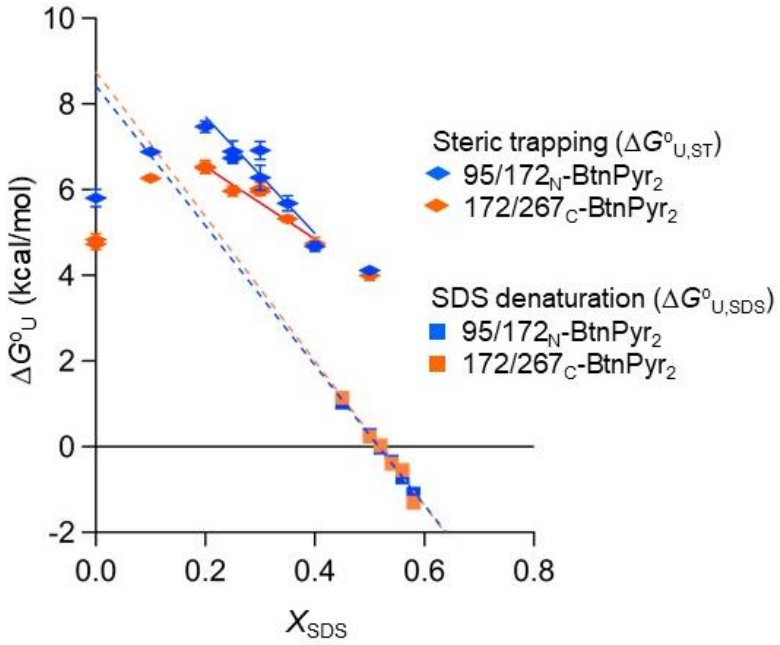


Figure 2.8 Dependence of thermodynamic stability of GlpG on SDS mole fraction. The plot containing $\Delta G^{o}_{U,ST}$'s (diamonds) obtained by steric trapping and $\Delta G^{o}_{U,SDS}$'s (squares) obtained by SDS denaturation as a function of SDS mole fraction (X_{SDS}) for 95/172$_N$-BtnPyr$_2$ and 172/267$_C$-BtnPyr$_2$. To fit $\Delta G^{o}_{U,ST}$, we accounted for the changes in the biotin affinity of mSA$_{DAB}$ variants which depended on X_{SDS}. Errors in $\Delta G^{o}_{U,ST}$ denote ± s. d. from fitting. Solid lines are the linear-regression fits of $\Delta G^{o}_{U,ST}$ in the range of X_{SDS}=0.2–0.4 and dashed lines indicate the extrapolation lines of $\Delta G^{o}_{U,SDS}$ to zero X_{SDS} from equilibrium SDS denaturation. The slope in the ΔG^{o}_{U} *vs* X_{SDS} plot represents the *m*-value. For 95/172$_N$-BtnPyr$_2$, *m*=16±3 (blue dashed line) from SDS denaturation and *m*=14±2 (blue solid line) from steric trapping. For 172/267$_C$-BtnPyr$_2$, *m*=17±2 (red dashed line) from SDS denaturation and *m*=8±1 (red solid line) from steric trapping. Errors in the *m*-values denote ± s. d. from fitting.

For water-soluble proteins, the m value is correlated with the hydrophobic surface area exposed upon unfolding (103). Although the physical meaning of the m value in SDS denaturation is still under debate (104), it is most likely related to the difference in the affinity of SDS for different states of the protein and hence to the degree of exposure of buried stabilizing interactions upon unfolding (105). Therefore, from the different denaturant sensitivities of the two double-

biotin variants, we conclude that trapping of the unfolded state with the biotin pair $95/172_N$-$BtnPyr_2$ led to substantial exposure of the buried surfaces throughout the protein, whereas trapping with the biotin pair $172/267_C$-$BtnPyr_2$ occurred mainly through subglobal unfolding that exposed less buried surface area. Steric trapping of $172/267_C$-$BtnPyr_2$ detects transient separation between TM3 and TM6 to which biotin labels are conjugated, and TM6 contains a biotin label (V267C-BtnPyr) as well as His254 of the catalytic dyad. Thus, subglobal unfolding should directly involve disruption of the active site.

X_{SDS}	mSA_{DAB}-W79M $K_{d,biotin}$ (nM)	mSA_{DAB}-S45A $K_{d,biotin}$ (nM)	mSA_{DAB}-S27A $K_{d,biotin}$ (nM)
0	79 ± 24	9.0 ± 4.3	1.4 ± 0.9
0.1	99 ± 27	5.5 ± 2.1	0.3 ± 0.1
0.2	100 ± 78	4.3 ± 2.7	0.4 ± 0.1
0.3	180 ± 50	5.8 ± 3.0	
0.4	260 ± 100	2.9 ± 0.6	
0.5	270 ± 70	5.8 ± 2.1	
0.6	560 ± 210	3.9 ± 2.0	

Table 2.2 Determination of biotin affinity ($K_{d,biotin}$) of mSA variants by FRET. Summary of $K_{d,biotin}$'s for mSA_{DAB}-W79M determined by direct binding assays, and mSA_{DAB}-S45A and mSA_{DAB}-S27A determined by FRET-based competition assays in various mole fractions of SDS (X_{SDS}). All experiments were done using GlpG G172C-BtnPyr variant in 5 mM DDM, 20 mM Na_2HPO_4 (pH 7.5), 200 mM NaCl, 0.25 mM TCEP solution. Error bars denote mean ± s. d. from fitting.

Subglobal unfolding has frequently been observed in HDX studies of water-soluble proteins (76, 106, 107) but has not been reported for membrane proteins. Besides the different m values, subglobal unfolding of GlpG is further supported by the lower stability of $172/267_C$-$BtnPyr_2$ measured by steric trapping (**Fig. 2.7(c)**) and the reproducibly higher refolding yield of $172/267_C$-$BtnPyr_2$ (>90%) as compared to $95/172_N$-$BtnPyr_2$ (50–70%), which imply that they have different unfolded states. We also note that $\Delta G^o_{U,ST}$ was larger than $\Delta G^o_{U,SDS}$ after it crossed the extrapolation lines at $X_{SDS} \approx 0.1$, and this discrepancy became increasingly pronounced up to 2.8 kcal/mol at $X_{SDS} = 0.4$ (**Fig. 2.8**). We reason that the larger $\Delta G^o_{U,ST}$ was primarily due to the conformational difference between the sterically trapped unfolded state and the SDS-induced unfolded state. Our DEER result supports this argument (**Fig. 2.5**). The sterically trapped unfolded state exhibited larger interspin distances, on average, than the SDS-induced unfolded state. Thus, steric trapping appears to induce more unraveled conformations than SDS, at least for the interactions between the biotinylated sites. However, we are cautious with this direct comparison because the compactness of the SDS-induced unfolded state may change as a function of X_{SDS} as a result of the effects of SDS on the size and shape of mixed micelles (108).

2.4.6. Strategy to identify cooperative interactions

The higher stability and more substantial unfolding obtained with $95/172_N$-$BtnPyr_2$ indicates that the native interactions between this biotin pair in the N-terminal region are critical to the conformational integrity of the whole protein. At the same time, the lower stability and subglobal unfolding obtained with $172/267_C$-$BtnPyr_2$ indicates that the C-terminal region possesses differential folding properties from the N-terminal region. This result suggests complex energetic coupling between different regions in GlpG. To clarify this complexity, we developed a

method to identify cooperative and localized interactions that contribute to the protein stability at a side chain level of resolution (**Fig. 2.9**).

First, we dissected GlpG into two subdomains: (i) the more stable N-terminal subdomain I encompassing TM1-L1-TM2-TM3-$L3_{198}$ (ending at residue 198 in L3 (loop 3)), whose unfolding was trapped with $95/172_N$-$BtnPyr_2$, and (ii) the less stable C-terminal subdomain II consisting of $L3_{199}$-TM4-TM5-L5-TM6 (starting from residue 199), whose subglobal unfolding was trapped with $172/267_C$-$BtnPyr_2$ (see ref. (88) for detailed dissection procedures). The uncertainty of the division point was ± 20–30 residues. Second, we made a single mutation (typically to alanine) in either subdomain to perturb a specific side chain interaction in the background of $95/172_N$-$BtnPyr_2$ and $172/267_C$-$BtnPyr_2$. We referred to these background double-biotin variants as 'wild type' (WT) because the wild-type native interactions were equally preserved in both, as shown by SDS denaturation (**Fig. 2.7(b)**). We referred to two double-biotin variants possessing the same mutation as 'mutants' (Mut). Next, we probed the stability changes induced by the mutation with two different biotin pairs using steric trapping. We quantified the differential effects of the same mutation on the stability of each subdomain ($\Delta\Delta\Delta G^o_U$) using **equation (8)** containing the stabilities of four variants:

$$\begin{aligned}\Delta\Delta\Delta G^o_U = &[(\Delta G^o_{U,95/172_N\text{-}BtnPyr_2}(WT) - \Delta G^o_{U,95/172_N\text{-}BtnPyr_2}(Mut)] \\ &-[\Delta G^o_{U,172/267_C\text{-}BtnPyr_2}(WT) - \Delta G^o_{U,172/267_C\text{-}BtnPyr_2}(Mut)] \\ =&\Delta\Delta G^o_{U,95/172_N\text{-}BtnPyr_2}(WT\text{-}Mut) - \Delta\Delta G^o_{U,172/267_C\text{-}BtnPyr_2}(WT\text{-}Mut) \qquad \textbf{(8)}\end{aligned}$$

$\Delta\Delta G^o_{U,95/172N\text{-}BtnPyr2}(WT\text{-}Mut)$ and $\Delta\Delta G^o_{U,172/267C\text{-}BtnPyr2}(WT\text{-}Mut)$ designate the stability changes caused by the same mutation in the backgrounds of $95/172_N$-$BtnPyr_2$ and $172/267_C$-$BtnPyr_2$, respectively. Thus, $\Delta\Delta\Delta G^o_U$ represents the difference in the stability changes that are probed with two different biotin pairs upon the same mutation.

If a mutation causes a similar degree of destabilization for both double-biotin variants with a difference within thermal fluctuation energy ($|\Delta\Delta\Delta G^{o}_{U}| \leq RT = 0.6$ kcal/mol, where R is the gas constant and T = 298 K), the mutated site engages in a 'cooperative' interaction. That is, the perturbation by the mutation propagates similarly to both subdomains. Among the cases in which $|\Delta\Delta\Delta G^{o}_{U}| > RT$, if a mutation preferentially destabilizes the subdomain containing it, we classified the perturbed interactions as being 'localized' within that subdomain. If mutation of a residue that makes its side chain contacts only with the subdomain containing it preferentially destabilizes the other subdomain, we classified the perturbation as 'over-propagated'.

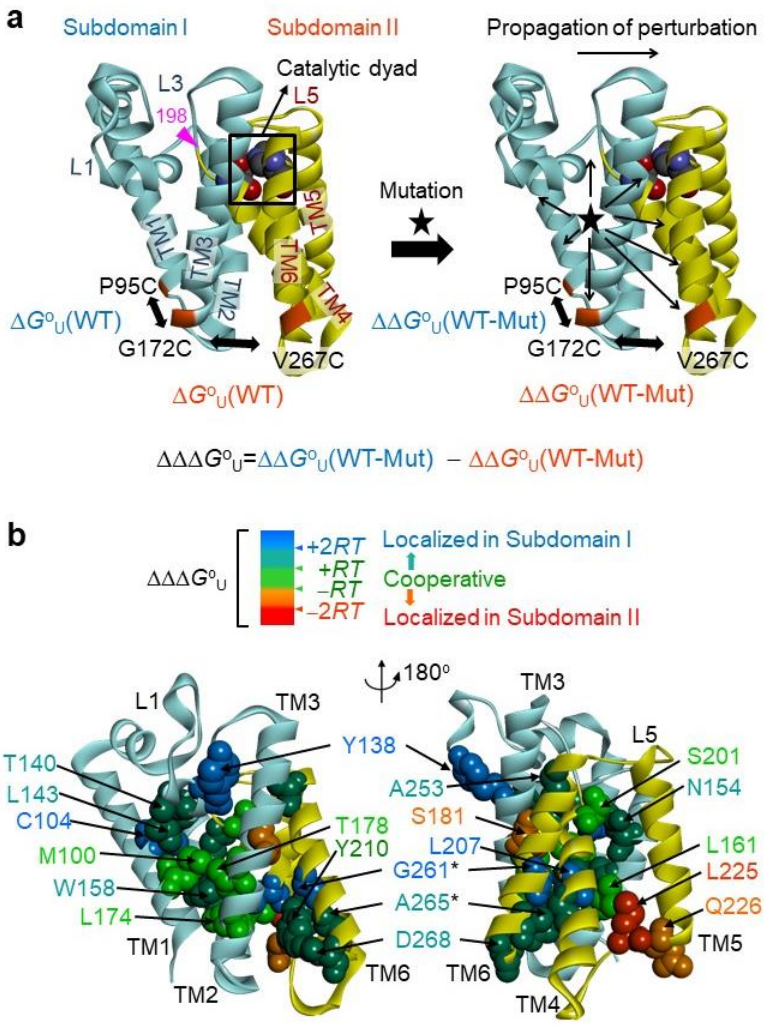


Figure 2.9 Cooperativity map reveals a network of clustered cooperative and localized interactions for the stability of GlpG under a native micellar condition. (**a**) Scheme for quantifying the cooperativity of interactions of a specific side chain. The stability changes ($\Delta\Delta G^{o}_{U}$) induced by the same mutation (black star) were probed with two biotin pairs, $95/172_{N}$-$BtnPyr_2$ and $172/267_{C}$-$BtnPyr_2$ located in the N- and C-terminal regions, respectively, and compared to each other to yield $\Delta\Delta\Delta G^{o}_{U}$ using **equation (8)**. The cyan-backbone region designates subdomain I (TM1-L1-TM2-TM3-$L3_{198}$), which ends at residue 198 in the L3 loop (marked with a magenta wedge) and the yellow-backbone region ($L3_{199}$-TM4-TM5-L5-TM6) indicates subdomain II. Catalytic dyad composed of Ser201/His254 is shown as spheres. (**b**) Cooperativity map at a side-chain resolution. The map shows the "cooperative" (green, $|\Delta\Delta\Delta G^{o}_{U}| \leq RT$=0.6 kcal/mol) and "localized" ($|\Delta\Delta\Delta G^{o}_{U}| > RT$) side-chain interactions. Localized interactions were further divided using additional cut-off energy values, $2RT \geq |\Delta\Delta\Delta G^{o}_{U}| > RT$ ("moderately-localized" interactions) and $|\Delta\Delta\Delta G^{o}_{U}| > 2RT$ ("highly-localized" interactions). Each side chain was color-coded based on these criteria for $\Delta\Delta\Delta G^{o}_{U}$ as shown in the figure. Interactions mediated by residues G261 and A265 (denoted with stars) were "over-propagated". Errors in individual ΔG^{o}_{U} were ±0.1–±0.2 kcal/mol (mean ± s. d. from fitting) and errors in $\Delta\Delta\Delta G^{o}_{U}$ ranged from ±0.1–±0.4 kcal/mol, which were calculated using the propagation of errors in ΔG^{o}_{U} (**Table 2.3**).

2.4.7. Cooperativity network in GlpG

We targeted 20 residues covering key packing regions (57) and analyzed their roles in the folding cooperativity of GlpG (**Table 2.3**). The stability changes upon mutation $\Delta\Delta G^{o}_{U}$ (WT-Mut) obtained by steric trapping were reasonably well correlated with the changes in melting temperature ΔT_m(WT-Mut) (57) upon corresponding mutation, which validated our approach. 20 $\Delta\Delta\Delta G^{o}_{U}$ values were distributed over a wide range from −1.8 to 2.0 kcal/mol, and their individual errors ranged from ±0.1 to ±0.4 kcal/mol, smaller than RT. We applied four cutoff values, $\Delta\Delta\Delta G^{o}_{U}$ = −2RT, −RT, RT and 2RT (i.e., five sets of the cooperativity profile), to account for the wide distribution of $\Delta\Delta\Delta G^{o}_{U}$ as well as to more precisely resolve the degree of cooperativity of each side chain interaction.

We mapped the effects of mutations onto the structure, which we called the 'cooperativity map' (**Fig. 2.9(b)**). Surprisingly, we observed clustering of cooperative and localized interactions in defined regions in the GlpG structure and divided their spatial distributions into four distinct groups. First, cooperative interactions of five residues, Met100, Leu161, Leu174, Thr178 and Ser201, clustered in the buried region that was surrounded by subdomain I and the subdomain interface near the center of the membrane. This cooperative cluster overlapped with one of the key packing regions previously identified and partially overlapped with the folding nucleus formed between TM1 and TM2 in the folding transition state (58).

Second, all tested residues located in the folded L1 loop (Tyr138, Thr140 and Leu143) and the residue packed against L1 (Cys104) in subdomain I engaged in moderately (RT < $\Delta\Delta\Delta G^{o}_{U}$ ≤ 2RT) or highly (2RT < $\Delta\Delta\Delta G^{o}_{U}$) localized interactions in subdomain I. This region is known to form non-native interactions in the folding transition state (58). Third, Leu225 ($\Delta\Delta\Delta G^{o}_{U}$ < −2RT) and Gln226 (−2RT ≤ $\Delta\Delta\Delta G^{o}_{U}$ < −RT) in TM5 in subdomain II, which were located at the

subdomain interface and exposed to the water-micelle interface, respectively, were both classified as localized in subdomain II. TM5 is not tightly packed against the rest of the protein and does not contribute much to the thermostability (57).

In the fourth cluster, interestingly, mutation of residues at the TM4-TM6 interface (Ala253, Gly261, Ala265 and Asp268) in subdomain II preferentially destabilized subdomain I, not the subdomain containing them. In particular, Gly261 and Ala265 make their side chain contacts entirely with the residues in subdomain II, but perturbing these interactions exerted larger impacts on the stability of subdomain I. Thus, we classified these residues as over-propagated. The TM4-TM6 interface harbors the catalytic dyad and is pivotal to both the stability and the function of GlpG (57). In particular, Gly261 and the dyad are absolutely conserved among rhomboid proteases (109). Our result suggests that these conserved residues are also critical to the energetic coupling between different regions of GlpG. Breakage of the interactions near the C terminus and the propagation of the breakage toward the N terminus are known to be the primary mechanism of the force-induced unfolding of GlpG (60).

It should be noted that 5 of the 20 tested mutations completely inactivated GlpG (**Table 2.3**). Thus, our steric trapping strategy allowed stability measurements of not only functional but also nonfunctional variants, which had been difficult to achieve with the original steric trapping framework. Although the two double-biotin variants bearing the same mutation exhibited differential stability change in DDM micelles, they possessed the same global stability as a result of SDS denaturation (**Fig. 2.10**). Therefore, we conclude that the networked side chain interactions revealed in this work are a novel phenomenon that occurs under native conditions.

Mutation	$95/172_N$-$BtnPyr_2$ $\Delta\Delta G^o_{U,95/172N\text{-}BtnPyr2}$ (WT-Mut)	Activity	$172/267_C$-$BtnPyr_2$ $\Delta\Delta G^o_{U,\ 172/267C\text{-}BtnPyr2}$ (WT-Mut)	Activity	$\Delta\Delta\Delta G^o_U$	Location
Cooperative interactions						
M100A	2.8±0.2	0.95 ±0.03	2.5±0.5	0.89 ±0.02	0.3±0.5	Subdomain I[a] TM1[b]/Interface[c]
L161A	1.9±0.2	0.13 ±0.02	1.8±0.4	0.10 ±0.01	0.1±0.4	Subdomain I TM2/Interface
L174A	3.8±0.2	0.23 ±0.04	3.3±0.2	0.14 ±0.05	0.5±0.2	Subdomain I TM3/Interface
T178A	0.7±0.1	1.35 ±0.05	0.3±0.1	1.63 ±0.05	0.5±0.1	Subdomain I TM3/Interface
S201T	1.0±0.2	0.04 ±0.02	1.0±0.3	0.03 ±0.02	0.0±0.4	Subdomain II TM4/Interface
Localized interactions in Subdomain I						
C104A	2.2±0.3	0.81 ±0.02	0.2±0.2	1.30 ±0.04	2.0±0.3	Subdomain I TM1/interface
Y138F	1.9±0.2	0.59 ±0.03	0.6±0.2	1.48 ±0.04	1.3±0.4	Subdomain I L1
T140A	1.7±0.1	1.39 ±0.05	0.7±0.2	1.19 ±0.04	0.9±0.2	Subdomain I L1
L143A	2.4±0.2	0.96 ±0.02	1.3±0.2	1.26 ±0.03	1.1±0.2	Subdomain I L1
N154A	1.3±0.2	0.07 ±0.01	0.4±0.3	0.09 ±0.01	0.9±0.2	Subdomain I TM2/interface
W158F	1.1±0.2	1.41 ±0.05	0.1±0.2	1.27 ±0.04	1.0±0.3	Subdomain I TM2/interface
L207A	4.1±0.1	0.08 ±0.01	2.7±0.1	0.08 ±0.02	1.4±0.1	Subdomain II TM4/interface
Y210F	2.0±0.2	1.15 ±0.04	1.3±0.1	0.68 ±0.02	0.7 ±0.2	Subdomain II TM4/Interface

Table 2.3 Stability changes induced by single substitutions and activities of singly-substituted variants. Stabilities were measured by steric trapping in pH 7.0 Na_2HPO_4, 200 mM NaCl, 0.25 mM TCEP, and 5 mM DDM solution. To calculate the stability change for each substitution, $\Delta G^o_{U,95/172N\text{-}BtnPyr2}$=5.8±0.2 kcal/mol and $\Delta G^o_{U,172/267C\text{-}BtnPyr2}$=4.7±0.1 kcal/mol were used as wild-type stabilities. Energy values are in kcal/mol. Activity values are relative to wild-type GlpG. ΔG^o_U is defined as $\Delta G^o_{U,\ 95/172N\text{-}BtnPyr2} - \Delta G^o_{U,\ 172/267C\text{-}BtnPyr2}$. Errors denote propagated s. d. calculated from s. d. of individual ΔG^o_U values.

Table 2.3 (cont'd)

Mutation	$95/172_N$-BtnPyr2 $\Delta\Delta G^o_{U,95/172N\text{-}BtnPyr2}$ (WT-Mut)	Activity	$172/267_C$-BtnPyr2 $\Delta\Delta G^o_{U,172/267C\text{-}BtnPyr2}$ (WT-Mut)	Activity	$\Delta\Delta\Delta G^o_U$	Location
Localized interactions in Subdomain II						
L225A	–0.6±0.2	0.28 ±0.05	1.2±0.4	0.33 ±0.04	–1.8±0.4	Subdomain II TM5/Interface
Q226A	–0.2±0.2	1.42 ±0.06	0.8±0.4	1.81 ±0.09	–1.0±0.3	Subdomain II TM5
S181A	–0.5±0.2	1.30 ±0.04	0.6±0.2	1.49 ±0.06	–1.1±0.2	Subdomain I TM3/Interface
Localized interactions in Subdomain I at the TM4/TM6 interface						
A253V	1.7±0.2	0.04 ±0.01	0.8±0.2	0.06 ±0.01	0.9 ±0.3	Subdomain II TM6/Interface
G261A*	4.1±0.2	0.06 ±0.05	2.7±0.2	0.00 ±0.05	1.4 ±0.2	Subdomain II TM6
A265V*	2.4±0.2	0.40 ±0.05	1.3±0.2	0.22 ±0.05	1.1 ±0.3	Subdomain II TM6
D268A	2.5±0.2	0.17 ±0.02	1.3±0.1	0.44 ±0.02	1.2 ±0.2	Subdomain II TM6/Interface

[a]Subdomain in which a mutated residue is located.
[b]Secondary structural elements in which a mutated residue is located.
[c]If a mutated residue is making more than one side-chain contacts with residues in both subdomains, the residue is designated to be located at the subdomain interface.
* Over-propagated interactions

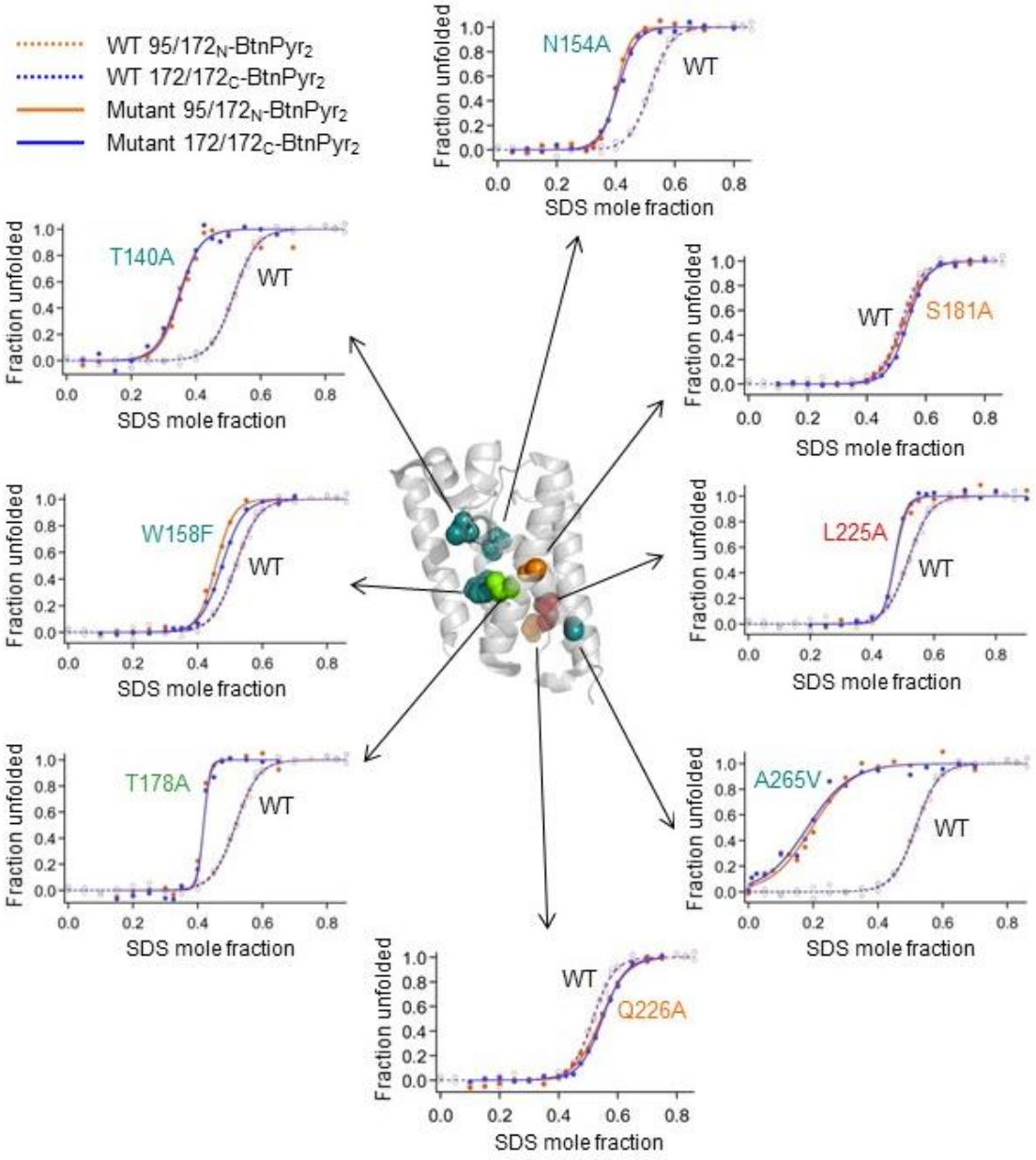


Figure 2.10 SDS denaturation of mutants. Mutations of 8 residues with various cooperativity profiles measured by steric trapping were chosen. $\Delta G°_{U,SDS}$ was obtained for both 95/172$_N$-BtnPyr$_2$ (blue filled circles) and 172/267$_C$-BtnPyr$_2$ (red filled circles) backgrounds. "WT" refers to 95/172$_N$-BtnPyr$_2$ (blue open circles) and 172/267$_C$-BtnPyr$_2$ (red open circles) with no other mutation. Mutated residues were color-coded according to their cooperativity profile (**Fig. 2.9(b)**). See Materials and Methods for detailed procedures.

2.5. Conclusion

Here we presented a new steric trapping strategy to investigate the thermodynamic stability of membrane proteins and the conformations of their unfolded states under native conditions by employing novel thiol-reactive biotin tags. Fluorescent BtnPyr allowed precise determination of the thermodynamic stability of GlpG through high-quality binding isotherms obtained by FRET. Paramagnetic BtnRG enabled characterization of the unfolded state based on the distance measurements using DEER. Because this combined strategy is not limited by either target-specific unfolding readout or specific lipid environments, it is applicable to other types of membrane proteins, including nonfunctional and misfolded variants whose folding is difficult to characterize under native conditions.

The unfolded state of proteins has gained substantial interest because it determines thermodynamic stability with the folded state, directs folding mechanisms and serves as a target for chaperoning and degradation (110). However, conformations of the unfolded states of membrane proteins are difficult to study under native conditions because of their transient nature, which prevents biophysical analysis. By combining DEER and steric trapping, we have elucidated a largely unraveled dynamic and heterogeneous conformational ensemble of the unfolded state of GlpG in non-denaturing micellar solution. The extent to which steric trapping could affect the protein conformation beyond the region containing the biotin pair is still an open question. Investigating the conformation of the unfolded state in a lipid bilayer, which provides a more defined hydrophobic environment than micelles, will be a crucial future task to elucidate the thermodynamics and mechanisms of membrane protein folding in cell membranes.

We identified subglobal unfolding of the C-terminal region, which encompasses the active site. This asymmetric stability profile of GlpG is analogous to the highly polarized folding

transition state possessing a compact folding nucleus in TM1–TM2 and largely unstructured TM3–TM6 (58). A single-molecule magnetic tweezers study has also identified TM3–TM6 or TM5–TM6 as a flexible region (60). Although we defined the region that underwent subglobal unfolding as the approximate C-terminal half TM4–TM6, it would be more reasonable to interpret that unfolding as an ensemble-averaged event that involved unfolding of a varied number of the C-terminal helices. Our work is unique in that we demonstrated partial unfolding even under non-denaturing conditions, which reflects intrinsic conformational malleability of the region that encompasses the active site. Although it is not clear whether subglobal unfolding is necessary during the catalytic cycle of GlpG, we speculate that this malleability is adequate for the conformational changes required for substrate interaction and catalytic mechanism. Further supporting this idea, disordering of the L5 loop, partial unfolding of TM5 and tilting of TM6 have been observed from crystal structures of GlpG in apo- and inhibitor-bound forms (92, 111, 112).

Our unprecedented cooperativity analysis suggests that the helical-bundle architecture of GlpG is maintained through a network of cooperative and localized interactions. Although the cooperativity network and its role in protein stability and function have been analyzed for water-soluble proteins (71, 74, 76, 106, 113), these have not been investigated for membrane proteins. Our experimentally determined cooperativity map indicates that the degree of cooperativity was largest for the buried residues near the center of the membrane and faded out toward the lipid- and water-contacting regions. This positional dependence of the cooperativity profile suggests that the complex environmental constraints for stabilizing membrane proteins—that is, protein-protein, protein-lipid and protein-water interactions—play an important role in the organization of the interaction network. Our general steric trapping strategy and steric trapping–based approaches will

serve as powerful tools for exploring the folding energy landscape of membrane proteins in native lipid bilayers.

2.6. Acknowledgements

This work was supported by startup funds from Michigan State University (H.H.), a Hunt for a Cure grant for basic cystic fibrosis research (H.H.), US National Cancer Institute R01CA149451 (X.H.), US National Eye Institute R01EY005216 (W.L.H.) and the Jules Stein Professor endowment (W.L.H.). The authors thank J. McCracken for general advice on electron paramagnetic resonance (EPR) and for helping with continuous-wave EPR measurements and C. Altenbach for the DEER analysis program. We thank E. Deans and C. Arthur for helping with protein preparations.

CHAPTER 3

Dissecting the Side Chain Interaction Energies of the Active Site Hydrogen Bond Network in the Rhomboid Protease GlpG

3.1. Summary

Intramembrane rhomboid proteases are of particular interest because of their ability to hydrolyze a substrate peptide bond buried in the membrane. The crystal structures of bacterial rhomboid GlpG have revealed the catalytic dyad Ser/His and one oxyanion Asn surrounded by the protein matrix and contacting a narrow water channel. Although multiple crystal structures have been solved, the catalytic mechanism of GlpG is not well understood. Because it is a serine protease, hydrogen bonding interactions among the active site residues are thought to play a critical role in the catalytic cycle. Here, we dissected the interaction energies among the active site residues His254, Ser201 and Asn154 of *E. coli* GlpG, which form a hydrogen bonding network. To achieve this, we employed the double-mutant cycle analysis combined with stability measurement using steric trapping. Surprisingly, in mild detergents, the active site residues were all weakly coupled with the interaction energies ($\Delta\Delta G_{\text{Inter}}$) of –1.4 kcal/mol between His254 and Ser201 and –0.2 kcal/mol between Ser201 and Asn154. Also, by analyzing the propagation of single mutations of the active site residues, we find that these residues are not only important for function but also for the folding cooperativity of GlpG. Our result suggests that the weak hydrogen bonds in the active site are sufficient to carry out the proteolytic function of rhomboid proteases, possibly providing the flexibility for bond breakage and reformation necessary for catalysis.

3.2. Introduction

Rhomboid proteases are a unique class of enzymes which mediate site-specific proteolysis of integral membrane proteins. They play a key role in a variety of biological processes by releasing membrane-bound effector proteins such as growth factors, transcription factors or enzymes, which leads to their activation (46, 47, 114–116). Mutational and inhibitory studies of

Rhomboid-1 in *Drosophila* suggested that rhomboids are serine proteases possessing a unique serine-histidine catalytic dyad, rather than the canonical triad found in other serine proteases (43). Crystallographic studies of the rhomboid GlpG from *E. coli* and *H. influenzae* have revealed the active site buried ~10 Å below the membrane surface and directly contacts a narrow water-filled cavity (**Figure 3.1**) (53, 117–119) Although the molecular details of how rhomboid proteases carry out proteolysis has not been confirmed, a mechanism has been proposed on the basis of inhibitory and crystallographic studies along with comparison to canonical serine proteases: The hydrogen bond between the catalytic histidine and serine (His254 and Ser201, numbering based on *E. coli* GlpG; $d_{His,N\varepsilon 2\cdots Ser,O\gamma} = 2.6$ Å) activates serine for a nucleophilic attack on the carbonyl carbon of the substrate peptide bond to create the first anionic tetrahedral intermediate (120). Emerging early as an important residue that may stabilize the intermediate was the conserved asparagine (Asn154) that presumably forms a weak hydrogen bond with the backbone amide group of Ser201 ($d_{AsnO\delta 1\cdots SerN} = 3.3$ Å) (55). Crystal structures with peptide inhibitors confirmed the tetrahedral intermediate is stabilized by a unique oxyanion triad by interactions with the conserved Asn154, His 150, and the backbone of the catalytic Ser201 (56). Next, the intermediate is collapsed, resulting in the formal cleavage of the peptide bond. His254 may then activate a water molecule to initiate the formation of the second tetrahedral intermediate and active site regeneration (119).

In canonical serine proteases, a catalytic triad (Ser-His-Asp) forms a tight hydrogen bond network, which coordinates a charge relay necessary for catalysis (121). In chymotrypsin, the unusually strong low-barrier hydrogen bond (LBHB) between His57 and Asp102 facilitates the nucleophilic attack of Ser195 on the substrate peptide bond and stabilizes the doubly protonated form of His57 (122). However, it has been argued whether strong hydrogen bonds between Ser195 and His57 as well as between His57 and Asp102 are necessary for catalysis, or weak hydrogen bonds are sufficient (123, 124). Therefore, measuring the hydrogen bond strengths in the active site network has been a focus of numerous studies (125, 126). Rhomboid proteases lack aspartate, which implies His254 alone should be sufficient to carry out the activation of Ser201 and the charge relay as a general base. It is an open question if rhomboids require strong hydrogen bonds to carry out peptide hydrolysis.

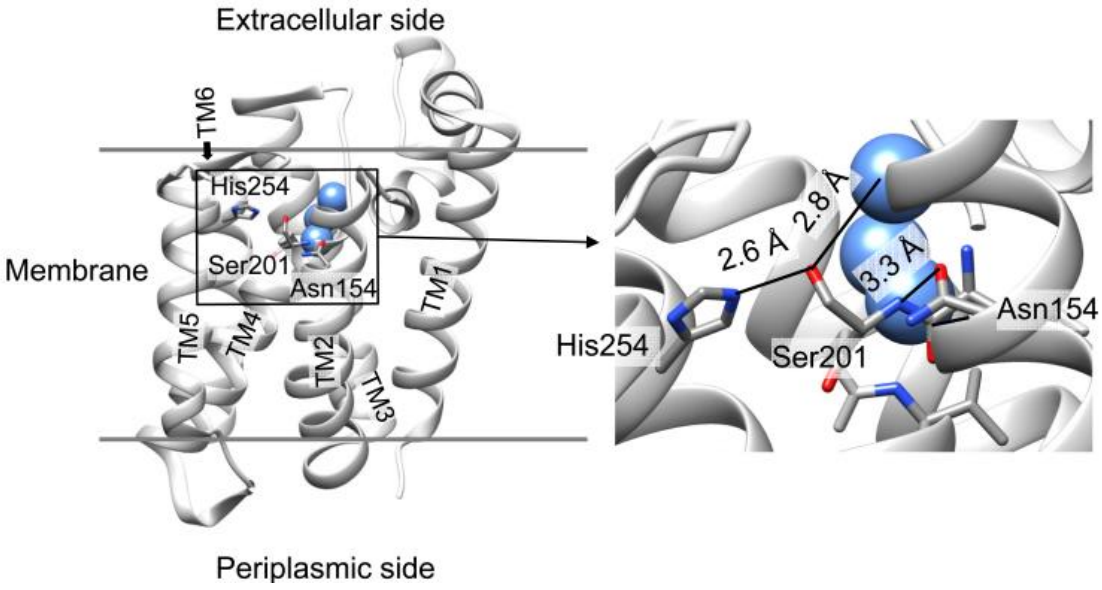


Figure 3.1. **Hydrogen bond network in the active site of the rhomboid protease GlpG of *E. coli*.** (*Left*) Structure of GlpG (PDB: 3B45 (51)) showing the location of the active site and the crystallographically identified water cavity. Ser201 and His254 form a catalytic dyad. Asn154, which forms a hydrogen bond with the backbone NH group of serine, is a part of the oxyanion hole (119). (*Right*) The hydrogen bond network in the active site.

3.3. Materials and Methods

3.3.1. Expression, purification and labelling of GlpG

The TM domain of GlpG (residue 87 –276) encoded by pET15b vector was expressed in *E. coli* BL21(DE3) RP strain with an N-terminal His_6-tag, for purification, as described in Chapter 2. Site-directed mutagenesis was carried out using the QuikChange Site-Directed Mutagenesis Kit (Agilent). GlpG was labeled with the thiol-reactive biotin derivative possessing pyrene fluorophore, BtnPyr-IA, and purified as described in Chapter 2. Typically, the labeling efficiency of BtnPyr-IA ranged from 1.5–2.2 as estimated from the concentration of BtnPyr determined by pyrene absorbance (ε_{346nm} = 43,000 $M^{-1}cm^{-1}$) and the concentration of GlpG determined by DC protein assay (Bio-Rad).

3.3.2. Preparation of monovalent streptavidin

Wild type monovalent streptavidin (mSA-WT), and its variants mSA-S27A and mSA-S45A, in which mutations were made on the active subunit of tetrameric mSA, was prepared as described previously (127). Each variant additionally contained a single-cysteine mutation S83C, to which the thiol-reactive dabcyl quencher (DABCYL Plus™ C2 maleimide, Anaspec) was labeled for binding assay between mSA and GlpG-$BtnPyr_2$ using FRET.

3.3.3. Expression and purification of GlpG substrate SN-LYTM2

For functional assay of GlpG, we used the second TM domain of the lactose permease of *E. coli* fused to staphylococcal nuclease (SN-LYTM2) as a model substrate. The protein was expressed, purified and labeled using the protocol described in Chapter 2.

3.3.4. Construction of binding isotherm to determine thermodynamic stability of GlpG using steric trapping

Thermodynamic stability of GlpG in DDM micelles (20 mM) was determined by measuring the attenuated second binding of mSA labeled with dabcyl quencher (mSA_{DAB}) to GlpG doubly labeled with BtnPyr ($95/172_N$-$BtnPyr_2$ or $172/267_C$-$BtnPyr_2$) at room temperature. mSA_{DAB} binding was monitored by quenching of pyrene fluorescence from BtnPyr labels by (FRET). 1 μM of $95/172_N$-$BtnPyr_2$ or $172/267_C$-$BtnPyr_2$ was titrated with mSA_{DAB} variant possessing a reduced biotin binding affinity, mSA_{DAB}-S45A ($K_{d,biotin}$ = 9.0 ± 4.3 nM) or mSA_{DAB}-S27A ($K_{d,biotin}$ = 1.4 ± 0.9 nM) in 20 mM DDM, 0.25 mM TCEP, 20 mM Na_2HPO_4 and 200 mM NaCl (pH 7.5). The use of mSA mutants was necessary to achieve the reversibility of the second mSA binding to obtain the thermodynamic stability. The titrated samples were transferred to a 96-well UV-compatible microplate, sealed with a polyolefin tape, and incubated for 5 days (for $95/172_N$-$BtnPyr_2$) or 2 days (for $172/267_C$-$BtnPyr_2$) at room temperature. Quenching of pyrene-monomer fluorescence at 390 nm was monitored with an excitation wavelength of 345 nm on SpectraMax M5e plate reader. Data were averaged from three readings. Non-specific FRET between pyrene and dabcyl was negligible.

3.3.5. Fitting of binding isotherm to determine thermodynamic stability of GlpG

Fitting equation to obtain thermodynamic stability of GlpG using steric trapping was described in Chapter 2. After obtaining the fitted K_U, the thermodynamic stability was calculated using the equation $\Delta G^o_{U,ST} = -RT\ln K_U$.

3.3.6. Double mutant cycle analysis

To measure the pairwise interaction energies of the active site residues, Ala-scanning double-mutant cycle analysis was employed. A double-mutant cycle involves wild type protein (WT), two single mutants and the corresponding double mutant. If the change in thermodynamic stability (ΔG^{o}_{U}) upon the double mutation ($\Delta\Delta G^{o}_{U,XY\text{-}AY}+\Delta\Delta G^{o}_{U,AY\text{-}AA}$) differs from the sum of the changes due to the single mutations ($\Delta\Delta G^{o}_{U,XY\text{-}XA} + \Delta\Delta G^{o}_{U,XY\text{-}AY}$), the two residues in WT are coupled and the magnitude of the difference (interaction energy: $\Delta\Delta G^{o}_{Inter}$) is related to the strength of interaction between them.

$$\Delta\Delta G^{o}_{Inter} = -\ [\Delta\Delta G^{o}_{U,XY\text{-}XA}+\Delta\Delta G^{o}_{U,XY\text{-}AY}] - [\Delta\Delta G^{o}_{U,XY\text{-}AY}+\Delta\Delta G^{o}_{U,AY\text{-}AA}]$$

$$= -\ [\Delta\Delta G^{o}_{U,XY\text{-}XA}+\Delta\Delta G^{o}_{U,XY\text{-}AY}] - [\Delta\Delta G^{o}_{U,XY\text{-}XA}+\Delta\Delta G^{o}_{U,XA\text{-}AA}] \qquad \textbf{(9)}$$

, where X and Y denotes wild type residues of interest and A designates Ala.

3.3.7. Cooperativity profiling

This is the method that we have developed to identify cooperative and localized side chain interactions that contribute to the protein stability. To apply this method to GlpG, we first made a single mutation (typically to alanine) to perturb a specific side-chain interaction in the background of double biotin variants of GlpG, $95/172_{N}$-$BtnPyr_2$ and $172/267_{C}$-$BtnPyr_2$. Next, using steric trapping, the stability changes induced by the mutation are measured with two different biotin pairs that are located in the N- and C-subdomain, respectively. The differential effect of the same mutation on the stability of each subdomain ($\Delta\Delta\Delta G^{o}_{U}$) is quantified using **equation (8)** described in Chapter 2.

3.4. Results and Discussion

Here, we determined the strengths of the hydrogen bonds between the active site residues of *E. coli* GlpG (His254, Ser201 and Asn154) using double-mutant cycle analysis combined with measurement of mutation-induced stability changes in mild dodecylmaltoside (DDM) micelles.

3.4.1. Mild destabilization by single alanine mutations in active site residues

To calculate the interaction energies ($\Delta\Delta G_{Inter}$) using double-mutant cycle analysis, we measured the thermodynamic stabilities (ΔG^{o}_{U}'s) of WT, single- and double-Ala mutants of GlpG. To measure ΔG^{o}_{U}, we employed steric trapping, which couples transient unfolding of a doubly-biotinylated protein to double binding of bulky monovalent streptavidin (mSA, 52 kD) (**Fig. 3.2(a)**) (83). Compared to conventional stability measurements using chemical denaturants, this method is advantageous because protein stability can be directly measured under native solvent and lipid conditions. Previously, we have identified optimal sites of thiol-specific biotinylation for steric trapping, P95C/G172C ($95/172_{N}$: "N" indicates the N-terminal subdomain where a biotin pair is located) (**Figure 3.2(b)**) and shown that the unfolded state trapped with this biotin pair is globally denatured (89). To measure ΔG^{o}_{U} of GlpG, we obtain a binding isotherm between doubly-biotinylated GlpG and mSA by employing a thiol-reactive biotin derivative with a pyrene fluorophore (BtnPyr) and mSA labeled with dabcyl quencher (mSA_{DAB}) (89). When a mSA_{DAB} variant with a reduced biotin binding affinity is used, the binding isotherm monitored by quenching of pyrene fluorescence displays two-phase mSA binding, in which the second attenuated binding is coupled to GlpG unfolding (89). ΔG^{o}_{U} of GlpG is determined by fitting the second binding phase to Eq. (3) (Chapter 2).

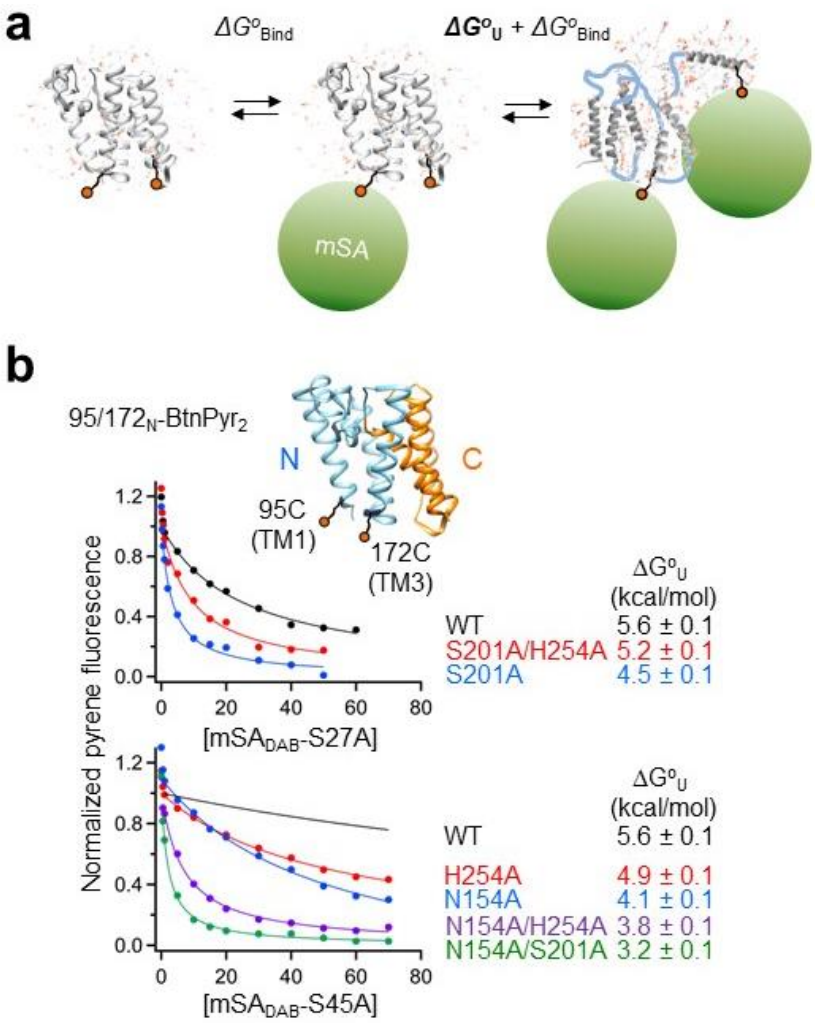


Figure 3.2 Measuring thermodynamic stability of GlpG using steric trapping. (a) Principle of steric trapping. When biotin tags are conjugated to two specific residues that are spatially close in the folded state but distant in the amino acid sequence, the first monovalent streptavidin (mSA) binds either biotin label with intrinsic binding affinity (ΔG^{o}_{Bind}). Because of steric hindrance, the second mSA binds only when native tertiary contacts are unraveled by transient unfolding. Hence, binding of the second mSA is attenuated depending on the stability of the target protein ($\Delta G^{o}_{Bind} + \Delta G^{o}_{U}$). By adjusting the biotin affinity of mSA by mutation, unfolding and binding reactions can be reversibly controlled, and ΔG^{o}_{U} of the target protein can be obtained by monitoring binding of the second mSA or protein denaturation. **(b)** Binding isotherms between double biotin variant of GlpG ($95/172_{C}$-$BtnPyr_2$) and mSA_{DAB} variants with a reduced biotin binding affinity. The backbone in cyan: N-subdomain (residues 87–198); the backbone in orange: C-subdomain (residues 199–276). Binding was monitored by quenching of pyrene fluorescence. Fluorescence intensity was normalized to the total intensity change of the second mSA binding phase. Errors denote standard deviations from fitting.

The binding isotherms using weaker biotin-binding mSA variants (mSA_{DAB}-S27A, $K_{d,biotin}$ = 1.4 × 10^{-9} M or mSA_{DAB}-S45A, $K_{d,biotin}$ = 9.0 × 10^{-9} M) (89) are shown in **Figure 3.2(b)**. Among the active site residues, Ser201, His254 and Asn154 are largely buried with the fractions of accessible side-chain surface area of 0.04, 0.015 and 0.00, respectively. The single Ala mutations at these residues abolished GlpG activity (**Fig. 3.3**). $\Delta G°_U$ of the double-biotin variant without additional Ala mutation (WT) was 5.6 ± 0.1 kcal/mol. Single mutations S201A and H254A were mildly destabilizing with $\Delta\Delta G°_U$'s of 1.1 ± 0.1 and 0.7 ± 0.1 kcal/mol, respectively. Ala mutation at completely buried Asn154 (N154A) induced larger destabilization with $\Delta\Delta G°_U$ = 1.5 ± 0.1 kcal/mol. Therefore, mutations in the active site induced mild destabilization relative to other previously characterized mutations in the buried region, for which $\Delta\Delta G°_U$ can be as large as ~4 kcal/mol (89). The mild destabilization by the single active site mutations obtained by steric trapping agrees reasonably well with previous studies using SDS-induced or thermal denaturation (58, 128).

3.4.2. Weak interaction energy between catalytic dyad of GlpG

Next, the stabilities of double-Ala mutants were measured. Interestingly, the double mutation on the catalytic dyad (S201A/H254A) yielded a smaller decrease in the stability ($\Delta\Delta G°_U$ = 0.4 ± 0.2 kcal/mol) than individual single mutations. The double mutations of N154A/H254A and N154A/S201A induced larger destabilization ($\Delta\Delta G°_U$ = 1.8–2.4 kcal/mol) than individual single mutations, implying an additive effect of the single mutations. Next, we determined the interaction strengths between the active site residue pairs using double-mutant cycles. If the change in $\Delta G°_U$ due to the double mutation ($\Delta\Delta G°_{U,XY\text{-}XA} + \Delta\Delta G°_{U,XA\text{-}AA}$; X and Y denote WT residues) differs from the sum of the changes due to the single mutations ($\Delta\Delta G°_{U,XY\text{-}XA} + \Delta\Delta G°_{U,XY\text{-}AY}$), the

two residues are coupled and the magnitude of the difference ($\Delta\Delta G_{Inter}$) is related to the strength of interaction between them (**Figure 3.4**) (129). From this analysis, Ser201 and His254, which form the catalytic dyad and are engaged in a close hydrogen bond, favorably interacted ($\Delta\Delta G_{Inter}$ = –1.4 ± 0.2 kcal/mol), whereas the interaction between Ser201 and Asn154, which form a more distant side chain-backbone hydrogen bond, was not significant ($\Delta\Delta G_{Inter}$ = –0.2 ± 0.2 kcal/mol).

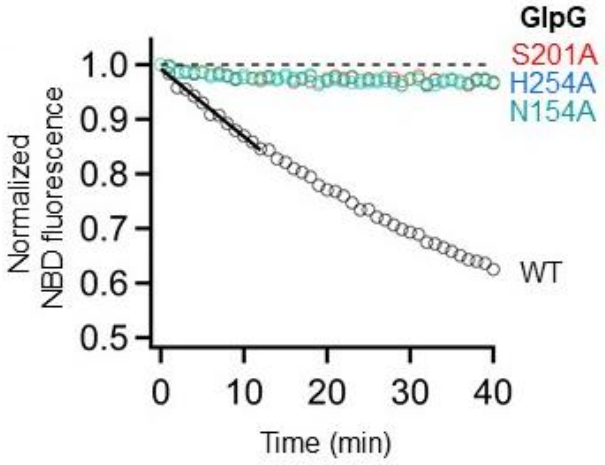


Figure 3.3 Activity assay of GlpG variants. Activity of GlpG wild type and single active-site mutants (His254A, S201A and N154A). The initial slope of the changes in NBD fluorescence versus time indicates the proteolytic activity of GlpG. This result shows that all active-site mutants were essentially inactive.

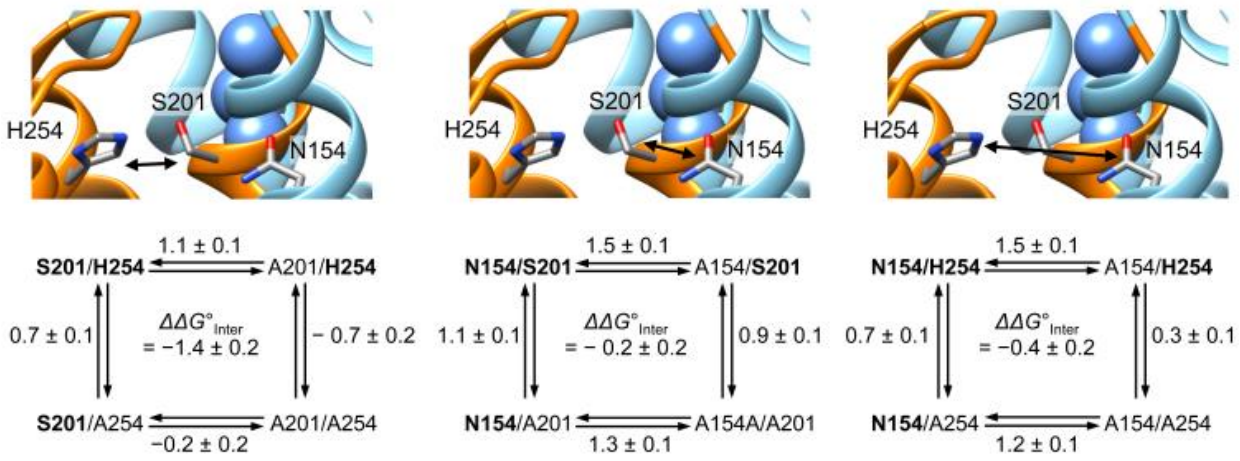


Figure 3.4 Double-mutant cycle analysis to measure the side chain interaction energies in the active site of GlpG. All values are in kcal/mol. Errors denote s. d. from fitting.

$\Delta\Delta G_{Inter}$ between His254 and Asn154, which are apparently not engaged in any interaction with each other by the crystal structure, was not significant either (–0.4 ± 0.2 kcal/mol), confirming the validity of our analysis. Hydrogen bonds can be categorized by the strength of their interaction: weak or conventional (2–12 kcal/mol), strong or low-barrier (12–24 kcal/mol), and very strong or single-well (>24 kcal/mol (122). Although favorable, the measured hydrogen bond strength of the His-Ser catalytic dyad of GlpG ($\Delta G_{Inter} \approx -1.4$ kcal/mol) is regarded as "weak". This interaction is substantially weaker than the His57-Asp102 interaction in the active site of chymotrypsin, which has been suggested to form a LBHB, as well as weaker than the Ser195-His57 interaction in the same protein classified as "moderately strong' at a low pH (122, 130). The weak hydrogen bond between His254 and Ser201 determined for GlpG must be sufficient to activate Ser201 for the nucleophilic attack on the peptide bond. In addition, the negligible interaction between Asn154 and Ser201 implies that Asn154 may easily gain the flexibility to be adapted to a conformation that can stabilize the oxyanion intermediate at the initial stage of the proteolysis mechanism.

3.4.3. Active site residues are involved in cooperative interactions

Finally, we analyzed the contribution of each active site residue to the folding cooperativity of GlpG using the steric trapping-based cooperativity profiling (**Figure 3.5**) (89). This method is based on the principle that steric trapping captures the transient opening of the tertiary interactions of the specific region to which a biotin pair is conjugated. Thus, we can measure the local stability of a protein and how the local sequence perturbation caused by a mutation is propagated throughout the protein structure. Briefly, the effect of a specific mutation on the stability ($\Delta\Delta G^{o}_{U}$) is measured with two biotin pairs located in different regions. If the difference in the measured stability changes ($\Delta\Delta\Delta G^{o}_{U}$) is smaller than thermal fluctuation energy (*i.e.*, $|\Delta\Delta\Delta G^{o}_{U}| \leq RT = 0.6$ kcal/mol), it indicates that the side chain perturbation by the mutation is propagated evenly throughout the protein and the mutated side chain is engaged in "cooperative" interactions. If the mutation preferentially destabilizes the subdomain that includes the mutation site ($|\Delta\Delta\Delta G^{o}_{U}| > RT$), the mutated side chain is engaged in "localized" interactions. If the mutation preferentially destabilizes the subdomain that does not include the mutation site with $|\Delta\Delta\Delta G^{o}_{U}| > RT$, the mutated side chain is engaged in "over-propagated" interactions.

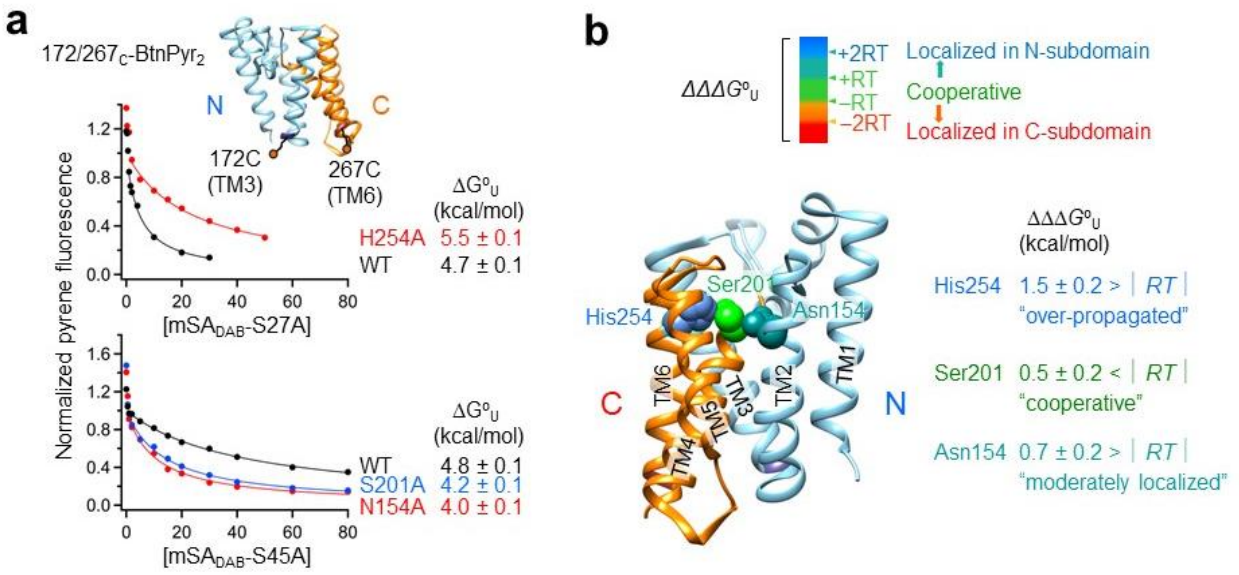


Figure 3.5 Cooperativity profiling of the active site residues of GlpG. (**a**) Binding isotherms between double biotin variants of GlpG ($172/267_C$-$BtnPyr_2$) and mSA_{DAB} variants to measure ΔG^o_U of the C-subdomain. (**b**) The cooperativity profiles of the three active site residues.

To apply this method to the active site residues of GlpG, the stability change upon each single alanine mutation was measured at the biotin pairs $95/172_N$-$BtnPyr_2$ (**Fig. 3.2(b)**) and $172/267_C$-$BtnPyr_2$ ("C" indicates the C-terminal subdomain in which the biotin pair is located) (**Figure 3.5(a)**). We have shown that these double biotin variants have the same global stability (89). S201A mutation at the subdomain interface similarly destabilized N- and C-subdomains ($\Delta\Delta G^o_U$ = 1.1 ± 0.1 kcal/mol and 0.6 ± 0.2 kcal/mol, respectively), yielding $|\Delta\Delta\Delta G^o_U|$ = 0.5 ± 0.2 $<RT$. Thus, Ser201 was classified as cooperatively engaged (**Figure 3.5(b)**). Interestingly, H254A mutation in C-subdomain induced larger destabilization of N-subdomain ($\Delta\Delta G^o_U$ = 0.7 ± 0.1 kcal/mol) than C-subdomain containing the mutated site ($\Delta\Delta G^o_U$ = –0.8 ± 0.2 kcal/mol), yielding $|\Delta\Delta\Delta G^o_U|$ = 1.5 ± 0.2 kcal/mol ($>2RT$). Thus, we assign His254 interactions as highly over-propagated. We reason that the stabilization of C-subdomain by H254A is due to the local

structural reorganization induced by the large changes in the side-chain volume and polarity. This reorganization appears to be compensated by destabilization of N-subdomain. Mutation at N154A preferentially destabilized N-subdomain, where the mutation resides. The resulting $|\Delta\Delta\Delta G^{o}_{U}|$ of 0.7 ± 0.2 kcal/mol was slightly larger than *RT*. Thus, we assigned Asn154 interactions are moderately localized. Our analysis indicates that the absolutely conserved catalytic dyad Ser201-His254 is not only critical for function but also highly communicative with their environment to maintain the folding cooperativity of GlpG.

3.5. Conclusion

Although strong hydrogen bonds have been implicated in catalysis by a variety of enzymes, and how the bilayer environment affects the hydrogen bond interaction energies is unclear, we revealed that the hydrogen bond network in the active site of GlpG is organized by weak interactions for its assembly and function. Notably, the strengths of the side chain hydrogen bonds obtained in this study fall into the range of those measured in various helical membrane proteins (0–2.0 kcal/mol) (131). Whereas previous studies have been mainly concerned with "structural" hydrogen bonds which are not directly involved in function, we present an example of measuring the strengths of "functional" hydrogen bonds conserved in the rhomboid protease family (109). Although maintaining the hydrogen bond network is necessary for function, the interaction remains mild, possibly to allow for bond breakage and reformation which are necessary to carry out catalysis. Our result provides important physical insights into the initial step of the proteolysis mechanism by rhomboids. We have quantitatively confirmed that the catalytic dyad Ser201 and His254 do form a favorable hydrogen bond. This bond is necessary for the activation of Ser201 for nucleophilic attack. Although favorable, the hydrogen bond interaction energy between this

pair can be categorized as “weak”. This may facilitate the bond breakage and reformation throughout the catalytic cycle. Also, the negligible hydrogen bond strength between Asn154 and the backbone NH of Ser201 may facilitate the stabilization of the first tetrahedral intermediate by Asn154. Granting, this study was carried out in detergent micelles, and these hydrogen bond strengths may be increased the 2-dimensional environment of the lipid bilayer.

3.6. Acknowledgements

We thank Ruiqiong Guo for generously providing the reagent BtnPyr-IA. Thank you to undergraduate J. Cho for supporting protein expression. This work is supported by NIH grant R01GM118685 and the Hunt for a Cure foundation to H.H.

CHAPTER 4

The Denatured State Ensemble of a Helical Membrane Protein is Largely Expanded Under Native Lipid Bilayer and Solvent Conditions

Kristen Gaffney[1,4], Ruiqiong Guo[2,4], Michael Bridges[3], Miyeon Kim[2], Wayne Hubbell[3], and Heedeok Hong[1,2]

[1]Department of Biochemistry & Molecular Biology and [2]Department of Chemistry, Michigan State University, East Lansing, MI, 48824, USA

[3]Jules Stein Eye Institute and Department of Chemistry and Biochemistry, University of California, Los Angeles, CA, 90095, USA

[4]These authors contributed equally to this work.

This work was carried out in close collaboration with Ruiqiong Guo and I, resulting in us being listed as co-first authors. Dr. Bridges carried out DEER-EPR measurements in the lab of Wayne Hubbell. Dr. Kim helped to prepare EPR samples. Dr. Guo carried out the vesicle incorporation and bicelle and vesicle mixing assays (Fig. 4.2) . I designed and optimized the liposome fusion, fluorescence activity assay by PEG (Fig. A1(c & e)). I carried out the proteinase digestion in various lipid environments (Fig. 4.4). Guo and I together carried out the transfer optimization and activity analysis in bicelles and liposomes (Fig. 4.3). Guo and I together prepared the unfolded states for DEER-EPR analysis in Fig. 4.5.

4.1. Summary

Membrane proteins fold under the physical constraints of the quasi-two-dimensional lipid bilayer with defined hydrophobic thickness. While studies of membrane proteins are primarily concerned with the native states, their denatured states are not well understood. Here we investigated the conformational features of the denatured state ensemble (DSE) of a stable helical-bundle membrane protein GlpG of *E. coli* under native bilayer and solvent conditions. The DSE was first prepared in non-denaturing micellar solution using steric trapping, which couples spontaneous unfolding of a doubly biotin-tagged protein to competitive binding of bulky monovalent streptavidin. The DSE was then transferred to *E. coli* lipid vesicles which provided the native bilayer environment. Our novel paramagnetic biotin derivative conjugated to GlpG enabled measurement of the interspin distances (d_{Inter}) between two specific biotinylated sites in the sterically trapped DSE by double electron-electron resonance spectroscopy. In bilayers, the average d_{Inter} increased from ~25 Å in the native state to ~55 Å in the DSE and the distribution was substantially broadened relative to that of the native state. Despite the physical constraints, the lipid bilayer did not impose compaction of the DSE in bilayers relative to micelles with loose topological constraints. Also, the DSE was highly susceptible to proteolysis by proteinase K, indicating unfolding of interhelical loops and protection of transmembrane helices. Our distance data agree well with the "Θ" solvent scaling behavior based on the polymer model, displaying that the membrane bilayer provides at the minimum, a good environment for the denatured state of a membrane protein. This suggests a delicate balance between protein-protein and protein-lipid interactions in maintaining the denatured state in the bilayer. This work illustrates that the quasi-two-dimensional environment provided by the lipid bilayer is indeed a good solvent for the denatured state of membrane proteins, and that the DSE can be widely expanded and dynamic

while embedded in the membrane. This provides valuable information about the early folding mechanisms and the conformation of the denatured state of membrane proteins in their native environment.

4.2. Introduction

The denatured states of proteins are as important as the native states because they determine the thermodynamic stability of a protein with its native state, direct early folding mechanisms, and serve as targets for chaperoning, degradation and membrane translocation (110, 132–134). Therefore, understanding the conformational nature of the denatured states has been one of the key subjects in protein folding studies over the past 50 years (135, 136). For the denatured states of globular proteins or intrinsically disordered proteins, a consensus is being made that they are an ensemble of fast-interconverting conformations largely expanded in water (137–139). In contrast, the denatured state is poorly understood for membrane proteins which account for 20–30% of all genes in most genomes (3). So far, the denatured states of helical membrane proteins have been mainly studied using chaotropic agents including anionic detergent SDS and polar organic solutes, urea and GdnHCl in micellar solution (80, 98, 99, 108, 140, 141). These studies indicate that the denatured states are heterogeneous with disrupted native interactions and nearly intact transmembrane (TM) helical segments. It has also been shown that the degree of expansion upon denaturation depends on the choice of denaturant as well as its concentration (108, 141).

The folding of helical membrane proteins can be divided into two thermodynamically distinct stages (24): In stage I, individual hydrophobic segments in a polypeptide chain insert into the bilayer to form stable TM helices, and in stage II, inserted TM helices fold into a compact native structure through side-to-side interactions. Thus, based on the findings, the denatured states

of helical membrane proteins could be described as an ensemble of conformations formed by the TM helices and probably unfolded interhelical loops before folding into the native state (i.e., the denatured state ensemble, DSE). Nonetheless, the current approaches using chaotropic agents in micellar solution cannot recapitulate the native lipid-protein and water-protein interactions with which the DSE's are associated with the cell membranes. Therefore, to understand the folding of membrane proteins, it is necessary to define the conformational features of the DSE's in the native lipid environments.

This chapter will describe the successful reconstitution of the on-pathway DSE of a stable six-helical bundle membrane protein GlpG of *E. coli* in the native lipid bilayer and solvent environments, and defined its conformation and compactness using double electron-electron resonance spectroscopy (DEER) and limited proteolysis. By applying the solvent-scaling models from polymer theory, we will evaluate that the degree of expansion and determine is correlation with the "Θ-solvent model", examining if the DSE's of helical membrane proteins can reasonably be accommodated by the lipid bilayers with balanced protein-protein and protein-lipid interactions.

4.3. Materials and Methods

4.3.1. GlpG expression, purification and labeling.

The TM domain of GlpG (residue 87 –276) encoded by pET15b vector was expressed in *E. coli* BL21(DE3) RP strain with an N-terminal His_6-tag, for purification, as described in Chapter 2. Site-directed mutagenesis was carried out using the QuikChange Site-Directed Mutagenesis Kit (Agilent). GlpG was labeled with the thiol-reactive biotin derivative possessing pyrene fluorophore, BtnPyr-IA, and purified as described in Chapter 2, with the following modification: for labeling of 95C/172C, SDS was added to a final SDS mole fraction ([SDS]/ ([SDS] + [DDM]))

of 0.8 to facilitate labeling and incubated for 30 min. For 172C/267C, the labeling reaction was proceeded without addition of SDS. SDS was removed after labeling by extensive washing with DDM. Typically, the labeling efficiency of BtnPyr-IA ranged from 1.5–2.2 as estimated from the concentration of BtnPyr determined by pyrene absorbance (ε_{346nm} = 43,000 $M^{-1}cm^{-1}$) and the concentration of GlpG determined by DC protein assay (Bio-Rad).

4.3.2. Bicelle preparation

15% (w/v) stock of DMPC (1,2-dimyristoyl-sn-glycero-3-phosphocholine)/DMPG (1,2-dimyristoyl-sn-glycero-3-phospho-(1'-rac-glycerol))/CHAPS (3-[(3-Cholamidopropyl) dimethylammonio]-1-propanesulfonate) (lipid-to-detergent molar ratio, q = 2.8) bicelles were prepared by hydrating DMPC/DMPG (molar ratio = 3:1) lipids with water. 20% (w/v) CHAPS was added to reach the desired q value. Bicelle samples were homogenized through three cycles of freeze-thaw using liquid nitrogen and a water bath at 42 ºC. Bicelle stocks were kept at -20 ºC prior to use.

4.3.3. Transfer of native and denatured GlpG to bicelles

GlpG doubly labeled with BtnPyr or BtnRG in DDM was incubated with a 5 times molar excess of mSA at room temperature until maximum denaturation was reached. The extent of denaturation was monitored using GlpG activity, monitored every 24 hours. Maximum denaturation was reached within 48 hours for $95/172_N$-$BtnRG_2$ and 24 hours for $172/267_C$-$BtnRG_2$. Native and denatured GlpG were directly injected into preformed bicelles to the final concentrations of 5 μM GlpG, 25 μM mSA, and 3% (w/v) DMPC/DMPG/CHAPS bicelles in 20 mM Na_2HPO_4, 40 mM NaCl, pH 7.5 and incubated overnight at room temperature.

4.3.4. Measuring the incorporation of native and denatured GlpG into bicelles

7.5% bicelle containing dabcyl-DOPE (quencher-labeled lipid, Avanti Polar Lipids) at 1% lipid-to-lipid molar ratio was prepared in 20 mM HEPES buffer (pH 7.5). GlpG variants were doubly labeled with BtnPyr-IA as described above. The incorporation of native or denatured GlpG into the bicelles was measured using quenching of pyrene fluorescence from GlpG by dabcyl label localized in the lipid region in bicelles.

As a negative control (i.e., no incorporation to bicelles), highly water-soluble mSA-WT labeled with pyrene was used. mSA was labeled using the following procedures: 1 mL of 30 μM mSA-WT in 20 mM HEPES buffer (pH 8.0) was incubated with a 10 times molar excess of amine-reactive pyrene (1-pyrenebutyric acid N-hydroxysuccinimide ester) solubilized in DMSO for 2 h at room temperature. The reaction was quenched with 0.1 mL 1.5 M hydroxylamine hydrochloride, which had been freshly dissolved in water at pH 8.5 (adjusted with sodium hydroxide) for 30 min. Excess free labels were removed on a desalting column equilibrated with 20 mM HEPES buffer (pH 7.5). The labeling efficiency of pyrene was ~3 labels per tetramer as determined by comparing the concentration of pyrene measured by UV-Vis absorbance (ε_{Molar} = 43,000 M^{-1} cm^{-1}) to the concentration of mSA measured by DC protein assay (Bio-Rad). To the final pyrene-labeled mSA-WT stock, DDM was added to a final concentration of 5 mM to match the DDM concentration of the experimental GlpG samples (see below).

To be used as a positive control (i.e., full incorporation in bicelles), GlpG labeled with pyrene was first reconstituted in DMPC/DMPG liposomes using the following procedures: Mixed dried lipid ([DMPC]: [DMPG] = 3:1) was dispersed in 20 mM HEPES buffer (pH 7.5) to a final lipid concentration of 4% (w/v). The lipid suspension was homogenized by three cycles of freeze-thaw and then extruded through 0.2 μm pore-size polycarbonate membrane (Whatman). DDM was

added to the liposome suspension to a final concentration of 40 mM and incubated for 30 min. Then, GlpG labeled with BtnPyr from stock was added to a final concentration of 10 μM. The lipid-protein-detergent mixture was incubated for 30 min. Three portions of Bio-Beads (Bio-Rad) were added (20 mg/mL for each) stepwise to remove detergent DDM. In each step, the mixture was gently stirred for 2 hr. In the first removal step, the samples were incubated at 4 ºC for 2 h and then moved to room temperature in the subsequent removal steps. The resulting proteoliposomes were extruded again using 0.2 μm pore size membrane. The total phospholipid concentration was determined using an organic phosphate assay. Based on the measured total lipid concentration, desired amount of CHAPS was added to form bicelles with q = 2.8. Then, the 7.5% bicelle stock containing dabcyl-labeled lipid (see above) was added to the final bicelle concentration of 3%, during which the bicelle constituents (labeled and unlabeled lipids and GlpG) are homogeneously mixed.

In the samples for negative and positive controls, the final pyrene and dabcyl concentrations were matched to those of experimental samples (see below). To be used as experiment, native or sterically trapped denatured GlpG in DDM was directly injected into preformed 7.5% bicelles containing DOPE-dabcyl at the final concentrations of 3% bicelles and the final pyrene concentration is around 5 μM as measured by UV-Vis absorbance at 346 nm (ε_{Molar} = 43,000 $M^{-1}cm^{-1}$). After mixing, the samples were equilibrated overnight at room temperature. Pyrene fluorescence of these samples was measured in 96-well plate using SpectraMax M5e plate reader (Molecular Devices) with excitation and emission wavelengths of 345 nm and 390 nm, respectively. The ratio of the pyrene fluorescence intensities for the experimental and positive control samples to the intensity for the negative control sample was used as a measure of GlpG incorporation in the bicelles.

4.3.5. Preparation of empty *E. coli* liposomes

Dried *E. coli* lipid (Avanti Polar Lipids) film was hydrated with 20 mM Na_2HPO_4 (pH 7.5), 40 mM NaCl buffer to a final lipid concentration of 10 mM. The lipid suspension was homogenized by three cycles of freeze-thaw and then extruded through 0.2 μm pore size polycarbonate membrane (Whatman).

4.3.6. Transfer of native and denatured GlpG into *E. coli* liposomes

25 μM GlpG variant $172/267_C$–$BtnPyr_2$ or $172/267_C$–$BtnRG_2$ in DDM was incubated with a 5 times molar excess of mSA-WT at room temperature overnight to obtain the sterically trapped denatured state. DDM was added to 10 mM empty *E. coli* liposomes to a final concentration of 10 mM and incubated for 30 min. Native or denatured GlpG was added to a final concentration of 5 μM. The lipid-protein-detergent mixture was incubated for 30 min. For detergent removal, three portions of Bio-Beads (Bio-Rad) were added (20 mg/mL for each) stepwise. In each step, the mixture was gently stirred for 1–2 h at room temperature. In the first removal step, the samples were incubated at 4 ºC for 2 h and then moved to room temperature in the subsequent removal steps. The resulting proteoliposomes were extruded using 0.2 μm pore size membrane.

Because of the high kinetic unfolding barrier, GlpG variant $95/172_N$–$BtnPyr_2$ or $172/267_C$–$BtnRG_2$ was first denatured with SDS. SDS was added to GlpG stock to the final SDS mole fraction of 0.9 and the final GlpG concentration of 25 μM and incubated at room temperature overnight. Then 5 times molar excess WT-mSA was added and incubated for 1 h to trap the denatured state. For native GlpG, no mSA was added. Then, DDM was added to lower the SDS mole fraction to 0.1 to bring denatured GlpG back to the native condition and incubated for 1 h. Then GlpG samples

in detergent was mixed with empty liposome. The following steps were the same as those for 172/267C variants described above.

4.3.7. Flotation assay of liposome samples

Pyrene labeled GlpG ($95/172_N$–$BtnPyr_2$ or $172/267_C$–$BtnPyr_2$) was reconstituted in *E. coli* liposome containing rhodamine-labeled lipid (DPPE-Rho, 1% lipid-to-lipid molar ratio, Avanti Polar Lipids). The proteoliposomes containing GlpG (50 μL) were mixed well with 60% (w/v) sucrose in 20 mM HEPES (pH 7.5) (50 μL). The mixture was loaded at the bottom of the centrifuge tube (Beckman Coulter polycarbonate tubes, 1 mL capacity). The sample was flash-frozen with liquid nitrogen after each step of adding 100 μL of sucrose solution at a lower concentration (20%, 10%, 5% and 2.5%). The tube was centrifuged at 35,000 rpm at 4 ºC for 2 h in a fixed angle rotor 50.4 Ti (Beckman Coulter Optima XE- 90 ultracentrifuge) with the acceleration and deceleration levels of 7. The tubes were taken out carefully and each ~50 μL fraction was taken from top to bottom. The fractions were solubilized in 2% β-OG. Rhodamine and pyrene fluorescence in each fraction was measured at the excitation wavelength of 560 nm and 345 nm and at the emission wavelength at 583 nm and 390 nm, respectively. The protein content in each fraction was also analyzed using by SDS-PAGE. 25 μL sample was taken out from each fraction and solubilized with 2% β-OG and SDS sample loading buffer.

4.3.8. Sodium carbonate extraction

There were three liposome samples for each GlpG variant: native GlpG in *E. coli* liposome, sterically trapped denatured GlpG in *E. coli* liposome and empty *E. coli* liposome mixed with water-soluble mSA-WT as a reference. 50 μL of each sample was incubated with 500 μL of pre-

chilled 0.1 M Na_2CO_3 buffer (pH 11.0) for 30 min on ice. Then the mixture was subject to ultracentrifugation at 4 ºC for 30 min at 90,000 g in Beckman polycarbonate tubes (4 mL tube capacity) in a 50.4 Ti rotor. Separated supernatants and pellets were incubated in 2.5 mL or 0.5 mL of 12.5% (w/w) trichloroacetic acid for at least 15 min on ice to precipitate all the protein content, followed by centrifugation for 30 min at 28,000 g at 4 ºC in a fixed angle rotor 50.4 Ti (Beckman Coulter Optima XE- 90 ultracentrifuge). All the pellets after the last centrifugation were first solubilized in 3% (w/v) β-OG, followed by the addition of SDS sample buffer and loaded on SDS-PAGE. In the gel (**Fig. A1**), S stands for the final pellet of the supernatant after the first centrifugation; P stands for the final pellet of the pellet after the first centrifugation. As a reference, the total (T) sample, which was the proteoliposome sample (25 μL) that had not been treated by with sodium carbonate, was directly solubilized with β-OG as a reference, followed by the addition of SDS sample loading buffer.

4.3.9. Expression and Purification of GlpG Substrate SN-LYTM2

For GlpG activity assays, we used the second transmembrane segment of lactose permease (LacYYTM2) as the model substrate of GlpG. LacYTM2 was, fused to staphylococcal nuclease (SN) connected by a linker with a TEV protease recognition site, as described in Chapter 2. For measuring GlpG activity in DDM micelles and DMPC/DMPG/CHAPS bicelles, SN-TEV-LacYTM2-His_6 were labelled with environment-sensitive fluorophore iodoacetyl-7-nitrobenz-2-oxa-1,3-diazol (IA-NBD amide, Setareh Biotech). To measure GlpG activity in *E. coli* liposomes (The principle of the assay is described in **Fig. A1 (c)**), SN-TEV-LacYTM2-His_6 were labeled with either with 5-(iodoacetamido) fluorescein (Sigma Aldrich) or 4-dimethylaminophenylazophenyl-4'-maleimide (DABMI, Setareh Biotech). After labeling, TEV-protease was added to cleave off

the SN domain at the TEV cleavage site, leaving LacYTM2-His_6 with conjugated fluorophore. TEV protease and SN were removed by Ni^{2+}-NTA chromatography.

4.3.10. Monitoring proteolytic activity of GlpG in micelles, bicelles and liposomes

The activity assay in micelles or bicelles was initiated by addition of a 10 times molar excess of the model substrate, NBD-labeled SN-LacYTM2 to GlpG in 20 mM Na_2HPO_4 (pH 7.5), 40 mM NaCl. Time-dependent decrease of NBD fluorescence, which is a measure of proteolytic activity, was monitored in 96-well plate using SpectraMax M5e plate reader (Molecular Devices) with excitation and emission wavelengths of 485 nm and 535 nm, respectively. Fluorescence change was normalized to a control sample containing NBD-SN-LacYTM2 alone. For activity measurement in bicelles, both SN-LacYTM2 and GlpG were pre-incorporated into 3% DMPC/DMPG/CHAPS bicelles.

To measure GlpG activity in liposomes, LacYTM2 labeled with fluorescein and DABMI were incorporated into liposomes composed of *E. coli* phospholipids (Avanti Polar Lipids) at a 1:1 molar ratio with total protein concentration of 50 μM and total lipid concentration of 5 mM. The reconstitution was performed using the following procedures: 5 mM preformed *E. coli* liposomes were incubated with 5 mM DDM at room temperature for 30 minutes. Then 25 μM $LacYTM2_{DAB}$ and 25 μM $LacYTM2_{FL}$ where added while vortexing, following an incubation at room temperature for 30 minutes. For detergent removal, three portions of BioBeads (Bio-Rad) were added (200 mg/mL for each) stepwise. In each step, the mixture was gently stirred for 1–2 h at room temperature. The resulting proteoliposomes were extruded using 0.2 μM pore size membrane.

For activity assay, the proteoliposomes containing LacYTM2 (10 μL) were mixed with the proteoliposomes (5 μL) containing 5 μM GlpG and 18.5 μL of buffer (20 mM Na_2HPO_4, 40 mM NaCl, pH 7.5). Fusion of proteoliposomes was initiated by addition of 16.5 μL 36% PEG_{3350}, 365 mM NaCl. Time-dependent change of fluorescein fluorescence was monitored at 37 ºC in 96-well plate using SpectraMax M5e plate reader (Molecular Devices) with excitation and emission wavelengths of 494 nm and 520 nm, respectively. Fluorescence increase, which is caused by dequenching of fluorescein fluorescence upon cleavage, was normalized to a control sample containing the proteoliposomes containing LacYTM2 mixed with the liposomes without GlpG.

4.3.11. Liposome fusion assay induced by PEG

This assay was used for obtaining the time scale of mixing between the enzyme GlpG and the substrate LacYTM2, which forms a basis for our GlpG activity assay in liposomes. We employed a FRET-based lipid mixing assay. To prepare the proteoliposomes containing the substrate, Cys-less LacYTM2 was reconstituted in *E. coli* liposomes containing 0.02 molar fraction of {N-(7-nitro-2,1,3-benzoxadiazol-4-yl)(ammonium salt) dipalmitoylphosphatidylethanolamine} (DPPE-NBD, FRET donor) and 0.02 molar fraction of quenching lipid {N-(lissamine rhodamine B sulfonyl)(ammonium salt) dipalmitoylphosphatidylethanolamine} (DPPE-Rho, FRET acceptor) to the final substrate concentration of 50 μM and the final lipid concentration of 5 mM. GlpG was reconstituted in *E. coli* liposomes without fluorescent label to the final protein concentration of 5 μM and the final lipid concentration of 5 mM. All the samples were prepared in 20 mM HEPES (pH 7.5) and 200 mM NaCl. The protein/lipid molar ratio was adjusted to mimic that in the activity assays described above.

PEG-induced liposome fusion was detected upon lipid mixing between fluorescently labeled (20 μL) and unlabeled liposomes (9.45 μL) which led to dequenching of NBD-fluorescence caused by separation of NBD and Rho. The fusion reaction was initiated upon addition of 11% (v/v, final concentration) PEG_{3350}. Total volume was 1.4 mL in a Hellma florescence cuvette. NBD fluorescence was detected with an excitation wavelength at 467 nm and an emission wavelength at 530 nm as a function of time with a 5 sec interval (PTI QW4 fluorimeter) with constant stirring at 37 ºC. As a negative control that represents no fusion, no PEG was added. As a positive control for a homogeneously mixing state, 12 μL of 100% Triton X-100 was added to a final concentration of 0.08% (w/v) to solubilize the liposomes.

4.3.12. Proteinase K digestion

5 μM GlpG ($95/172_N$-$BtnRG_2$ or $172/267_C$-$BtnRG_2$) in the absence and presence of 25 μM mSA was prepared in 10 mM DDM, 10 mM DMPC/DMPG/CHAPS bicelles and 10 mM *E. coli* liposomes, as described above. 2 mM $CaCl_2$ was added to enhance the stability of proteinase K (Sigma). Proteolysis was initiated by addition of 0.14 μg/mL proteinase K. An aliquot of each sample was taken at a specified time, and the reaction was quenched by addition of 10 mM permethylsulfoxide. For post-proteolysis removal of bound mSA molecules that had been added to trap the denatured state of GlpG, 4 mM DTT (dithiothreitol) was added to cleave the disulfide bond that links BtnRG label bound with mSA to cysteine. For GlpG samples reconstituted in *E. coli* liposomes, 2% (w/v) β-OG was added to first solubilize the proteoliposomes before addition of SDS sample buffer. Proteolysis reaction by proteinase K was monitored by SDS-PAGE.

4.3.13. Sample preparation for DEER

To obtain the sterically-trapped denatured state in DDM micelles, 120 μL of GlpG variants $95/172_N$–$BtnRG_2$ or $172/267_C$–$BtnRG_2$ (25 μM) was incubated with a 5 times molar excess of mSA-WT in 40 mM DDM, 20 mM Na_2HPO_4 (pH 7.5), 40 mM NaCl at room temperature for three days ($95/172_N$–$BtnRG_2$) or overnight ($172/267_C$–$BtnRG_2$). Then the samples were concentrated to about ~50 μL using Amicon Ultra 0.5 mL (MWCO = 10kD, Millipore Sigma). Glycerol was added to a final 10% (v/v). Native GlpG samples were obtained in the same way but without addition of mSA-WT.

The native and sterically trapped denatured states of $95/172_N$-$BtnRG_2$ and $172/267_C$-$BtnRG_2$ (5 μM GlpG without or with 25 μM mSA-WT) were prepared in 20 mM and 3 % (w/v) DMPC/DMPG/CHAPS as described above (see the subsection, Transfer of native and denatured GlpG to bicelles). Samples were then concentrated using 0.5 mL Amicon centrifugal concentration filter unit (MWCO = 10 kD) and diluted in 20 mM Na_2HPO_4 (pH 7.5), 40 mM NaCl, 10% (v/v) glycerol. Final concentrations of the GlpG variants were typically 40~70 μM.

The native and denatured states of $95/172_N$-$BtnRG_2$ and $172/267_C$-$BtnRG_2$ with 5 μM GlpG and 25 μM mSA were first prepared in micelles and transferred to *E. coli* liposomes as described above (see the subsection, Transfer of native and denatured GlpG into *E. coli* liposomes). To suppress the unwanted inter-molecular dipolar coupling between spin-labeled GlpG in DEER measurements, the lipid concentration was doubled to 20 mM and a 3- or 6-molar excess of Cys-less GlpG (S201A) was mixed with spin-labeled GlpG in DDM prior to addition to the *E. coli* liposomes for reconstitution. After detergent removal by BioBeads and extrusion, samples were concentrated by spinning down the proteoliposomes using a fixed angle rotor 50.4 Ti (Beckman Coulter Optima XE- 90 ultracentrifuge) at 35,000 rpm for 2 h. The resulting pellets were

resuspended in 20 mM Na_2HPO_4 (pH 7.5), 40 mM NaCl, 10% (v/v) glycerol. Final spin-labeled GlpG concentrations were typically 40~60 μM. All samples were flash frozen in liquid nitrogen and stored at -80 ºC.

4.4. Results and Discussion

4.4.1. Reconstitution of the on-pathway DSE in lipid bilayers

In general, under native conditions, detailed biophysical characterization of the denatured states of stable proteins is difficult because of its low population and short lifetime (142, 143). We overcame this difficulty by employing steric trapping (81, 88), which couples unfolding of a doubly-biotinylated protein to competitive binding of bulky monovalent streptavidin (mSA) (**Fig. 4.1**). Using this approach, we were able to trap denatured GlpG in a large quantity without disrupting native lipid-protein and protein-water interactions. Previously, we have identified two pairs of biotinylation sites in GlpG, Pro95/Gly172 and Gly172/Val267, which are optimal for steric trapping (Chapter 1) (88). After substitution of each pair with cysteine residues, GlpG was doubly labeled with a thiol-reactive biotin derivative possessing nitroxide spin label (BtnRG-thiopyridine) or fluorescent pyrene (BtnPyr-IA) (88). With the resulting biotin pair, the denatured states are trapped by mSA approximately at the N-terminal half ($95/172_N$) or the C-terminal half ($172/267_C$). The BtnPyr label serves as a convenient fluorescent marker to detect GlpG. The paramagnetic BtnRG label allows for trapping of the denatured states and measurement of the inter-spin distances between biotinylated sites using DEER. DEER is adequate for measuring the dimension of the denatured states because both long-range distance (15–60 Å) and distribution can be obtained (97).

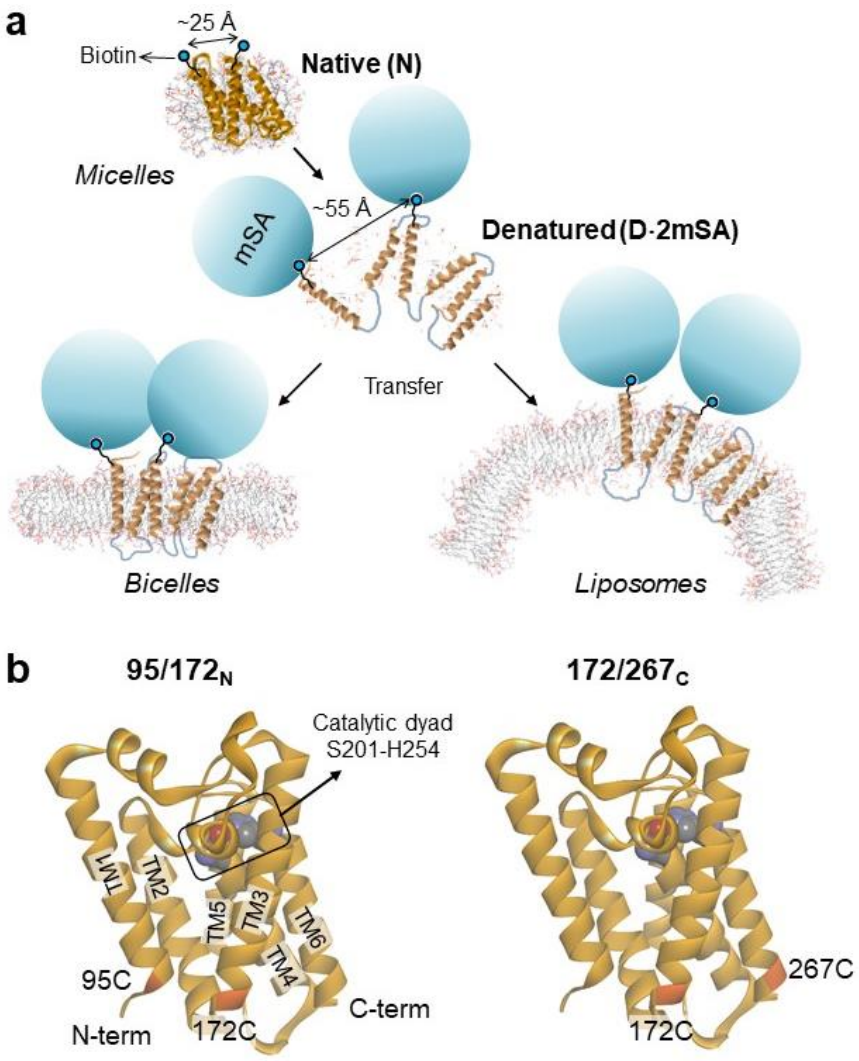


Figure 4.1 Steric trapping strategy to reconstitute denatured GlpG (D·2mSA) in the lipid bilayer. (a) Doubly-biotinylated GlpG was first denatured using a steric trapping in DDM micelles. For reconstitution in bicelles, denatured GlpG was directly injected into preformed DMPC/DMPG/CHAPS (molar ratio=4:1:2) bicelles. For reconstitution in liposomes, the liposomes composed of *E. coli* phospholipids were pre-saturated with detergent DDM. After transfer of denatured GlpG, detergents were removed by polystyrene beads. (b) Two double cysteine variants employed for steric trapping. In each variant, designated cysteine residues are conjugated to a thiol-reactive biotin derivative possessing a fluorescent or paramagnetic group.

The sterically trapped DSEs of double-biotin variants of GlpG were first prepared in DDM micelles upon addition of excess mSA-WT that tightly binds to biotin labels ($K_{d,biotin} \approx 10^{-14}$ M; $k_{off,biotin} \approx$ weeks) (**Fig. 4.1(a)**) (85, 95, 144). Next, the DSE's were reconstituted in two lipid bilayer environments: (1) Phospholipid bicelles, which are discoidal planar bilayer fragments edge-stabilized by detergent. The DSE's were directly injected to the large negatively charged DMPC/DMPG/CHAPS bicelles (molar ratio = 4:1:1.4; lipid-to-detergent molar ratio, q= 2.5; disk diameter ≈ 30 nm (145)) that mimicked the negatively charged cell membranes; (2) The large unilamellar liposomes composed of *E. coli* phospholipids (diameter = 150 nm), which provided the native lipid environment for *E. coli* GlpG. Liposomes were first pre-saturated with DDM and, after transfer of the DSE's, DDM was removed by polystyrene beads.

To test incorporation of the DSE's into the bilayered region of bicelles, we employed fluorescence quenching using GlpG labeled with fluorescent BtnPyr ($95/172_N$-$BtnPyr_2$ and $172/267_C$-$BtnPyr_2$) and the bicelles containing the quencher (dabcyl)-labeled lipid (DOPE-dabcyl) (**Figure 4.2(a)**). Pyrene fluorescence from the DSE's of both double biotin variants was substantially quenched after injection into the preparation of bicelles close to the levels of full incorporation, indicating partitioning of the DSE's into the bilayered region. Incorporation of the DSE's into *E. coli* liposomes was tested using a liposome floatation assay (**Figure 4.2(b) and Fig. A1(a)**). After centrifugation in a sucrose gradient, a majority of denatured GlpG labeled with BtnPyr co-floated with the liposomes containing fluorescently labeled lipids (DPPE-rhodamine). Also, the DSE's reconstituted in liposomes were completely resistant to sodium carbonate extraction, indicating membrane integration (see **Fig. A1(b)**).

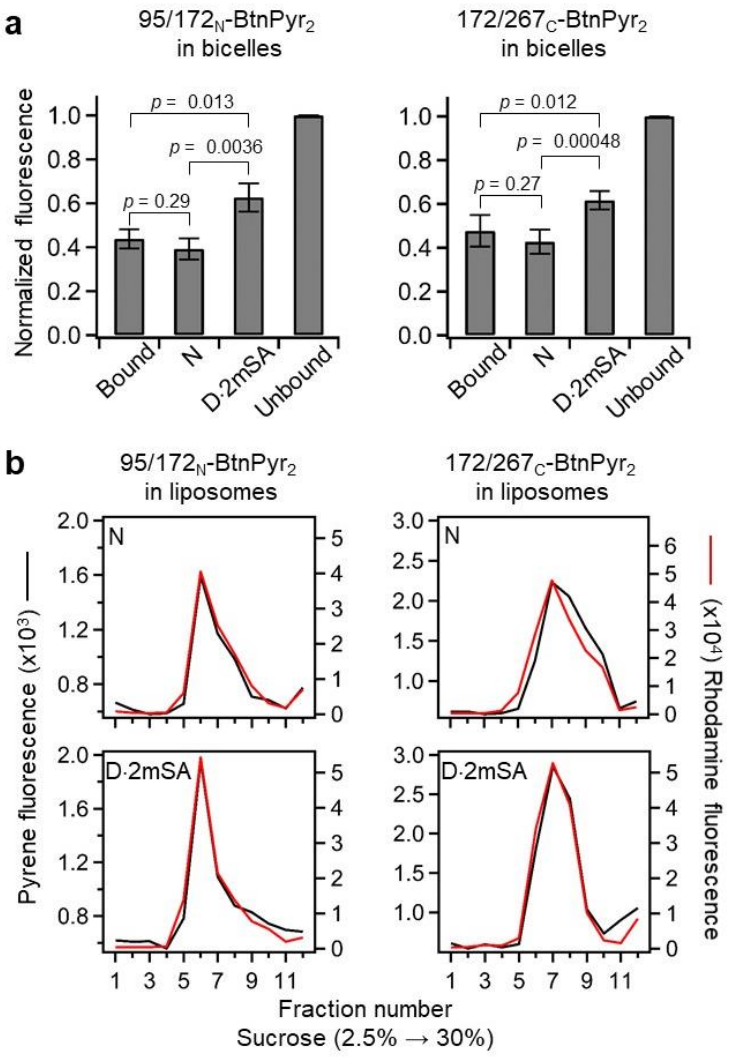


Figure 4.2 Reconstitution of denatured GlpG in the native lipid and solvent environments. **(a)** Fluorescence quenching assay to measure bicelle-association of native (N) and denatured (D·2mSA) GlpG. Binding of pyrene-labeled GlpG (double biotin variants, $95/172_N$-BtnPyr$_2$ and $172/267_C$-BtnPyr$_2$) to dabcyl (quencher)-labeled bicelles induced quenching of pyrene fluorescence. Pyrene-labeled mSA, which is highly soluble in water, was used as a negative control (unbound). Native GlpG, which was first reconstituted in DMPC/DMPG liposomes and then solubilized by CHAPS to form bicelles was used as a positive control (bound). **(b)** Liposome floatation assay in a sucrose gradient to measure membrane-association of native (N) and denatured (D·2mSA) GlpG. Pyrene-labeled native and denatured GlpG (double biotin variants, $95/172_N$-BtnPyr$_2$ and $172/267_C$-BtnPyr$_2$) co-floated with rhodamine-labeled liposomes.

To ensure that the sterically trapped DSE's initially prepared in micelles retain its denaturation status after reconstitution in the bilayers, we measured GlpG activity as a folding

indicator before and after reconstitution (**Fig. 4.3**). In this assay we used GlpG labeled with BtnRG-thiopyridine (95/172$_N$-BtnRG$_2$ and 172/267$_C$-BtnRG$_2$), whose disulfide linkage to cysteine can be reversibly broken by addition of a reducing agent. In both bicelles and liposomes, the activity levels of the DSE's in micelles were maintained after reconstitution. We further examined if the trapped DSE's reconstituted in the bilayers would refold after the steric repulsion was relieved by dissociation of bound mSA. Upon addition of a reducing agent DTT that re-leased BtnRG labels with bound mSA, the activity was regained to >90% of the native level, indicating refolding. Therefore, the sterically trapped DSE's reconstituted in the bilayers are on-pathway in the folding energy landscape of GlpG.

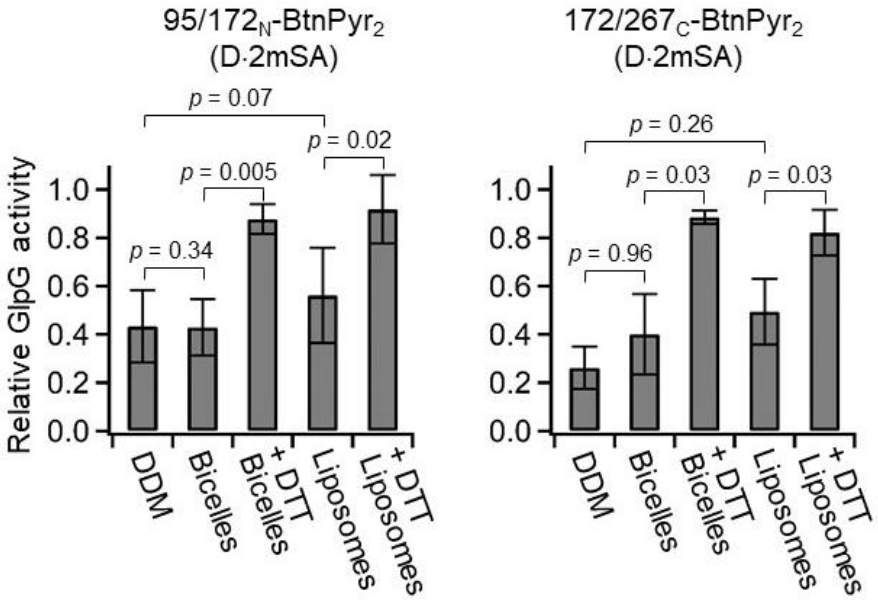


Figure 4.3 Transferred unfolded state is retained and on-pathway. The proteolytic activity of denatured GlpG (95/172$_N$-BtnRG$_2$ and 172/267$_C$-BtnRG$_2$) in micelles, bicelles and liposomes to test the retainment of the sterically trapped denatured state in the bilayers. DTT was added to initiate refolding by releasing the mSA-bound biotin labels from denatured GlpG. GlpG activity in the presence of mSA was normalized to that in the absence of mSA. Error bars denote ± SEM. (n = 3). P values were obtained using Student's t-test.

4.4.2. The global flexibility of the DSE measured by proteolysis is higher in the bilayer

To understand the conformational features of the DSE under native conditions, we first tested limited proteolysis by proteinase K (ProK) in micelles, bicelles and liposomes (**Fig. 4.4**). ProK is a robust nonspecific endopeptidase known to proteolyze water-exposed flexible regions in a protein, but not the regions with stable secondary structure including TM helical segments (146). Time-dependent proteolysis was measured for the DSE's trapped at two different biotin pairs ($95/172_N$-$BtnRG_2$ and $172/267_C$-$BtnRG_2$) using SDS-PAGE (**Fig. 4.4**). In this data, a reducing agent dithiothreitol (DTT) was added after termination of proteolysis reaction to break the linkage between BtnRG label bound with mSA and GlpG. Thus, we can directly observe the digestion of GlpG on SDS-PAGE. Because the fraction of doubly biotinylated GlpG was ~50%, if the sterically trapped denatured state is partially or fully denatured, we expected that ~50% of GlpG would be digested or fragmentized. By image analysis using ImageJ software (147), we estimate ~50% of the total native GlpG fraction was partially or fully digested by ProK.

Combined with the activity data (**Fig. 4.3**), this result illustrates that double binding of mSA induced an increase in conformational flexibility, demonstrating protein denaturation by steric trapping. In micelles, the DSE's trapped at the different biotin pairs displayed clearly different proteolysis patterns: the DSE trapped at the N-terminal half ($95/172_N$-$BtnRG_2$) were proteolyzed yielding only smaller fragments (< 8 kDa), whereas the DSE trapped at the C-terminal half ($172/267_C$-$BtnRG_2$) yielded three larger fragments (17, 13 and 11 kDa) (**Figure 4.4**). Previously, we have shown that, in micelles, the state trapped at the N-terminal biotin pairs in $95/172_N$-$BtnRG_2$ is globally denatured, while the state trapped at the C-terminal biotin pair in $172/267_C$-$BtnRG_2$, is partially denatured (88). The proteolysis to multiple larger fragments

observed for 172/267$_C$-BtnRG$_2$ suggests a partially denatured state with heterogeneous conformations with varied degrees of compactness, supporting our previous finding. In bicelles,

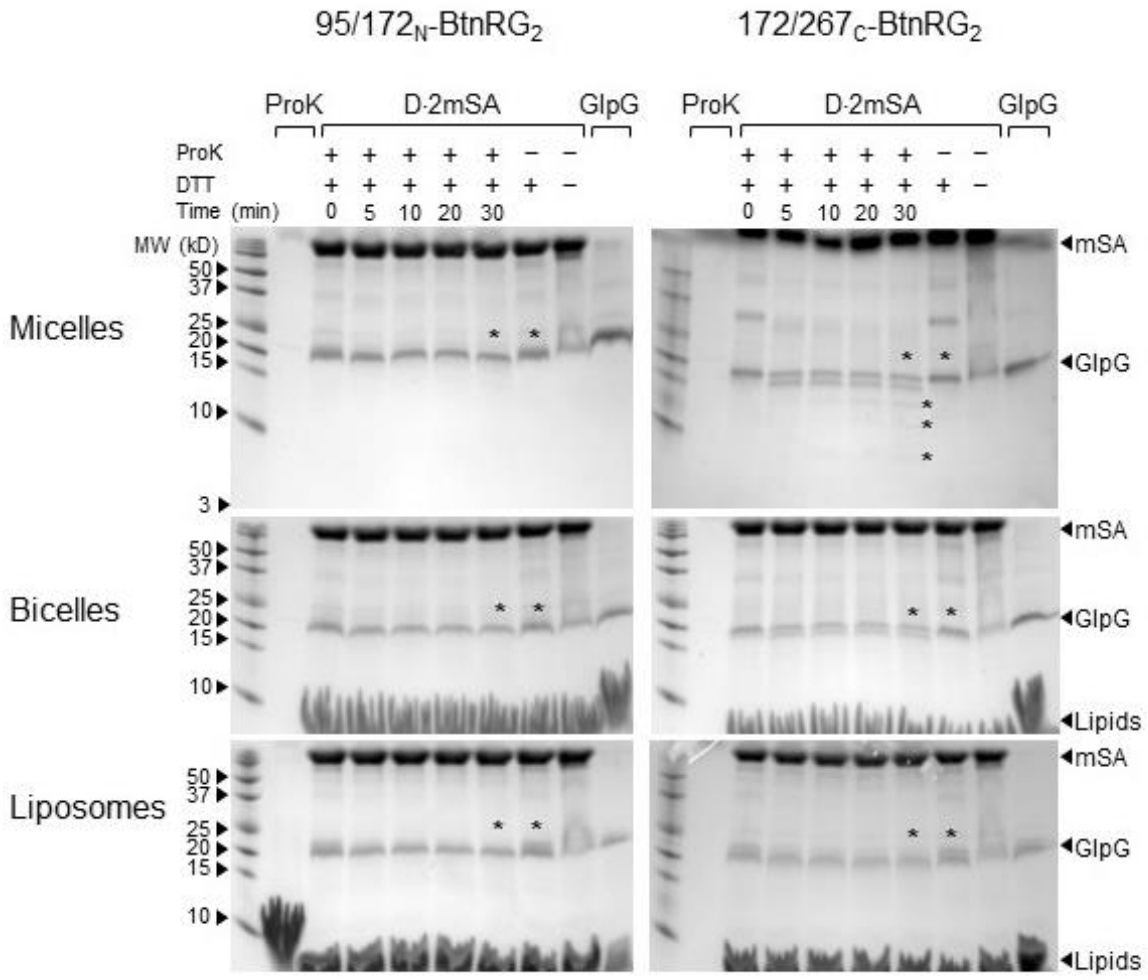


Figure 4.4 Limited proteolysis of denatured GlpG (D·2mSA, 95/172N-BtnRG2 and 172/267C-BtnRG2) by proteinase K (ProK) in (top) DDM micelles, (middle) DMPC/DMPG/CHAPS bicelles, and (bottom) *E. coli* liposomes. After termination of proteolysis reactions, DTT was added to release bound mSA from GlpG. Compare the intensities of GlpG bands (asterisk marks in each gel) in the absence and presence of ProK to confirm proteolysis of GlpG. GlpG are not completely proteolyzed because biotinylation reactions of double cysteine variants are not complete. Singly labeled and unlabeled GlpG are not subject to steric trapping and thus not denatured. These species remain folded and are protected from ProK.

ProK induced maximal proteolysis (i.e., proteolysis to only smaller fragments with <8kDa) for the denatured state trapped at $95/172_N$-BtnRG$_2$, while yielding one larger fragment (~19 kDa) and maximally proteolyzed fragments for $172/267_C$-BtnRG$_2$. In liposomes, the DSE's were maximally proteolyzed regardless of the location of the biotin pair.

Surprisingly, the increase of the lipid content increased the susceptibility to proteolysis for the DSE trapped at the C-terminal half ($172/267_C$-BtnRG$_2$), and eventually in detergent-free liposomes, the DSE's trapped at different biotin pairs became indistinguishable with regard to their proteolysis patterns. This result suggests the conformational flexibility of the DSEs of membrane proteins strongly depends on the amphiphilic environment, and the lipid bilayers keep the DSE in more globally flexible conformations compared to micelles and bicelles. We would be able to obtain more precise dynamic profile of the whole polypeptide chains in the DSE's by analyzing their proteolyzed fragments by ProK. Currently, mass analysis and peptide mapping of the proteolysis products are in progress (data not shown).

4.4.3. The DSE is expanded in the lipid bilayers

Next, we quantified the degree of expansion of the DSE's under native bilayer and solvent conditions. Distances between the two paramagnetic biotin la-bels ($95/172_N$-BtnRG$_2$ or $172/267_C$-BtnRG$_2$) were measured for the native state and sterically trapped DSEs in bicelles and liposomes using DEER. For native GlpG in micelles, BtnRG reports a slightly longer interspin distance by 2−4 Å than widely used spin label R1 (88). We have shown that upon denaturation by steric trapping, the median interspin distance (d_{Med}) increased from 28 Å to 49 Å for $95/172_N$-BtnRG$_2$ and from 26 Å to 51 Å for $172/267_C$-BtnRG$_2$ (1.7-2.0 times expansion relative to the native state). Here we pursued answering two specific questions: (1) How much is the DSE of GlpG expanded

relative to the native state in the bilayers? (2) Does the quasi-two dimensional physical constraint of the bilayers induce compaction of the DSE relative to micelles with looser topological constraints?

Because the Tikhonov regularization to fit the time-dependent dipolar evolution data yielded highly heterogeneous interspin distances for the DSE's without a dominant distance component, we chose to fit the data for the DSE's assuming that the distance distribution conforms to a single Gaussian function. In the native state, the most probable interspin distances (d_{Prob}) in the bilayer environments were overall similar to those in micelles (d_{Prob} = 27−28 Å for 95/172$_N$-BtnRG$_2$ and d_{Prob} =24−30 Å for 172/267$_C$-BtnRG$_2$, **Fig. 4.5**). In bicelles, the DSE's exhibited broad distributions over the entire distance range detectable by DEER (15−60 Å) and significant expansion. The d_{Prob}'s increased from 28 Å in the native state to 35 Å in the DSE for 95/172$_N$-BtnRG$_2$ and from 30 Å to 47 Å for 172/267$_C$-BtnRG$_2$, i.e., the d_{Prob}'s increased by 1.3 and 1.6 fold in the DSE relative to the native state, respectively. Nonetheless, relative to micelles, bicelles did not induce a large expansion of the DSE's. In *E. coli* liposomes, we expected that the DSE's would expand to a similar degree to those in bicelles.

Interestingly, however, we observed larger expansion in liposomes: The d_{Prob}'s increased from 27 Å in the native state to 43 Å in the DSE for 95/172$_N$-BtnRG$_2$ and from 24 Å to 52 Å for 172/267$_C$-BtnRG$_2$. These distance increases correspond to 1.6–2.2 fold relative to the native state and are similar to those in micelles. Surprisingly, despite the quasi-two dimensional constraints of the native lipid bilayers, the lipid bilayers did not impose significant compaction of the DSE's relative to those in micelles with looser topological constraints. Although we highly diluted spin labeled GlpG in liposomes (lipid-to-protein molar ratio, L/P >7,000), the co-localization of multiple spin-labeled GlpG in liposomes may cause unwanted intermolecular dipolar coupling,

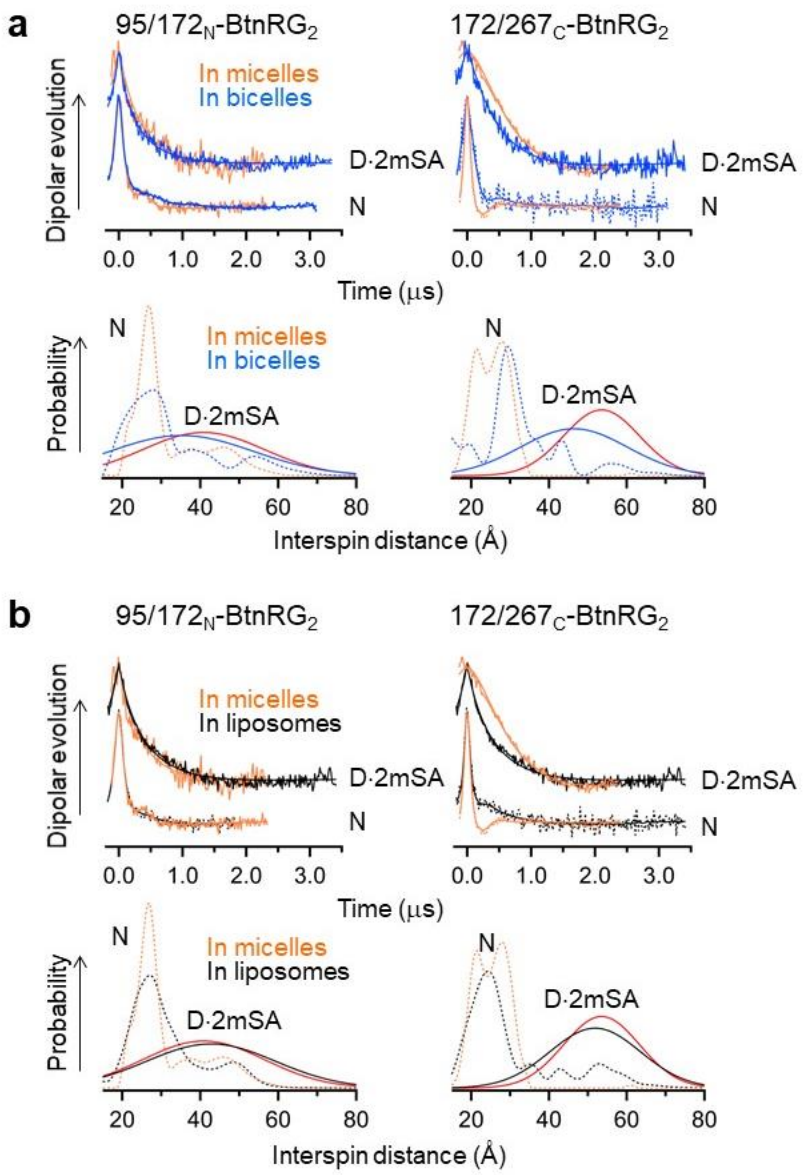


Figure 4.5 Distance distributions in the denatured states of GlpG measured by DEER. (a–b) (*Top*) Background-subtracted dipolar evolution data and their fits and (*Bottom*) interspin distance distributions for native (N) and sterically trapped denatured (D·2mSA) states of GlpG ($95/172_N$-$BtnRG_2$ and $172/267_C$-$BtnRG_2$). The fitting was performed under the assumption that the probabilities of interspin distances conform to a Gaussian distribution. (a) Comparison of DEER data in micelles and bicelles. (b) Comparison of DEER data in micelles and liposomes. The approximate upper limit of the reliable mean distance was ~60 Å.

leading to an overestimation of interspin distances. To test this possibility, we further increased L/P up to 12,000 or co-incorporated an inactive variant of unlabeled GlpG at various molar

excesses relative to spin-labeled GlpG. Under all tested conditions, the overall interspin distances in the DSE did not significantly change, demonstrating that the observed distance distributions mainly originated from intra-molecular dipolar coupling.

Finally, we quantitatively evaluated the ability of the amphiphilic environments tested in this study for solubilizing the DSE based on the distance information obtained from DEER. From polymer theory, the solvents in which a given type of long chain homopolymer is dissolved can be classified into three types, "good", "theta (Θ)" and "poor", depending on the relative strengths between intra-chain and chain-solvent interactions (148, 149). In a "good" solvent, the solvent-chain interaction is more favorable than the intra-chain interaction, and consequently the polymer chain is highly expanded. In a "Θ" solvent, the long-range intrachain and solvent-chain interactions are balanced so that the chain contracts to the degree that cancels out the chain expansion caused by excluded volume. Notably, in the "Θ" solvent, the chain conformations are governed by local forces and random-flight statistics. In a "poor" solvent, the intra-chain interaction overwhelms the solvent-chain interaction, leading to the collapse of the polymer into overall compact conformations. Experimentally, the solvent "quality" can be identified by measuring the ensemble-averaged molecular dimension (radius of gyration, RG) as a function of the number of monomeric units in a polymer chain (i.e., number of amino acids in a polypeptide chain). In the case of a polypeptide chain in three-dimensional space, RG is described using the following equation (150):

$$R_g = R_o N_{AA}^{\nu} \quad \textbf{(10)}$$

, where R_o = 1.98 Å, a constant related to the persistence length of a polypeptide chain, N_{AA} denotes the number of amino acids in a polypeptide chain, and ν is a characteristic exponent defining the solvent quality. $\nu = 0.6$ for a "good solvent", $\nu = 0.5$ for a "Θ solvent", and $\nu = 0.33$ for a "poor solvent" (148). Alternatively, when an end-to-end distance between a residue pair in a

polypeptide chain is measured, equation (10) can be modified into the following equation (150, 151):

$$\left(< R^2 >\right)^{1/2} = \left(6\right)^{1/2} R_o N_{AA}{}^{v} \qquad \textbf{(11)}$$

, where $(<R2>)^{1/2}$ is a root-mean-square distance (RMSD) for a residue pair between which a distance is measured, and N_{AA}: indicates that the number of residues between the residue pair. However, the denatured state of a helical membrane protein is confined in a quasi-two dimensional lipid bilayer with a defined hydrophobic thickness (D = ~30 Å). To establish a prediction model for the degree of expansion for the denatured state of a membrane protein, we employed the model formulated by Daoud and de Gennes for describing the behavior of macromolecular chains in a "good" solvent confined into a flat slit with a defined width (D) (152):

$$\left(< R^2 >\right)^{1/2} = \left(6\right)^{1/2} \left(R_o{}^5 / D\right)^{1/4} N_{AA}{}^{v} \qquad \textbf{(12)}$$

For a good solvent, ν = 0.75. Interestingly, the equation for a "Θ solvent" condition under the quasi-two dimensional constraints collapses into the same equation as equation 11 under the three-dimensional condition with the same characteristic exponent (i.e., ν = 0.5). By assuming that the denatured state is a random-coiled polypeptide chain, we constructed a series of prediction curves describing inter-residue distances as a function of the residue separation under the hypothetical solvent conditions with varying quality (**Fig. 4.6**). According to the two stage model and previous experimental results (24, 99), the denatured state of membrane proteins embedded in the bilayer would possess a significant helical content. However, a MD simulation study indicates that the hypothetical denatured states of helical globular proteins with intact secondary structures display apparently the same inter-residue distance distributions as the completely random-coiled denatured states (150). Intriguingly, our experimental interspin RMSD values determined by DEER for the

DSE in micelles, bicelles and liposomes fell into the range close to the predicted values for the 2D or 3D "Θ−solvent" model (**Fig. 4.6**).

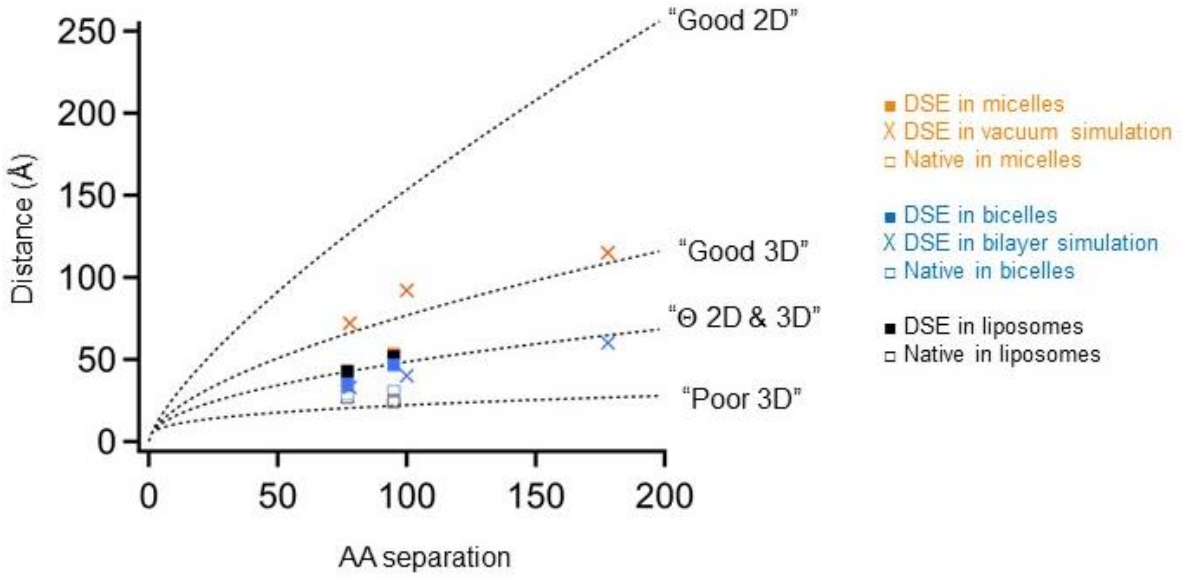


Figure 4.6 The values of the intrachain RMSDs as a function of residue separation obtained from DEER. Those values correspond to the most probable distance from a single-Gaussian fit of the time-dependent dipolar evolution data. The dashed lines indicate the predicted RMSDs from the solvent scaling theories based on the random-coiled polymer models. For the polymers that freely diffuse in three-dimension (Fitzkee and Rose **2004** *PNAS* 101, 12497), the prediction lines were calculated using the equation, $(<R^2>)^{1/2} = (6)^{1/2} R_o N_{AA}^{n}$, where $(<R^2>)^{1/2}$: root-mean-square distance (RMSD); R_o = 1.98 Å, a constant related to persistence length; N_{AA}: the number of residues between the spin labeled sites; n: solvent-scaling exponent characteristic to the solvent quality. In 3D, n = 0.6 for a "good solvent", n = 0.5 for "a Θ solvent", and n = 0.33 for "a poor solvent". For the polymers confined in quasi-2D space under "good solvent" condition, we used the formulation derived by Doud and de Gennes (**1977** *J. Physique* 38, 85), $(<R^2>)^{1/2} = (6)^{1/2}(R_o^5/D)^{1/4} N_{AA}^{n}$, where D: the height of the slit in which the polymer is confined, D = 30 Å (the hydrophobic thickness of a bilayer). For the polymers in quasi-2D space under a "Θ solvent" condition, the equation and the solvent-scaling exponent are the same as those in 3D. The MD simulation was performed by the Wolynes group (Schafer *et al.* **2016** *PNAS* 113, 2098) for the thermally denatured states in vacuum and an implicit bilayer.

Under our assumptions, the peptide segments in the denatured state are allowed to freely move in all directions within the bilayer. However, in the real denatured state, the membrane topology of the hydrophobic segments connected with hydrophilic loops is likely to be fixed

because of the high energetic cost of crossing the hydrophilic loops across the bilayer. Previously, the Wolynes group has performed MD simulation of the thermally denatured state of GlpG in vacuum and an implicit bilayer (59). In their study, although the denatured states retain a small fraction of the native contacts, the inter-residue distances in vacuum simulation agree well with the predicted values from the 3D "good" solvent model (**Fig. 4.6**). Because the fraction of native contacts that they used for their bilayer simulation is similar to that found by their in vacuum simulation, their inter-residue distances in bilayer simulation may represent a more accurate prediction for the 2D "good" solvent model than our random coil model. Interestingly, their simulation result agreed very well with our experimental values, suggesting that bicelles and liposomes may behave as "good" solvents for the denatured state of GlpG.

Taken together, our DEER and limited-proteolysis data as well as available theoretical and computational data strongly suggest that the lipid bilayers "at worst" exhibit the Θ-solvent behavior for the denatured state of GlpG, implying that the lipid-protein interactions are balanced with the protein-protein interactions. Therefore, upon synthesis and membrane insertion, the expanded denatured states of membrane proteins would not nonspecifically collapse into misfolded forms, but fold into their compact native states through specific intramolecular interactions. Although intriguing, this suggestion would be better supported by more physically relevant simulation studies that could provide more accurate reference distance information under each solvent condition. For example, MD simulation could be performed in an explicit bilayer mimicking *E. coli* membrane, and the lipid solvation strength for the denatured state could be changed to modulate the extent of the native contacts for modeling different solvent qualities.

4.5. Conclusion

In this study, for the first time, we investigated the conformational features of the denatured state of a membrane protein under native lipid and solvent conditions. The most striking finding of this study is that despite the quasi-2D constraints of the lipid bilayers, the denatured state is expanded (**Fig. 4.5**) and exhibits global flexibility (**Fig. 4.4**) and possibly "good" solvent behavior (**Fig. 4.6**). This finding implies that the cell membranes are reasonably good at keeping the denatured states of membrane proteins intact, preventing intramolecular or intermolecular collapse. Therefore, under normal physiological conditions, the biogenesis of membrane proteins would occur without an overwhelming burden for handling misfolded membrane proteins by molecular chaperones and degradation machines. However, these quality control mechanisms would be necessary because certain intrinsically unstable membrane proteins would be subject to misfolding and aggregation in membranes crowded with other membrane proteins. Also, environmental stresses such as heat or oxidation would in-crease the risk of misfolding. Overall, our study provides fundamental insights into the physical properties of cellular membranes as media for the folding of membrane proteins.

4.6. Acknowledgements

This work is supported by NIH grant R01GM118685 and the Hunt for a Cure foundation to H.H.

CHAPTER 5

Concluding Remarks

The protein folding problem has been a focus of numerous studies over the last ~60 years, with a lack of attention to membrane proteins. Despite the importance of the protein folding problem of membrane proteins, the progress of the field has been slow due to lack of available methods to control their reversible folding (9). In recent years, steric trapping has emerged as a promising tool to study the folding properties of membrane protein directly under native lipid and solvent conditions (84, 85, 153). Steric trapping couples the spontaneous unfolding of a doubly-biotinylated protein to the binding of mSA, which traps the protein in its unfolded state (**Fig. 2.1**). One of the major downfalls of the previously applied steric trapping approach was the reliance on a protein-specific unfolding readout, such as enzymatic activity or intrinsic chromophores. This work highlights the application of novel tripartite chemical probes, which combine all the necessary components of steric trapping into one thiol-reactive compound (**Fig. 2.2**). These probes are sensitive to mSA binding, by a FRET-based system, or protein unfolding; by DEER-EPR.

In Chapter 2, these novel chemical probes were utilized to study the folding of the intramembrane protease GlpG from *E. coli* in DDM micelles. It was shown that the double binding of mSA specifically induced protein unfolding, and that the unfolding is reversible by the addition of excess free biotin which released bound mSA (**Fig. 2.4**). By monitoring the attenuated second mSA binding coupled to unfolding, the free energy of unfolding of GlpG under non-denaturing conditions was quantified to be ~5 kcal/mol. GlpG possesses moderate stability, similar to that found for soluble proteins. The stability value was quite different than that determined by the classical SDS denaturation method (ΔG^{o}_{U} ~8 kcal/mol) using linear extrapolation to the zero SDS condition. A similar discrepancy has been reported for bR (82). I demonstrated that the discrepancy stemmed from the nonlinear dependence of GlpG stability on SDS mole-fraction X_{SDS} (**Fig. 2.7**).

By design, steric trapping probes the local conformational change of the region which contains the biotin probes. By varying the position of the biotin probes to cover the N- and C-terminal halves of GlpG, I was able to trap different subsets of the unfolded states. This was evident by the lower stability of the C-terminal region as well as the smaller surface area exposure upon protein unfolding. Interestingly, this subglobal unfolding in the C-terminal region exactly encompassed the active site of GlpG, located on the TM4-TM6 interface. It has been suggested that GlpG may require conformational flexibility in this region to facilitate substrate entry and catalysis (112, 154), although the specific mechanism of substrate entry is not yet clear (56).

To determine the extent of the folding cooperativity between the two subdomains of GlpG, a novel cooperativity profiling was developed. I showed the propagation of the energetic effects of mutation and the networked nature of the protein interactions. Using the mutations of 26 residues throughout GlpG, it was shown that most were involved in moderate or strong cooperative interactions to maintain GlpG stability. Interestingly, mutations on the TM4-TM6 interface, which solely make contacts with the C-terminal domain of GlpG, were highly cooperative and over-propagated, preferentially destabilizing the N-terminal domain. These helices have previously been shown to have tight packing interactions which may facilitate proper alignment of the catalytic dyad, Ser201 and His254, which reside on the ends of TM4 and TM6, respectively (57). It is possible that the evolutionary pressure to maintain the active site architecture of GlpG while still allowing for conformational flexibility near the active site, lead to asymmetric stability profile of the TM4-TM6 interface, with residues located only in the C-terminal half of GlpG over-propagating their stability contributions to the N-terminal folding nucleus of GlpG.

This leads me to question the hydrogen bond strength within the catalytic dyad of GlpG, which constitutes the active site. Previous studies of serine proteases, as well as other enzymes

which require hydrogen bonds for catalysis, have suggested that strong hydrogen bond networks may be a feature of the active site that enables the efficient catalytic reaction (126, 155, 156). The catalytic mechanism by GlpG would require breakage and formation of the hydrogen bonds within the membrane, which is predicted to be energetically costly (17). The new steric trapping strategies developed in this work are powerful because of their ability to quantify protein stability without the necessity of monitoring protein activity, allowing for stability analysis of active-site mutants (i.e., inactive variants). Combining stability measurement using steric trapping with double-mutant cycle analysis, I found that the hydrogen bond network in GlpG is relatively weak, in which the His-Ser catalytic hydrogen bond strength amounts to ~1.4 kcal/mol. This result strongly suggests that the proteolysis reaction by rhomboid proteases does not require strong hydrogen bonding that activates Ser for nucleophilic attack of the substrate peptide bond. This study was carried out in DDM micelles, which would not fully represent the lipid bilayer environment. Therefore, it would be beneficial to test the hydrogen bond strength in bicelles or a lipid bilayer.

Upon attempts to push the steric trap strategy towards lipid bilayer systems, I observed enormous kinetic stability of GlpG, displaying no protein unfolding within a reasonable time-scale (~weeks, data not shown). Reconstitution of steric trapping in DMPC/DMPG/CHAPS bicelles yielded unfolding rates between 2~4 days, which is ~40 times slower unfolding in bicelles than in non-denaturing DDM micelles (157). This raises the question, can the unfolded state of GlpG or other membrane proteins reasonably exist in a cellular environment after they initially fold upon release from ribosomes? Using DEER-EPR, we were able to measure the distance between our biotin probes possessing a pair of spin-label on our GlpG in both the folded and unfolded states. In DDM, we measured a widely expanded unfolded state upon denaturation with mSA (Chapter 2, **Fig. 2.5**) with the interspin distances up to ~55 Å. Amazingly, upon transfer of the unfolded

state in DDM to the two-dimensional DMPC/DMPC/CHAPS bicelles or *E. coli* liposomes, the degree of expansion was maintained near ~55 Å, and displayed very heterogeneous distance distributions (Chapter 4, **Fig. 4.5**). Proteolysis of the unfolded state revealed increasing conformational dynamics in the bilayers (**Fig. 4.4**), and mapping of the regions that are protected from digestion in various bilayer environments will give abundant information about the structure of the unfolded state of a membrane protein. Interestingly, despite the two-dimensional environment, the lipid bilayer remained a good solvent for the unfolded state of GlpG. The detailed mechanisms of how the physical and chemical properties of various lipid environments will affect the kinetic and thermodynamic properties of membrane protein folding remains a field of great importance.

APPENDIX

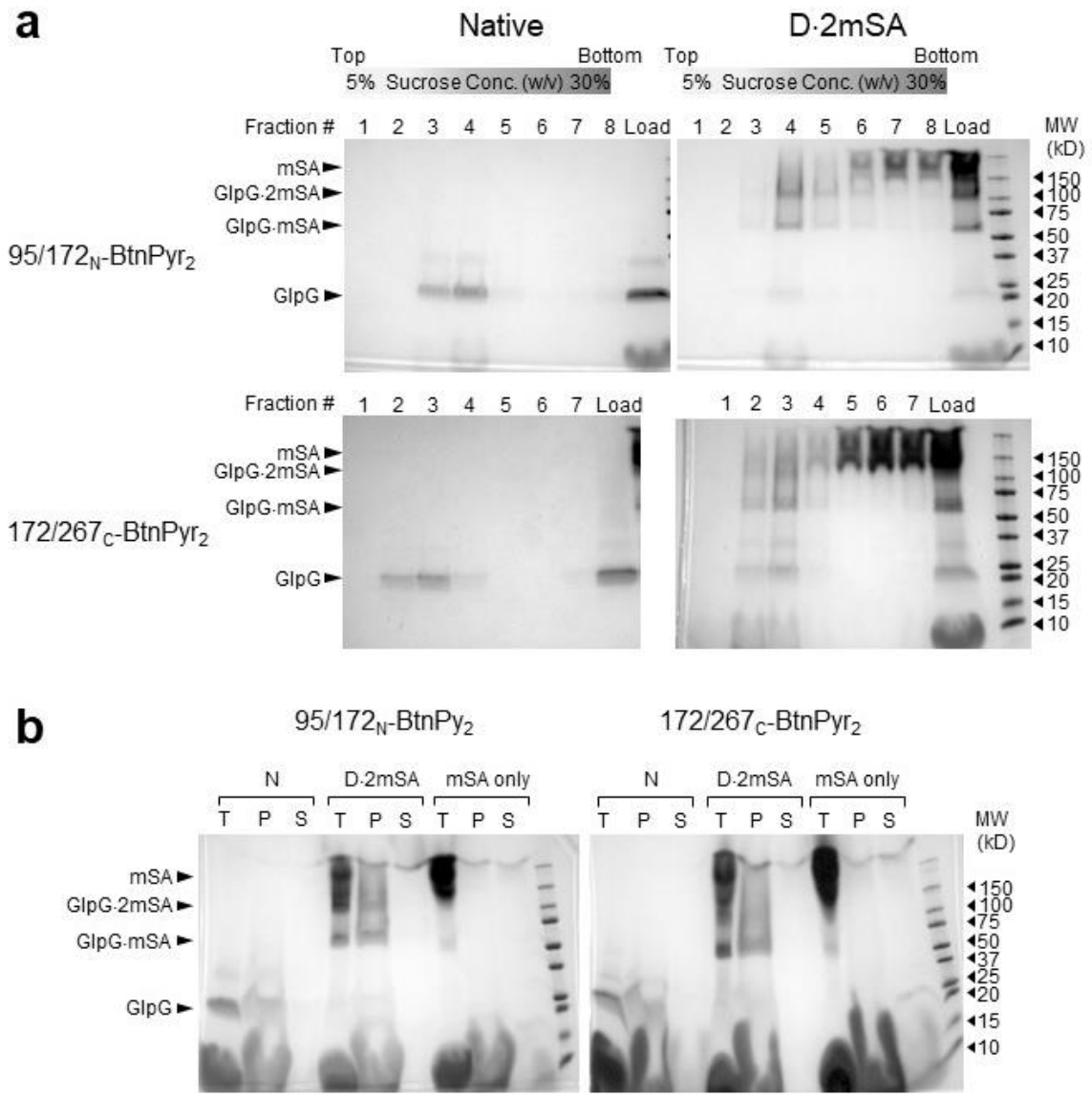


Figure A1 Reconstitution of the denatured states in the lipid bilayer environments.
(a–b) Integration of sterically trapped denatured GlpG into *E. coli* liposome.
(a) Liposome floatation assay for (*left*) native or (*right*) sterically trapped denatured GlpG (D·2mSA) reconstituted in liposomes. Sucrose concentration (w/v) increased from 5% (top layer, Fraction 1) to 30% (Fraction 8, bottom layer). For the native state of $172/267_N$-BtnPyr$_2$, Fraction 7 had 30% sucrose. The GlpG samples incubated in 30% sucrose solution were placed at the bottom. Floatation of GlpG or GlpG ·2mSA to the lower sucrose concentration zones indicates the association of the proteins with liposomes. mSA does not migrate according to MW due to tetrameric structure maintained in the presence of SDS.
(b) Sodium carbonate extraction of native (N) and sterically trapped denatured GlpG (D·2mSA) reconstituted in *E. coli* liposomes. T: total samples without carbonate extraction; P: pellet; S: supernatant. Both native and denatured GlpG were partitioned into the pellet, indicating

transmembrane integration. (c–e) A fluorescence-based assay for measuring GlpG activity in liposomes.

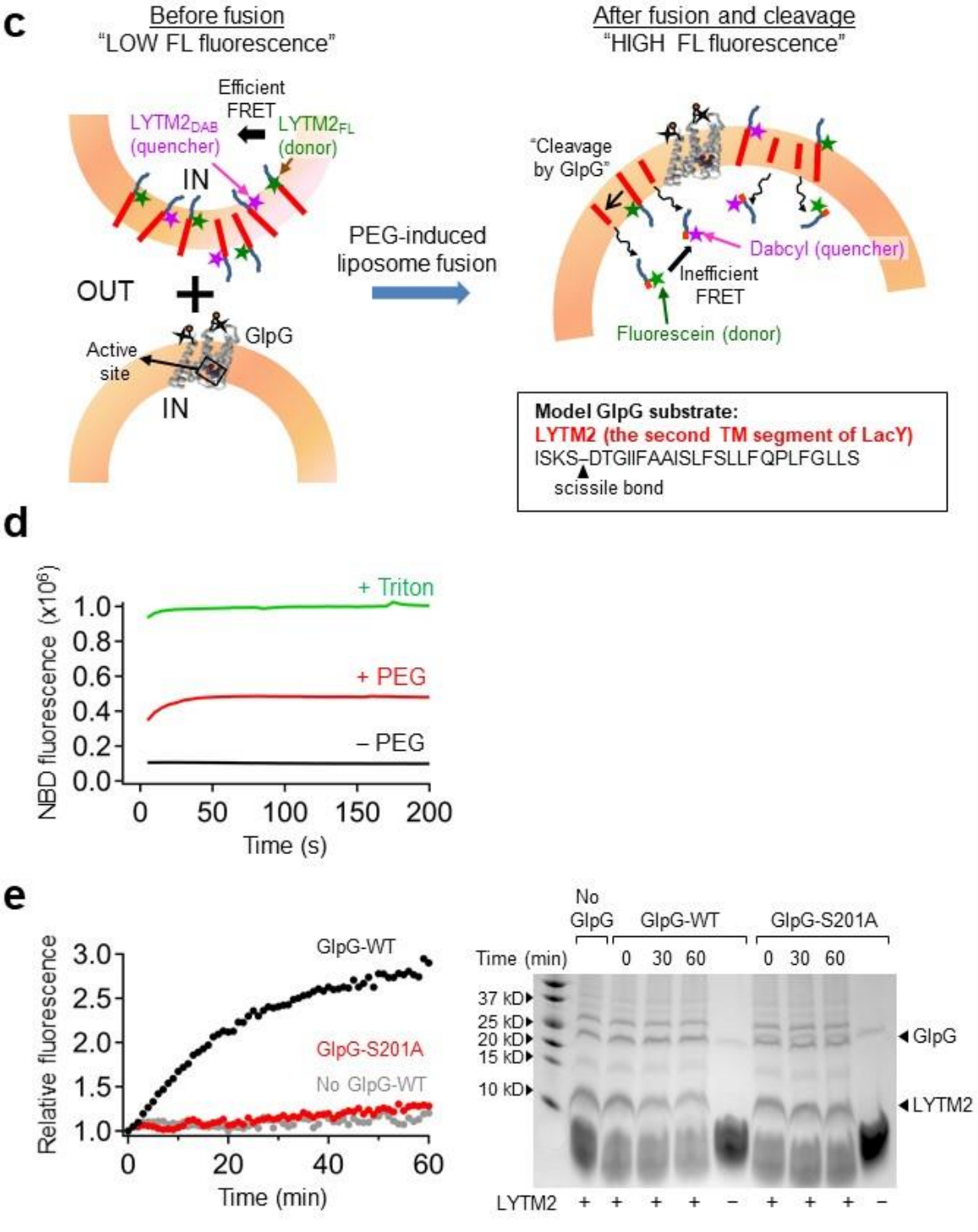
Here we developed a convenient assay for precise measurement of GlpG activity in liposomes employing polyethylene glycol (PEG 3,500)-induced liposome fusion.

(c) The principle of the assay. First, we prepared two types of vesicles, one containing GlpG and the other containing a mixture of its model TM substrate LYTM2 (the second TM domain of lactose permease, LacY) labeled with two different chromophores, fluorescein ($LYTM2_{FL}$, FRET donor) and non-fluorescent quencher dabcyl ($LYTM2_{DAB}$, FRET acceptor). Next, the vesicles are mixed in the presence of PEG to induce liposome fusion. Before fusion, fluorescein fluorescence is highly quenched due to efficient FRET between $LYTM2_{FL}$ and $LYTM2_{DAB}$ which are confined in the same vesicle. After fusion, mixing of GlpG and LYTM2 induces the cleavage of LYTM2, releasing the peptide fragments possessing chromophores into the aqueous phase. Diffusion of FRET pairs into the larger aqueous space causes inefficient FRET, leading to an increase of fluorescein fluorescence, the rate of which is indicative of the proteolytic activity of GlpG.

(d) Kinetics of PEG-induced liposome fusion to induce the enzyme-substrate mixing. Fusion of two types proteoliposomes composed of *E. coli* phospholipids: the liposomes containing NBD (FRET-donor)- and rhodamine (FRET-acceptor)-labeled dipalmitoyl-phosphatidyl-enthanolamine and the liposomes containing unlabeled wild-type GlpG and LacYTM2) at 37^{o}C was monitored by dequenching of NBD fluorescence at 535 nm with the excitation at 467 nm. Dead time of mixing was ~15 sec. This result indicates that liposome fusion for the enzyme-substrate mixing occurs with 1 min, which is much faster than the time scale of the cleavage reaction (~tens of minutes, see **Fig. A1(e)**).

(e) Time-dependent dequenching of fluorescein fluorescence depends on the proteolytic activity of GlpG. (*Left*) After the addition of PEG to the liposome samples, fluorescein (FL) fluorescence was monitored over time. In the presence of wild type GlpG, FL fluorescence increases. Inactivating mutant GlpG-S201A induces no change in fluorescence, as does the addition of empty vesicles. (*Right*) Time-dependent proteolysis of LacYTM2 in liposomes monitored by SDS-PAGE. We observe time-dependent loss of LacYTM2 band in the presence of GlpG-WT overtime, but not in the presence of GlpG-S201A. Therefore, dequenching of FL fluorescence is indicative of cleavage of LacYTM2 by GlpG. See Chapter 4 Materials and Methods for detailed description.

Figure A1 (cont'd)



REFERENCES

REFERENCES

1. Nicolson, G. L. (2014) The Fluid - Mosaic Model of Membrane Structure: Still relevant to understanding the structure, function and dynamics of biological membranes after more than 40 years. *Biochim. Biophys. Acta - Biomembr.* **1838**, 1451–1466

2. Cournia, Z., Allen, T. W., Andricioaei, I., Antonny, B., Baum, D., Brannigan, G., Buchete, N. V., Deckman, J. T., Delemotte, L., del Val, C., Friedman, R., Gkeka, P., Hege, H. C., Hénin, J., Kasimova, M. A., Kolocouris, A., Klein, M. L., Khalid, S., Lemieux, M. J., Lindow, N., Roy, M., Selent, J., Tarek, M., Tofoleanu, F., Vanni, S., Urban, S., Wales, D. J., Smith, J. C., and Bondar, A. N. (2015) Membrane Protein Structure, Function, and Dynamics: a Perspective from Experiments and Theory. *J. Membr. Biol.* **248**, 611–640

3. Krogh, A., Larsson, B., Von Heijne, G., and Sonnhammer, E. L. L. (2001) Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *J. Mol. Biol.* **305**, 567–580

4. Bross, P., Andresen, B., Corydon, T., and Gregersen, N. (1999) Protein misfolding and degradation in genetic disease. *eLS*. 10.1002/9780470015902.a0006016.pub2

5. Bryngelson, J. D., Onuchic, J. N., and Wolynes, P. G. (1994) Funnels, Pathways and the Energy Landscape of Protein Folding: A Synthesis. *Proteins*. **21**, 167–195

6. Dill, K. a, and MacCallum, J. L. (2012) The protein-folding problem, 50 years on. *Science*. **338**, 1042–6

7. Hartl, F. U., Bracher, A., and Hayer-Hartl, M. (2011) Molecular chaperones in protein folding and proteostasis. *Nature*. **475**, 324–32

8. Dill, K. A. (1990) Dominant Forces in Protein Folding. *Biochemistry*. **29**, 7133–7155

9. Bowie, J. U. (2005) Solving the membrane protein folding problem. *Nature*. **438**, 581–9

10. Rowe, S. M., and Clancy, J. P. (2006) Advances in cystic fibrosis therapies. *Curr. Opin. Pediatr.* **18**, 604–613

11. Sanders, C. R., and Nagy, J. K. (2000) Misfolding of membrane proteins in health and disease : the lady or the tiger ?

12. Arora, K., Moon, C., Zhang, W., Yarlagadda, S., Penmatsa, H., Ren, A., Sinha, C., and Naren, A. P. (2015) Stabilizing Rescued Surface-Localized Δ F508 CFTR by Potentiation of Its Interaction with Na + /H + Exchanger Regulatory Factor 1. *Biochemistry*. **53**, 4169–4179

13. Schlebach, J. P., Narayan, M., Alford, C., Mittendorf, K. F., Carter, B. D., Li, J., and Sanders, C. R. (2015) Conformational Stability and Pathogenic Misfolding of the Integral Membrane Protein PMP22. *J. Am. Chem. Soc.* **137**, 8758–8768

14. Kim, B. L., Schafer, N. P., and Wolynes, P. G. (2014) Predictive energy landscapes for folding α-helical transmembrane proteins. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 11031–6

15. Akopian, D., Shen, K., Zhang, X., and Shan, S. (2013) Signal Recognition Particle: An essential protein targeting machine. *Ann. Rev. Biochem*. **82**, 693–721

16. Johnson, A. E., and van Waes, M. A. (1999) The Translocon: A Dynamic Gateway at the ER Membrane. *Annu. Rev. Cell Dev. Biol.* **15**, 799–842

17. Ben-tal, N., Ben-shaul, A., Nicholls, A., and Honig, B. (1996) Free-Energy Determinants of a-Helix Insertion into Lipid Bilayers. *Biophys. J.* **70**, 1803–1812

18. Skowronek, M. H., Rotter, M., and Haas, I. G. (1999) Molecular characterization of a novel mammalian DnaJ-like Sec63p homolog. *Biol. Chem.* **380**, 1133–1138

19. Sanders, S. L., Whitfield, K. M., Vogel, J. P., Rose, M. D., and Schekman, R. W. (1992) Sec61p and BiP directly facilitate polypeptide translocation into the ER. *Cell*. **69**, 353–365

20. Tatu, U., and Helenius, A. (1997) Interactions between newly synthesized glycoproteins, calnexin and a network of resident chaperones in the endoplasmic reticulum. *J. Cell Biol.* **136**, 555–565

21. Van Den Berg, B., Clemons, W. M., Collinson, I., Modis, Y., Hartmann, E., Harrison, S. C., and Rapoport, T. A. (2004) X-ray structure of a protein-conducting channel. *Nature*. **427**, 36–44

22. Wimley, W. C. (2003) The versatile β-barrel membrane protein. *Curr. Opin. Struct. Biol.* **13**, 404–411

23. Dalbey, R. E., Wang, P., and Kuhn, A. (2011) Assembly of bacterial inner membrane proteins. *Annu. Rev. Biochem.* **80**, 161–87

24. Popot, J. L., and Engelman, D. M. (1990) Membrane Protein Folding and Oligomerization: The Two-Stage Model. *Biochemistry*. **29**, 4031–4037

25. Wimley, w. c., and White, S. H. (1996) Experimentallly determined hydrophobicity scale for proteins at membrane interfaces. *Nat. Struct. Biol.* **3**, 842–848

26. Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., Nilsson, I., White, S. H., and Von Heijne, G. (2005) Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature*. **433**, 377–381

27. Moon, C. P., and Fleming, K. G. (2011) Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. *Proc. Natl. Acad. Sci.* **108**, 10174–10177

28. Koehler, J., Woetzel, N., Staritzbichler, R., and Sanders, C. R. (2010) NIH Public Access. **76**, 13–29

29. Lin, M., Gessmann, D., Naveed, H., and Liang, J. (2016) Outer Membrane Protein Folding and Topology from a Computational Transfer Free Energy Scale. *J. Am. Chem. Soc.* **138**, 2592–2601

30. Kyte, J., and Doolittle, R. F. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105–132

31. Gray, A. N., and Henderson-frost, J. M. (2011) Unbalanced Charge Distribution as a Determinant for Dependence of. 10.1128/mBio.00238-11.Editor

32. Dalbey, R. E., and Chen, M. (2004) Sec-translocase mediated membrane protein biogenesis. *Biochim. Biophys. Acta - Mol. Cell Res.* **1694**, 37–53

33. Shimohata, N., Nagamori, S., Akiyama, Y., Kaback, H. R., and Ito, K. (2007) SecY alterations that impair membrane protein folding and generate a membrane stress. *J. Cell Biol.* **176**, 307–17

34. Oberai, A., Joh, N. H., Pettit, F. K., and Bowie, J. U. (2009) Structural imperatives impose diverse evolutionary constraints on helical membrane proteins

35. London, E., and Khorana, H. G. (1982) Denaturation and renaturation of bacteriorhodopson. *J. Biol. Chem.* **257**, 7003–7011

36. Hunt, J. F., Earnest, T. N., Bousche, O., Kalghatgi, K., Reilly, K., Horva, O. C., Rothschild, K. J., and Engelman, D. M. (1997) A Biophysical Study of Integral Membrane Protein Folding. *Biochemistry*. **36**, 15156–15176

37. Farady, C. J., and Craik, C. S. (2010) Mechanisms of Macromolecular Protease Inhibitors. *ChemBioChem*. **11**, 2341–2346

38. Wolfe, M. S. (2009) Intramembrane-cleaving proteases. *J. Biol. Chem.* **284**, 13969–13973

39. Manolaridis, I., Kulkarni, K., Dodd, R. B., Ogasawara, S., Zhang, Z., Bineva, G., O'Reilly, N., Hanrahan, S. J., Thompson, A. J., Cronin, N., Iwata, S., and Barford, D. (2013) Mechanism of farnesylated CAAX protein processing by the intramembrane protease Rce1. *Nature*. **504**, 301–305

40. Wasserman, J. D., and Freeman, M. (1997) Control of EGF receptor activation in Drosophila. *Trends Cell Biol.* **7**, 431–436

41. Mayer, U., and Nusslein-Volhard, C. (1988) A group of genes required for pattern formation in the ventral ectoderm of the Drosophila embryo. *Genes Dev.* **2**, 1496–1511

42. Wasserman, J. D., Urban, S., and Freeman, M. (2000) A family of rhomboid -like genes : Drosophila rhomboid-1 and roughoid / rhomboid-3 cooperate to activate EGF receptor signaling. 10.1101/gad.14.13.1651

43. Urban, S., Lee, J. R., and Freeman, M. (2001) Drosophila Rhomboid-1 defines a family of putative intramembrane serine proteases. *Cell*. **107**, 173–182

44. Urban, S., and Wolfe, M. S. (2005) Reconstitution of intramembrane proteolysis in vitro reveals that pure rhomboid is sufficient for catalysis and specificity. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 1883–8

45. Dowse, T. J., Pascall, J. C., Brown, K. D., and Soldati, D. (2005) Apicomplexan rhomboids have a potential role in microneme protein cleavage during host cell invasion. *Int. J. Parasitol.* **35**, 747–756

46. Stevenson, L. G., Strisovsky, K., Clemmer, K. M., Bhatt, S., Freeman, M., and Rather, P. N. (2007) Rhomboid protease AarA mediates quorum-sensing in *Providencia stuartii* by activating TatA of the twin-arginine translocase. *Proc. Natl. Acad. Sci.* **104**, 1003–1008

47. McQuibban, G. A., Saurya, S., and Freeman, M. (2003) Mitochondrial membrane remodelling regulated by a conserved rhomboid protease. *Nature*. **423**, 537–541

48. Strisovsky, K., Sharpe, H. J., and Freeman, M. (2009) Sequence-Specific Intramembrane Proteolysis: Identification of a Recognition Motif in Rhomboid Substrates. *Mol. Cell*. **36**, 1048–1059

49. Urban, S., and Freeman, M. (2003) Substrate specificity of rhomboid intramembrane proteases is governed by helix-breaking residues in the substrate transmembrane domain. *Mol. Cell*. **11**, 1425–34

50. Akiyama, Y., and Maegawa, S. (2007) Sequence features of substrates required for cleavage by GlpG, an Escherichia coli rhomboid protease. *Mol. Microbiol.* **64**, 1028–37

51. Wang, Y., Maegawa, S., Akiyama, Y., and Ha, Y. (2007) The Role of L1 Loop in the Mechanism of Rhomboid Intramembrane Protease GlpG. *J. Mol. Biol.* **374**, 1104–1113

52. Ben-Shem, A., Fass, D., and Bibi, E. (2007) Structural basis for intramembrane proteolysis by rhomboid serine proteases. *Proc. Natl. Acad. Sci.* **104**, 462–466

53. Lemieux, M. J., Fischer, S. J., Cherney, M. M., Bateman, K. S., and James, M. N. G. (2007) The crystal structure of the rhomboid peptidase from *Haemophilus influenzae* provides insight into intramembrane proteolysis. *Proc. Natl. Acad. Sci. U. S. A.* **104**, 750–4

54. Maegawa, S., Koide, K., Ito, K., and Akiyama, Y. (2007) The intramembrane active site of GlpG, an E. coli rhomboid protease, is accessible to water and hydrolyses an extramembrane peptide bond of substrates. *Mol. Microbiol.* **64**, 435–47

55. Xue, Y., and Ha, Y. (2012) Catalytic mechanism of rhomboid protease GlpG probed by 3,4-dichloroisocoumarin and diisopropyl fluorophosphonate. *J. Biol. Chem.* **287**, 3099–107

56. Cho, S., Dickey, S. W., and Urban, S. (2016) Crystal Structures and Inhibition Kinetics Reveal a Two-Stage Catalytic Mechanism with Drug Design Implications for Rhomboid Proteolysis. *Mol. Cell*. **61**, 329–340

57. Baker, R. P., and Urban, S. (2012) Architectural and thermodynamic principles underlying intramembrane protease function. *Nat. Chem. Biol.* **8**, 759–768

58. Paslawski, W., Lillelund, O. K., Kristensen, J. V., Schafer, N. P., Baker, R. P., Urban, S., and Otzen, D. E. (2015) Cooperative folding of a polytopic α-helical membrane protein involves a compact N-terminal nucleus and nonnative loops. *Proc. Natl. Acad. Sci.* **112**, 7978–7983

59. Schafer, N. P., Truong, H. H., Otzen, D. E., Lindorff-Larsen, K., and Wolynes, P. G. (2016) Topological constraints and modular structure in the folding and functional motions of GlpG, an intramembrane protease. *Proc. Natl. Acad. Sci.* **113**, 2098–2103

60. Min, D., Jefferson, R. E., Bowie, J. U., and Yoon, T. (2015) Mapping the energy landscape for second-stage folding of a single membrane protein. *Nat. Chem. Biol.* **11**, 981–987

61. Spector, A. A., and Yorek, M. A. (1985) Membrane lipid composition and cellular function. **26**, 1015–1035

62. Kučerka, N., Liu, Y., Chu, N., Petrache, H. I., Tristram-Nagle, S., and Nagle, J. F. (2005) Structure of fully hydrated fluid phase DMPC and DLPC lipid bilayers using x-ray scattering from oriented multilamellar arrays and from unilamellar vesicles. *Biophys. J.* **88**, 2626–2637

63. Lee, A. G. (2004) How lipids affect the activities of integral membrane proteins. *Biochim. Biophys. Acta*. **1666**, 62–87

64. Lee, A. G. (2003) *Lipid-protein interactions in biological membranes: A structural perspective*, 10.1016/S0005-2736(03)00056-7

65. Press, E. B., Simmonds, A. C., East, J. M., Jones, O. T., Rooney, E. K., Mcwhirter, J., Lee, A. G., and East, B. C. (1982) Annular and Non-Annular Binding Sites on the (Ca 2+ + Mg 2+)-ATPase. *Biochim. Biophys. Acta*. **693**, 398–406

66. Webb, R. J., East, J. M., Sharma, R. P., and Lee, A. G. (1998) Hydrophobic mismatch and

the incorporation of peptides into lipid bilayers: A possible mechanism for retention in the Golgi. *Biochemistry*. **37**, 673–679

67. Killian, J. A. (1998) Hydrophobic mismatch between proteins and lipids in membranes. *Biochim. Biophys. Acta - Rev. Biomembr.* **1376**, 401–415

68. Bondar, A.-N., del Val, C., and White, S. H. (2009) Rhomboid protease dynamics and lipid interactions. *Structure*. **17**, 395–405

69. Oliveberg, M., and Wolynes, P. G. (2005) The experimental survey of protein-folding energy landscapes. *Q. Rev. Biophys.* **38**, 245–288

70. Dill, K. A., and Chan, H. S. (1997) From Levinthal to pathways to funnels. *Nature*. **4**, 10–19

71. Chamberlain, A. K., Handel, T. M., and Marqusee, S. (1996) Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nature*. **3**, 782–787

72. Bai, Y., Sosnick, T., Mayne, L., and Englander, S. (1995) Protein folding intermediates: native-state hydrogen exchange. *Science (80-.).* **269**, 192–197

73. Sekhar, A., and Kay, L. E. (2013) NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. *Proc. Natl. Acad. Sci.* **110**, 12867–12874

74. Park, C., and Marqusee, S. (2004) Probing the High Energy States in Proteins by Proteolysis. *J. Mol. Biol.* **343**, 1467–1476

75. Sharon, M., and Robinson, C. V. (2007) The Role of Mass Spectrometry in Structure Elucidation of Dynamic Protein Complexes. *Annu. Rev. Biochem.* **76**, 167–193

76. Bai, Y., and Englander, S. W. (2012) Future Directions in Folding: The Multi-State Nature of Protein Structure. *Proteins*. **151**, 201–209

77. Cui, Q., and Karplus, M. (2008) Allostery and cooperativity revisited. *Protein Sci.* **17**, 1295–1307

78. Nussinov, R., and Tsai, C. (2013) Review Allostery in Disease and in Drug Discovery. *Cell*. **153**, 293–305

79. le Coutre, J., Kaback, H. R., Patel, C. K. N., Heginbotham, L., and Miller, C. (1998) Fourier transform infrared spectroscopy reveals a rigid □ -helical assembly for the tetrameric Streptomyces lividans K □channel. *Proc. Natl. Acad. Sci.* **95**, 6114–6117

80. Joh, N. H., Min, A., Faham, S., Whitelegge, J. P., Yang, D., Woods Jr, V. L., and Bowie, J.

U. (2008) Modest stabilization by most hydrogen-bonded side-chain interactions in membrane proteins. *Nature*. **453**, 1226–1272

81. Blois, T. M., Hong, H., Kim, T. H., and Bowie, J. U. (2009) Protein Unfolding with a Steric Trap. *J Am Chem Soc*. **131**, 13914–13915

82. Chang, Y.-C., and Bowie, J. U. (2014) Measuring membrane protein stability under native conditions. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 219–24

83. Hong, H., Blois, T. M., Cao, Z., and Bowie, J. U. (2010) Method to measure strong protein-protein interactions in lipid bilayers using a steric trap. *Proc. Natl. Acad. Sci. U. S. A*. **107**, 19802–7

84. Hong, H., and Bowie, J. U. (2011) Dramatic Destabilization of Transmembrane Helix Interactions by Features of Natural Membrane Environments. *J. Am. Chem. Soc.* **133**, 11389–11398

85. Jefferson, R. E., Blois, T. M., and Bowie, J. U. (2013) Membrane Proteins Can Have High Kinetic Stability. *J. Am. Chem. Soc.* **135**, 15183–15190

86. Lee, J. R., Urban, S., Garvey, C. F., and Freeman, M. (2001) Regulated intracellular ligand transport and proteolysis control EGF signal activation in Drosophila. *Cell*. **107**, 161–171

87. Stevenson, L. G., Strisovsky, K., Clemmer, K. M., Bhatt, S., Freeman, M., and Rather, P. N. (2007) Rhomboid protease AarA mediates quorum-sensing in Providencia stuartii by activating TatA of the twin-arginine translocase

88. Guo, R., Gaffney, K., Yang, Z., Kim, M., Sungsuwan, S., Huang, X., Hubbell, W. L., and Hong, H. (2016) Steric trapping reveals a cooperativity network in the intramembrane protease GlpG. *Nat. Chem. Biol.* **12**, 353–360

89. Guo, R., Gaffney, K., Yang, Z., Kim, M., Sungsuwan, S., Huang, X., Hubbell, W. L., and Hong, H. (2016) Steric trapping reveals a cooperativity network in the intramembrane protease GlpG. *Nat. Chem. Biol.* **12**, 353–360

90. Santoro, M. M., and Bolen, D. W. (1988) Unfolding Free Energy Changes Determined by the Linear Extrapolation Method. 1. Unfolding of Phenylmethanesulfonyl a-Chymotrypsin Using Different Denaturants. *Biochemistry*. **27**, 8063–8068

91. Vosyka, O., Vinothkumar, K. R., Wolf, E. V., Brouwer, A. J., Liskamp, R. M. J., and Verhelst, S. H. L. (2013) Activity-based probes for rhomboid proteases discovered in a mass spectrometry-based assay. *Proc. Natl. Acad. Sci.* **110**, 2472–2477

92. Zoll, S., Stanchev, S., Began, J., Lep, M., Peclinovská, L., Majer, P., and Strisovsky, K. (2014) Substrate binding and specificity of rhomboid intramembrane protease revealed by substrate – peptide complex structures. *EMBO J.* **33**, 2408–2421

93. Sherratt, A. R., Blais, D. R., Ghasriani, H., Pezacki, J. P., and Goto, N. K. (2012) Activity-based protein profiling of the escherichia coli GlpG rhomboid protein delineates the catalytic core. *Biochemistry*. **51**, 7794–7803

94. Arutyunova, E., Panwar, P., Skiba, P. M., Gale, N., Mak, M. W., and Lemieux, M. J. (2014) Allosteric regulation of rhomboid intramembrane proteolysis. *EMBO J.* **33**, 1–13

95. Howarth, M., Chinnapen, D. J., Gerrow, K., Dorrestein, P. C., Grandy, M. R., Kelleher, N. L., El-husseini, A., and Ting, A. Y. (2006) A monovalent streptavidin with a single femtomolar biotin binding site. *Nature*. **3**, 267–273

96. Klumb, L. A., Chu, V., and Stayton, P. S. (1998) Energetic Roles of Hydrogen Bonds at the Ureido Oxygen Binding Pocket in the. *Biochemistry*. **37**, 7658–7663

97. Jeschke, G. (2012) DEER Distance Measurements on Proteins. *Annu. Rev. Phys. Chem.* **63**, 419–446

98. Dockter, C., Volkov, A., Bauer, C., Polyhach, Y., Jeschke, G., Paulsen, H., Dockter, C., Volkovb, A., Bauerb, C., Polyhachc, Y., Joly-lopezd, Z., Jeschkes, G., and Paulsen, H. (2009) Refolding of the integral membrane protein light-harvesting complex II monitored by pulse EPR. *Proc. Natl. Acad. Sci.* **106**, 18485–18490

99. Krishnamani, V., Hegde, B. G., Langen, R., and Lanyi, J. K. (2012) Secondary and Tertiary Structure of Bacteriorhodopsin in the SDS Denatured Sate. *Biochemistry*. **51**, 1051–1060

100. Hubbell, W. L., Cafiso, D. S., and Altenbach, C. (2000) Identifying conformational changes with site-directed spin labeling. *Nature*. **7**, 735–740

101. Mok, Y., Kay, C. M., Kay, L. E., and Forman-kay, J. (1999) NOE Data Demonstrating a Compact Unfolded State for an SH3 Domain Under Non-denaturing Conditions. *J. Mol. Biol.* **289**, 619–638

102. Byler, D. M., and Susi, H. (1986) Examination of the secondary structure of proteins by deconvolved FTIR spectra. *Biopolymers*. **25**, 469–487

103. Myers, J. K., Pace, C. N., and Scholtz, J. M. (1995) Denaturant m values and heat capacity changes : Relation to changes in accessible surface areas of protein unfolding. *Protein Sci.* **4**, 2138–2148

104. Renthal, R. (2006) An unfolding story of helical transmembrane proteins. *Biochemistry*. **45**, 14559–14566

105. Otzen, D. E. (2003) Folding of DsbB in Mixed Micelles : A Kinetic Analysis of the Stability of a Bacterial Membrane Protein. *J. Mol. Biol.* **330**, 641–649

106. Bédard, S., Mayne, L. C., Peterson, R. W., Wand, A. J., and Englander, S. W. (2008) The Foldon Substructure of Staphylococcal Nuclease. *J. Mol. Biol.* **376**, 1142–1154

107. Llinás, M., Gillespie, B., Dahlquist, F. W., and Marqusee, S. (1999) The energetics of T4 lysozyme reveal a hierarchy of conformations. *Nat. Struct. Biol.* **6**, 1072–1078

108. Dutta, A., Kim, T., Moeller, M., Wu, J., Alexiev, U., and Klein-Seetharaman, J. (2010) Characterization of Membrane Protein Non-native States. 2. The SDS-Unfolded States of Rhodopsin. *Biochemistry*. **49**, 6329–6340

109. Lemberg, M. K., and Freeman, M. (2007) Functional and evolutionary implications of enhanced genomic analysis of rhomboid intramembrane proteases. *Genome Res.* **17**, 1634–1646

110. Dill, K. A., and Shortle, D. (1991) Denatured states of proteins. *Annu. Rev. Biochem.* **60**, 795–825

111. Xue, Y., and Ha, Y. (2013) Large lateral movement of transmembrane helix S5 is not required for substrate access to the active site of rhomboid intramembrane protease. *J. Biol. Chem.* **288**, 16645–16654

112. Wu, Z., Yan, N., Feng, L., Oberstein, A., Yan, H., Baker, R. P., Gu, L., Jeffrey, P. D., Urban, S., and Shi, Y. (2006) Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry. *Nat. Struct. Mol. Biol.* **13**, 1084–1091

113. Liu, T., Whitten, S. T., Hilser, V. J., Liu, T., Whitten, S. T., and Hilser, V. J. (2007) Functional residues serve a dominant role protein ensemble in mediating the cooperativity of the. *Proc. Natl. Acad. Sci.* **104**, 4347–4352

114. Bier, E., Jan, L. Y., and Jan, Y. N. (1990) rhomboid, a gene required for dorsoventral axis establishment and peripheral nervous system development in Drosophila melanogaster. *Genes Dev.* **4**, 190–203

115. Cipolat, S., Rudka, T., Hartmann, D., Costa, V., Serneels, L., Craessaerts, K., Metzger, K., Frezza, C., Annaert, W., D'Adamio, L., Derks, C., Dejaegere, T., Pellegrini, L., D'Hooge, R., Scorrano, L., and De Strooper, B. (2006) Mitochondrial Rhomboid PARL Regulates Cytochrome c Release during Apoptosis via OPA1-Dependent Cristae Remodeling. *Cell*. **126**, 163–175

116. Brossier, F., Jewett, T. J., Sibley, L. D., and Urban, S. (2005) A spatially localized rhomboid protease cleaves cell surface adhesins essential for invasion by Toxoplasma. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 4146–4151

117. Wang, Y., Zhang, Y., and Ha, Y. (2006) Crystal structure of a rhomboid family intramembrane protease. *Nature*. **444**, 179–183

118. Vinothkumar, K. R., Strisovsky, K., Andreeva, A., Christova, Y., Verhelst, S., and Freeman, M. (2010) The structural basis for catalysis and substrate specificity of a rhomboid protease. *EMBO J.* **29**, 3797–809

119. Ben-shem, A., Fass, D., and Bibi, E. (2007) Structural basis for intramembrane proteolysis by rhomboid serine proteases. *Proc. Natl. Acad. Sci.* **104**, 462–466

120. Lemberg, M. K., Menendez, J., Misik, A., Garcia, M., Koth, C. M., and Freeman, M. (2005) Mechanism of intramembrane proteolysis investigated with purified rhomboid proteases. *Eur. Mol. Biol. Organ. J.* **24**, 464–472

121. Hedstrom, L. (2002) Serine protease mechanism and specificity. *Chem. Rev.* **102**, 4501–4523

122. Frey, P. A., Whitt, S. A., and Tobin, J. B. (1994) A Low-Barrier Hydrogen Bond in the Catalytic Triad of Serine Proteases. *Science (80-.).* **264**, 1927–1930

123. Warshel, a, and Papazyan, A. (1996) Energy considerations show that low-barrier hydrogen bonds do not offer a catalytic advantage over ordinary hydrogen bonds. *Proc. Natl. Acad. Sci.* **93**, 13665–13670

124. A low-barrier hydrogen bond in the catalytic triad of serine proteases? Theory versus experiment (1997)

125. Lin, J., Cassidy, C. S., and Frey, P. A. (1998) Correlations of the basicity of His 57 with transition state analogue binding, substrate reactivity, and the strength of the low-barrier hydrogen bond in chymotrypsin. *Biochemistry*. **37**, 11940–11948

126. Frey, P. A. (2004) Strong hydrogen bonding in chymotrypsin and other serine proteases. *J. Phys. Org. Chem.* **17**, 511–520

127. Howarth, M., Chinnapen, D. J., Gerrow, K., Dorrestein, P. C., Grandy, R., Kelleher, N. L., El-husseini, A., and Ting, A. Y. (2006) A monovalent streptavidin with a single femtomolar biotin binding site. *Nat. Methods*. **3**, 267–273

128. Baker, R. P., and Urban, S. (2012) Architectural and thermodynamic principles underlying intramembrane protease function. *Nat. Chem. Biol.* **8**, 759–68

129. Horovitz, A. (1996) Double-mutant cycles: a powerful tool for analyzing protein structure and function. *Fold. Des.* **1**, R121–R126

130. Markley, J. L., and Westler, W. M. (1996) Protonation-state dependence of hydrogen bond strengths and exchange rates in a serine protease catalytic triad: Bovine chymotrypsinogen A. *Biochemistry*. **35**, 11092–11097

131. Bowie, J. U. (2011) Membrane protein folding: How important are hydrogen bonds? *Curr.*

Opin. Struct. Biol. **21**, 42–49

132. Saibil, H. (2013) Chaperone machines for protein folding, unfolding and disaggregation. *Nat. Rev. Mol. cell Biol.* **14**, 630–642

133. Matouschek, A. (2003) Protein unfolding — an important process in vivo? *Curr. Opin. Struct. Biol.* **13**, 98–109

134. Sauer, R. T., and Baker, T. a (2011) AAA+ proteases: ATP-fueled machines of protein destruction. *Annu. Rev. Biochem.* **80**, 587–612

135. Tanford, C. (1968) Protein Denaturation. *Adv. Protein Chem.* **23**, 121–282

136. Dill, K. A. (1985) Theory for the Folding and Stability of Globular Proteins. *Biochemistry*. **24**, 1501–1509

137. Riback, J. A., Bowman, M. A., Zmyslowski, A. M., Knoverek, C. R., Jumper, J. M., Hinshaw, J. R., Kaye, E. B., Freed, K. F., Clark, P. L., and Sosnick, T. R. (2017) Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. *Science (80-.).* **358**, 238–241

138. Meng, W., Luan, B., Lyle, N., Pappu, R. V., and Raleigh, D. P. (2013) The denatured state ensemble contains significant local and long-range structure under native conditions: Analysis of the N-terminal domain of ribosomal protein L9. *Biochemistry*. **52**, 2662–2671

139. Aznauryan, M., Delgado, L., Soranno, A., Nettels, D., Huang, J., Labhardt, A. M., Grzesiek, S., and Schuler, B. (2016) Comprehensive structural and dynamical view of an unfolded protein from the combination of single-molecule FRET, NMR, and SAXS. *Proc. Natl. Acad. Sci.* **113**, E5389–E5398

140. Dutta, A., Tirupula, K. C., Alexiev, U., and Klein-Seetharaman, J. (2010) Characterization of membrane protein non-native states. 1. Extent of unfolding and aggregation of rhodopsin in the presence of chemical denaturatns. *Biochemistry*. **49**, 6317–6328

141. Jacso, T., Bardiaux, B., Broecker, J., Fiedler, S., Baerwinkel, T., Mainz, A., Fink, U., Vargas, C., Oschkinat, H., Keller, S., and Reif, B. (2013) The mechanism of denaturation and the unfolded state of the a-helical membrane-associated protein Mistic. *J. Am. Chem. Soc.* **135**, 18884–18891

142. Burton, R. E., Huang, G. S., Daugherty, M. A., Calderone, T. L., and Oas, T. G. (1997) The energy landscape of a fast-folding protein mapped by Ala→Gly substitutions. *Nat. Struct. Biol.* **4**, 305–310

143. Gillespie, J. R., and Shortle, D. (1997) Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels. *J. Mol. Biol.* **268**, 158–169

144. Srisa-art, M., Dyson, E. C., Andrew, J., and Edel, J. B. (2008) Monitoring of Real-Time Streptavidin - Biotin Binding Kinetics Using Droplet Microfluidics Monitoring of Real-Time Streptavidin - Biotin Binding Kinetics Using Droplet Microfluidics. *Anal. Chem.* **80**, 7063–7067

145. Glover, K. J., Whiles, J. A., Wu, G., Yu, N. J., Deems, R., Struppe, J. O., Stark, R. E., Komives, E. A., and Vold, R. R. (2001) Structural evaluation of phospholipid bicelles for solution-state studies of membrane-associated biomolecules. *Biophys. J.* **81**, 2163–2171

146. Wu, C. C., MacCoss, M. J., Howell, K. E., and Yates III, J. R. (2003) A method for the comprehensive proteomic analysis of membrane proteins. *Nat. Biotechnol.* **21**, 532–538

147. Schneider, C. A., Rasband, W. S., and Eliceiri, K. W. (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods*. **9**, 671–675

148. Flory, P. J. (1953) *Principles of Polymer Chemistry*

149. Chan, H. S., and Dill, K. A. (1991) Polymer principles in protein structure and stability. *Annu. Rev. Biophys. Biophys. Chem.* **20**, 447–490

150. Fitzkee, N. C., and Rose, G. D. (2004) Reassessing random-coil statistics in unfolded proteins. *Proc. Natl. Acad. Sci.* **101**, 12497–12502

151. Fehr, N., García-Rubio, I., Jeschke, G., and Paulsen, H. (2016) Early folding events during light harvesting complex II assembly in vitro monitored by pulsed electron paramagnetic resonance. *Biochim. Biophys. Acta - Bioenerg.* **1857**, 695–704

152. Daoud, M., and De Gennes, P. G. (1977) Statistics of macromolecular solutions trapped in small pores. *J. Phys.* **38**, 85–93

153. Blois, T. M., Hong, H., Kim, T. H., and Bowie, J. U. (2009) Protein unfolding with a steric trap. *J. Am. Chem. Soc.* **131**, 13914–13915

154. Baker, R. P., and Urban, S. (2015) Cytosolic extensions directly regulate a rhomboid protease by modulating substrate gating. *Nature*. **523**, 101–105

155. Neidhart, D., Wei, Y., Cassidy, C., Lin, J., Cleland, W. W., and Frey, P. A. (2001) Correlation of low-barrier hydrogen bonding and oxyanion binding in transition state analogue complexes of chymotrypsin. *Biochemistry*. **40**, 2439–2447

156. Wu, C. H., Ito, K., Buytendyk, A. M., Bowen, K. H., and Wu, J. I. (2017) Enormous Hydrogen Bond Strength Enhancement through π-Conjugation Gain: Implications for Enzyme Catalysis. *Biochemistry*. **56**, 4318–4322

157. Yang, Y., Guo, R., Gaffney, K., Kim, M., Tian, W., Wang, B., Liang, J., and Hong, H. (2018) Folding-degradation relationship of a membrane protein mediated by the

universally conserved ATP-dependent protease FtsH. *J Am Chem Soc*. **140**, 4656–4665