

INVESTIGATION OF FOLDING, STABILITY AND FUNCTION OF α -HELICAL
MEMBRANE PROTEINS UNDER NATIVE CONDITIONS

By

Ruiqiong Guo

A DISSERTATION

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

Chemistry–Doctor of Philosophy

2018

ABSTRACT

INVESTIGATION OF FOLDING, STABILITY AND FUNCTION OF α -HELICAL MEMBRANE PROTEINS UNDER NATIVE CONDITIONS

By

Ruiqiong Guo

Membrane proteins count for 25~30% of all proteins and carry out a variety of critical biological processes, such as nutrient transport, signal transduction, catalysis and generation of metabolic energy. Despite the importance, understandings of membrane protein folding lag far behind those of water-soluble proteins. The knowledge gap stems from inherent difficulties in controlling the reversible folding of membrane proteins in lipid bilayers, which is necessary for thermodynamic analysis of driving forces and mechanisms of folding. Steric trapping is a promising tool to reversibly control membrane protein folding. It utilizes the strong binding affinity between the biotin affinity tag and the bulky tag-binding protein streptavidin. In my Ph.D. research, I developed an array of novel methods by synthesizing a set of novel biotinylated protein probes, advancing the steric trapping method for a general application. Applying those methods to studying the folding of a helical-bundle membrane protein, rhomboid protease GlpG in detergent micelles, I mapped its folding energy landscape by revealing subglobal unfolding of the region encompassing the active sites, and quantifying a network of cooperative and localized interactions to maintain the stability. Combining computational methods, I elucidated the role of packing interactions in the stability and function of GlpG, showing that the advanced steric trap method can be used for studying driving forces in membrane protein folding. By using the novel biotinylated spin label, I was able to determine the inter-spin distance between the two biotinylated sites in the sterically trapped denatured state by double electron-electron resonance spectroscopy in the native lipid bilayer environments. These novel steric trapping methods can be applied to investigate a variety

of problems in the folding and stability of membrane proteins directly under native lipid and solvent conditions.

Dedicated to my loving Mom Yan Feng and Dad Jianping Guo,
who are an endless font of love and support

ACKNOWLEDGMENTS

I would like to thank my advisor and friend Dr. Heedeok Hong for his support and guidance throughout my entire Ph.D. years. Dr. Hong is a supportive advisor and he is always available to help me solve the problems I have during research. He is a very open-minded advisor. We had a lot of inspiring discussions about our research, which is the most important guidance during my Ph.D. He is always so patient and encouraging, even at negative experimental results. Heedeok is also an awesome friend to me. His encouraged me to go through some frustrating moments by believing that everything would work out well in the end. And it turned out the everything worked out in its best way in the end. It is my great honor to be the first graduate in Hong lab. It has been an amazing journey to be in this lab for the past five and half years. In addition, I would like to thank my committee members, Dr. David Weliky, Dr. Xuefei Huang and Dr. John McCracken for their guidance and help with my research.

I'm grateful to have such wonderful lab members here. I would like to acknowledge Dr. Miyeon Kim for teaching me all the biochemical and molecular biological techniques when I first joined the lab. And Miyeon always helps us organize the lab to create a good working environment. I also appreciated all the sweet birthday and holiday gifts Miyeon prepared for us. I especially thank Kristen Gaffney for collaborating with me on the projects described in Chapter 2 and 4 and for all the discussions and efficient teamwork. I'd like to thank Yiqing Yang for her friendship and the relaxing tea-time we have for the past few years. In addition, I'd like to thank Shaima Muhammednazaar and Duwage Gunasekara for the nice discussions and talks. I wish good luck to Jiaqi Yao and Zhen Li who recently joined our lab. I would like to thank Erin Dean, the former undergraduate student, who worked with me for the Chapter 3.

I would like to thank all the collaborators who I worked with. Firstly, I thank Professor Xuefei Huang and his former Graduate student Suttipun Sungsuwan (Berm). Berm taught me the organic synthesis and helped me a lot during the synthesis of the biotin probes. I also appreciate the help from other Huang lab members Dr. Weizhun Yang, Dr. Peng Wang, Zeren Zhang, Jicheng Zhang, Seyedmehdi Hossaini Nasr and Dr. Changxin Huo. I'd like to thank Professor Wayne Hubbell and his former postdoc Zhongyu Yang and his current postdoc Mike Bridges. Zhongyu and Mike helped us do the extensive DEER measurement and data analysis. I'd like to thank Professor Guowei Wei and his graduate students Zixuan Cang and Menglun Wang for helping us do the structural homology and cavity analysis. I thank Dr. Seung-Gu Kang for the extensive MD simulation.

I would like to extend my thanks to Professor Lisa Lapidus, Professor John McCracken and Professor Dana Spence for allowing me to use the instruments. Dr. Lijun Chen, Dr. Anthony Schillmiller and Professor Daniel Jones helped me a lot with the mass spectrometry measurements. I would like to thank Professor Liangliang Sun and his graduate student Xiaojing Shen for helpful suggestions on the mass experiments. I also would like to thank Professor Gary Blanchard for being my reference and Professor Remi Beaulac for giving me suggestions when I chose the research lab to join. I would also like to appreciate the help from staffs in Department of Chemistry. I would like to thank all my dear friends here in East Lansing. Because of their companion, I lived a life full of pure happiness and freedom in the past six years. I feel so lucky to meet with so many talented and considerate friends.

I dedicate this dissertation to my parents for their unconditional love and support. I could not get the Ph.D. degree without their support and encouragement. I love them very much.

When I looked back at the past six years, everything happened in a perfect way: I chose to come to MSU, which was a nice and relaxing place to live; I chose to join Hong lab, which equipped me with excellent scientific training and outstanding publication records; I was lucky to land on a job as I planned right after graduation. I feel grateful to have such a happy and fruitful graduate life here. I will always remember the great people I met here and great things happened here.

TABLE OF CONTENTS

LIST OF TABLES	x
LIST OF FIGURES	xi
Chapter 1 Introduction to membrane protein folding problems.....	1
Protein folding, a legacy of basic science	2
Membrane protein folding problems	5
Two-stage model for membrane proteins.....	7
Hydrophobicity scales for understanding the energetics of membrane insertion and accurate prediction of transmembrane.....	11
Driving forces in the second stage of membrane protein folding	13
Current methods to study membrane protein folding.....	17
SDS denaturation	17
Urea and GdnHCl denaturation	18
Steric trapping.....	20
Single molecule force spectroscopy	22
Advanced steric trapping and its applications.....	24
REFERENCES.....	27
Chapter 2 General steric trapping strategy reveals an intricate cooperativity network in the intramembrane protease GlpG under native conditions.....	32
Summary	33
Introduction	33
Results	36
Design and synthesis of new steric-trapping probes	36
Development of high-throughput activity assay for GlpG	38
Steric trapping controls reversible folding of GlpG	39
Steric-trapped unfolded state is widely unraveled.....	42
Thermodynamic stability of GlpG determined by steric trapping.....	46
Subglobal unfolding of GlpG near the active site	51
Steric trapping to measure spontaneous unfolding rate of GlpG	56
Strategy to identify cooperative and localized interactions.....	58
Cooperativity network in GlpG	60
Discussion	64
Materials and Methods	68
Synthesis of BtnPyr-IA and BtnRG-TP.....	68
Preparation of GlpG DNA constructs.....	71
Expression of GlpG	71
Labeling of GlpG and determination of labeling efficiency using SDS-PAGE gel shift assay	72
Fluorescence-based high-throughput activity assay for GlpG	73
Double electron-electron resonance EPR spectroscopy (DEER-EPR)	74
Construction of binding isotherms to determine thermodynamic stability	74

REFERENCES.....	77
Chapter 3 Role of packing defects in the stability and function of an integral membrane protein	81
Summary	82
Introduction	82
Results	84
Identification of conserved cavities in the rhomboids.....	84
Mutational effects of the cavity-filling mutants evaluated by computational methods	88
The effect of cavity-filling mutants on stability and activity of GlpG	96
The interaction free energy between three stabilizing mutants	100
Cooperativity interactions of the stabilizing mutants	102
Discussion	102
Materials and Method.....	103
Homology Modeling and Molecular Dynamic (MD) Simulation of Human Rhomboid Protease RHBDL2.	103
Identification of Common Cavities among Three Rhomboid proteases	104
Double mutant cycle analysis	105
REFERENCES.....	107
Chapter 4 Is the lipid bilayer a good solvent for the denatured state of membrane proteins?	110
Summary	111
Introduction	112
Results	113
Reconstitution of the on-pathway DSE in the lipid bilayers	113
The global flexibility of the DSE measured by proteolysis is higher in the bilayers	122
The DSE is expanded in the lipid bilayers	124
The lipid bilayers exhibit “ Θ -solvent” behavior for the denatured state of GlpG.....	130
Discussion	133
Materials and Methods	135
Bicelle preparation.....	135
Transfer of native and denatured GlpG to bicelles.....	135
Measuring the incorporation of native and denatured GlpG into bicelles.....	136
Preparation of empty <i>E. coli</i> liposomes.....	138
Transfer of native and denatured GlpG into <i>E. coli</i> liposomes	138
Flotation assay of liposome samples	139
Sodium carbonate extraction	140
Monitoring proteolytic activity of GlpG in micelles, bicelles and liposomes.....	140
Liposome fusion assay induced by PEG	141
Proteinase K digestion	143
Sample preparation for DEER.....	143
REFERENCES.....	146
Chapter 5 Concluding remarks	149

LIST OF TABLES

Table 2.1 Stability changes by single substitutions and activities of singly-substituted variants in the backgrounds of double-biotin GlpG variants 95C/172C-BtnPyr ₂ and 172C/267C-BtnPyr ₂ . .	61
Table 3.1 List of Cavities (voids and pockets) in <i>E. coli</i> GlpG identified on the CASTp server (http://sts.bioe.uic.edu/castp/) using a probe radius of 1.4 Å.....	85
Table 3.2 Statistics of the volume fluctuation of the cavities during MD simulation.	92
Table 4.1 Statistical parameters of the interspin distance distributions in the native and sterically trapped denatured states of GlpG in micelles, bicelles and liposomes.	125
Table 4.2 Statistical parameters of the inter-spin distance distributions of the sterically trapped denatured states at an increasing molar excess of unlabeled native GlpG in <i>E. coli</i> liposomes.	128
Table 4.3 Statistical parameters of the interspin distance distributions of the SDS-induced denatured states in the presence and absence of bound mSA.	128

LIST OF FIGURES

Figure 1.1 Funnel-like protein energy landscape.....	4
Figure 1.2 Cumulative unique membrane protein structures.....	7
Figure 1.3 Two-stage model for α -helical membrane protein folding.....	10
Figure 2.1 Principle of steric trapping and steric-trapping probes developed in this study.	37
Figure 2.2 New high-throughput assay for measuring the proteolytic activity of GlpG.	39
Figure 2.3 GlpG reversibly unfolds by double-binding of mSA.	41
Figure 2.4 DEER suggests steric trapping induce wide separation of two biotinylated sites.....	44
Figure 2.5 Characterization of activity and reversibility of GlpG labeled with paramagnetic BtnRG for steric trapping and DEER measurements.....	46
Figure 2.6 Thermodynamic stability of GlpG using steric trapping and SDS denaturation.....	48
Figure 2.7 Dependence of thermodynamic stability of GlpG on SDS mole fraction.	52
Figure 2.8 Cooperativity map reveals a network of clustered cooperative and localized interactions for the stability of GlpG under a native micellar condition.	54
Figure 2.9 Steric trapping of GlpG to measure the spontaneous unfolding rate k_U	57
Figure 2.10 Energy landscape of GlpG in DDM.	58
Figure 2.11 Synthesis scheme of BtnPry-IA.....	68
Figure 2.12 Synthesis scheme of BtnRG-TP.	70
Figure 3.1 Cavities in rhomboid proteases.	86
Figure 3.2 The common cavities in the structures of rhomboid proteases, <i>E. coli</i> GlpG, <i>H. influenzae</i> GlpG and <i>H. sapiens</i> RHBDL2.	89
Figure 3.3 MD simulation result of <i>E. coli</i> GlpG and its cavity-filled variants in explicit bilayer (POPE/POPG, molar ratio = 3:1) and water.	90
Figure 3.4 Analysis of cavity volume and packing on the snapshot structures from MD simulation.	93
Figure 3.5 Impacts of single cavity-filling mutations on the stability and activity of GlpG.	97

Figure 3.6 Additivity and cooperativity of stabilizing cavity-filling mutations.	101
Figure 4.1 Steric trapping strategy to reconstitute denatured GlpG (D·2mSA) in the lipid bilayers.	115
Figure 4.2 Reconstitution of denatured GlpG in the native lipid and solvent environments.....	117
Figure 4.3 Reconstitution of the denatured states in the lipid bilayer environments.....	119
Figure 4.4 Limited proteolysis of denatured GlpG (D·2mSA, 95/172 _N -BtnRG ₂ and 172/267 _C - BtnRG ₂) by proteinase K (ProK) in (<i>top</i>) DDM micelles, (<i>middle</i>) DMPC:DMPG:CHAPS bicelles, and (<i>bottom</i>) <i>E. coli</i> liposomes.	123
Figure 4.5 Distance distributions in the denatured states of GlpG measured by DEER.	127
Figure 4.6 Effects of sample reconstitutions on DEER measurements.	129
Figure 4.7 The values of the intrachain RMSDs as a function of residue separation obtained from DEER.	134
Figure 5.1 Conclusion and outlook of the dissertation research.	152

Chapter 1 Introduction to membrane protein folding problems

Protein folding, a legacy of basic science¹

Proteins are the workhorses of life. A protein's biological function is determined by its three-dimensional (3D) native structure, which in turn is encoded in its one-dimensional (1D) amino acid sequence. Protein folding is a process by which the 1D amino acid sequence folds into its functional 3D structures¹. Misfolded and unfolded proteins have the tendency to aggregate in cells and to induce toxicity unless they are refolded with the aid of molecular chaperones or degraded by proteases. Several neurodegenerative diseases, such as Alzheimer's disease, Parkinson's disease, Creutzfeldt–Jakob disease², are caused by the accumulation of amyloid fibrils resulting from aggregation of misfolded proteins.

In 1958, Kendrew and Perutz published the first structure of a globular protein, myoglobin, at 6 Å resolution³, which set up the foundation of structural biology. Researchers were also astonished by the complexity of the structure lacking symmetry and regularity. The structural features of proteins raised up the fundamental question in molecular biology called “the protein folding problem”, *i.e.*, what are the physical principles that construct such complex and irregular protein structures? Three major specific questions constitute “the protein folding problem”: (1) How is the 3D native structure of a protein determined by the physicochemical properties that are encoded in its 1D amino acid sequence? (2) How can proteins fold so fast considering that there is an almost astronomical number of possible conformations? (3) Can the native 3D structure be predicted from the amino acid sequence for a protein? Furthermore, can new proteins with desired functions be designed?

In 1969, Cyrus Levinthal performed a thought experiment⁴: For example, a polypeptide of 100 residues will have 99 peptide bonds, and therefore 198 different ϕ and ψ angles. If each of these bond angles has three stable rotational isomers, the protein would fold into a maximum of 3^{198}

different conformations. Even if proteins can search 10 billion ways in one second, it would take 10^{85} s, that is 10^{67} times longer of the age of the universe. While most proteins fold in milliseconds to seconds. Levinthal, therefore, proposed that proteins fold through a pathway with conformational limitations to guide the protein folding. This experiment is called Levinthal's paradox.

Researchers now agree that proteins fold fast because of the rapid formation of local stable conformations, forcing proteins to fold in a funnel-like energy landscape⁵. Furthermore, the unfolded proteins may not be considered as a completely random coil. Because the hydrophobic groups tend to collapse together, it is possible that the unfolded proteins form a more compact conformation than a true random coil. Each of the unfolded protein molecules is at the different positions of higher energy states in the energy landscape and finds its own way down to the global free energy minimum, which is usually the native folded structure (**Figure 1.1**). As proteins approach the bottom of the energy landscape, they will gradually follow similar routes because of the narrow funnel shape near the bottom, although in details they follow different routes on the way. According to this model, individual protein molecules will have different transition states. Thus, the transition from one conformation to another also needs to be described as a set of pathways. Rather than thinking of protein as folding in a defined path from the unfolded state to the native state, one should think of it as a bundle of different conformations moving collectively from one location to another. Native protein can be treated as a single defined conformation, but in dynamics, it exists as an ensemble of conformations, a small proportion of which can be quite different from the others. And the fraction of “different” conformation depends on its thermodynamic stability relative to others within the ensemble.

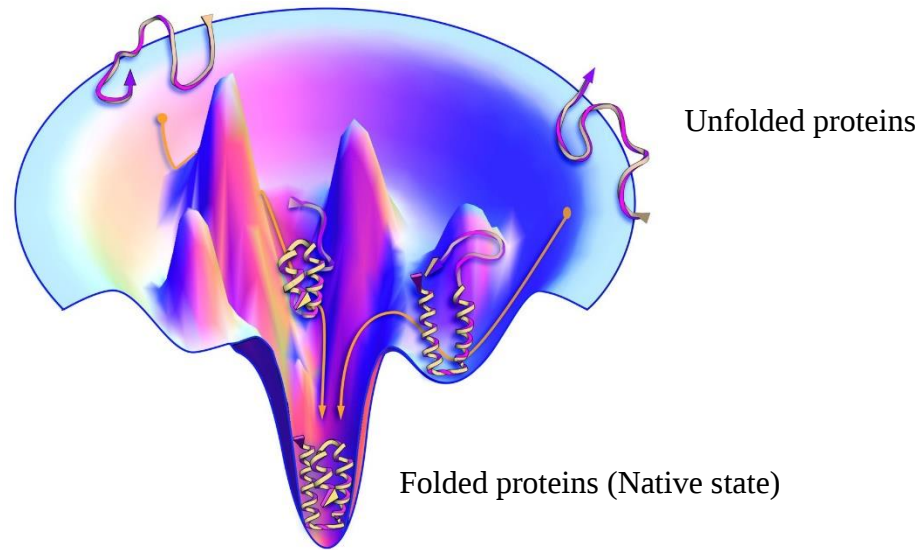


Figure 1.1 Funnel-like protein energy landscape. Proteins have a funnel-like energy landscape with many high-energy, unfolded structures and only a few low-energy, folded structures. Reprint from Dill *et al.* ¹(license number: 4351080226126)

Proteins can be denatured by chemicals such as urea or guanidine, or under harsh conditions such as extreme pH or heat. Protein folding is often studied experimentally by a sudden dilution of denaturants or a sudden change in pH or temperature. Hydrophobicity distribution of amino acids in the buried and solvent-exposed regions provides clues for the driving forces of protein folding. For water-soluble proteins, the interior of proteins is enriched with nonpolar amino acids and the hydrophilic residues are dominant at the protein surface. This is the key evidence that hydrophobic effect is the major driving force in the folding of water-soluble proteins⁶. The hydrophobic interaction is entropic in origin⁷. A hydrophilic group can form hydrogen bonds with water molecules nearby, while hydrophobic groups cannot. Thus, hydrophobic groups in water lead to a decrease in the hydrogen bonds with nearby water molecules, putting them into an energetically unfavorable state. This leads water molecules to form a more ice-like structure to form more hydrogen bonds with themselves. The hydrophobic effect is usually illustrated with the model of dropping oil into water. Oil drops will cluster together, not because of the attraction of oil

molecules but because of the reorientation of waters to minimize the exposure of hydrophobic groups, thus minimize the loss of energy. Overall, the addition of hydrophobic groups to water will result in an increase in enthalpy and a significant loss of entropy. Hydrophobic interaction is not a force but rather than a decrease of an unfavorable energy.

Membrane protein folding problems

The physical codes of protein folding include hydrogen bonding, van der Waals interactions, backbone angle preferences, electrostatic interaction and hydrophobic interactions between chemical groups within a polypeptide chain. For water-soluble proteins, the major driving force is the hydrophobic effect inducing the burial of nonpolar residues in the protein interior. However, besides the proteins that fold in water, there is another major class of proteins that fold in a membrane. Membrane proteins count for 25~30% of all proteins⁸. Unlike water-soluble proteins that fold in an isotropic aqueous environment, membrane proteins fold in an anisotropic membrane environment with chemical and physical complexities.

Cellular membrane is one of the major components in cells. It is composed of lipid bilayers. They are not only a permeability barrier dividing life and environment but also carry out a variety of biological processes. These processes are mainly carried out by membrane proteins. Membrane proteins ferry nutrients across the membrane, receive chemical signals from outside the cell and activate the intracellular action. Membrane proteins are also directly involved in the generation of metabolic energy. Enzymes in the membrane can do the same thing they do in the cytoplasm of a cell. For membrane proteins to perform these critical functions, they have to correctly fold within a lipid bilayer. Misfolded or aggregated of membrane protein can lead to severe diseases such as Alzheimer's disease (aggregation of A β peptides derived from membrane protein amyloid

precursor protein), cystic fibrosis (excessive degradations of cystic fibrosis transmembrane regulator protein bearing missense mutations), and blindness (misfolding of rhodopsin)⁹.

Despite the importance, studying the structure and folding of membrane proteins is challenging. The lipid bilayers provide important environmental constraints for shaping membrane proteins, while their chemical and physical complexity increases the difficulties in obtaining high-quality crystals for crystallography, resolving heterogeneous spectra for NMR spectroscopy and reversibly controlling the folding and unfolding reaction for folding studies. The first structure of membrane proteins was solved about 30 years later than that of water-soluble proteins. Since then, the number of solved membrane protein structures did not grow at an expected exponential rate predicted 20 years after the first structure, as shown in **Figure 1.2**. For water-soluble proteins, the expected exponential increasing rate was reached 20 years after the first structure was solved and it turned out that this prediction was quite accurate¹⁰. However, the increasing rate of membrane proteins slowed down after 20 years. Therefore, the number of available structures of membrane proteins lags far behind that of water-soluble proteins. This lack of structural information prevented the advancement of the knowledge-based force fields for accurate structure prediction of membrane proteins as well as identification of structural folds and motifs that are critical for their assembly^{11,12,13,14}. In the subsequent sections, the conceptual framework of membrane protein folding as well as the current challenges and progress in the field are discussed.

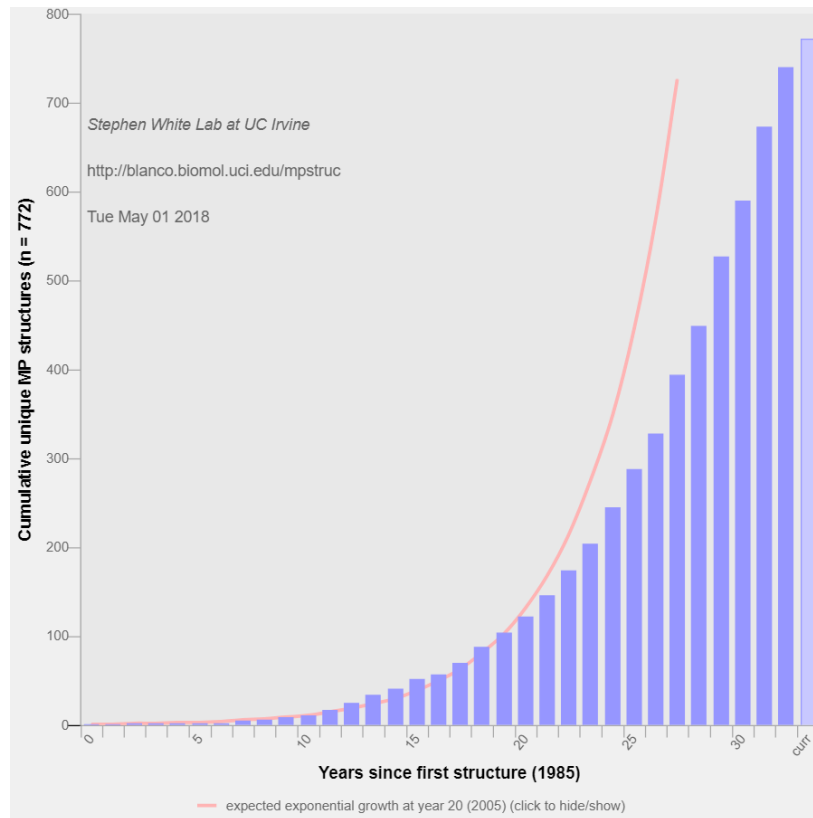


Figure 1.2 Cumulative unique membrane protein structures. The red plot is the expected growth curve at year 20 (2005). However, the growth after that lagged far behind than the expected rate. Unique proteins in the database are 772; Coördinate files in the database are 2506; Published reports of membrane protein structures in database is 1391. <http://blanco.biomol.uci.edu/mpstruc/>

Two-stage model for membrane proteins

The environment of a membrane protein, *i.e.*, a lipid bilayer is a molecular assembly formed by the hydrophobic effect. It can be divided into two regions, the interfaces composed of lipid headgroups and associated water, and the hydrophobic core. The overall bilayer structure is maintained by the complex lateral pressure profile¹⁵. The pressure profile includes: (1) the line tension (*i.e.*, negative lateral pressure) that induces the water-bilayer separation by the cohesive hydrophobic effect; (2) The repulsion between the headgroups (*i.e.*, positive pressure); (3) The repulsion between the hydrocarbon chains (*i.e.*, positive pressure). At equilibrium, the line tension is balanced with the chain and headgroup repulsion. For correct localization within the membrane,

polypeptide chains have to favorably interact with the membranes. At the same time, for folding, they have to make favorable interactions with themselves overcoming the interaction with the environment¹⁶. It is not clearly understood how membrane proteins fold and are stabilized in the bilayer overcoming their presumably favorable interactions with the complex environments. In 1980's, Popot and Engelman proposed the “two-stage model” for membrane protein folding¹⁷. This model was derived mainly from the experimental studies of bacteriorhodopsin (bR). bR is a proton pump activated by the *cis-trans* isomerization of covalent bound retinal upon exposure to light. In their experiment, chemically denatured (NH₂OH) or SDS denatured bR refolded into mild detergents or lipid vesicles¹⁸. The conformation of the refolded bR was the same as the native proteins. In addition, after treated with chymotrypsin, the cleaved N-terminal segment with two TM helices and the C-terminal segment with seven TM helices can reassemble to form the native structure¹⁷.

In the first stage, individual hydrophobic segments of MPs are laterally inserted into the membrane as stable helices. In cells, this step is mediated by a protein conducting channel called translocon and occurs cotranslationally. In the second stage, inserted helices fold into the final 3D structure by lateral interactions. Further modification can happen such as binding of prosthetic groups, folding of the loop regions and oligomerization¹⁹.

The first step is largely driven by the hydrophobic effect. As the nascent peptide chain emerges from the ribosome, if it contains a stretch of nonpolar residues (10–20 amino acids), the segment serves as a membrane-targeting signal that can be recognized by a signal recognition particle (SRP), the SRP binds to the nonpolar peptide segment chain as well as to the ribosome, halting translation. The ribosome-SRP complex then binds to an SRP receptor on the ER membrane (for bacteria, on the cytoplasmic membrane). The SRP receptor is weakly associated with the membrane-integrated

protein-conduction channel, a translocon. Binding to the receptor triggers the hydrolysis of GTP bound both to SRP and the receptor. This event detaches the SRP from the receptor, opens up the plug blocking the channel lumen of the translocon in the rest state, and resumes protein translation⁷. During translation, the translocon integrates the polypeptide segments into the membrane or translocates them across depending on the hydrophobicity of the segment. Importantly, although elongation of a polypeptide chain is driven by GTP hydrolysis, the partition of a translocating polypeptide chain between the translocon and the membrane is known to be in equilibrium²⁰.

However, it is not clear whether the interactions between the transmembrane helices start to form during the insertion or after the completion of insertion. Cymer and Heijne suggested a folding pathway for polytopic membrane proteins in which at least the early folding steps occur cotranslationally²¹. By in vitro experiment in *E. coli*, they found that the C-terminal transmembrane helices could already “sense” the presence of the N-terminal transmembrane helices when they were about to exit the translocon. Thus, tertiary interactions already start to form during the insertion step. This study suggests that membrane insertion does not only depend on the hydrophobicity of the transmembrane segment but also is related to the tertiary interactions with other transmembrane helices.

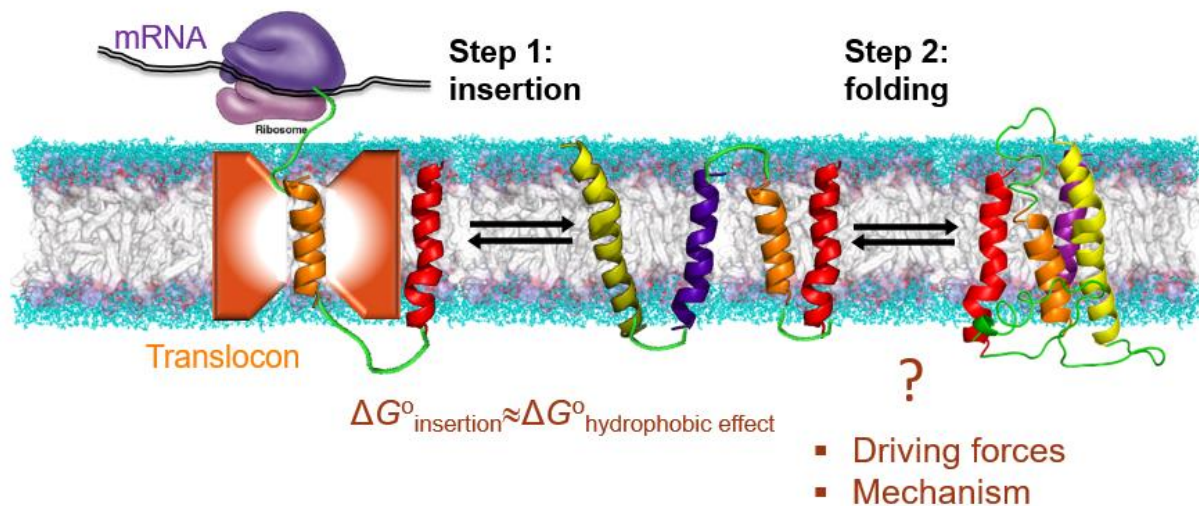


Figure 1.3 Two-stage model for α -helical membrane protein folding. The two-stage models that represent the biogenesis of α -helical membrane protein (PDB: 2HI7 for protein, PDB file for POPC lipid is from Tieleman's group page <http://cmb.bio.uni-goettingen.de/cholmembranes.html>)

Despite the intriguing mechanistic insights provided by the insertion-folding coupling described earlier, the insertion stage is believed to be thermodynamically controlled in general. The key evidence supporting this statement is the “predictability” of the TM segments using the hydropathy plot. In the insertion stage, the hydrophobic regions composed of about 20 contiguous amino acids can be recognized as transmembrane spans by the translocon. Then, it would be possible to predict TM helical regions by scanning the amino acid sequence of a whole polypeptide chain with a ~ 20 amino acid window and obtaining the average hydrophobicity at each scanning step. A hydropathy plot is a graph of amino acid hydrophobicity against amino acid sequence⁷. The sufficiently long continuous hydrophobic regions have a high probability of forming transmembrane helices. The hydropathy plots have been enormously successful in the prediction of TM segments in a membrane protein²². They also have served as the key evidence that supports the thermodynamic partitioning of TM segments into the membrane because the plot uses the hydrophobicity, which is a thermodynamic quantity, for prediction.

Hydrophobicity scales for understanding the energetics of membrane insertion and accurate prediction of transmembrane

To predict the TM domain from the hydropathy plot, it is important to obtain an accurate hydrophobicity value for each amino acid. Early efforts include using the transfer free energy of amino acids from water to an organic solvent (*e.g.*, octanol), which served as a mimic of the hydrophobic core of the membrane²³. Later, Hessa and von Heijne developed another hydrophobicity scale in the biological context involving an *in vitro* translation system, the ER-derived microsomes containing a translocon, and the model TM segments with various sequences²⁴. This method utilizes the differential degree of glycosylation of the two flanking regions of a model TM segment, which depends on their insertion. Briefly, if the TM segment is inserted into the membrane, the TM segment possesses the topology with one flanking region is located in the ER lumen and the other in the cytosol. Then, a single glycosylation occurs on the luminal region. On the other hand, if the TM segment is not inserted into the membrane, both flanking regions are located in the ER lumen and double-glycosylation occur. Then, the equilibrium constant of insertion was obtained by quantifying the bands of singly- and doubly-glycosylated TM segments on SDS-PAGE. By placing a test amino acid at the center of the model TM segments, they were able to obtain the hydrophobicity values (*i.e.*, translocon-bilayer partition free energies) of all amino acids in a biological membrane environment. The hydrophobicity scales from the water-octanol partition and the translocon-membrane partition agreed reasonably well except for tryptophan and proline, which are less likely to insert in the biological context. Probably, the difference in the chemical property between octanol and the center of the bilayer may have caused the discrepancy. For example, the bulky size and hydrogen bonding ability of the indole ring in Trp would make its partition into octanol more favorable than into the bilayer center lacking the

ability to solvate the polar groups. The Hessa-von Heijne hydrophobicity scale was determined between the translocon-to-bilayer transitions. Despite the biological relevance, this scale still cannot account for the partition between the lipid bilayers and a completely water-solvated state, which is essential for the physical description of the stability of membrane proteins. Fleming and Moon reported the first water-to-bilayer hydrophobicity scale, using thermodynamic folding measurements of the transfer of an amino acid side chain of outer membrane phospholipase A (OmpLA) from phospholipid bilayer to the bulk water phase²⁴. The main advantage of using this outer membrane protein is that it spontaneously folds and inserts into lipid membranes from a water-solubilized unfolded state. They claimed that both of those previous scales would undervalue the energetics of most amino acids if they were used to represent water-to-bilayer side-chain partitioning. Two exceptions were aspartic acid and glutamic acid, whose energetics were underrepresented in Moon-Fleming scale because the folding experiments were performed at pH 3.8, which is close to the pKa values for Asp and Glu side chains. They also found that the arginine side chain near the center of a lipid bilayer was accommodated with much less energetic cost than predicted by previous molecular dynamics simulations.

There are also notable computational methods for calculation of transfer free energy. Liang's group reported a computational approach to calculate the folding free energy of the transmembrane region of outer membrane β -barrel proteins by combining an empirical energy function with a reduced discrete state space model²⁵. The strength of this method is the derivation of the hydrophobicity of amino acid side chains at different membrane-depths, whereas the previous experimentally determined hydrophobicity scales were measured at the bilayer center, which is regarded as free of water. It is important to assess the transfer free energy of amino acids from solution into a lipid bilayer. It will help us better understand the energetics of membrane protein

folding and accurately predict the TM segments of membrane proteins based on amino acid sequences.

Driving forces in the second stage of membrane protein folding

For water-soluble proteins, the predominant distribution of nonpolar residues inside the protein provided a piece of evidence that the main driving force in folding is the hydrophobic effect. For membrane proteins, the free energy gain from the hydrophobic effect is largely consumed in the insertion step. Then, what drives the folding of inserted TM helices in the bilayer? Analysis of the amino acid distribution in the interior and exterior of membrane proteins of known structure could not provide a clear answer to this question. While the surface exposed residues of membrane proteins are dominantly nonpolar, whether the protein interior has a stronger hydrophobicity than the exterior cannot be concluded in a general trend. For example, Stevens and Arkin reported that the membrane protein interior and exterior had similar hydrophobicity by analyzing TM helices in the dataset of 9 structures²⁶. While Adamian showed a biased distribution of amino acids in the interior and exterior of membrane proteins²⁷. The possible reason could be that proteins evolve for both stability and functions. Therefore, the hydrophobicity would not be optimized purely for stability. For examples, protein channel requires polar residues in protein interior for transport of polar molecules. In the following paragraphs, I will discuss the current progress in understanding the driving forces of membrane protein folding.

Van der Waals interaction is the attraction and repulsion between instantaneous atomic dipoles that originate from electron fluctuation. Without significant contributions from the hydrophobic effect, van der Waals interactions significantly contribute to the integrity of the protein interior and can be dominant in determining the tertiary structure of TM domains. Tertiary folding of membrane proteins depends on the association of transmembrane helices. The contribution of van

der Waals interaction to membrane protein stability had been quantitatively studied using the model system of dimerization of the single-span glycoporphin A transmembrane domain (GpATM). The GpATM dimer features a glycine zipper GxxxG motif near the central region of the TM peptide. This motif enables the close packing of the neighboring residues, in which the helix-helix interface is formed by the small glycine residues²⁷. This inter-molecular packing happened not only for single-span helices but also in polytopic α -helical membrane proteins. Using sedimentation equilibrium analytical ultracentrifuge assay in detergent micelles, a wide array of mutants were tested to measure the packing contributions of individual amino acids^{28,29}. The mutational effects were not simply additive, which implied that the energetic coupling between distant interfacial residues. The study using steric trap (will be described later in this chapter) also revealed that the energetic contribution of each side chain to the dimer stability was significantly different in lipid bilayers compared to that in detergent micelles³⁰. This result indicated the importance of studying membrane protein folding in the native lipid environments.

Although the mean packing density of proteins is similar to that in the crystal of small organic molecules³¹, Kellis *et al.* found that making cavities in the protein interior would destabilize the protein³². The destabilization effect was larger than the energy difference of transferring the residue from water to nonpolar solvents. This fact implied that the destabilization effect not only stemmed from the loss of the hydrophobic effect, but also from the loss of van der Waals packing interactions. Eriksson estimated the free energy cost of forming a cavity in protein interior is 24-33 cal/mol/ $\text{\AA}^3$ or 20 cal/mol/ $\text{\AA}^2$, using water-soluble protein T4 lysozyme³³. This packing contribution is smaller but comparable to that from the hydrophobic effect (25 cal/mol/ $\text{\AA}^2$)³⁴. In membrane proteins, the free energy cost of forming a cavity in protein interior is 36 cal/mol/ $\text{\AA}^3$ or 18 cal/mol/ $\text{\AA}^2$, which is surprisingly similar to that of water-soluble proteins³⁵. Structural and

statistical analysis showed that membrane proteins tend to bury more fractional area of side chains than water-soluble proteins. Also, membrane proteins tend to bury more small side chains while water-soluble proteins accommodate large hydrophobic and aromatic residues better³⁶. That is, the residues in the interior membrane proteins can pack closely with each other. These results suggest that the energetic contribution of packing in membrane proteins is similar to that in soluble proteins, while membrane proteins utilize the packing interactions more extensively to achieve their stability.

Theoretical and experimental studies using model compounds showed that transferring non-hydrogen bonded polar groups in the aqueous phase to hydrogen bonded pair in the nonpolar solvent is highly unfavorable mainly because of the large desolvation cost³⁷. While many experimental studies with water-soluble proteins revealed a favorable stabilizing role of hydrogen bonding in the protein interior. For membrane proteins, it was expected that hydrogen bonding would play a larger stabilizing role in the low dielectric environment of the lipid bilayer. Interestingly, however, the experimental studies indicate that the hydrogen bonding in membrane proteins is not as strong as expected from the theoretical studies. Engelman³⁸ and DeGrado³⁹ groups studied the role of hydrogen bonding interaction in the association of single-span TM helices. They found that the energetic contribution of polar side chain was not much different from that of nonpolar side chains. An Ala scanning and SDS denaturation study showed that Ala mutations of buried polar residues destabilized bR to a similar degree of nonpolar residues⁴⁰. The thermodynamic contribution of inter-helical side chain hydrogen bonds was determined by a double-mutant cycle analysis, revealing a moderate stabilization energy of 0.6 kcal/mol on average⁴¹. A measurement using GlpG also reported the moderate strengths of inter-helical side chain hydrogen bonds⁴².

To explain the discrepancy between the theoretically predicted strong hydrogen bonds and experimentally determined weak hydrogen bonds, Bowie suggested that hydrogen-bonding potential prevalent in the polypeptide chain of membrane proteins should compete with the specific hydrogen bonding interactions, leading to a weaker net interaction. Thus, even though the solvent cannot compete for hydrogen bonds in the folded state, the membrane protein itself is a plentiful source of alternative hydrogen bonding partners. In addition, water molecules also exist in the lipid bilayer that can form competitive hydrogen bonds. So the dielectric is not as low as it is in pure organic solvent⁴³.

Salt-bridge interactions exist when two oppositely charged groups are located within 4 Å¹⁶. Salt-bridge interactions are prevalent in proteins, so it is considered as an important driving force in protein folding. Experiments with Arc repressor showed that salt-bridges stabilized the protein. However, replacing the charged residue with similar-sized nonpolar residues stabilized the protein even more⁴⁴. It suggests that the salt-bridge interactions are favorable but not as stabilizing as nonpolar residues in the protein interior.

It is difficult to study the salt-bridge interactions in membrane proteins mainly because of their rare frequency. There were studies showing the important role of salt-bridges in the function and stability of membrane proteins. β -barrel membrane protein OmpA has been used for the thermodynamic study of salt-bridge interactions. Double-mutant cycle analysis of the charged tetrad revealed that the strength of salt-bridge interactions varied a lot⁴⁵. The disruption of the main salt bridge by mutation dramatically reduced the open probability and the open rate, which suggests the channel opening is mediated by the salt-bridge switching. This result also suggests that the salt-bridge interactions can be dynamic such that if there are neighboring charged residues, the salt bridge switching may happen. The strong and dynamic salt-bridge interactions may act as

a molecular switch for conformational changes that are required for the function of membrane proteins¹⁶. However, the thermodynamic analysis of α -helical membrane proteins has not been measured experimentally.

Current methods to study membrane protein folding

The study of protein folding has to be conducted under the condition in which the folded state of proteins represents the most favored conformation being in equilibrium with the unfolded states⁴⁶. Under this condition, based on the thermodynamic principle of protein folding⁴⁶, the denatured proteins should be able to refold to the unique folded state under the native condition no matter how they are unfolded. Conventional methods for studying water-soluble protein folding such as chemical or thermal denaturation are not effective to reversibly unfold membrane proteins. Although challenging, the folding studies of membrane proteins have been carried out using several methods including SDS denaturation, urea and GdnHCl denaturation, steric trapping and single molecule force spectroscopy. The principles and key findings using these methods are described below.

SDS denaturation

This method is used for polytopic α -helical membrane proteins. In this method, the strong ionic denaturing detergent SDS is used as a denaturant. A folded protein solubilized in detergent micelles is denatured by increasing the mole fraction of SDS relative to the total detergent concentration, *i.e.* $X_{\text{SDS}} = [\text{SDS}] / ([\text{SDS}] + [\text{other mild detergents}])$. A folding readout such as Trp fluorescence, circular dichroism and absorbance of intrinsic chromophoric prosthetic groups is monitored as an increasing mole fraction of SDS. Reversibility is achieved by increasing the fraction of mild detergents. In many cases, the equilibrium unfolding curves representing the fraction of the unfolded state are well fitted with a two-state model. The free energy of unfolding

at different SDS mole fraction ($\Delta G_{\text{unfold}}^{\circ}$) was measured in the transition region and then extrapolated to a zero SDS mole fraction to obtain the free energy of unfolding under a native condition⁴⁷.

The folding and stability of several proteins have been studied using this method, including diacylglycerol kinase (DGK)⁴⁸, bR^{40,49}, DsbB⁵⁰, GlpG⁵¹, PMP22⁵², *etc.* Those studies provided important insights into the driving forces and transition states in membrane protein folding. However, the mechanism of SDS denaturation is not clearly understood yet so the validity of linear extrapolation is questionable⁵³. It is also not clear if the unfolded state induced by SDS fully lacks the tertiary interactions. Recent studies on the conformation of SDS induced unfolded states showed that the conformations depended on the concentration of SDS. In lower SDS concentration (0.05-0.3% (w/v), $X_{\text{SDS}} = 0.5-0.85$), the protein was largely opened up, while in higher SDS concentration (0.3-3% (w/v)), the protein-SDS complex becomes more compact than that in the non-denaturing micellar environment⁵³. Two studies using the steric trapping method (see below), including the one from our group, indicate that the linear extrapolation may not be valid at lower SDS mole fraction region^{54,55}. Furthermore, strong anionic SDS disrupts the bilayer structure of the membrane. Therefore, it cannot be applied to the lipid bilayer environment, which is a more real native environment for membrane proteins.

Urea and GdnHCl denaturation

Denaturants such as urea and GdnHCl have been widely used for the reversible control of folding of water-soluble proteins. In the membrane protein folding, those methods have been mainly applied to study the folding of β -barrel membrane proteins⁵⁶. Urea and GdnHCl are membrane compatible and can effectively solubilize less hydrophobic β -barrel membrane proteins. The

insertion and folding of β -barrel membrane proteins happen spontaneously from water to the bilayer, which is different from the two-stage model for α -helical proteins. However, urea and GdnHCl are generally not effective in inducing unfolding and refolding of α -helical membrane proteins. Probably, because of the hydrophobic nature of the helical membrane proteins, these highly polar denaturants are difficult to penetrate into the membrane-embedded region of the proteins for inducing unfolding. In addition, because of the same reason, these denaturants cannot effectively solubilize the unfolded state such that the unfolded proteins rather aggregate than refold into the native state. Recently, the Booth lab has developed a method for refolding a GPCR into *n*-decyl- β -D-maltoside (DM) micelles from a urea-denatured state on a solid support⁵⁷, demonstrating that urea can be still a useful tool for studying the folding of α -helical membrane proteins. Several examples have been reported showing that these chemical denaturation methods are effective in unfolding several major facilitator proteins (MFP), which are sugar transporters, in detergent micelles^{58,59}. Although micelles are easy to solubilize membrane proteins and can easily mix with denaturants, there are disadvantages to study folding in micelles. Unfolded states of membrane proteins in micelles are not constrained to a 2D lipid sheet, but rather in a 3D solution. Therefore, the increased entropy of the unfolded states can distort their free energy level relative to that in the lipid bilayers. Furthermore, detergents do not recapitulate the native lipid-protein interactions. The detergent molecules usually have different headgroups and only single hydrocarbon chain that is often shorter than a typical phospholipid typically possessing two fatty acyl chains. In addition, the lateral pressure profile of micelles is different from that of the lipid bilayer, which exerts a different force gradient on the protein.

Steric trapping

Steric trapping is the newest method originally developed in the Bowie lab⁶⁰. This approach utilizes the strong binding affinity between biotin and streptavidin. The membrane protein of interest is labeled with two biotins at exposed sites that are close in distance but far apart in the amino acid sequence. In the folded state, only one streptavidin can bind to one of the biotins while the second binding is prohibited by the steric hindrance. When the protein transiently unfolds under dynamic equilibrium, the second streptavidin can bind to trap the protein at the unfolded states. Refolding can be achieved by competing off the bound streptavidin with free biotin. This method does not disrupt the lipid bilayer structure because the biotinylated sites are at the flexible loop regions of the target membrane proteins. Depending on the placement of the biotin tags, the steric trap method can trap the unfolded states of different domains.

The method was first proved by a soluble protein dihydrofolate reductase (DHFR)⁶⁰. After doubly-labeled with biotin, DHFR loses activity depending on the biotinylation of both labeling sites and the activity loss is correlated with increasing concentrations of monovalent streptavidin (mSA). The unfolded state trapped by mSA was also probed by limited proteolysis, indicating that steric trapped DHFR is susceptible to proteolysis with the similar degree of chemically denatured proteins. The reversibility was hard to achieve with wild-type mSA because of the high binding affinity (10^{-14} M). Therefore, the usage of a mSA mutant with a reduced biotin affinity was necessary. Activity recovery after addition of excess free biotin demonstrated that steric trapped unfolding was reversible. The unfolding free energy measured by steric trap agreed well with that measured by urea denaturation.

The method was also demonstrated in a few membrane protein systems including GpATM^{30,61}, a trimeric membrane-bound enzyme DGK⁶², a light-driven proton pump bR⁵⁴, and an

intramembrane protease GlpG⁵⁵ in our lab, for measuring the strength of protein dimerization interactions in bilayers, free energies of unfolding, and unfolding kinetics.

In the study of GpATM dimer, each monomer was labeled with pyrene at the N-terminus. Biotin was attached to a biotin acceptor peptide fused to each monomer by an enzymatic reaction. The dissociation was monitored by the de-quenching of pyrene. The dissociation constant determined in detergent agreed well with previous measurements using dilution of GpATM in detergent micelles. Interestingly, the association of GpATM is greatly enhanced in POPC. Such high association affinity cannot be measured by denaturant dilution methods because of the detection limit using FRET or analytical ultracentrifugation³⁰. Therefore, steric trapping can serve as a method for measuring strong protein-protein interactions in the lipid bilayer environment. In addition, steric trap can be applied to various lipid bilayer environments. The follow-up study with other lipid compositions revealed that the dimerization affinity is weaker in the negatively charged bilayers because of the interaction between protein and lipid headgroups⁶¹. These results illustrated that membrane proteins are not necessarily more stable in membranes than in detergent.

In the study of bR, the absorbance of retinal was used as an unfolding readout because of its sensitivity to the conformational states⁵⁴. This study was done in bicelle environment which captures the interaction between lipids and proteins. The key finding of this study was to test the validity of the linear extrapolations from SDS unfolding measurements. The unfolding free energies measured at lower SDS mole fraction turned out to be curvy instead of being linear as a function of SDS mole fraction. The unfolding free energy measured by steric trap without adding SDS is ~11 kcal/mol. While if the linear extrapolation was used, the free energy of unfolding would be 26 kcal/mol.

In our lab, we further advanced the steric trapping method for general application by introducing novel biotin probes with a spectroscopic reporter group to achieve unfolding by steric trapping and detection of unfolding at the same time, without exploiting an intrinsic characteristic of a protein for a functional assay. The details of the method and key findings are described in Chapter 2.

Single molecule force spectroscopy

Mechanical forces have been applied to pulling the membrane proteins to force them to unfold. Atomic Force Microscope (AFM) tip is typically used to pull a single protein molecule, one end of each is attached on the surface. In the study using membrane protein bR, purple membrane was adsorbed onto a flat surface and the AFM tip was adsorbed to a protein molecule at the C-terminal. Then the protein was pulled out with an increasing force. The polypeptide chain starts to unfold and extend with increasing force, producing a force/extension curve. The amount of force required to unfold each segment decreases sequentially indicating that the removal of helices further destabilizes the remaining membrane-embedded segments. By attaching the pulling tag at different positions, the relative stability between helix pairs can be measured and a detailed energy landscape of helix-helix interactions in the unfolding pathway could be provided⁶³. However, the mode of unfolding orthogonally to the membrane is not a biologically relevant process. The pulled unfolded state is extended polypeptide chain without secondary structure. Thus, it cannot distinguish the second stage of folding (*i.e.*, folding) from the first one (*i.e.*, insertion). Besides, this is not a reversible process so unfolding free energy cannot be directly obtained. In a recent study from Perkin's group, they used optimized ultrashort cantilevers to improve spatiotemporal resolution⁶⁴. They revealed a highly detailed view of bR unfolding with many intermediates with differences of only a few amino acids, resolving small changes in the molecular conformations. Notably, they claimed that the mechanical unfolding of bR at standard stretching rates was at

equilibrium. The refolding was widespread but masked by experimental limitations when using conventional cantilevers. They suggested that in the previous single molecule force spectroscopy experiments, the elements of bR secondary structure likely unfolded and refolded but did so faster than the force probe could respond. Single-molecule force spectroscopy still relies upon pulling membrane proteins out of the bilayer, preventing the study of the second folding stage within the membrane sheet. For AFM pulling experiments, a part of the protein has to be tailored to the surface, so only refolding of all but one or two transmembrane segments can be observed.

In contrast to AFM, optical and magnetic tweezers can pull membrane proteins along the membrane plane. In the Bowie lab, the force spectroscopy using magnetic tweezers was employed to study unfolding of GlpG in bicelles⁶⁵. The magnetic tweezers exert force on a bead that is tethered to a surface via DNA handles and a single GlpG protein molecule reconstituted in a bicelle. Attachment of the DNA handles at the N and C-terminal of the transmembrane domain of GlpG exerts a mechanical unfolding force on GlpG. Cycles of unfolding and refolding in bicelles exhibit a hysteresis in the force-extension curve. The force-extension curve confirmed that GlpG could completely refold at a low force. During the refolding, the extended polypeptide chain formed helical structure first then refold in the lipid environment. The extension distance of the intermediates corresponded to the segment lengths that are roughly pairs of helices. Because magnetic tweezer pulls the protein from both terminals, to verify the terminus at which unfolding starts, destabilizing mutations were made at either end of the protein. By studying the effect of mutants on the unfolding rate from both terminals, they found that GlpG unfolds directionally from the C- to N- terminal under mechanical force. The pulling experiment was repeated for hundreds of times under different forces to obtain the stability data. The probability of unfolding was plotted against the applied force and the unfolding rate at zero tension was obtained by an extrapolation

of the probability curve. Extrapolating the refolded fraction yielded the folding rate. The free energy of unfolding calculated from the unfolding and folding rates at zero force is $6.54 k_B T$, which is fairly similar to the value obtained in our lab using steric trap in DDM. But the unfolding states under strong mechanical forces still remain unknown.

Advanced steric trapping and its applications

So far, the methods for studying membrane protein folding described above achieved the reversible unfolding in detergent micelles or bicelle environment. But the reversible control of membrane protein folding in the native environment— liposome has not been achieved. Steric trapping is a promising method for achieving this because it can be used in a mild, non-disrupting experimental setting. However, the previous steric trapping studies were limited to the proteins with intrinsic unfolding readout such as enzyme activity and absorbance of an intrinsic prosthetic chromophore. We believe that steric trapping method has not attained its full analytical capability.

In this dissertation research, I developed a series of the steric trapping methods for their general application to membrane proteins by synthesizing novel biotin probes processing spectroscopic groups that are sensitized by mSA binding or protein unfolding. The principles were proved using GlpG in DDM. By applying those methods to GlpG, 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, the properties that are difficult to study by other methods. This project will be described in details in Chapter 2.

By combining experimental characterization of GlpG by steric trapping with computational studies including molecular dynamics (MD) simulation, I elucidated the role of packing interaction in the stability and function of GlpG. We suggest that the packing defects are required for the functionally

important movement of the structural elements in GlpG. This project will be described in details in Chapter 3.

Membrane proteins fold under the physical constraints of the quasi-two-dimensional lipid bilayers with defined hydrophobic thickness. Understanding conformation of the denatured state is crucial to define thermodynamic stability and folding mechanisms of proteins, while most studies of membrane proteins focused on the native state. By using the novel spin-labeled biotin derivative conjugated to GlpG, we can measure the inter-spin distance between the two biotinylated sites in the sterically trapped denatured state by electron-electron double resonance spectroscopy in the native lipid bilayer environments. I successfully reconstituted the denatured state ensemble (DSE) in the native lipid bilayers using steric trapping and characterized its degree of expansion and dynamics using DEER spectroscopy and proteolysis by proteinase K. In Chapter 4, I demonstrate that the DSE in lipid bilayers are a largely expanded and dynamic conformational ensemble despite the physical constraints of the lipid bilayer. I conclude that the lipid bilayer is reasonably good at solubilizing the DSE of membrane proteins and this feature of bilayer may help membrane proteins to prevent the formation of collapsed misfolded states.

REFERENCES

REFERENCES

- (1) Dill, K. A.; MacCallum, J. L. The Protein-Folding Problem, 50 Years On. *Science* **2012**, 338, 1042–1046.
- (2) Selkoe, D. J. Folding Proteins in Fatal Ways. *Nature* **2003**, 426, 900–904.
- (3) Kendrew, J. C.; Bodo, G.; Dintzis, H. M.; Parrish, R. G.; Wyckoff, H.; Phillips, D. C. A Three-Dimensional Model of the Myoglobin Molecule Obtained by x-Ray Analysis. *Nature* **1958**, 181, 662–666.
- (4) Levinthal, C. How to Fold Graciously. *Mössbauer Spectrosc. Biol. Syst. Proc.* **1969**, 24, 22–24.
- (5) Bryngelson, J. D.; Nelson, J.; Nicholas, O.; Wolynes, P. G. Funnels , Pathways , and the Energy Landscape of Protein Folding : A Synthesis. *Proteins* **1995**, 21, 167–195.
- (6) Dill, K. A. Dominant Forces in Protein Folding. *Biochemistry* **1990**, 29, 7133–7155.
- (7) Williamson, M. *How Proteins Work*; Garland Science, 2012.
- (8) Wallin, E.; Heijne, G. Von. Genome-Wide Analysis of Integral Membrane Proteins from Eubacterial, Archaeal, and Eukaryotic Organisms. *Protein Sci.* **1998**, 7, 1029–1038.
- (9) Cohen, F. E.; Kelly, J. W. Therapeutic Approaches to Protein- Misfolding Diseases. *Nature* **2003**, 426, 905–909.
- (10) White, S. H. The Progress of Membrane Protein Structure Determination. **2005**, 1948–1949.
- (11) Baker, D. Protein Folding, Structure Prediction and Design. *Biochem. Soc. Trans.* **2014**, 42, 225–229.
- (12) Das, R.; Baker, D. Macromolecular Modeling with Rosetta. *Annu. Rev. Biochem.* **2008**, 77, 363–382.
- (13) Barth, P.; Wallner, B.; Baker, D. Prediction of Membrane Protein Structures with Complex Topologies Using Limited Constraints. *Proc. Natl. Acad. Sci.* **2009**, 106, 1409–1414.
- (14) Barth, P.; Schonbrun, J.; Baker, D. Toward High-Resolution Prediction and Design of Transmembrane Helical Protein Structures. *Proc. Natl. Acad. Sci.* **2007**, 104, 15682–15687.
- (15) Cantor, R. S. The Lateral Pressure Profile in Membranes: A Physical Mechanism of General Anesthesia. *Biochemistry* **1997**, 36, 2339–2344.
- (16) Hong, H. Toward Understanding Driving Forces in Membrane Protein Folding. *Arch. Biochem. Biophys.* **2014**, 564, 297–313.
- (17) Popot, J. L.; Engelman, D. M. Membrane Protein Folding and Oligomerization: The Two-

Stage Model. *Biochemistry* **1990**, 29, 4031–4037.

- (18) Huang, K. S.; Bayley, H.; Liao, M. J.; London, E.; Khorana, H. G. Refolding of an Integral Membrane Protein. Denaturation, Renaturation, and Reconstitution of Intact Bacteriorhodopsin and Two Proteolytic Fragments. *J. Biol. Chem.* **1981**, 256, 3802–3809.
- (19) Engelman, D. M.; Chen, Y.; Chin, C. N.; Curran, A. R.; Dixon, A. M.; Dupuy, A. D.; Lee, A. S.; Lehnert, U.; Matthews, E. E.; Reshetnyak, Y. K.; Senes, A.; Popot, J. L. Membrane Protein Folding: Beyond the Two Stage Model. *FEBS Lett.* **2003**, 555, 122–125.
- (20) Hessa, T.; Kim, H.; Bihlamaier, K.; Lundin, C.; Boekel, J.; Andersson, H.; Nilsson, I.; White, S.; Von, G. Recognition of Transmembrane Helices by the Endoplasmic Reticulum Translocon. *Nature* **2005**, 433, 377–381.
- (21) Cymer, F.; Heijne, G. Von. Cotranslational Folding of Membrane Proteins Probed by Arrest-Peptide – Mediated Force Measurements. *Proc Natl Acad Sci U S A* **2013**, 110, 14640–14645.
- (22) Snider, C.; Jayasinghe, S.; Hristova, K.; White, S. H. MPEx: A Tool for Exploring Membrane Proteins. *Protein Sci.* **2009**, 18, 2624–2628.
- (23) Wimley, W. C.; Creamer, T. P.; White, S. H. Solvation Energies of Amino Acid Side Chains and Backbone in a Family of Host–Guest Pentapeptides [†]. *Biochemistry* **1996**, 35, 5109–5124.
- (24) Hessa, T.; Meindl-Beinker, N. M.; Bernsel, A.; Kim, H.; Sato, Y.; Lerch-Bader, M.; Nilsson, I.; White, S. H.; Von Heijne, G. Molecular Code for Transmembrane-Helix Recognition by the Sec61 Translocon. *Nature* **2007**, 450, 1026–1030.
- (25) Lin, M.; Gessmann, D.; Naveed, H.; Liang, J. Outer Membrane Protein Folding and Topology from a Computational Transfer Free Energy Scale. *J. Am. Chem. Soc.* **2016**, 138, 2592–2601.
- (26) Stevens, T. J.; Arkin, I. T. Turning an Opinion Inside-Out : Rees and Eisenberg ’ s Commentary (Proteins 2000 ; 38 : 121 – 122) on “ Are Membrane Proteins ‘ Inside-Out ’ Proteins ? ” (Proteins 1999 ; 36 : 135 – 143). *Proteins* **2000**, 40, 463–464.
- (27) Adamian, L.; Nanda, V.; DeGrado, W. F.; Liang, J. Empirical Lipid Propensities of Amino Acid Residues in Multispan Alpha Helical Membrane Proteins. *Proteins Struct. Funct. Genet.* **2005**, 59, 496–509.
- (28) Fleming, K. G.; Ackerman, A. L.; Engelman, D. M. The Effect of Point Mutations on the Free Energy of Transmembrane α -Helix Dimerization. *J. Mol. Biol.* **1997**, 272, 266–275.
- (29) Fleming, K. G.; Engelman, D. M. Specificity in Transmembrane Helix-Helix Interactions Can Define a Hierarchy of Stability for Sequence Variants. *Proc. Natl. Acad. Sci.* **2001**, 98, 14340–14344.
- (30) Hong, H.; Blois, T. M.; Cao, Z.; Bowie, J. U. Method to Measure Strong Protein-Protein Interactions in Lipid Bilayers Using a Steric Trap. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, 107, 19802–19807.

- (31) Liang, J.; Dill, K. A. Are Proteins Well-Packed? *Biophys. J.* **2001**, *81*, 751–766.
- (32) Kellis, J. T.; Nyberg, K.; Fersht, A. R. Energetics of Complementary Side-Chain Packing in a Protein Hydrophobic Core. *Biochemistry* **1989**, *28*, 4914–4922.
- (33) Eriksson, A.; Baase, W.; Zhang, X.; Heinz, D.; Blaber, M.; Baldwin, E.; Matthews, B. Response of a Protein Structure to Cavity-Creating Mutations and Its Relation to the Hydrophobic Effect. *Science* (80-.). **1992**, *255*, 178–183.
- (34) Reynolds, J. A.; Gilbert, D. B.; Tanford, C. Empirical Correlation Between Hydrophobic Free Energy and Aqueous Cavity Surface Area. *Proc. Natl. Acad. Sci.* **1974**, *71*, 2925–2927.
- (35) Joh, N. H.; Oberai, A.; Yang, D.; Whitelegge, J. P.; Bowie, J. U. Similar Energetic Contributions of Packing in the Core of Membrane and Water-Soluble Proteins. *J. Am. Chem. Soc.* **2009**, *131*, 10846–10847.
- (36) Eilers, M.; Shekar, S. C.; Shieh, T.; Smith, S. O.; Fleming, P. J. Internal Packing of Helical Membrane Proteins. *Proc. Natl. Acad. Sci. U. S. A.* **2000**, *97*, 5796–5801.
- (37) Ben-Tal, N.; Sitkoff, D.; Topol, I. A.; Yang, A.-S.; Burt, S. K.; Honig, B. Free Energy of Amide Hydrogen Bond Formation in Vacuum, in Water, and in Liquid Alkane Solution. *J. Phys. Chem. B* **1997**, *101*, 450–457.
- (38) Zhou, F. X.; Cocco, M. J.; Russ, W. P.; Brunger, A. T.; Engelman, D. M. Interhelical Hydrogen Bonding Drives Strong Interactions in Membrane Proteins. *Nat. Struct. Biol.* **2000**, *7*, 154–160.
- (39) Choma, C.; Gratkowski, H.; Lear, J. D.; DeGrado, W. F. Asparagine-Mediated Self-Association of a Model Transmembrane Helix. *Nat. Struct. Biol.* **2000**, *7*, 161–166.
- (40) Faham, S.; Yang, D.; Bare, E.; Yohannan, S.; Whitelegge, J. P.; Bowie, J. U. Side-Chain Contributions to Membrane Protein Structure and Stability. *J. Mol. Biol.* **2004**, *335*, 297–305.
- (41) Joh, N. H. J.; Min, A.; Faham, S.; Whitelegge, J. P.; Yang, D.; Woods, V. L.; Bowie, J. U. Modest Stabilization by Most Hydrogen-Bonded Side-Chain Interactions in Membrane Proteins. *Nature* **2008**, *453*, 1266–1270.
- (42) Baker, R. P.; Urban, S. Architectural and Thermodynamic Principles Underlying Intramembrane Protease Function. *Nat. Chem. Biol.* **2012**, *8*, 759–768.
- (43) Bowie, J. U. Membrane Protein Folding: How Important Are Hydrogen Bonds? *Curr. Opin. Struct. Biol.* **2011**, *21*, 42–49.
- (44) Waldburger, C. D.; Schildbach, J. F.; Sauer, R. T. Are Buried Salt Bridges Important for Protein Stability and Conformational Specificity? *Nat. Struct. Biol.* **1995**, *2*, 122–128.
- (45) Hong, H.; Szabo, G.; Tamm, L. K. Electrostatic Couplings in OmpA Ion-Channel Gating Suggest a Mechanism for Pore Opening. *Nat. Chem. Biol.* **2006**, *2*, 627–635.
- (46) Anfinsen, C. B. Principles That Govern the Folding of Protein Chains. *Science* (80-.). **1973**, *181*, 223–230.

- (47) Hong, H.; Joh, N. H.; Bowie, J. U.; Tamm, L. K. *Chapter 8 Methods for Measuring the Thermodynamic Stability of Membrane Proteins*; 1st ed.; Elsevier Inc., 2009; Vol. 455.
- (48) Lau, F. W.; Bowie, J. U. A Method for Assessing the Stability of a Membrane Protein. *Biochemistry* **1997**, *36*, 5884–5892.
- (49) Curnow, P.; Booth, P. J. The Transition State for Integral Membrane Protein Folding. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 773–778.
- (50) Otzen, D. E. Folding of DsbB in Mixed Micelles: A Kinetic Analysis of the Stability of a Bacterial Membrane Protein. *J. Mol. Biol.* **2003**, *330*, 641–649.
- (51) Baker, R. P.; Urban, S. Architectural and Thermodynamic Principles Underlying Intramembrane Protease Function. *Nat. Chem. Biol.* **2012**, *8*, 759–768.
- (52) Schleich, J. P.; Peng, D.; Kroncke, B. M.; Mittendorf, K. F.; Narayan, M.; Carter, B. D.; Sanders, C. R. Reversible Folding of Human Peripheral Myelin Protein 22, a Tetraspan Membrane Protein. *Biochemistry* **2013**, *52*, 3229–3241.
- (53) Dutta, A.; Kim, T. Y.; Moeller, M.; Wu, J.; Alexiev, U.; Klein-Seetharaman, J. Characterization of Membrane Protein Non-Native States. 2. the SDS-Unfolded States of Rhodopsin. *Biochemistry* **2010**, *49*, 6329–6340.
- (54) Chang, Y.-C.; Bowie, J. U. Measuring Membrane Protein Stability under Native Conditions. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 219–224.
- (55) Guo, R.; Gaffney, K.; Yang, Z.; Kim, M.; Sungsuwan, S.; Huang, X.; Hubbell, W. L.; Hong, H. Steric Trapping Reveals a Cooperativity Network in the Intramembrane Protease GlpG. *Nat. Chem. Biol.* **2016**, *12*, 353–360.
- (56) Tamm, L. K.; Hong, H.; Liang, B. Folding and Assembly of β -Barrel Membrane Proteins. *Biochim. Biophys. Acta - Biomembr.* **2004**, *1666*, 250–263.
- (57) Di Bartolo, N.; Compton, E. L. R.; Warne, T.; Edwards, P. C.; Tate, C. G.; Schertler, G. F. X.; Booth, P. J. Complete Reversible Refolding of a G-Protein Coupled Receptor on a Solid Support. *PLoS One* **2016**, *11*, 1–16.
- (58) Findlay, H. E.; Rutherford, N. G.; Henderson, P. J. F.; Booth, P. J. Unfolding Free Energy of a Two-Domain Transmembrane Sugar Transport Protein. *Proc. Natl. Acad. Sci.* **2010**, *107*, 18451–18456.
- (59) Harris, N. J.; Findlay, H. E.; Simms, J.; Liu, X.; Booth, P. J. Relative Domain Folding and Stability of a Membrane Transport Protein. *J. Mol. Biol.* **2014**, *426*, 1812–1825.
- (60) Blois, T. M.; Hong, H.; Kim, T. H.; Bowie, J. U. Protein Unfolding with a Steric Trap. *J. Am. Chem. Soc.* **2009**, *131*, 13914–13915.
- (61) Hong, H.; Bowie, J. U. Dramatic Destabilization of Transmembrane Helix Interactions by Features of Natural Membrane Environments. *J. Am. Chem. Soc.* **2011**, *133*, 11389–11398.
- (62) Jefferson, R. E.; Blois, T. M.; Bowie, J. U. Membrane Proteins Can Have High Kinetic Stability. *J. Am. Chem. Soc.* **2013**, *135*, 15183–15190.

- (63) Oesterhelt, F.; Oesterhelt, D.; Pfeiffer, M.; Engel, A.; Gaub, H. E.; Müller, D. J. Unfolding Pathways of Individual Bacteriorhodopsins. *Science* (80-.). **2000**, *288*, 143–146.
- (64) Yu, H.; Siewny, M. G. W.; Edwards, D. T.; Sanders, A. W.; Perkins, T. T. Hidden Dynamics in the Unfolding of Individual Bacteriorhodopsin Proteins. *Science* (80-.). **2017**, *355*, 945–950.
- (65) Min, D.; Jefferson, R. E.; Bowie, J. U.; Yoon, T.-Y. Mapping the Energy Landscape for Second-Stage Folding of a Single Membrane Protein. *Nat. Chem. Biol.* **2015**, *11*, 981–987.

Chapter 2 General steric trapping strategy reveals an intricate cooperativity network in the intramembrane protease GlpG under native conditions

Ruiqiong Guo, Kristen Gaffney, Zhongyu Yang, Miyeon Kim, Suttipun Sungsuwan, Xuefei Huang, Wayne L Hubbell & Heedeok Hong

The content of this chapter was published in Nat. Chem. Biol. 2016, 12, 353–360.

*The energy landscape part in this chapter was published in Yang, Y.; Guo, R.; Gaffney, K.; Kim, M.; Muhammednazaar, S.; Tian, W.; Wang, B.; Liang, J.; Hong, H. J. Am. Chem. Soc. **2018**, 140, 4656–4665.*

I worked with Kristen Gaffney to publish the *Nat. Chem. Biol.* article for which we were listed as co-1st authors. Kristen optimized the GlpG expression and labeling sites and performed the experiments in **Figure 2.2**, **Figure 2.6c** and the determination of K_d values for the mSA mutants. Kristen also helped with the experiments in **Figure 2.7** and **Figure 2.8b**. I performed the experiment for all the rest figures in this chapter, with the help from Dr. Miyeon Kim. I also perform the probe design and synthesis under the help from Dr. Suttipun Sungsuwan and Professor Xuefei Huang. The DEER measurement was done by Dr. Zhongyu Yang and Professor Wayne L Hubbell.

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 a set of 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 those methods to an intramembrane protease GlpG, 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.

Introduction

Understanding the free energy landscape of protein folding requires determination of the free energy levels of states in equilibrium with the native folded state as well as analysis of the energy barriers to reach the native conformation¹. This task has been mainly carried out by equilibrium and kinetic folding studies using denaturants that can readily shift the population distribution between the folded and unfolded state². However, the overall shape of the folding energy landscape substantially changes in the presence of denaturant, and certain short-lived higher energy states may not be detected in denaturing conditions^{3,4}. 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 (HDX), NMR relaxation dispersion and

proteolysis have revealed the dynamic and multi-state nature of the native conformational ensemble⁴⁻⁸, which is crucial for protein function⁹⁻¹¹. 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^{12,13}.

Steric trapping is a promising tool for investigating membrane protein folding directly under native conditions. It couples unfolding of a target protein with two biotin tags to competitive binding of bulky monovalent streptavidin molecules (mSA, MW=52 kD)¹⁴⁻¹⁸ (**Figure 2.1a**). 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 either biotin label with intrinsic binding affinity ($\Delta G^{\circ}_{\text{Bind}}$), but due to steric hindrance, the second mSA binds only when the native tertiary contacts are unraveled by transient unfolding. This unfolding-binding coupling weakens the apparent affinity of the second mSA relative to that of the first mSA depending on the protein stability. Thus, thermodynamic stability (ΔG°_U) of the target protein can be determined by measuring the degree of attenuation of the second mSA binding.

Although promising, it is yet difficult to apply this method to various types of membrane proteins. Steric trapping requires two essential features: two site-specifically conjugated biotin labels on a target protein and a probe to monitor binding of mSA or protein unfolding. Site-specific biotinylation has been achieved by labeling of engineered cysteine residues with thiol-reactive biotin derivatives^{14,15,18}. For detection of unfolding, widely-used tools such as tryptophan fluorescence and circular dichroism cannot be used due to large signal interferences from mSA molecules. A method for direct detection of mSA binding has not been developed yet. Thus, the

application has been limited to proteins possessing convenient unfolding readouts such as conformation-sensitive retinal absorption of bacteriorhodopsin (bR) and enzymatic activities of diacylglycerol kinase (DGK) and dihydrofolate reductase (DHFR)^{14,15,18}. Therefore, to further advance the method for more general application, we have developed a set of new steric trapping strategies by utilizing novel thiol-reactive biotin probes containing spectroscopic reporter groups for sensitive detection of mSA binding and protein unfolding. These new strategies have been applied to analyze the thermodynamic stability, compactness of the unfolded state ensemble and folding cooperativity of a six-helical bundle intramembrane protease GlpG from *E. coli*.

GlpG is a member of the rhomboid protease family widely conserved in all kingdoms of life. Rhomboid proteases play a key role in diverse biological processes by activating membrane-bound signaling proteins or enzymes via cleavage of a specific peptide bond near the membrane¹⁹⁻²². Due to the functional importance of rhomboid proteases and detailed structural information available (28 PDB entries), GlpG has emerged as an important model for studying the folding of helical membrane proteins. Critical regions for the stability of *E. coli* GlpG have been identified using heat and SDS denaturation tests of 151 variants²³. A kinetic folding study using SDS as a denaturant has suggested a compact folding nucleus and multiple frustrated loops in the folding transition state²⁴. A more recent single-molecule pulling study has shown that, at a constant tension, GlpG largely unfolds cooperatively in ~60% of total forced-unfolding trajectories while a minor but significant portion (~40%) unfolds via one or two intermediate steps²⁵. Here, using steric trapping, we provide new insights into the folding energy landscape of GlpG in the absence of perturbants including heat, chemical denaturant or mechanical force. We have elucidated a largely 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.

Results

Design and synthesis of new steric-trapping probes

Our steric-trapping probes are highlighted by three key features that are integrated into one molecular tag (**Figure 2.1b**): (1) a biotin group for binding of mSA, (2) a thiol-reactive group for conjugation to engineered cysteine residues on a target protein, and (3) a fluorescent or paramagnetic reporter group whose spectroscopic signal is sensitized by mSA binding or protein unfolding. Each tripartite probe was synthesized by stepwise nucleophilic substitutions of building blocks possessing characteristic features to a lysine or cysteine template. The reporter groups employed in this study possessed versatile utilities. BtnPyr-IA (**Figure 2.1b, left**) is a pyrene-based fluorescent sensor to detect binding of mSA. When used to doubly label GlpG, pyrene (donor) fluorescence becomes remarkably sensitive to binding of quencher (acceptor)-labeled mSA by Förster resonance energy transfer (FRET). BtnRG-TP (**Figure 2.1b, right**) is a paramagnetic sensor to detect protein unfolding. When used to doubly label GlpG, the spin labels allow distance measurements in the native and steric-trapped unfolded state using double electron-electron resonance spectroscopy (DEER).

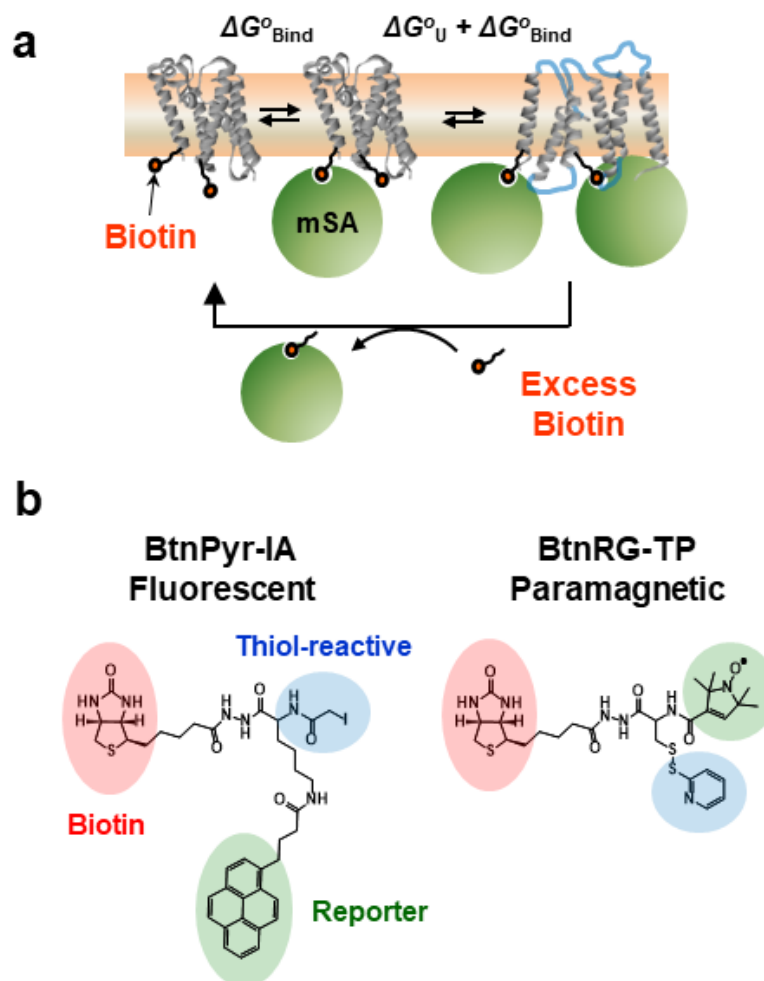


Figure 2.1 Principle of steric trapping and steric-trapping probes developed in this study. (a) Steric trapping principle for measuring thermodynamic stability (ΔG°_U) of proteins labeled with two biotin tags. First monovalent streptavidin (mSA) binds unhindered to either biotin label. Binding of the second mSA to the folded state is inhibited due to the steric hindrance with pre-bound mSA, but occurs when the protein is transiently unfolded. Coupling of mSA binding to unfolding leads to attenuation of the apparent binding affinity of the second mSA, whose degree is correlated with the stability. ΔG°_U is obtained by fitting of the second binding phase (**equation (3) in Materials and Methods**). Folding reversibility is tested upon addition of excess free biotin by which bound mSA molecules are released by competition. **(b)** Thiol-reactive biotin derivatives possessing a spectroscopic reporter group developed in this study. BtnPyr-IA: biotin (red shaded)-pyrene (green shaded)-iodoacetamide (blue shaded) conjugated to a lysine template, and BtnRG-TP: biotin (red shaded)-1-oxyl-2,2,5,5-tetramethylpyrroline spin label (green shaded)-thiopyridine (blue shaded) conjugated to a cysteine template.

Development of high-throughput activity assay for GlpG

To prove the principle of our steric trapping strategies employing the new probes, we used the intramembrane protease GlpG as a model and its proteolytic activity as a folding indicator. As a model substrate, the second transmembrane (TM) segment of the lactose permease of *E. coli* fused to water-soluble staphylococcal nuclease (SN-LacYTM2) was chosen due to its efficient cleavage by GlpG²⁶. So far, the activity of GlpG in detergents has been measured by quantifying the band intensity of the cleavage product of TM substrates on SDS-PAGE²³. We found it difficult to obtain a precise initial cleavage rate using this method. Thus, for more precise and efficient activity measurements, we developed a robust fluorescence-based assay that can be transformed into a high-throughput format. The environment-sensitive fluorophore nitrobenzodiazole (NBD) was conjugated to an engineered cysteine residue located upstream of the scissile bond (**Figure 2.2a**) in LacYTM2. Cleavage of SN-LacYTM2 by GlpG induced the release of the NBD-labeled water-soluble portion from micellar to aqueous phase, accompanied by a large decrease of NBD fluorescence. Time-dependent monitoring of fluorescence yielded a substrate half-life which agreed well with the SDS-PAGE assay (**Figure 2.2b**).

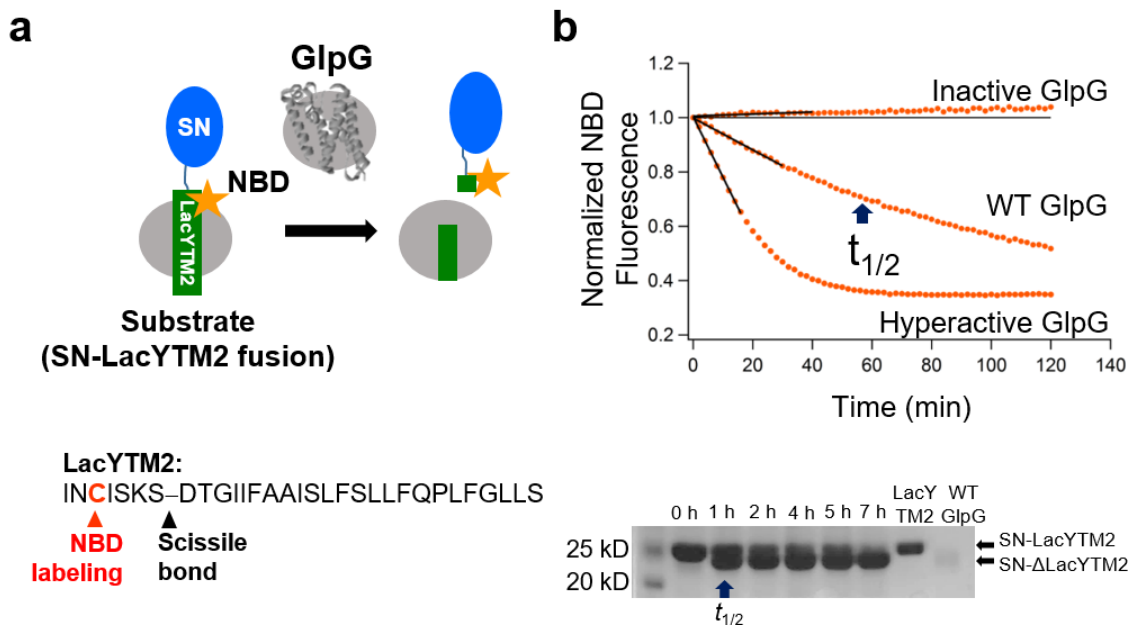


Figure 2.2 New high-throughput assay for measuring the proteolytic activity of GlpG. A new high-throughput activity assay allows for precise measurements of GlpG activity in a 96-well format. **(a)** Second transmembrane segment of the lactose permease of *E. coli* fused to staphylococcal nuclease domain (SN-LacYTM2). IA-NBD, a thiol-reactive environment-sensitive fluorophore is conjugated to an engineered cysteine in the P5 position from the scissile bond. Cleavage of LacYTM2 leads to a large decrease in the fluorescence intensity as NBD is transferred from the nonpolar micellar phase into the bulk aqueous phase. **(b) (Top)** Changes in the NBD fluorescence over time due to the proteolytic activity of GlpG. Addition of wild-type (WT) GlpG displays decreasing NBD fluorescence. In contrast, addition of inactive GlpG variant (S201T) displays negligible change in NBD fluorescence, and hyperactive GlpG variant (W236A) increases the rate of NBD fluorescence change relative to WT. **(Bottom)** In the conventional SDS-PAGE assay for GlpG activity, a lower molecular weight band appears, which corresponds to cleaved SN-LacYTM2 (SN-ΔLacYTM2). For WT GlpG, the half-life ($t_{1/2}$) of the substrate estimated by SDS-PAGE is similar to that measured by the NBD fluorescence change.

Steric trapping controls reversible folding of GlpG

Here all studies were performed in dodecylmaltoside (DDM) detergent, in which a majority of functional and folding studies of GlpG were carried out^{23,27-30}, and with the isolated TM domain (residues 87-276) for which all 28 structures of GlpG were solved. The activity of the TM domain tested with the substrate SN-LacYTM2 was indistinguishable from that of the full-length protein.

For steric trapping, we first identified optimal residue pairs for cysteine substitution to conjugate thiol-reactive biotin labels using the activity assay. After testing of multiple single- and combined double-cysteine variants, two double-cysteine variants, P95C/G172C and G172C/V267C were selected (**Figure 2.3a**). Individual single-cysteine variants P95C, G172C and V267C labeled with BtnPyr-IA maintained the wild-type activity level (**Figure 2.3b, top**) and this activity level was not significantly altered after binding of wild-type mSA (mSA-WT) to each biotin label. 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 led to a substantial loss of activity implying GlpG was trapped in the unfolded state (**Figure 2.3b, bottom**).

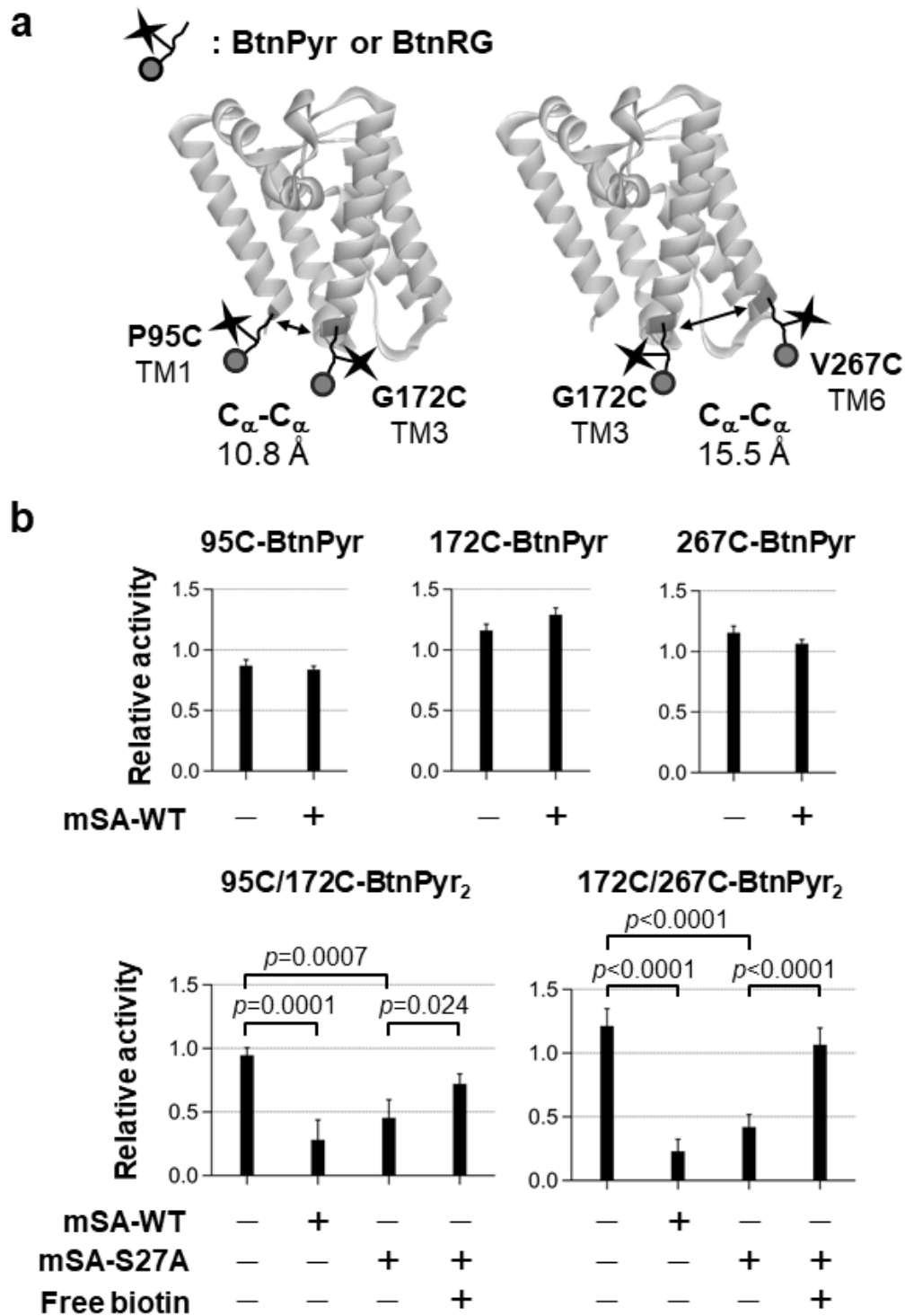


Figure 2.3 GlpG reversibly unfolds by double-binding of mSA. (a) Locations of two different biotin pairs for steric trapping in the structure of GlpG (PDB code: 3B45) and their C_{α} - C_{α} distances. **(b)** Proof of steric trapping principle using the activity of GlpG as a folding indicator. All activity levels were normalized relative to the activity of wild-type GlpG. Error bars denote mean $\pm$ s. d.

from 3-5 independent measurements. (*Top*) binding of wild-type mSA (mSA-WT, 20 μ M) to individual single-cysteine variant labeled with BtnPyr did not affect the activity. (*Bottom*) saturated binding of mSA-WT to each double-cysteine variant labeled with BtnPyr led to an inactivation of GlpG. Folding reversibility was tested using the following steps: Labeled GlpG variants were first inactivated with mSA-S27A variant (20 μ M) possessing a weaker biotin binding affinity ($K_{d, \text{biotin}}=1.4$ nM) for 2~5 days. 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. All *p*-values obtained from Student's *t*-test were smaller than the threshold significance level ($\alpha = 0.05$), indicating that the activity changes for the unfolding and refolding reactions were significant.

Next, we tested if the steric-trapped unfolded state can reversibly refold to the native state after dissociation of bound mSA. Wild-type mSA (mSA-WT) binds biotin with an enormously high affinity ($K_{d, \text{biotin}} \approx 10^{-14}$ M) and slow dissociation rate ($k_{\text{off}} \approx \text{days}$)³¹. Thus, to test the reversibility, we used an mSA-S27A variant with a weaker biotin affinity ($K_{d, \text{biotin}}=1.4 \times 10^{-9}$ M) to facilitate dissociation of bound mSA after addition of excess free biotin (**Figure 2.1a**)³². Double-biotin variants were first inactivated with excess mSA-S27A until the activity reached ~40% of the native state (**Figure 2.3b**, bottom). Dissociation of bound mSA induced by addition of excess free biotin led to the reactivation of both variants. For 95C/172C-BtnPyr₂, 50~70% of activity was regained while >90% was regained for 172C/267C-BtnPyr₂. This result demonstrates that steric trapping can control the reversible folding of GlpG simply by addition of mSA for unfolding and by addition of free biotin for refolding without using denaturants.

Steric-trapped unfolded state is widely unraveled

So far, protein unfolding by steric trapping has been tested by the loss of enzymatic activity (DHFR¹⁴ and DGK¹⁸), decrease of retinal absorbance (bR¹⁵), or increased susceptibility to proteolysis (DHFR¹⁴ and bR¹⁵). Although those features indicate unfolding, a possibility remains that the protein conformation with two bound mSA molecules is only locally distorted or compact with significant residual tertiary interactions. Therefore, to elucidate the physical dimensions of

the unfolded state induced by steric trapping as well as to gain insights into the unfolded state ensemble of membrane proteins under native conditions, we used a thiol-reactive biotin derivative possessing a 1-oxy-2,2,5,5-tetramethylpyrroline spin label (BtnRG-TP) (**Figure 2.1b**, right). By labeling of double-cysteine variants of GlpG with this probe, we have the advantages of both trapping the unfolded state and measuring distances between the spin labels using DEER. DEER allows measurements of long-range inter-spin distances (typically 15~60 Å)³³. An important feature of DEER is that it provides not only the most probable distance but also the distance distribution, which is of great interest in characterization of the unfolded state of proteins. This technique was used to characterize the SDS-induced unfolded states of bR and light harvesting complex II^{34,35}. Here, we obtained inter-spin distances for 95C/172C-BtnRG₂ and 172C/267C-BtnRG₂ of GlpG in their native state, SDS-induced unfolded state and steric-trapped unfolded state (**Figure 2.4**).

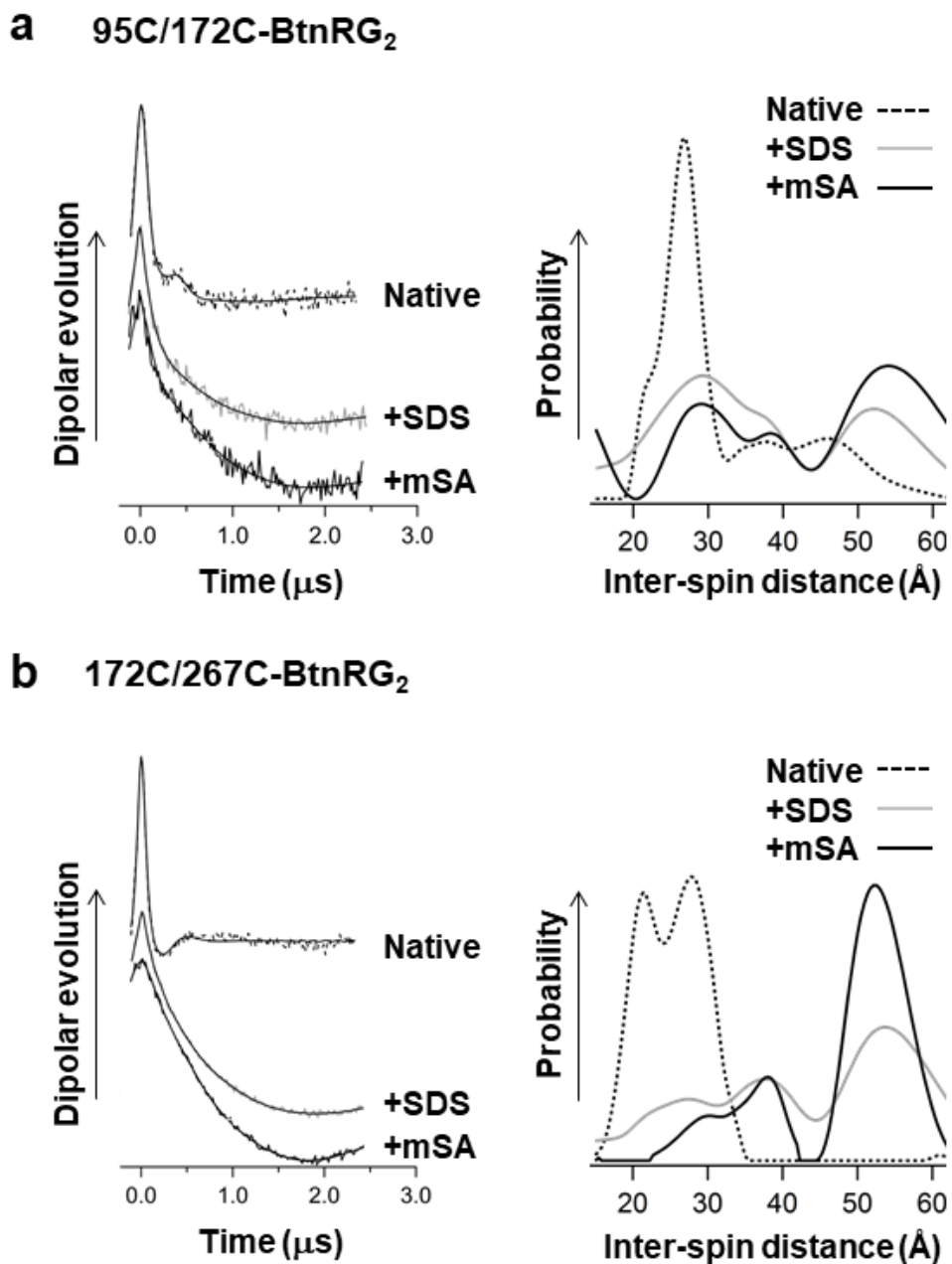


Figure 2.4 DEER suggests steric trapping induce wide separation of two biotinylated sites. Background-subtracted dipolar evolution data and their fits (*left*) and inter-spin distances (*right*) for the native (dashed lines), SDS-unfolded (gray solid lines) and steric-trapped (black solid lines) for (a) 95C/172C-BtnRG₂ GlpG and (b) 172C/267C-BtnRG₂ GlpG. The maximum dipolar evolution times were 2.3~2.5 μs and the approximate upper limit of the reliable mean distance was ~53 Å³³.

For both variants, SDS (SDS mole fraction = $[\text{SDS}]/([\text{DDM}]+[\text{SDS}]) > 0.8$, in which the unfolded fraction exceeded 0.9. **Figure 2.5c**) induced substantial broadening of the inter-spin distance distribution over the range from the native-like distances (15-35 Å) up to ~60 Å (**Figure 2.4**, right panels). The result indicates a heterogeneous conformational ensemble of the unfolded state in SDS. Interestingly, in non-denaturing micellar solution, the unfolded state induced by steric trapping also exhibited a broad inter-spin distance distribution, but the longer-distance components (45~60 Å) were even more populated relative to those in SDS (solid gray and black lines in **Figure 2.4**, right). The spin-labeled biotin pairs 95C/172C-BtnRG₂ and 172C/267C-BtnRG₂ cover approximately the N-terminal and C-terminal halves of the polypeptide chain, respectively. The increase of the most probable inter-spin distance from ~25 Å in the native state to ~55 Å in the steric-trapped unfolded state corresponds to a ~30 Å expansion of each half. Considering this increased dimension is comparable to the whole diameter of native GlpG, our DEER data for GlpG rules out a compact unfolded conformational ensemble under a non-denaturing condition, which has been observed for several water-soluble proteins³⁸. It should be noted that, because of the detection limit of DEER (15~60 Å), even longer distance components (>60 Å) may actually exist but have not been detected. Addition of dithiothreitol (DTT) to break the disulfide bond between GlpG and the mSA-bound BtnRG label led to a significant regain of the activity (>70%) indicating that a majority of the steric-trapped unfolded conformations were able to refold (**Figure 2.5**).

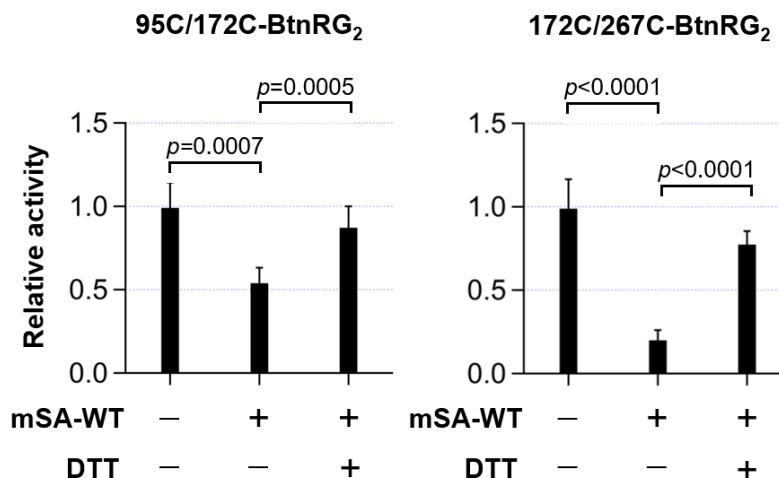


Figure 2.5 Characterization of activity and reversibility of GlpG labeled with paramagnetic BtnRG for steric trapping and DEER measurements. Saturated binding of mSA-WT to each double-cysteine variant labeled with BtnRG led to an inactivation of GlpG. Folding reversibility was tested by addition of 4 mM of dithiothreitol (DTT) followed by incubation for 4 h at room temperature. DTT induced the cleavage of the disulfide linkage between cysteine on the protein and BtnRG label with bound mSA, which then released the steric restraints by two mSA molecules. The activity levels were normalized relative to that of wild-type GlpG. Error bars denote mean $\pm$ s. d. from 3~5 independent measurements. All p -values from Student's t -test were lower than 0.05, the threshold confidence level, indicating that the results for unfolding/refolding reactions were statistically significant.

Our DEER and proteolysis results demonstrate that steric trapping induced a true unfolded state, which can be described as an ensemble of largely-expanded dynamic and heterogeneous conformations. This work also represents the first structural characterization of the unfolded state of a helical membrane protein under non-denaturing conditions.

Thermodynamic 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 (**Figure 2.1b**) should sensitively change upon either mSA binding or protein unfolding. Here we have achieved a highly sensitive detection of mSA binding by employing FRET between pyrene on BtnPyr label

as a donor and non-fluorescent chromophore DABCYL specifically labeled on mSA (mSA_{DAB}) as an acceptor (**Figure 2.6a**). Saturated binding of mSA_{DAB} to double-biotin GlpG variants led to a large decrease of donor fluorescence by ~75%. Nonspecific FRET was not significant up to 80 μ M mSA_{DAB} concentration (**Figure 2.6b**).

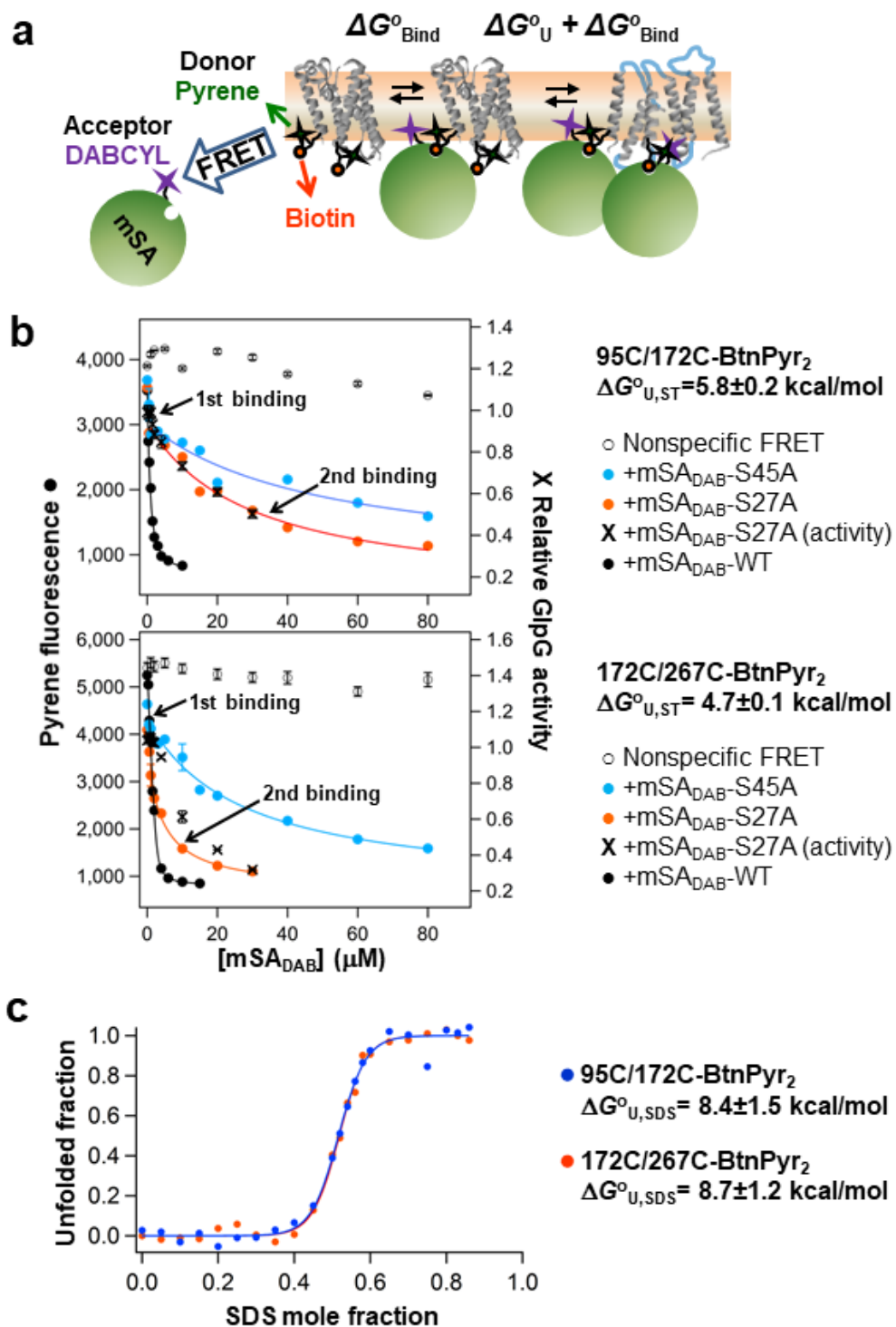


Figure 2.6 Thermodynamic stability of GlpG using steric trapping and SDS denaturation. (a) Steric trapping strategy using FRET between fluorescent BtnPyr doubly-labeled on GlpG (donor) and non-fluorescent quencher DABCYL (acceptor) labeled near the biotin binding pocket (Y83C)

of the active subunit of mSA. **(b)** Binding isotherms of 95C/172C-BtnPyr₂ and 172C/267C-BtnPyr₂ (1 μ M) with three mSA variants mSA_{DAB}-WT (black circles, $K_{d, \text{biotin}} \approx 10^{-14}$ M), mSA_{DAB}-S27A (red circles, $K_{d, \text{biotin}} = 1.4 \times 10^{-9}$ M) and mSA_{DAB}-S45A (blue circles, $K_{d, \text{biotin}} = 9.0 \times 10^{-9}$ M) obtained by FRET. At an increasing concentration of mSA_{DAB}-S27A, the activity change (crosses) agreed well with the fluorescence change in the second binding phase implying coupling of unfolding to mSA binding. The thermodynamic stability ($\Delta G^{\circ}_{U,ST}$) of each variant was obtained by fitting of the second mSA-binding curves to **equation (3)** in **Materials and Methods**. Nonspecific FRET (open circles) corresponds the fluorescence intensity of double-biotin GlpG variants which were pre-saturated with 10 μ M of the high-affinity mSA-WT (without DABCYL-label) at an increasing concentration of the lower-affinity variant mSA_{DAB}-S45A. Thus, mSA_{DAB}-S45A cannot compete for biotin label and only diffuses around in the solution. Errors in fluorescence denote mean $\pm$ s. d. from 4 independent measurements. Errors in activity denote $\pm$ s. d. from fitting. Errors in $\Delta G^{\circ}_{U,ST}$ values denote mean $\pm$ s. d. from 3 independent measurements. **(c)** Unfolding of GlpG variants 95C/172C-BtnPyr₂ and 172C/267C-BtnPyr₂ (0.4 μ M) as a function of SDS mole fraction measured by FRET between 11 Trp residues of GlpG and pyrene groups on BtnPyr labels. The stabilities ($\Delta G^{\circ}_{U,SDS}$) of the two variants extrapolated to zero-SDS mole fraction are similar to each other. Errors in $\Delta G^{\circ}_{U,SDS}$ values denote mean $\pm$ s. d. from fitting.

By design, steric trapping specifically captures transient unfolding of the tertiary interactions between two biotin labels. Thus, probing the stability with biotin pairs located in different regions (95C/172C-BtnPyr₂ and 172C/267C-BtnPyr₂, **Figure 2.3a**) provides a novel opportunity for testing the unfolding cooperativity of GlpG. To ensure the native tertiary interactions of GlpG were equally preserved in both double-biotin variants, we measured their stabilities by SDS denaturation (**Figure 2.6c**). Linear extrapolation of the denaturation data to zero-SDS mole fraction yielded the same thermodynamic stability ($\Delta G^{\circ}_{U,SDS}$) for 95C/172C-BtnPyr₂ (8.4 ± 1.5 kcal/mol) and 172C/267C-BtnPyr₂ (8.7 ± 1.2 kcal/mol), which was comparable to that of the full-length wild type GlpG measured by SDS denaturation and tryptophan fluorescence ($\Delta G^{\circ}_{U,SDS} = 7.4$ kcal/mol)²⁴.

Figure 2.6b shows binding isotherms obtained by FRET between each double-biotin variant and three DABCYL-labeled mSA variants with different biotin affinities. An essential element of steric trapping for determination of protein stability is choosing an mSA variant whose binding with the

biotin label ($\Delta G^{\circ}_{\text{Bind}}$) optimally competes with folding ($\Delta G^{\circ}_{\text{U}}$) to yield an attenuated second binding in a desired [mSA] range (**Figure 2.6a**). Among mSA variants tested, mSA_{DAB}-S27A yielded optimal binding isotherms, in which the first tight and second weaker binding phases were clearly separated (**Figure 2.6b**). The parallel activity measurement showed that, for each GlpG variant, the weaker-binding phase coincided with the activity loss (*i.e.* unfolding) supporting the unfolding-binding coupling in the steric trapping scheme (**Figure 2.6a**). The same coupling was observed when a high-affinity variant, mSA_{DAB}-WT was used, further confirming that the activity loss strictly depended on the second binding of mSA_{DAB}.

The high-quality binding isotherms obtained by our sensitive FRET strategy allowed for the precise determination of the thermodynamic stability ($\Delta G^{\circ}_{\text{U,ST}}$, ST stands for steric trapping) of GlpG. Fitting of the attenuated second binding phases yielded $\Delta G^{\circ}_{\text{U,ST}}=5.8\pm0.2$ kcal/mol for 95C/172C-BtnPyr₂ and $\Delta G^{\circ}_{\text{U,ST}}=4.7\pm0.1$ kcal/mol for 172C/267C-BtnPyr₂ (**Figure 2.6b**). Both $\Delta G^{\circ}_{\text{U,ST}}$'s measured directly in a non-denaturing DDM solution were significantly lower than the extrapolated stabilities from SDS denaturation ($\Delta G^{\circ}_{\text{U,SDS}} = 8.4\text{-}8.7$ kcal/mol) (**Figure 2.6c**) but higher than the stability of GlpG in bicelles ($6.5 k_{\text{B}}T$ equivalent to ~ 4 kcal/mol) obtained by extrapolation to zero force using single-molecule magnetic tweezers²⁵. Furthermore, if the packed TM helices of GlpG unfolded cooperatively, the same $\Delta G^{\circ}_{\text{U,ST}}$ would be expected regardless of the position of the biotin pair. However, while SDS denaturation of the two GlpG variants yielded the same global stability, $\Delta G^{\circ}_{\text{U,ST}}$'s obtained by steric trapping were comparable but significantly different by 1.1 ± 0.2 kcal/mol.

Subglobal unfolding of GlpG near the active site

To track down the origin of the discrepancy between the stability obtained by steric trapping ($\Delta G^{\circ}_{U,ST}$) in DDM micelles and the extrapolated stability obtained by SDS denaturation ($\Delta G^{\circ}_{U,SDS}$), we measured the stability of the two GlpG variants using steric trapping in the range of SDS mole fraction ($X_{SDS}=0\sim 0.4$), where the major fraction of GlpG existed in the folded state (folded fraction >0.9) (**Figure 2.6c**). The $\Delta G^{\circ}_{U,ST}$ vs X_{SDS} plot (**Figure 2.7**) revealed three major features that clearly deviated from the behavior predicted from the linear extrapolation of the SDS denaturation data (diamonds vs dashed lines). First, rather than following a linearly-decreasing trend, $\Delta G^{\circ}_{U,ST}$ of both variants exhibited an upward curvature as X_{SDS} increased. Second, while $\Delta G^{\circ}_{U,ST}$ of 95C/172C-BtnPyr₂ (blue diamonds) were overall larger than $\Delta G^{\circ}_{U,ST}$ of 172C/267C-BtnPyr₂ (red diamonds), they remarkably converged at $X_{SDS}\approx 0.4$, where the main melting transition by SDS began, and this convergence was maintained up to $X_{SDS}\approx 0.5$. This result further confirms that the two GlpG variants possessed the same global stability. Third, $\Delta G^{\circ}_{U,ST}$ was larger than $\Delta G^{\circ}_{U,SDS}$ after they crossed the extrapolation lines at $X_{SDS}\approx 0.1$ and this discrepancy became increasingly pronounced up to 2.8 kcal/mol at $X_{SDS}=0.4$.

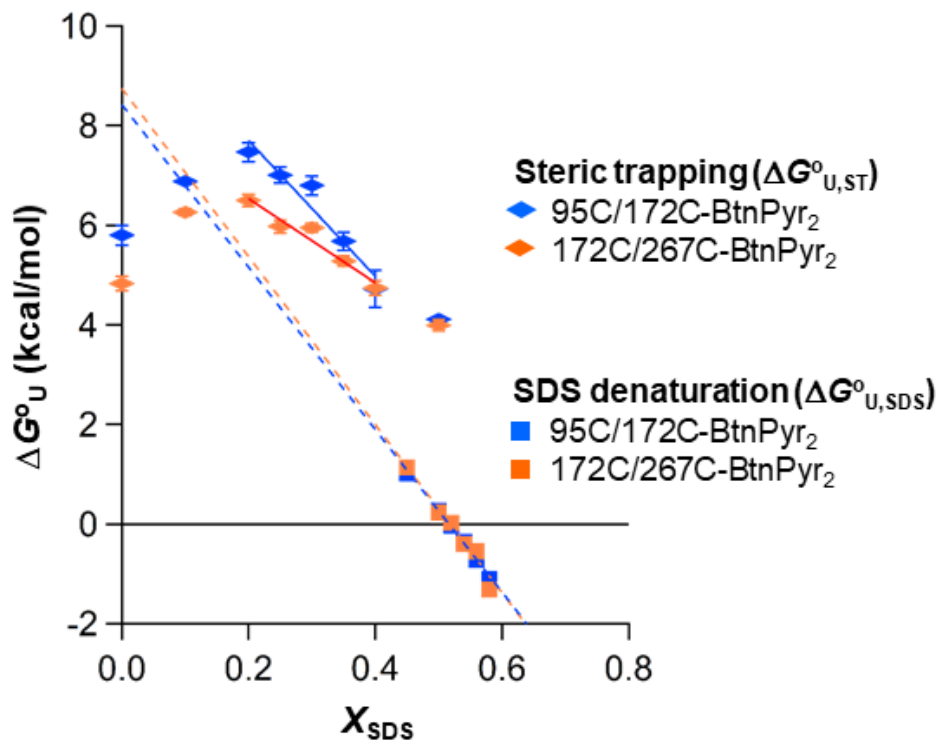


Figure 2.7 Dependence of thermodynamic stability of GlpG on SDS mole fraction. The plot containing $\Delta G^{\circ}_{U,ST}$ (diamonds) obtained by steric trapping and $\Delta G^{\circ}_{U,SDS}$ (squares) obtained by SDS denaturation as a function of SDS mole fraction (X_{SDS}) for 95C/172C-BtnPyr₂ and 172C/267C-BtnPyr₂. To fit $\Delta G^{\circ}_{U,ST}$, we accounted for the changes in the biotin affinity of mSADAB variants which depended on X_{SDS} . Errors in $\Delta G^{\circ}_{U,ST}$ denote mean $\pm$ s. d. from 2-3 independent measurements. Solid lines are the linear-regression fits of $\Delta G^{\circ}_{U,ST}$ in the range of $X_{SDS}=0.2\sim0.4$ and dashed lines indicate the extrapolation lines of $\Delta G^{\circ}_{U,SDS}$ to zero X_{SDS} from SDS denaturation. The slope in the ΔG°_U vs X_{SDS} plot represents the m -value. For 95C/172C-BtnPyr₂, $m=16\pm3$ (blue dashed line) from SDS denaturation and $m=14\pm2$ (blue solid line) from steric trapping. For 172C/267C-BtnPyr₂, $m=17\pm2$ (red dashed line) from SDS denaturation and $m=8\pm1$ (red solid line) from steric trapping. Errors in the m -values denote $\pm$ s. d. from linear-regression fits.

The nonlinear behavior of $\Delta G^{\circ}_{U,ST}$ at low SDS mole fractions implies a complex interaction between GlpG and DDM/SDS mixed micelles. A similar disagreement between steric trapping and SDS denaturation has been reported for bR in DMPC/CHAPSO/SDS mixed micelles¹⁵. In the case of GlpG, $\Delta G^{\circ}_{U,ST}$'s of both variants were maximized at $X_{SDS}\approx0.2$ but, at higher X_{SDS} , SDS persistently destabilized both proteins. Notably, in the range of $X_{SDS}=0.2\sim0.4$ (**Figure 2.7**, blue vs

red solid lines), the m -value of 95C/172C-BtnPyr₂ (14 ± 2 kcal/mol/ X_{SDS}), which represents the slope of $\Delta G^{\circ}_{\text{U}}$ against X_{SDS} , was significantly larger than that of 172C/267C-BtnPyr₂ (8 ± 1 kcal/mol/ X_{SDS}), but similar to those obtained by SDS denaturation (16 ± 3 and 17 ± 2 kcal/mol/ X_{SDS} , respectively; blue and red dashed lines). According to the theory of water-soluble protein folding, the m -value is correlated with the hydrophobic surface area exposed upon unfolding⁴⁰. Although the physical meaning of the m -value in SDS denaturation is still under debate⁴¹, it is most likely related to the difference in the affinity of SDS for different states of the protein, hence to the degree of exposure of buried structural elements upon unfolding⁴². Therefore, from the different denaturant sensitivities (m -value) of the two double-biotin variants, we suggest that trapping of the unfolded state with the biotin pair 95C/172C-BtnPyr₂ lead to substantial exposure of the buried surfaces throughout the protein while trapping with the biotin pair 172C/267C-BtnPyr₂ mainly occur through subglobal unfolding which involves exposure of the less buried-surface area. That is, depending on the position of a biotin pair, a different unfolded state ensemble may be trapped. The biotin pair 172C(TM3)/267C(TM6)-BtnPyr₂ and the catalytic dyad Ser201(TM4)/His254(TM6) are spatially close to each other in the C-terminal half of the polypeptide chain and TM6 harbors a biotin label (267C-BtnPyr) as well as His254 in the dyad (**Figure 2.8a**). Thus, subglobal unfolding should directly involve disruption of the active sites.

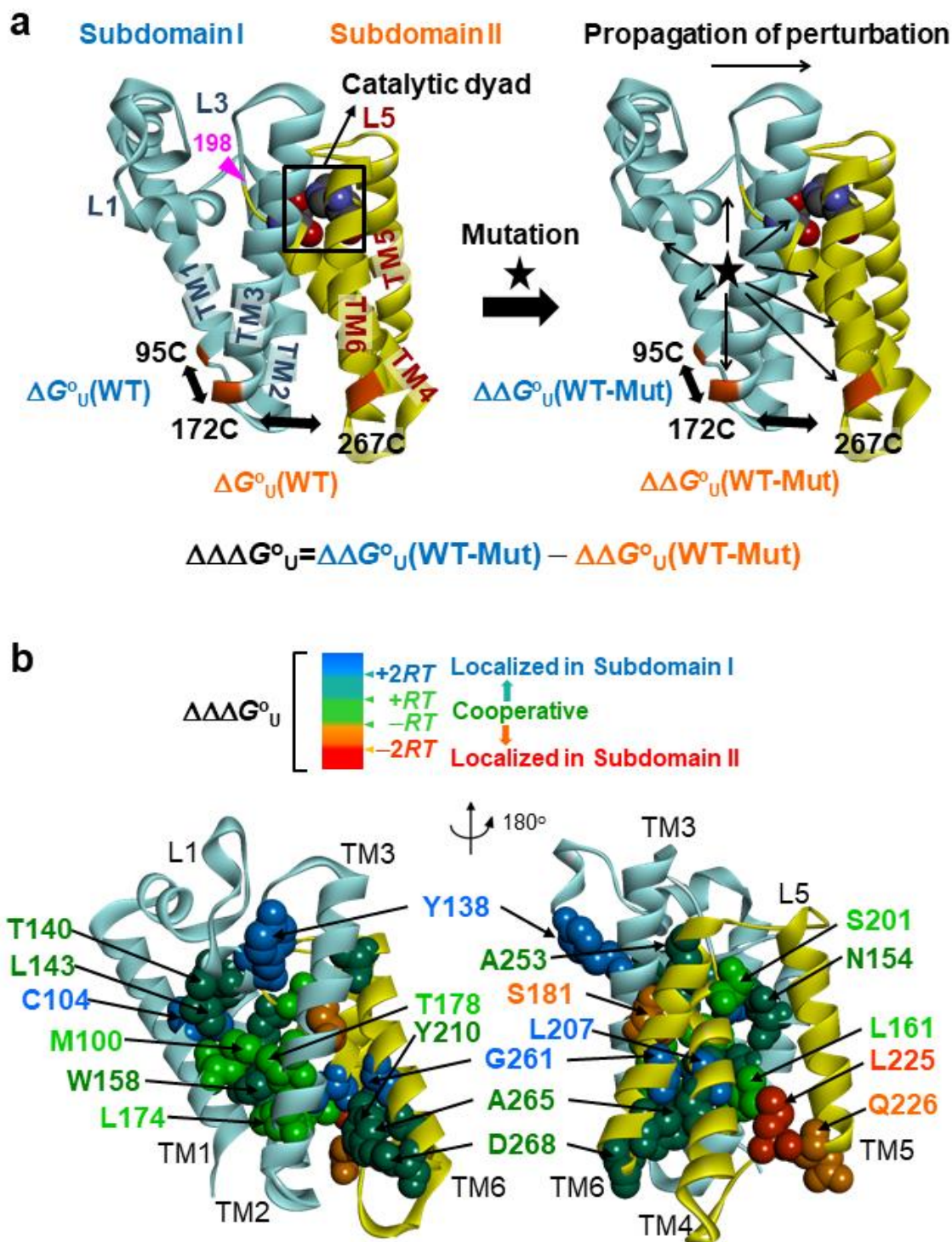


Figure 2.8 Cooperativity map reveals a network of clustered cooperative and localized interactions for the stability of GlpG under a native micellar condition. (a) Scheme for quantifying the cooperativity of interactions of a specific side chain. The stability changes ($\Delta \Delta G^{\circ}_U$) induced by the same mutation (black star) were probed with two biotin pairs, 95C/172C-BtnPyr₂ and 172C/267C-BtnPyr₂ located in different regions and compared to yield $\Delta \Delta \Delta G^{\circ}_U$ using

equation (1). The cyan-backbone region designates subdomain I (TM1-L1-TM2-TM3-L3₁₉₈), which ends at residue 198 in the L3 loop (marked with a magenta wedge) and the yellow-backbone region (L3₁₉₉-TM4-TM5-L5-TM6) indicates subdomain II. The uncertainty of the subdomain-division point is $\pm 20\sim 25$ residues around residue 198. **(b)** Cooperativity map at a side-chain resolution. The map shows the “cooperative” (green, $|\Delta\Delta\Delta G^0_U| \leq 0.6$ kcal/mol, thermal energy RT) and “localized” ($|\Delta\Delta\Delta G^0_U| > RT$) side-chain interactions. Localized interactions were further divided using additional cut-off energy values, $2RT \geq |\Delta\Delta\Delta G^0_U| > RT$ (“moderately-localized” interactions) and $|\Delta\Delta\Delta G^0_U| > 2RT$ (“highly-localized” interactions). Each side chain was color-coded based on these criteria for $\Delta\Delta\Delta G^0_U$ as shown in the figure. Errors in individual ΔG^0_U were $\pm 0.1\sim\pm 0.2$ kcal/mol (mean $\pm$ s. d. from fitting) and errors in $\Delta\Delta\Delta G^0_U$ ranged from $\pm 0.1\sim\pm 0.4$ kcal/mol, which were calculated using the propagation of errors in ΔG^0_U (see **Table 2.1**).

Subglobal unfolding has been frequently observed from HDX studies of water-soluble proteins, which is characterized by lower denaturant-sensitivity of the free energy of opening of specific secondary structural elements than that of global unfolding^{9,43,44}. Besides the different m -values, subglobal unfolding of GlpG is supported by the facts that we obtained a lower stability for 172C/267C-BtnPyr₂ using steric trapping (**Figure 2.6b**) and its trapped unfolded state reproducibly refolded with a higher yield (>90%) than that of 95C/172C-BtnPyr₂ (50~70%) (**Figure 2.3b**, bottom panels), implying different unfolded states.

We reason that the larger $\Delta G^0_{U,ST}$ than $\Delta G^0_{U,SDS}$ at $X_{SDS} > 0.1$ (**Figure 2.7**, solid vs dashed lines) was primarily due to the conformational difference between the steric-trapped unfolded state and the SDS-induced unfolded state. Our DEER result supports this argument (gray and black solid lines in the right panels of **Figure 2.4**). Although both unfolded states can be described as expanded heterogeneous conformational ensembles, the steric-trapped unfolded state on average exhibited larger expansion than the SDS-unfolded state. Thus, steric trapping appears to induce more aggressively unfolded conformations than SDS at least for the interactions between the biotinylated sites. However, we are cautious with this direct comparison because the compactness

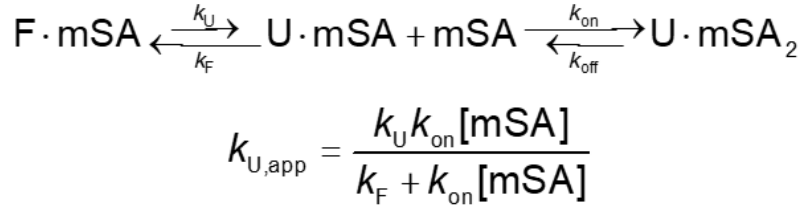
and conformation of the SDS-induced unfolded state may change as a function of X_{SDS} due to the effects of SDS on the size and shape of mixed micelles⁴⁵.

Steric trapping to measure spontaneous unfolding rate of GlpG

Steric trapping can also be applied for unfolding kinetic measurements. To verify the subglobal unfolding scenario, we did the unfolding kinetic measurement with N- and C- terminal biotin pairs. Using steric trapping, k_U can be determined by shifting the reaction flux dominantly towards unfolding upon addition of a molar excess of mSA-WT with a high biotin-binding affinity (**Figure 2.9a**). The apparent unfolding rate ($k_{U,\text{app}}$) is asymptotic as a function of mSA concentration, the maximum value of which corresponds to k_U (**Figure 2.9b**). We chose to use 20 times molar excess of mSA to unfold GlpG, as indicated with arrow in Figure 2.9b.

By measuring the k_U at different temperature, we were able to describe the unfolding energy landscape in DDM micelles using ΔG^0_U , k_U and $E_{a,U}$ (**Figure 2.10**). Notably, we obtained an highly asymmetric unfolding energy landscape in micelles, *i.e.*, the N-subdomain possessed higher kinetic ($\Delta E_{a,U} \approx 8$ kcal/mol) and thermodynamic ($\Delta \Delta G^0_U \approx 1$ kcal/mol) stability than the C-subdomain. This asymmetry further proved our hypothesis of subglobal unfolding.

a



b

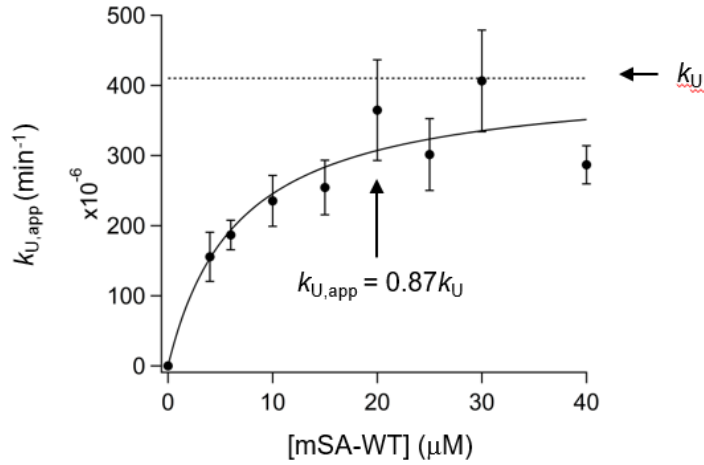


Figure 2.9 Steric trapping of GlpG to measure the spontaneous unfolding rate k_U . (a) Steric trapping for measuring k_U was achieved by shifting the reaction flux towards the unfolding direction using wild type monovalent streptavidin (mSA-WT) possessing high-affinity to biotin (mSA-WT; $K_{d,biotin} \approx 10^{-14}$ M; $k_{on,biotin} \approx 10^7$ M⁻¹ s⁻¹; $k_{off,biotin} \approx$ weeks). Under the steady state condition, in which k_U (unfolding rate) $\ll k_F$ (folding rate) and $k_{off, biotin}$ (off-rate of mSA from biotin) $\ll k_{on, biotin} \cdot [mSA-WT]$ (on-rate of mSA to biotin), the apparent unfolding rate ($k_{U,app}$) can be approximated to an asymptotic equation. At high mSA-WT concentration, $k_{U,app}$ approaches k_U . (b) Dependence of the apparent unfolding rate ($k_{U,app}$) on the concentration of mSA-WT. The unfolding rates were measured for the double-biotin variant of GlpG, 95/172_N-BtnPyr₂ (1 mM, BtnPyr) at different mSA concentrations in 20 mM sodium phosphate (pH 7.5), 200 mM NaCl and 5 mM dodecylmaltoside (DDM). GlpG activity was used as an unfolding readout in the unfolding kinetic measurement at each mSA-WT concentration. The data were fit to the steady-state kinetic equation shown in Figure S5a. In the subsequent unfolding kinetic study, the mSA-to-GlpG molar ratio of 20 was used, at which $k_{U,app}$ was close to k_U (upward arrow). Errors designate $\pm$ STD from fitting.

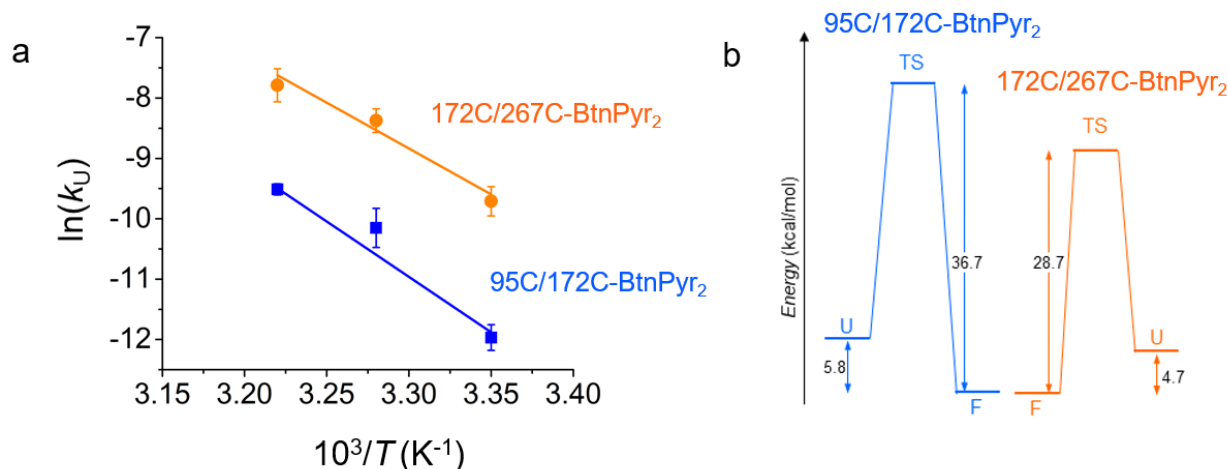


Figure 2.10 Energy landscape of GlpG in DDM. (a) Arrhenius plot for measuring the activation energy of unfolding ($E_{a,U}$) of GlpG. Spontaneous unfolding rates (k_U) were measured using steric trapping in DDM at various temperatures. **(b)** Unfolding energy landscape of GlpG in DDM including the thermodynamic stability (ΔG_U) and $E_{a,U}$. F, U and TS denote the folded state, unfolded state and transition state, respectively.

Strategy to identify cooperative and localized interactions

The higher stability and more substantial unfolding ($\Delta G_{U,ST}^0 = 5.8 \pm 0.2$ kcal/mol and $m = 14 \pm 2$ kcal/mol/ X_{SDS}) obtained with 95C/172C-BtnPyr₂ implies that the native tertiary interactions between this biotin pair in the N-terminal region are critical to the conformational integrity of the whole GlpG. On the other hand, the lower stability and subglobal unfolding ($\Delta G_{U,ST}^0 = 4.7 \pm 0.1$ kcal/mol and $m = 8 \pm 1$ kcal/mol/ X_{SDS}) obtained with 172C/267C-BtnPyr₂ indicates that the C-terminal region containing this biotin pair possesses differential folding properties from the N-terminal region. Overall, this result suggests complex energetic coupling among different structural regions in GlpG. To clarify this complexity, we have developed a method to identify cooperative and localized interactions that contribute to protein stability at a side-chain resolution. Our basic strategy is to quantify how a structural perturbation induced by mutating a specific side

chain would be propagated to other regions within the protein using steric trapping and stability analysis (**Figure 2.8**).

First, we dissected GlpG into two subdomains: (1) the more stable N-terminal subdomain I encompassing TM1-L1-TM2-TM3-L3₁₉₈ (ending at residue 198 in the L3 loop, cyan backbone in **Figure 2.8a**) whose unfolding is trapped with the biotin pair 95C/172C-BtnPyr₂, and (2) the less stable C-terminal subdomain II consisting of L3₁₉₉-TM4-TM5-L5-TM6 (starting from residue 199 in the L3 loop, yellow backbone) whose unfolding is trapped with the biotin pair 172C/267C-BtnPyr₂. By this dissection, we defined subdomain II as the region that underwent subglobal unfolding and were able to register each residue of interest to either subdomain to analyze the influence of perturbing its side-chain interactions on the stability of each subdomain.

Second, we made a single mutation (typically to alanine) in either subdomain in the background of 95C/172C-BtnPyr₂ and 172C/267C-BtnPyr₂. Those two double-biotin variants were referred to as “wild type (WT)” because the native tertiary interactions were equally preserved in both as shown by SDS denaturation (**Figure 2.6c**). The two double-biotin variants possessing the same mutation were referred to as “mutants” (Mut). Next, the stability changes induced by the mutation were probed with the two different biotin pairs using steric trapping and compared to each other. We quantified the differential effect of each mutation on the stability of each subdomain ($\Delta\Delta G^{\circ}_U$) using **equation (1)** containing the stabilities of four variants ($\Delta G^{\circ}_{U,95C/172C-BtnPyr_2}(WT)$, $\Delta G^{\circ}_{U,95C/172C-BtnPyr_2}(Mut)$, $\Delta G^{\circ}_{U,172C/267C-BtnPyr_2}(WT)$ and $\Delta G^{\circ}_{U,172C/267C-BtnPyr_2}(Mut)$):

$$\begin{aligned}\Delta\Delta G^{\circ}_U &= [(\Delta G^{\circ}_{U,95C/172C-BtnPyr_2}(WT) - \Delta G^{\circ}_{U,95C/172C-BtnPyr_2}(Mut))] \\ &\quad - [\Delta G^{\circ}_{U,172C/267C-BtnPyr_2}(WT) - \Delta G^{\circ}_{U,172C/267C-BtnPyr_2}(Mut)] \\ &= \Delta\Delta G^{\circ}_{U,95C/172C-BtnPyr_2}(WT-Mut) - \Delta\Delta G^{\circ}_{U,172C/267C-BtnPyr_2}(WT-Mut)\end{aligned}\tag{1}$$

Here, $\Delta\Delta G^{\circ}_{U,95C/172C-BtnPyr_2}(WT-Mut)$ and $\Delta\Delta G^{\circ}_{U,172C/267C-BtnPyr_2}(WT-Mut)$ designate the stability changes caused by the same mutation in the backgrounds of 95C/172C-BtnPyr₂ and 172C/267C-BtnPyr₂, respectively. Thus, $\Delta\Delta\Delta G^{\circ}_U$ represents the difference in the stability changes that were 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 ($|\Delta\Delta\Delta G^{\circ}_U| \leq RT = 0.6 \text{ kcal/mol}$ at 25 °C, R : gas constant; T : absolute temperature), the mutated site is engaged in a “cooperative” interaction. That is, the perturbation by the mutation is similarly propagated to both subdomains. If a mutation preferentially destabilizes the subdomain containing it ($|\Delta\Delta\Delta G^{\circ}_U| > RT$), the perturbed interactions are “localized” within that subdomain. If mutation of a residue, which belongs to one subdomain and makes its entire tertiary contacts only with the subdomain containing it, preferentially destabilizes the other subdomain, the perturbation is regarded as “over-propagated”.

Cooperativity network in GlpG

We targeted 20 residues covering key packing regions that had been previously identified using thermal denaturation²³, and assessed their role in the folding cooperativity of GlpG (**Table 2.1**). The stability changes upon mutation ($\Delta\Delta G^{\circ}_U$ (WT-Mut)) measured by steric trapping were reasonably well correlated with the changes in melting temperature ΔT_m (WT-Mut)²³ for common 17 mutations, which validated our steric trapping approach. Our 20 $\Delta\Delta\Delta G^{\circ}_U$ values were distributed over a wide range from -1.8 to 2.0 kcal/mol and their individual errors ranged from ± 0.1 to $\pm 0.4 \text{ kcal/mol}$ ($\pm$ s. d.), smaller than RT .

95C/172C-BtnPyr ₂			172C/267C-BtnPyr ₂		$\Delta\Delta\Delta G^{\circ}_U$	Location
Mutation	$\Delta\Delta G^{\circ}_{U,95C/172C-BtnPyr_2}$ (WT-Mut)	Activity	$\Delta\Delta G^{\circ}_{U,172C/267C-BtnPyr_2}$ (WT-Mut)	Activity		
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

Table 2.1 Stability changes by single substitutions and activities of singly-substituted variants in the backgrounds of double-biotin GlpG variants 95C/172C-BtnPyr₂ and 172C/267C-BtnPyr₂. Stabilities were measured by steric trapping in pH 7.0 sodium phosphate, 200 mM NaCl, 0.25 mM TCEP and 5 mM DDM solution. In calculating the stability change for each substitution, $\Delta G^{\circ}_{U,95C/172C-BtnPyr_2}=5.8\pm0.2$ kcal/mol and $\Delta G^{\circ}_{U,172C/267C-BtnPyr_2}=4.7\pm0.1$ kcal/mol were used as wild-type stabilities. All energy values are in kcal/mol. Activity values are relative to wild-type GlpG. $\Delta\Delta\Delta G^{\circ}_U$ is defined as $\Delta G^{\circ}_{U,95C/172C-BtnPyr_2}-\Delta G^{\circ}_{U,172C/267C-BtnPyr_2}$. Errors denote propagated s. d. calculated from s. d. of individual ΔG°_U values.

Table 2.1 (cont'd)

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

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

^aSubdomain in which a mutated residue is located (see **Figure 2.8b**).^bSecondary structural elements in which a mutated residue is located (see **Figure 2.8a**).

^cIf 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

The effects of mutations were mapped onto the structure which we called the “cooperativity map” (**Figure 2.8b**). We applied four cut-off values $\Delta\Delta\Delta G^0_U = -2RT$, $-RT$, RT and $2RT$ (*i.e.* five sets of the cooperativity profile) to account for the wide distribution of $\Delta\Delta\Delta G^0_U$ spanning ~ 4 kcal/mol, as well as to more precisely resolve the degree of cooperativity of each side-chain interaction. Surprisingly, cooperative and localized interactions were significantly clustered into defined regions in the GlpG structure, whose spatial distribution can be divided into four distinct groups. First, cooperative interactions ($|\Delta\Delta\Delta G^0_U| \leq RT$, green in **Figure 2.8b**) of five residues, Met100(TM1), Leu161(TM2), Leu174 (TM3), Thr178(TM3) and Ser201(TM4), were clustered in the packing region buried in the interior of subdomain I and the subdomain interface near the center of the membrane. Interestingly, Ser201, a part of the catalytic dyad Ser201/His254, was engaged in cooperative interactions. Moderately localized interactions by Trp158 (teal, $RT < \Delta\Delta\Delta G^0_U \leq 2RT$) and highly localized interactions by Leu207 (blue, $2RT < \Delta\Delta\Delta G^0_U$) were also present in this region. This cooperative cluster overlaps with one of the key packing regions²³ and also partially with the folding nucleus in TM1 and TM2 identified from the ϕ -value analysis²⁴. Second, all tested residues located in the folded L1 loop (Tyr138, Thr140 and Leu143) and the residue packed against L1 (Cys104) were engaged in moderately or highly localized interactions in subdomain I. Based on the same ϕ -value analysis, it was suggested that this region is frustrated in the folding transition state²⁴. Third, Leu225 (red, $\Delta\Delta\Delta G^0_U < -2RT$) and Gln226 (orange, $-2RT \leq \Delta\Delta\Delta G^0_U < -RT$) in TM5 of subdomain II were located at the subdomain interface and exposed to the water-micelle interface, respectively. They were both classified as localized interactions in subdomain II in varying degrees. It is known that most residues in TM5 are not tightly packed against the rest of the protein and do not significantly contribute to the thermostability²³.

The fourth cluster is of particular interest. Mutations of the residues at the TM4/TM6 interface (Ala253, Gly261, Ala265 and Asp268), which belonged to subdomain II, preferentially destabilized subdomain I. Particularly, the side chains of Gly261 and Ala265 were not making any tertiary contacts with the residues in subdomain I, but perturbing their interactions exerted larger impacts on the stability of subdomain I. Thus, we classified these two residues as “over-propagated”. This interface harbors the catalytic dyad Ser201/His254 and plays a pivotal role in both stability and function of GlpG²³. Especially, Gly261 and the catalytic dyad are absolutely conserved among rhomboid proteases⁴⁶. Our result suggests that these conserved residues are critical not only to the conformational and functional integrity of GlpG but also to the energetic coupling between different structural regions of GlpG. The single-molecule magnetic tweezers study suggested that the breakage of the interactions near the C-terminal and its propagation towards the N-terminal is the primary mechanism of the force-induced unfolding of GlpG²⁵.

It should be noted that 5 among 20 tested mutations completely inactivated GlpG (**Table 2.1**). Thus, our new steric trapping strategy allowed for the stability measurements of not only functional but also non-functional variants, which had not been possible under the original steric trapping framework. While the two double-biotin variants bearing the same mutation exhibited differential stability change in non-denaturing DDM solution, they possessed the same global stability when their stabilities were measured using SDS denaturation. Therefore, we conclude that the networked nature of the side-chain interactions that we have revealed (**Figure 2.8b**) is a novel phenomenon that occurs under native conditions.

Discussion

Here we presented a general steric-trapping strategy to investigate thermodynamic stability of membrane proteins and conformation of their unfolded state under native conditions. Our strategy

utilizes versatile thiol-reactive molecular tags, in which the essential features for steric trapping, a biotin to bind mSA and a probe to monitor mSA binding or protein unfolding are integrated. The fluorescent probe BtnPyr allowed for determination of the thermodynamic stability of GlpG through high-quality binding isotherms conveniently obtained by FRET. The paramagnetic probe BtnRG enabled characterization of the unfolded state ensemble of GlpG based on the distance distributions obtained by DEER. Because this combined strategy does not require either target-specific unfolding readout or specific lipid environments, it is broadly applicable to various types of membrane proteins including nonfunctional and misfolded variants whose folding characterization under native conditions is difficult.

The unfolded state of proteins has gained significant attention because it determines the thermodynamic stability with the folded state, directs early folding mechanisms, and serves as a target for chaperoning and degradation⁴⁷. However, the physical dimension of the unfolded state ensemble of membrane proteins was difficult to measure in native micellar or bilayer environments due to its transient nature preventing detailed biophysical analysis. By applying DEER to the steric-trapped unfolded state, for the first time, we elucidated a largely-unraveled dynamic and heterogeneous conformational ensemble of the unfolded state of GlpG in a native micellar environment. It should be taken into account that the conformational ensemble in the steric-trapped unfolded state may have been influenced by steric repulsion between two bound mSA molecules, representing a subset of the true ensemble. However, in the selection process for optimal biotinylation sites, we found that saturated binding of mSA to the biotin pairs conjugated to 94C/172C and 172C/271C, whose C_{α} - C_{α} distances were similar to those of 95C/172C and 172C/267C in this study, completely retained the native activity level without inducing unfolding. This result implies that two bound mSA molecules are allowed to coexist within a close range

probably also in the steric-trapped unfolded state. Therefore, steric repulsion between bound mSA molecules may not be the sole reason for the largely expanded unfolded state. It is still an open question to what extent trapping would affect the protein conformation beyond the region containing a biotin pair. It would be also an intriguing future study to test if a similar degree of expansion and heterogeneity of the unfolded state would be recapitulated in the lipid bilayer which provides a more rigid and defined hydrophobic environment than micelles.

By probing the stability of GlpG with the biotin pairs located in different regions, we identified subglobal unfolding in the C-terminal region which contained the active site. This asymmetric stability profile of GlpG is analogous to the highly-polarized folding transition state obtained by the ϕ -value analysis in DDM/SDS mixed micelles, which was described as possessing a compact folding nucleus in the N-terminal TM1~TM2 and largely unstructured TM3~TM6²⁴. In the single-molecule pulling study, Min *et al.* suggested that TM3~TM6 or TM5~TM6 are more flexible regions²⁵. Although we have defined the region that undergoes subglobal unfolding as the continuous secondary structural elements L3₁₉₉-TM4-TM5-L4-TM6 (**Figure 2.8a**), it would be more reasonable to interpret subglobal unfolding as an ensemble-averaged event, which involves unfolding of a various number of C-terminal TM helices. Our work is unique in that we demonstrated the existence of the differential unfolding behavior along the polypeptide chain under native conditions, in the absence of chemical denaturants or pulling force. This partial unfolding behavior may reflect intrinsic conformational malleability of the region near the active site of GlpG. Although it is not clear if subglobal unfolding is necessary during the catalytic cycle of GlpG, we speculate that this malleability may be adequate for conformational changes required for substrate interaction and catalytic mechanism. Further supporting this idea, significant

disordering of the L5 loop, partial unfolding of TM5 and tilting of TM6 have been observed from multiple crystal structures of GlpG in *apo* and inhibitor-bound forms^{28,48,49}.

Our unprecedented cooperativity analysis suggests that, although apparently tightly packed, the helical-bundle architecture of GlpG is maintained through an intricate network of cooperative and localized interactions. Although the concept of a cooperativity network and its role in protein stability and function has been established for water-soluble proteins^{4,7,9,43,50}, such aspects have not been explored for membrane proteins. Our experimentally-determined cooperativity map (**Figure 2.8b**) indicates that the degree of cooperativity is the largest for the buried residues near the center of the membrane and gradually fades out toward the lipid- and water-contacting regions. This positional dependence of the cooperativity profile suggests that complex environmental constraints that contribute to stabilization of membrane proteins, *i.e.* protein-protein, protein-lipid and protein-water interactions, also play an important role in the organization of the interaction network. Therefore, our general steric trapping strategy and steric trapping-based approaches to evaluate the stability, unfolded state ensemble and cooperativity will serve as powerful tools for exploring the folding energy landscape of membrane proteins in native cell membranes, which remains as a far-reaching goal.

Materials and Methods

Synthesis of BtnPyr-IA and BtnRG-TP

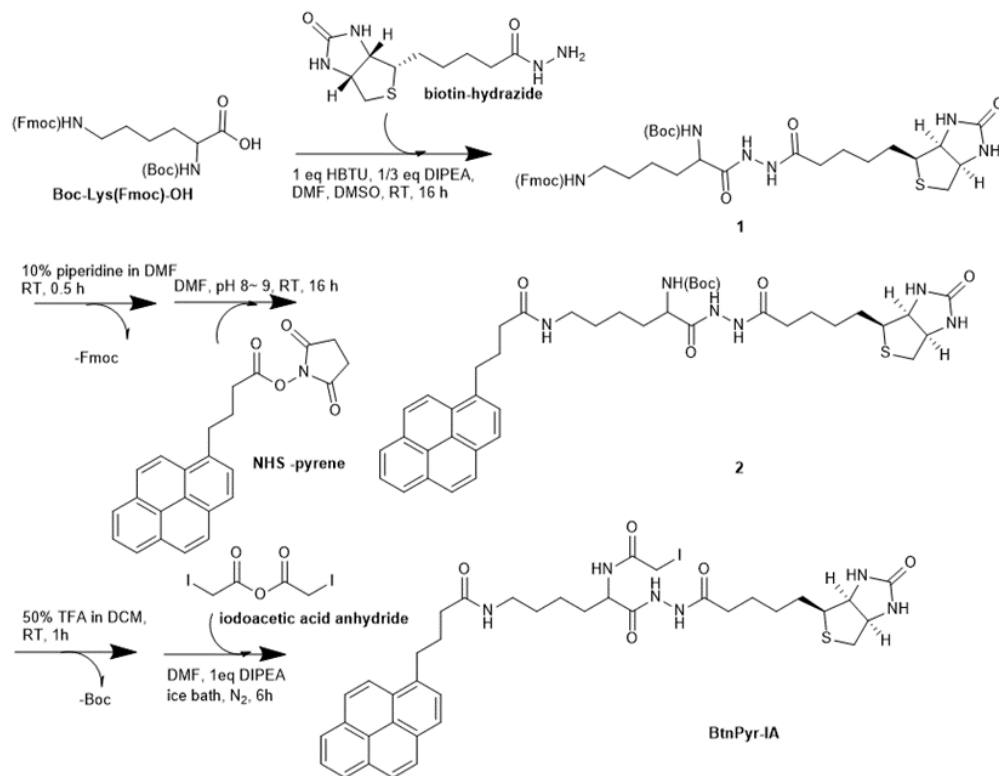


Figure 2.11 Synthesis scheme of BtnPry-IA

Boc-Lys(Fmoc)-OH (Chem-Impex International) was used as a template. Iodoacetamide group, biotin and pyrene groups were conjugated to the template step by step. 362 mg (0.78 mmol) of Boc-Lys(Fmoc)-OH was activated by addition of 439 mg (1.16 mmol) of *O*-(Benzotriazol-1-yl)-*N,N,N',N'*-tetramethyluronium hexafluorophosphate (HBTU) (Chem-Impex International) and 50 μ L (0.28 mmol) of *N,N*-Diisopropylethylamine (DIPEA) (Sigma) in 2 mL dimethylformamide (DMF) (Sigma). After incubation for 20 min, 100 mg (0.39 mmol) of biotin-hydrazide (Santa Cruz Biotechnology) dissolved in 1 mL of dimethyl sulfoxide (DMSO) (Avantor Performance Materials) was added dropwise and stirred at room temperature overnight. Then, excess diethyl ether (Sigma) was added to the reaction to make separated two layers. After removing the upper layer, ethyl

acetate (Sigma) was added to the yellow lower layer to precipitate out **1**. The precipitation was washed several times with diethyl ether. Deprotection of Fmoc was performed with a 30-min incubation of **1** in 10% v/v piperidine (Sigma) in DMF followed by precipitation with diethyl ether. The yield at this point was ~50%. Then the product (Boc-Lys-biotin) was dissolved in 2 mL DMF and diethylamine (Sigma) solution (pH~8). After addition of 1.5 times molar excess of 1-Pyrenebutyric acid *N*-hydroxysuccinimide ester (NHS-pyrene), the mixture was stirred overnight at room temperature in dark. The following steps were carried out in dark. **2** was precipitated out with diethyl ether. Deprotection of Boc group was performed by incubation of **2** in 50% trifluoroacetic (TFA) acid (Sigma) in dichloromethane (DCM) (Sigma) for 30 min. After precipitation with diethyl ether, the product was dissolved in DMF. The solution was put in an ice bath and one equivalent of DIPEA and 1.5 times molar excess of iodoacetic acid anhydride (Sigma) dissolved in DMF were slowly added. The reaction mixture was protected with N₂ gas and incubated for 4 h. The final product was precipitated out as light yellow powder with diethyl ether and analyzed by electron spray ionization (ESI) mass spectrometry (Xevo G2-S QTof, Waters). The exact mass of M+H is 825.2287 and the actual peak was at 825.2308 .Total yield was ~10%.

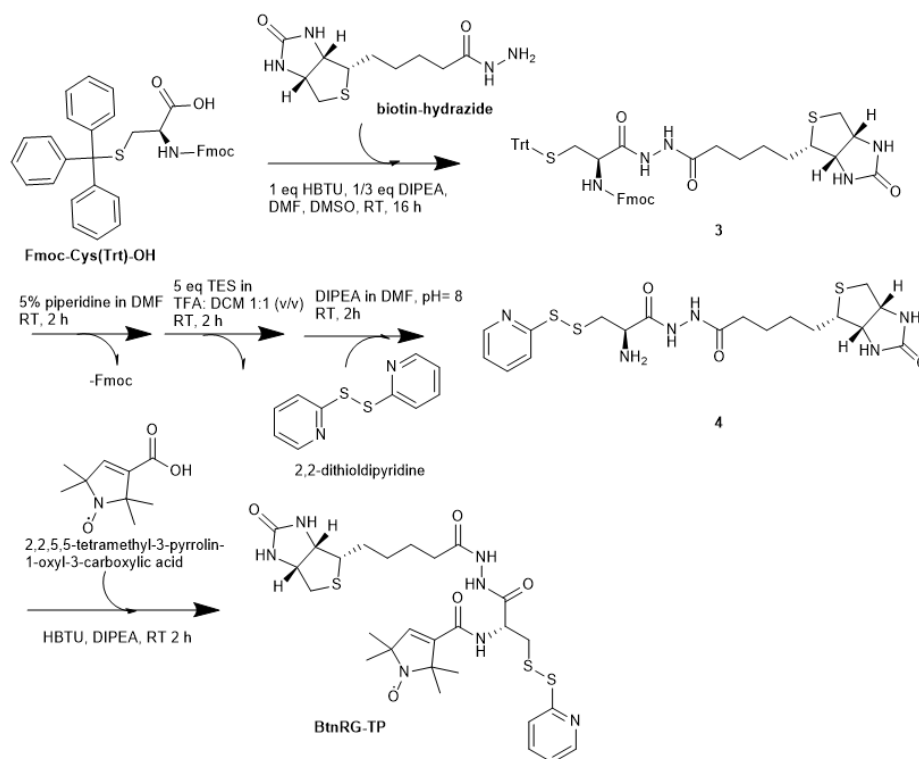


Figure 2.12 Synthesis scheme of BtnRG-TP.

Fmoc-Cys(Trt)-OH (Chem-Impex International) was used as a template. Iodoacetamide, biotin and spin label (2,2,5,5-Tetramethyl-3-pyrrolin-1-oxyl) groups were conjugated step by step. 340 mg (0.78 mmol) of Fmoc-Cys(Trt)-OH was activated by adding 439 mg (1.16 mmol) of HBTU and 50 μ L (0.28 mmol) of DIPEA in 2 mL DMF. After incubating for 20 min, 100 mg (0.39 mmol) of biotin-hydrazide dissolved in 1 mL DMSO was added dropwise. The reaction mixture was stirred at room temperature overnight. Then, excess diethyl ether (Sigma) was added to the reaction to make separated two layers. After removing the upper layer, ethyl acetate (Sigma) was added to the yellow lower layer to precipitate out **3**. Deprotection of Fmoc was performed by incubation of **3** for 30 min in 10% v/v piperidine in DMF followed by precipitation using diethyl ether. Then, Trt group was removed by incubation of the precipitation from last step in TFA/DCM (1:1) mixture in the presence of 5 molar equivalent of triethylsilane (TES) (Sigma) for 1 h. The product (NH₂-

Cys-biotin) was precipitated out using diethyl ether and washed at least five times. The washed product was dissolved in DMF, and DIPEA was added until pH reaches at ~8. 1.5 times molar excess of 2,2-dithioldipyridine (Sigma) dissolved in DMF was added. The reaction mixture was stirred for 2 h at room temperature and **4** was precipitated out using diethyl ether. **4** was dissolved in DMF and added slowly to 1.5 times molar excess of 2,2,5,5-tetramethyl-3-pyrrolin-1-oxyl-3-carboxylic acid (Acros Organics, NJ) whose carboxylic group was activated with 1.5 times molar excess of HBTU and 1 molar equivalent of DIPEA. The reaction was stirred at room temperature for 3 h. The final product, BtnRG-TP was precipitated out as white powder using diethyl ether and washed several times. BtnRG-TP was analyzed by electron spray ionization (ESI) mass spectrometry (Model: Xevo G2-S QToF, Waters). Exact Mass of M+H is 637.2167 and the actual peak was at 637.2178. Total yield was about 20%.

Preparation of GlpG DNA constructs

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 pET15b vector containing an N-terminal His₆-tag. Site-directed mutagenesis for introducing amino acid mutations was performed using the QuikChange Site-Directed Mutagenesis Kit (Agilent).

Expression of GlpG

GlpG was expressed in *E. coli* BL21(DE3) RP strain. Cells were grown at 37 °C until OD₆₀₀ = 0.6 was reached. Protein expression was induced with 0.5 mM IPTG, 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 Ni²⁺-NTA affinity chromatography (Qiagen) after solubilization with 1% dodecylmaltoside (DDM, Anatrace).

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 μ M and incubated with a ten-fold molar excess Tris(2-carboxyethyl)phosphine hydrochloride (TCEP-HCl, Pierce) for 1 h at room temperature. 40 times molar excess of BtnPyr-IA or BtnRG-TP dissolved in DMSO (~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 Ni²⁺-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 after concentration. Typically, the labeling efficiency of BtnPyr-IA and BtnRG-TP ranged from 1.5~2.2 as estimated from SDS-PAGE gel shift assay or comparison of the concentration of BtnPyr determined by pyrene absorbance ($\epsilon_{346\text{nm}}=43,000 \text{ M}\cdot\text{cm}^{-1}$) and the concentration of GlpG determined by DC protein assay (Bio-Rad). 10 μ l of 5 μ M of labeled GlpG was incubated with 10 μ l 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:WT-mSA 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 dissociation of WT-mSA 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. 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^{31,51}.

Fluorescence-based high-throughput activity assay for GlpG

As a folding indicator for GlpG, we used the proteolytic activity of GlpG 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 domain¹⁶, TEV protease recognition site, and C-terminal His₆-tag (SN-TEV-LacYTM2-His₆). In the LacYTM2 region, the position which was five residue upstream from the scissile bond (P5 position) was substituted with cysteine for labeling with thiol-reactive, environment-sensitive fluorophore 7-Nitrobenz-2-oxa-1,3-diazol (NBD). SN-TEV-LacYTM2-His₆ containing the substituted cysteine was expressed in BL21(DE3) RP *E. coli* strain. Cells were grown in Terrific Broth at 30 °C for 16 h, until OD₆₀₀ reached 1.5~2.0, and then the protein expression in an inclusion body form was induced with 1 mM IPTG. Cells were further grown at 37 °C for 3 h. Isolated inclusion bodies were solubilized with 1.5% n-Decyl-β-D-maltoside (DM). After centrifugation, the supernatant was applied to Ni²⁺-NTA affinity chromatography for purification. Bound SN-TEV-LacYTM2-His₆ was eluted with Tris buffer (0.5% DM, 50 mM TrisHCl, 200 mM NaCl, 200 mM imidazole, pH 8.0). After removing imidazole using EconoPac 10DG desalting column (BioRad), the protein was concentrated and labeled with IANBD amide (Setareh biotech). 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 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.

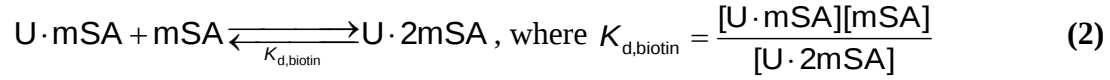
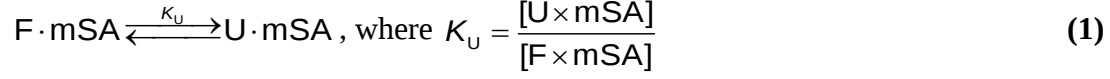
Double electron-electron resonance EPR spectroscopy (DEER-EPR)

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 prior to data collection. For data collection, 36-ns π -pump pulse was applied to the low field peak of the nitroxide absorption spectrum, and the observer $\pi/2$ (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 ($\pi/2$) 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 inter-spin distance range was from ~ 15 to ~ 60 Å. DEER data were analyzed using a model-free maximum-entropy analysis approach developed by Christian Altenbach (<http://www.chemistry.ucla.edu/directory/hubbell-wayne-l>).

Construction of binding isotherms to determine thermodynamic stability

1 μM of GlpG labeled with BtnPyr was titrated with mSA specifically labeled with DABCYL (AnaSpec) at Y83C-position of the active subunit (mSA_{DAB}) in 5 mM DDM, 0.25 mM TCEP, 20 mM sodium phosphate 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 95C/172C-BtnPyr₂) or 2 days (for 172C/267C-BtnPyr₂) 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.

Our fitting equation for obtaining thermodynamic stability of GlpG using steric trapping was based on the following reaction scheme¹⁴:



Fitting equation for the second mSA binding phase was:

$$F = \frac{1}{[1 + \left(K_{d,biotin} + \frac{K_{d,biotin}}{K_U} \right) \frac{1}{[mSA]}]} (F_{\infty} - F_0) + F_0 \quad (3)$$

, where F is 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 of unfolding of GlpG. After obtaining the fitted K_U , the thermodynamic stability obtained by steric trapping was calculated using the equation $\Delta G^0_{U,ST} = -RT \ln K_U$.

REFERENCES

REFERENCES

- 1 Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins* **1995**, *21*, 167-195.
- 2 Oliveberg, M. & Wolynes, P. G. The experimental survey of protein-folding energy landscapes. *Q Rev Biophys* **2005**, *38*, 245-288.
- 3 Dill, K. A. & Chan, H. S. From Levinthal to pathways to funnels. *Nat Struct Biol* **1997**, *4*, 10-19.
- 4 Chamberlain, A. K., Handel, T. M. & Marqusee, S. Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nat Struct Biol* **1996**, *3*, 782-787.
- 5 Bai, Y., Sosnick, T. R., Mayne, L. & Englander, S. W. Protein folding intermediates: native-state hydrogen exchange. *Science* **1995**, *269*, 192-197.
- 6 Sekhar, A. & Kay, L. E. NMR paves the way for atomic level descriptions of sparsely populated, transiently formed biomolecular conformers. *Proc Natl Acad Sci U S A* **2013**, *110*, 12867-12874.
- 7 Park, C. & Marqusee, S. Probing the high energy states in proteins by proteolysis. *J Mol Biol* **2004**, *343*, 1467-1476.
- 8 Sharon, M. & Robinson, C. V. The role of mass spectrometry in structure elucidation of dynamic protein complexes. *Annu Rev Biochem* **2007**, *76*, 167-193.
- 9 Bai, Y. & Englander, S. W. Future directions in folding: the multi-state nature of protein structure. *Proteins* **1996**, *24*, 145-151.
- 10 Cui, Q. & Karplus, M. Allostery and cooperativity revisited. *Protein Sci* **2008**, *17*, 1295-1307.
- 11 Nussinov, R. & Tsai, C. J. Allostery in disease and in drug discovery. *Cell* **2013**, *153*, 293-305.
- 12 le Coutre, J., Kaback, H. R., Patel, C. K., Heginbotham, L. & Miller, C. Fourier transform infrared spectroscopy reveals a rigid alpha-helical assembly for the tetrameric *Streptomyces lividans* K⁺ channel. *Proc Natl Acad Sci U S A* **1998**, *95*, 6114-6117 (1998).
- 13 Joh, N. H. *et al.* Modest stabilization by most hydrogen-bonded side-chain interactions in membrane proteins. *Nature* **2008**, *453*, 1266-1270.
- 14 Blois, T. M., Hong, H., Kim, T. H. & Bowie, J. U. Protein unfolding with a steric trap. *J Am Chem Soc* **2009**, *131*, 13914-13915.
- 15 Chang, Y. C. & Bowie, J. U. Measuring membrane protein stability under native conditions. *Proc Natl Acad Sci U S A* **2014**, *111*, 219-224.
- 16 Hong, H., Blois, T. M., Cao, Z. & Bowie, J. U. Method to measure strong protein-protein interactions in lipid bilayers using a steric trap. *Proc Natl Acad Sci U S A* **2010**, *107*, 19802-19807.

- 17 Hong, H. & Bowie, J. U. Dramatic destabilization of transmembrane helix interactions by features of natural membrane environments. *J Am Chem Soc* **2011**, *133*, 11389-11398.
- 18 Jefferson, R. E., Blois, T. M. & Bowie, J. U. Membrane proteins can have high kinetic stability. *J Am Chem Soc* **2013**, *135*, 15183-15190.
- 19 Lee, J. R., Urban, S., Garvey, C. F. & Freeman, M. Regulated intracellular ligand transport and proteolysis control EGF signal activation in *Drosophila*. *Cell* **2001**, *107*, 161-171 (2001).
- 20 McQuibban, G. A., Saurya, S. & Freeman, M. Mitochondrial membrane remodelling regulated by a conserved rhomboid protease. *Nature* **2003**, *423*, 537-541.
- 21 Stevenson, L. G. *et al.* Rhomboid protease AarA mediates quorum-sensing in *Providencia stuartii* by activating TatA of the twin-arginine translocase. *Proc Natl Acad Sci U S A* **2007**, *104*, 1003-1008.
- 22 Urban, S., Lee, J. R. & Freeman, M. *Drosophila* rhomboid-1 defines a family of putative intramembrane serine proteases. *Cell* **2001**, *107*, 173-182 (2001).
- 23 Baker, R. P. & Urban, S. Architectural and thermodynamic principles underlying intramembrane protease function. *Nat Chem Biol* **2012**, *8*, 759-768,.
- 24 Paslawski, W. *et al.* Cooperative folding of a polytopic alpha-helical membrane protein involves a compact N-terminal nucleus and nonnative loops. *Proc Natl Acad Sci U S A* **2015**, *112*, 7978-7983.
- 25 Min, D.; Jefferson, R. E.; Bowie, J. U.; Yoon, T.-Y. Mapping the energy landscape for second-stage folding of a single membrane protein. *Nat. Chem. Biol.* **2015**, *11*, 981–987.
- 26 Akiyama, Y. & Maegawa, S. Sequence features of substrates required for cleavage by GlpG, an *Escherichia coli* rhomboid protease. *Mol Microbiol* **2007**, *64*, 1028-1037.
- 27 Vosyka, O. *et al.* Activity-based probes for rhomboid proteases discovered in a mass spectrometry-based assay. *Proceedings of the National Academy of Sciences of the United States of America* **2013**, *110*, 2472-2477.
- 28 Zoll, S. *et al.* Substrate binding and specificity of rhomboid intramembrane protease revealed by substrate-peptide complex structures. *Embo Journal* **2014**, *33*, 2408-2421.
- 29 Sherratt, A. R., Blais, D. R., Ghasriani, H., Pezacki, J. P. & Goto, N. K. Activity-Based Protein Profiling of the *Escherichia coli* GlpG Rhomboid Protein Delineates the Catalytic Core. *Biochemistry* **2012**, *51*, 7794-7803.
- 30 Arutyunova, E. *et al.* Allosteric regulation of rhomboid intramembrane proteolysis. *Embo Journal* **2014**, *33*, 1869-1881.
- 31 Howarth, M. *et al.* A monovalent streptavidin with a single femtomolar biotin binding site. *Nat Methods* **2006**, *3*, 267-273.
- 32 Klumb, L. A., Chu, V. & Stayton, P. S. Energetic roles of hydrogen bonds at the ureido oxygen binding pocket in the streptavidin-biotin complex. *Biochemistry* **1998**, *37*, 7657-7663.
- 33 Jeschke, G. DEER distance measurements on proteins. *Annu Rev Phys Chem* **2012**, *63*, 419-446.

- 34 Dockter, C. *et al.* Refolding of the integral membrane protein light-harvesting complex II monitored by pulse EPR. *Proc Natl Acad Sci U S A* **2009**, *106*, 18485-18490.
- 35 Krishnamani, V., Hegde, B. G., Langen, R. & Lanyi, J. K. Secondary and tertiary structure of bacteriorhodopsin in the SDS denatured state. *Biochemistry* **2012**, *51*, 1051-1060.
- 36 Hubbell, W. L., Cafiso, D. S. & Altenbach, C. Identifying conformational changes with site-directed spin labeling. *Nat Struct Biol* **2000**, *7*, 735-739.
- 37 Hirst, S. J., Alexander, N., Mchaourab, H. S. & Meiler, J. RosettaEPR: An integrated tool for protein structure determination from sparse EPR data. *Journal of Structural Biology* **2011**, *173*, 506-514.
- 38 Mok, Y. K., Kay, C. M., Kay, L. E. & Forman-Kay, J. NOE data demonstrating a compact unfolded state for an SH3 domain under non-denaturing conditions. *J Mol Biol* **1999**, *289*, 619-638.
- 39 Byler, D. M. & Susi, H. Examination of the secondary structure of proteins by deconvolved FTIR spectra. *Biopolymers* **1986**, *25*, 469-487.
- 40 Myers, J. K., Pace, C. N. & Scholtz, J. M. Denaturant m values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci* **1995**, *4*, 2138-2148.
- 41 Renthal, R. An unfolding story of helical transmembrane proteins. *Biochemistry* **2006**, *45*, 14559-14566.
- 42 Otzen, D. E. Folding of DsbB in mixed micelles: A kinetic analysis of the stability of a bacterial membrane protein. *Journal of Molecular Biology* **2003**, *330*, 641-649.
- 43 Bedard, S., Mayne, L. C., Peterson, R. W., Wand, A. J. & Englander, S. W. The foldon substructure of staphylococcal nuclease. *J Mol Biol* **2008**, *376*, 1142-1154.
- 44 Llinas, M., Gillespie, B., Dahlquist, F. W. & Marqusee, S. The energetics of T4 lysozyme reveal a hierarchy of conformations. *Nat Struct Biol* **1999**, *6*, 1072-1078.
- 45 Dutta, A. *et al.* Characterization of membrane protein non-native states. 2. The SDS-unfolded states of rhodopsin. *Biochemistry* **2010**, *49*, 6329-6340.
- 46 Lemberg, M. K. & Freeman, M. Functional and evolutionary implications of enhanced genomic analysis of rhomboid intramembrane proteases. *Genome Res* **2007**, *17*, 1634-1646.
- 47 Dill, K. A. & Shortle, D. Denatured states of proteins. *Annu Rev Biochem* **1991**, *60*, 795-825.
- 48 Xue, Y. & Ha, Y. Large lateral movement of transmembrane helix S5 is not required for substrate access to the active site of rhomboid intramembrane protease. *J Biol Chem* **2013**, *288*, 16645-16654.
- 49 Wu, Z. *et al.* Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry. *Nat Struct Mol Biol* **2006**, *13*, 1084-1091.
- 50 Liu, T., Whitten, S. T. & Hilser, V. J. Functional residues serve a dominant role in mediating the cooperativity of the protein ensemble. *Proc Natl Acad Sci U S A* **2007**, *104*, 4347-4352.

- 51 Hong, H., Chang, Y. C. & Bowie, J. U. Measuring transmembrane helix interaction strengths in lipid bilayers using steric trapping. *Methods Mol Biol* **2013**, 1063, 37-56.

Chapter 3 Role of packing defects in the stability and function of an integral membrane protein

Ruiqiong Guo, Seung-Gu Kang, Zixuan Cang, Erin Deans, Guowei Wei and Heedeok Hong

This chapter will be submitted as an article in Journal of American Chemistry Society.

I would like to acknowledge Dr. Seung-Gu Kang and Zixuan Cang for their effort on modeling of protein structures and MD simulation. Dr. Seung-Gu Kang generated the data in **Figure 3.3** and **3.4** and **Table 3.2**. Zixuan Cang generate the structures in **Figure 3.1**. Erin Deans helped with the steric trapping experiment in determining the thermodynamic stability of some mutants. I would also like to thank Professor Guowei Wei for advice on the manuscript and experimental design.

Summary

Packing interaction is a critical driving force in the folding of membrane proteins. Despite of the importance, packing defects are prevalent in membrane proteins, and the role of the defects in the stability and function of membrane proteins is not well understood. Here we tackled this problem using the intramembrane protease GlpG of *E. coli* as a model by testing two hypotheses: 1) improving packing by cavity-filling mutations generally increase the protein stability; 2) if packing defects are critical for function, it would be possible to lock the protein conformation into either inactive or constitutively active state by modifying the size and distribution of the cavities. We designed 12 cavity-filling mutations and examined their impacts on the stability and activity. Despite improved packing, we only found three stabilized variants, the activity of which was substantially reduced. Interestingly, all stabilizing mutations were mapped onto the regions that are distant from the active site and possess conformational plasticity. We suggest that the packing defects facilitate functionally important movements of GlpG. Perhaps, packing in membrane proteins has evolved to delicately balance the stability and flexibility that is necessary for achieving optimal function.

Introduction

Globular proteins are efficiently packed to minimize the size of packing defects (*i.e.*, cavities including voids and pockets)¹. Indeed, the protein interior has a mean packing density (~ 0.74) similar to the crystals of small organic molecules (0.70–0.78) as well as the close-packed hard sphere model (~ 0.74)². On the other hand, protein structures are remarkably tolerant to amino acid substitutions and the size distribution of packing defects in the interior agrees well with that expected from the random-packed sphere model, suggesting their liquid-like nature³⁻⁵. While creating a cavity incurs the free energy cost of 25–30 cal/mol/Å³,⁶⁻⁸ increasing the protein size by

100 residues adds up the number of cavities by ~ 15 (7–8 voids and 7–8 pockets) on average and their sizes are broadly distributed from a few $\text{\AA}^3$ up to $\sim 1,000 \text{\AA}^3$ ⁵. Despite the unfavorable contribution to the protein stability, why are the packing defects so prevalent? The energetic penalty associated with cavity formation largely stems from the loss of van der Waals packing interactions^{6, 8-10}. Because globular proteins fold majorly driven by the hydrophobic effect rather than by packing, cavities may be allowed to form randomly as a consequence of folding⁵. In contrast, a number of studies indicate that certain cavities are strictly conserved for ligand binding, catalysis, allostery, transport or conformational changes that are necessary for function¹¹⁻¹⁵. With regard to the impact of cavities on protein stability, helical membrane proteins may serve as an important counter-example because, in contrast to globular proteins, packing interaction is known as a critical driving force in their folding^{7, 16-17}. In the bilayer where water molecules are scarce, the hydrophobic effect cannot drive folding of transmembrane (TM) helices into a compact native structure. Thus, to yield net stabilization, packing interaction between TM helices needs to overcome competing van der Waals interaction with lipids^{7, 16-17}. Thus, compared to globular proteins, the cavity formation in membrane proteins can compromise the stability and may be more tightly associated with function during evolution. Despite the importance, we notice that the mean packing densities of membrane proteins are not much different from those of globular proteins and this fact motivated us to systemically investigate the role of packing defects in the stability and function of membrane proteins.

We chose the six-helical bundle membrane protein GlpG of *E. coli*, an enzyme that belongs to the universally conserved rhomboid intramembrane protease family, as a model. So far, the contribution of packing to protein stability and function has been studied mainly by creating cavities using deletion mutations^{7, 17-18}. In this study, we approached this problem in an opposite

way, *i.e.* the role of each cavity was probed using small-to-large mutations which reduce the cavity size. The impact of cavity-filling mutations on the stability and activity of GlpG were examined by testing two hypotheses: 1) improving packing by small-to-large mutations generally stabilizes a protein, and 2) if packing defects are critical for function, it is possible to lock the protein conformation into either inactive or constitutively active state by modifying the cavity sizes.

Results

Identification of conserved cavities in the rhomboids

The mean interior packing density of GlpG is 0.722¹⁹, which falls in the typical density range of well-packed globular proteins^{5, 20} but is lower than that of other membrane proteins of known structure²¹. With a 1.4 Å-radius probe, we found a total of 24 cavities in GlpG on the CASTp server (**Table 3.1**)²². This number (13.4 cavities/100 residues) was similar to the average number of cavities in globular proteins (~15 cavities/100 residues) obtained using the same method⁵. Next, we searched for the cavities that are potentially conserved in the rhomboid family. Since only two rhomboid structures of distinct origins are available (GlpG of *E. coli* and GlpG of *H. Influenzae*, 40.1% sequence identity)²³⁻²⁴, we built the structural model of a distant homolog, human RHBDL2 (26.2% sequence identity relative to *E. coli* GlpG) using homology modeling and MD simulation (**Figures 3.1**). In the three rhomboid structures, cavities were distributed throughout the proteins and their size distribution was highly heterogeneous ranging from 3.9 Å³ to 523.6 Å³ (**Figure 3.1c**)⁵. To identify the cavities that are located in spatially similar regions, we superimposed the three backbone coordinates and performed cavity analysis directly on the superimposed structures using CASTp. The identified cavities would correspond to the free volumes common to all structures, or stem from the increased surface roughness by superposition. The fictitious cavities in the latter case were filtered out by mapping the identified cavities back to the structure of *E. coli*

GlpG (**Materials and Methods**). Out of the total 24 cavities in *E. coli* GlpG, 13 cavities were assigned as “common”.

ID	Area (Å ²)	Volume (Å ³)
1	118.237	141.676
2	127.691	150.600
3	75.831	75.817
4	23.502	30.178
5	84.111	55.615
6	23.367	22.337
7	106.600	82.963
8	59.469	47.280
9	1.701	3.862
10	26.503	17.464
11	16.545	11.181
12	26.404	16.067
13	1.928	1.989
14	37.510	20.431
15	39.751	21.496
16	38.255	16.155
17	42.874	17.331
18	30.012	15.225
19	19.778	8.330
20	29.572	14.613
21	27.025	13.189
22	26.376	12.708
23	25.574	11.868
24	14.992	6.692

Table 3.1 List of Cavities (voids and pockets) in *E. coli* GlpG identified on the CASTp server (<http://sts.bioe.uic.edu/castp/>) using a probe radius of 1.4 Å. The area and volume was calculated on the molecular surface of each cavity.

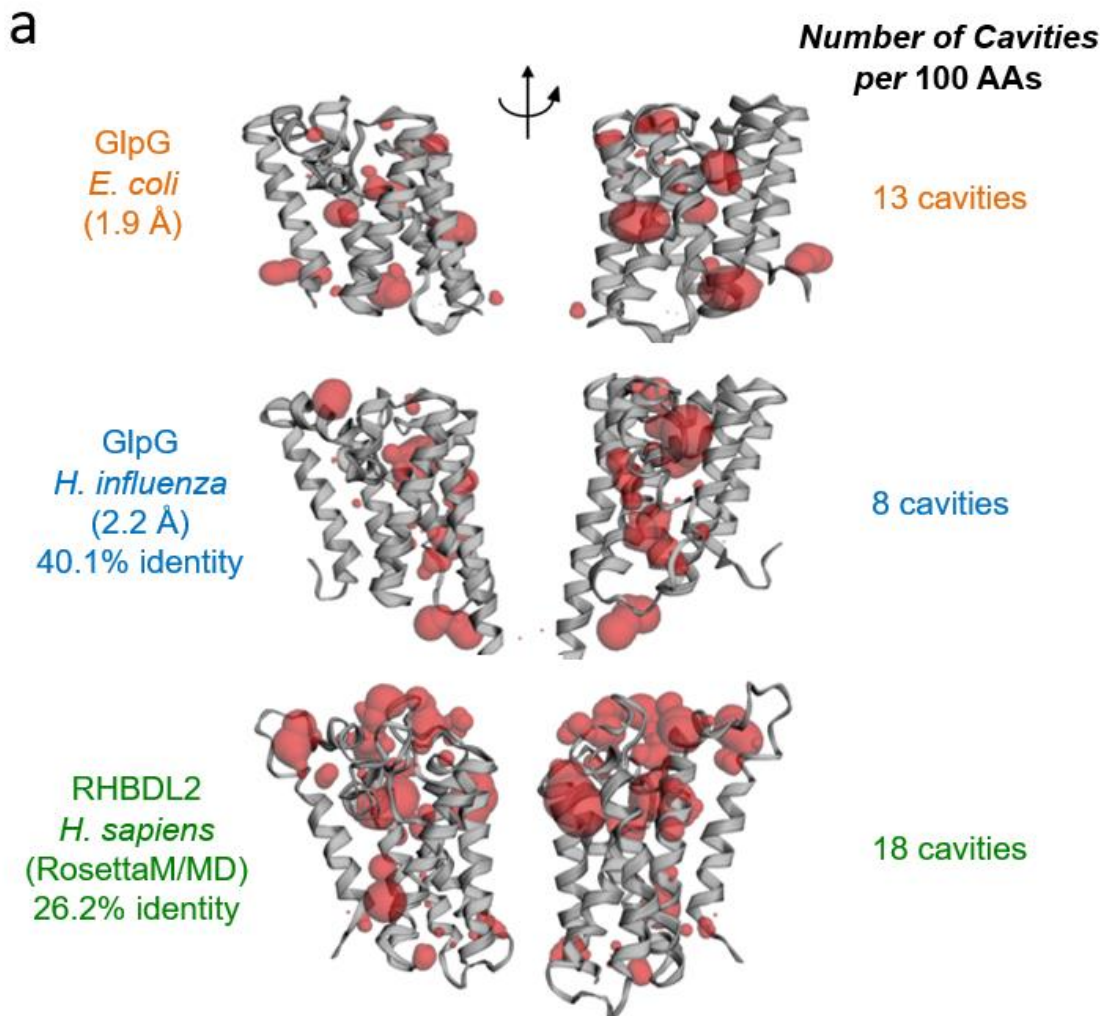
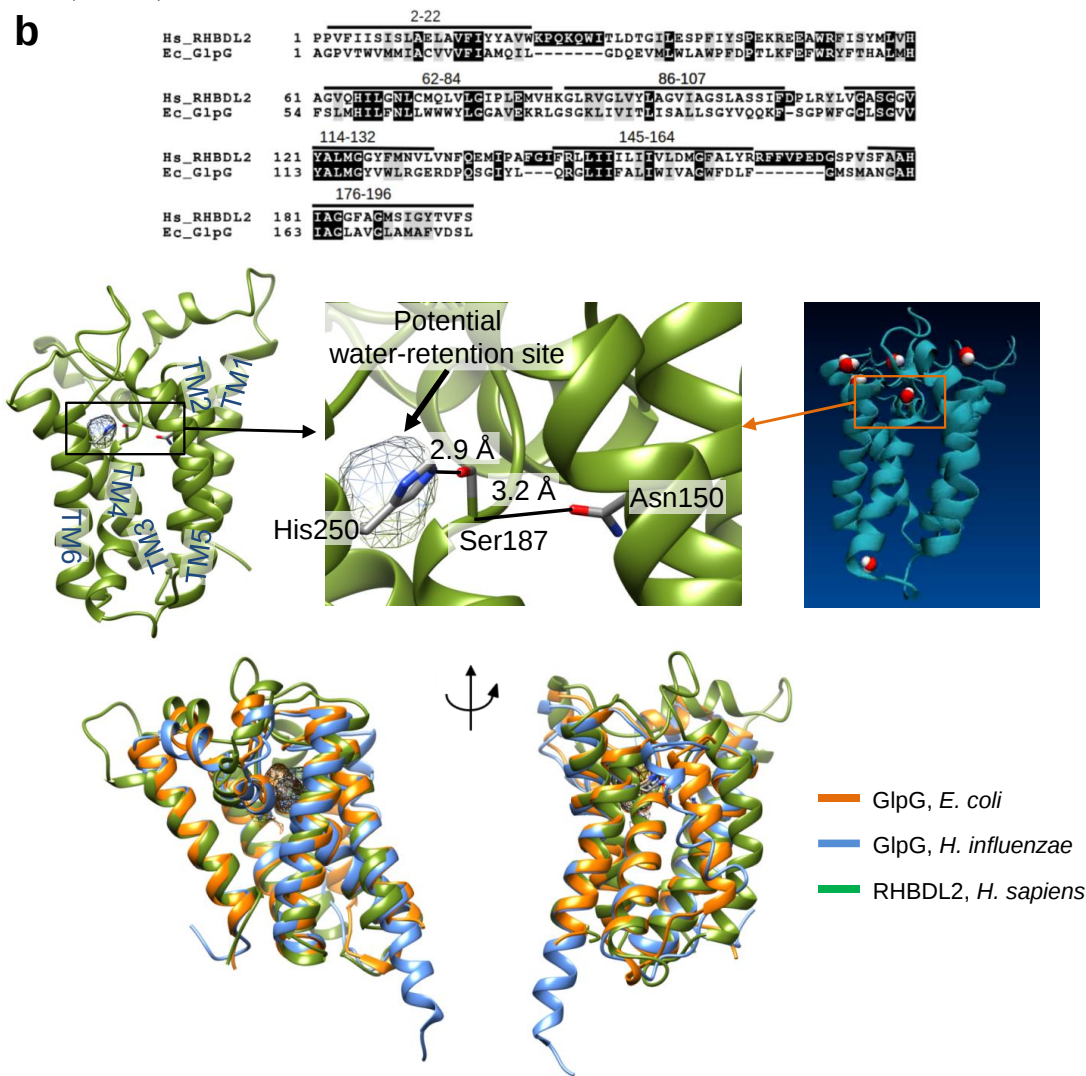


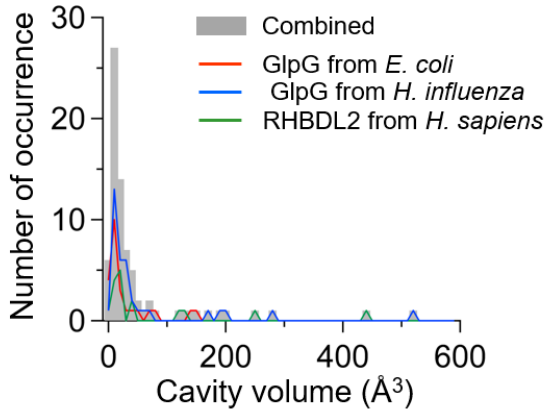
Figure 3.1 Cavities in rhomboid proteases. (a) Spatial distributions of the cavities in *E. coli* GlpG (3B45), *H. influenzae* GlpG (2NR9) and *H. sapiens* RHBDL2 (modeled, see **Figure 3.1b**) obtained using the CASTp server. (b) (Top) the sequence alignment of human rhomboid protease RHBDL2 and *E. coli* GlpG. The predicted six transmembrane helices are marked with their residue numbers. (Middle) structure of RHBDL2 obtained by homology modeling using Rosetta Membrane and MD simulation in a POPC bilayer. The validity of the structure was evaluated in three aspects: 1) the spatial proximity among the catalytically important residues including the Ser-His catalytic dyad (Ser187-His250) and a putative oxanionhole (Asn150); 2) the presence of a potential water-retention site near the dyad that are commonly found in the crystal structures; 3) the penetration and residence of several water molecules near the catalytic dyad during MD simulation. A structural snapshot with penetrated water molecules are shown in right, supporting its proteolytic activity involving water. (Bottom) Superimposed structures of *E. coli* GlpG (PDB code: 3B45), *H. Influenzae* GlpG (2NR9) and *H. sapiens* RHBDL2. The RMSD of the C α pairs between 3B45 and 2NR9 was 1.171 Å, and the RMSD between 2NR9 and RHBDL2 was 1.056 Å. (c) The size distributions of the cavities identified from *E. coli* GlpG (3B45), *H. influenzae* GlpG (2NR9) and *H. sapiens* RHBDL2 (modeled). The histograms include the distributions of the cavity volumes identified from individual structures as well as their combined distribution.

Figure 3.1 (cont'd)

b



c



Mutational effects of the cavity-filling mutants evaluated by computational methods

Among the 13 common cavities, we selected five cavities whose volume was larger than 40 Å³ (Cavities I–V, **Figure 3.2**) and designed 11 small-to-large (*i.e.*, cavity-filling) mutations for the residues contacting these cavities. This cut-off volume corresponds to the approximate volume difference between Ala and Val²⁵ (25–50 Å³) allowing room for cavity-filling mutations. The effects of each substituted residue on the structure and dynamics of GlpG were examined using all-atomic prolonged MD simulations (up to ~150 ns) under explicit solvent and bilayer conditions. The backbones of the cavity-filled variants fluctuated within the RMSD of 0.5–1.5 Å relative to that of crystal structure 2IC8 (**Figure 3.3a**), indicating the compatibility of substituted residues with the template structure. The RMSF's of MD simulated WT and variants exhibited large fluctuations in the loop regions and the relative rigidity of the TM helices (**Figure 3.3b**). Large amplitude movements of TM helices was not observed during simulation.

Does each cavity-filling mutation truly induce the volume reduction of the targeted cavity? To answer this question, we took two approaches using the structural snapshots of all variants obtained from MD simulation: (1) measuring the cavity volumes at all time frames and (2) evaluating the degree of packing around the cavity-contacting residues using the occluded surface packing (OSP) analysis²⁰. We observed large volume fluctuations for all targeted cavities with the standard deviations of 20–100% relative to the average volumes (**Table 3.2** and **Figure 3.4a**). The mutations modified the volumes of not only the cavities where the mutations were made, but also other cavities where mutations were not made. Notably, the large volume fluctuations mainly stemmed from transient connections and separations among adjacent cavities, making the clear description of the mutational effects on the cavity size difficult. On the other hand, the OSP analysis provided direct information on the mutational effect on local packing (**Figure 3.4b**). In

each variant, the OSP values consistently increased for the residues surrounding the cavity targeted for mutation, indicating improved packing. However, the OSP values also changed for other residues throughout the proteins in a random manner. Our MD simulation combined with volume and packing analysis shows fundamental insights into the nature of the protein interior and mutational effects on the structure and dynamics of proteins: (1) The protein interior is highly fluidic so that the connection, separation, appearance and disappearance of the packing defects consistently occur; (2) a mutation appears to bear a desired effect on the local region, but the perturbation appears randomly propagated to other regions of the protein.

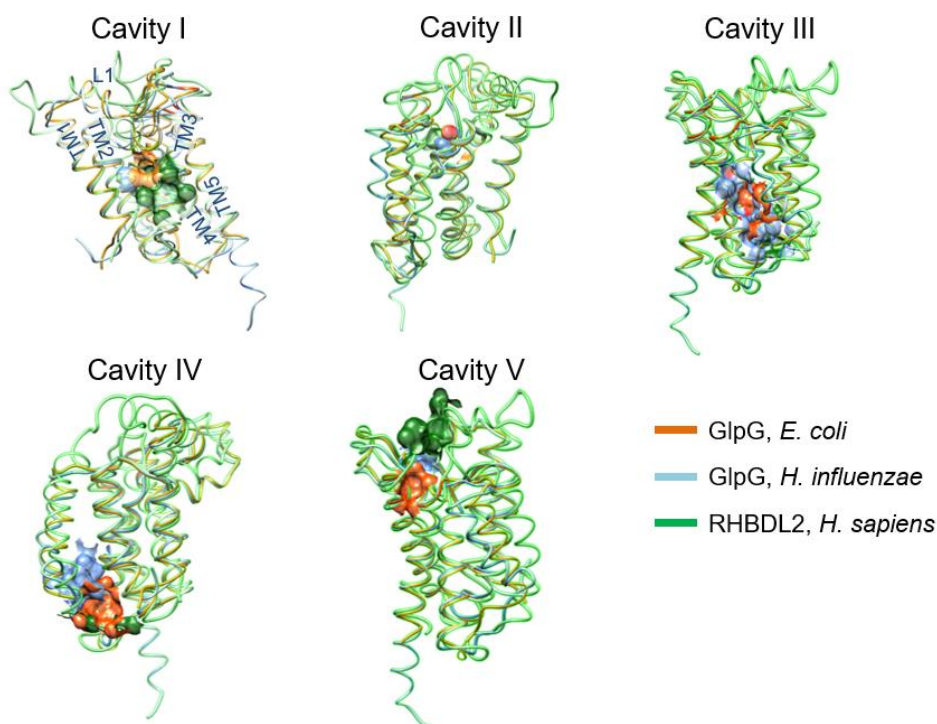


Figure 3.2 The common cavities in the structures of rhomboid proteases, *E. coli* GlpG, *H. influenzae* GlpG and *H. sapiens* RHBDL2. The larger cavities ($V_{ms} > 40 \text{ \AA}^3$) in *E. coli* GlpG and the cavities in the other two structures that are found in spatially similar regions are depicted. Each cavity is displayed by the surface presentation of the residues surrounding the cavity.

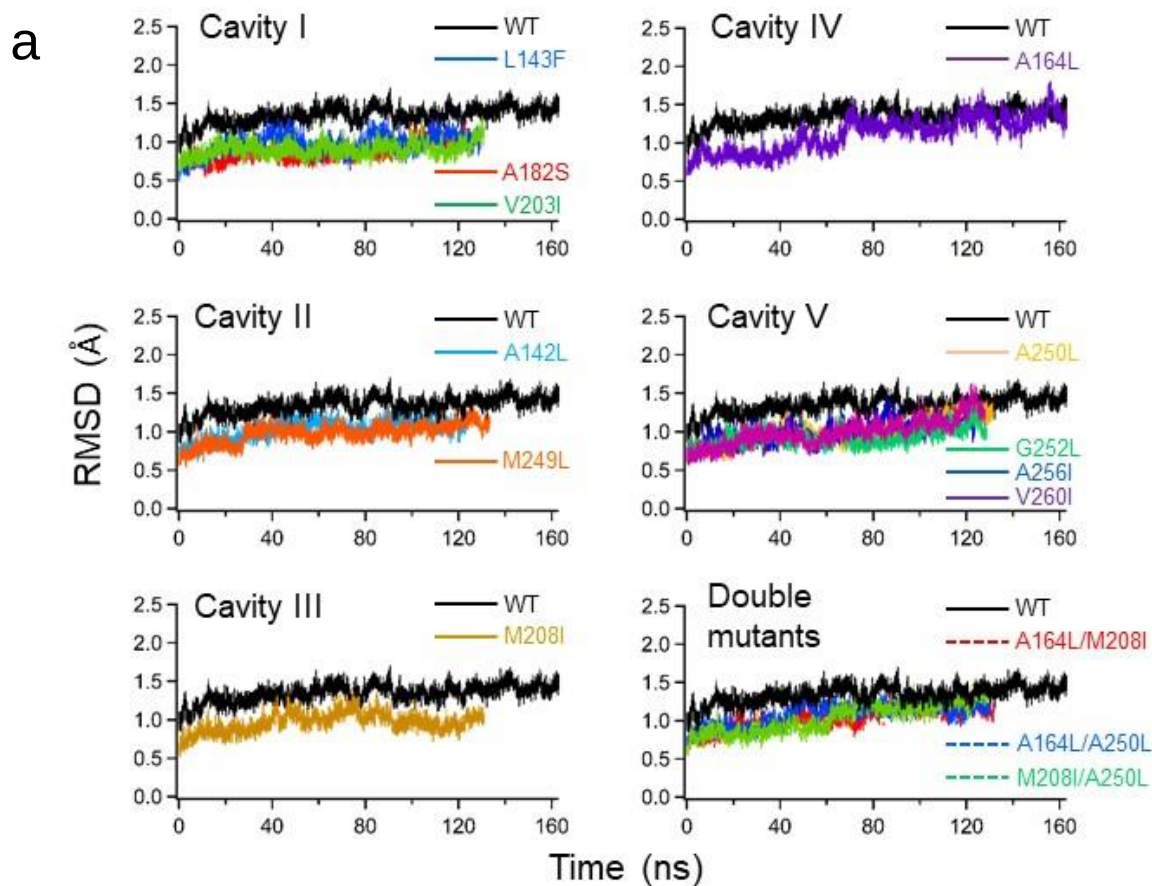
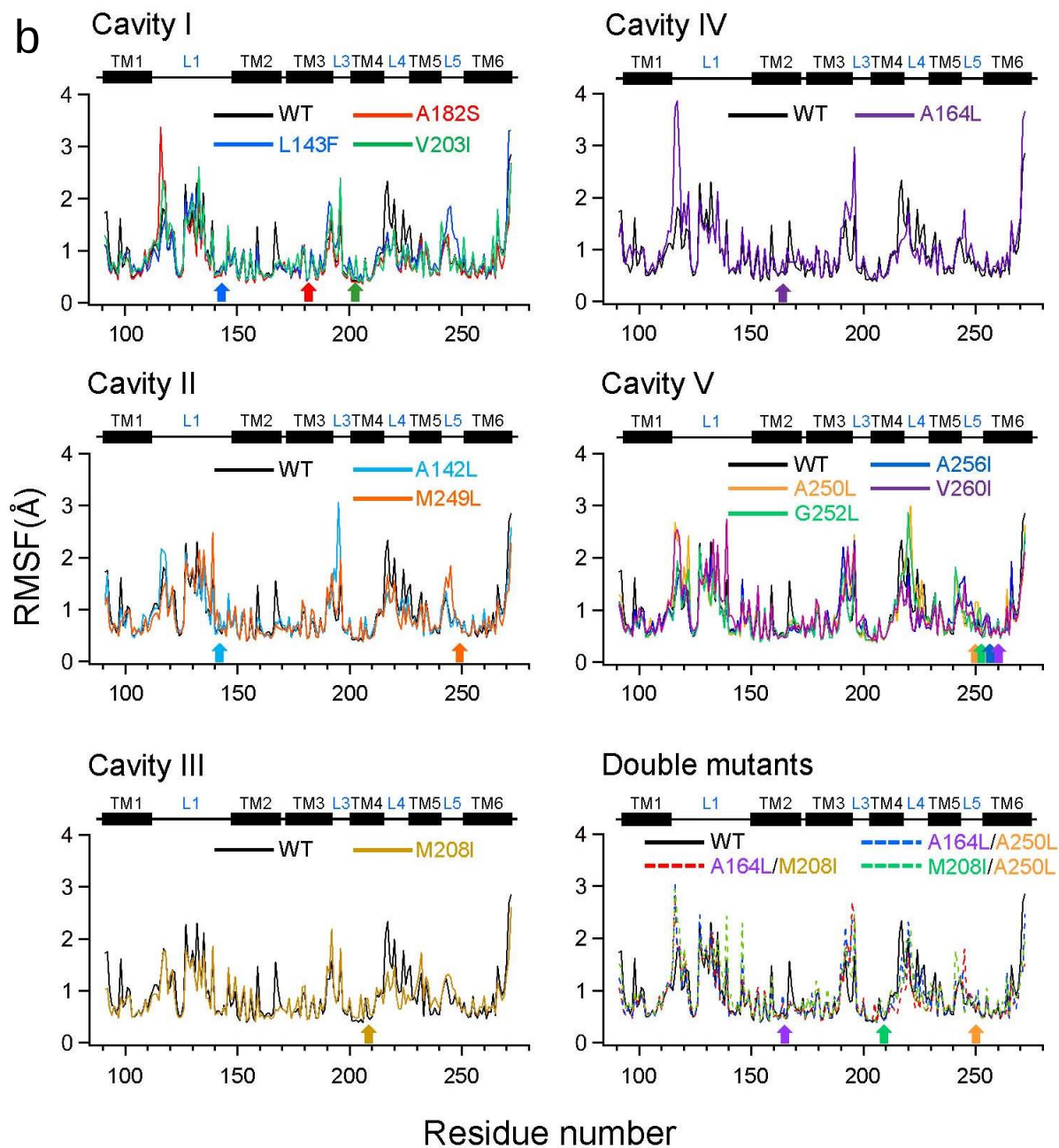


Figure 3.3 MD simulation result of *E. coli* GlpG and its cavity-filled variants in explicit bilayer (POPE/POPG, molar ratio = 3:1) and water. (a) The root-mean-square-displacements (RMSD's) of the C_{α} atoms in each GlpG variant were calculated relative to those in the wild type crystal structure (2IC8) during simulation. **(b)** The root-mean-square-fluctuations (RMSF's) of the residues in each GlpG variant were calculated relative to those in the wild type crystal structure (2IC8).

Figure 3.3 (cont'd)



Cavity targeted by mutation	Mutation	Cavity I		Cavity II		Cavity III		Cavity IV		Cavity V	
		<Vol>	σ_{vol}	<Vol>	σ_{vol}	<Vol>	σ_{vol}	<Vol>	σ_{vol}	<Vol>	σ_{vol}
	WT	63.5	15.7	15	8.1	85.2	19.7	63	26.6	16.7	7.7
Cavity I	L143F	81.9	18.8	9.5	5.9	116.9	22.9	77.2	19.4	12.3	7.3
	A182S	70.8	18.1	10.1	5.3	122.8	21.2	75.9	18.4	10.0	6.5
	V203I	78.1	19.2	10.2	6.1	124.8	22.1	79.5	18.1	13.4	8.2
Cavity II	A142L	58.8	16.2	4.4	4.6	118.4	22.8	76.3	19.6	13.2	8.3
	M249L	80.7	25.6	8.4	5	106.7	20.5	77.9	18.2	7.9	6.9
Cavity III	M208I	76.7	20.5	6.3	5.3	84.5	19.4	81.6	18.3	11.4	6.8
Cavity IV	A164L	85.5	21.2	7.0	6.1	111.0	21.1	68.9	18.1	12.1	7.2
Cavity V	A250L	72.6	20.4	5.0	4.9	128.1	22.7	72.6	19.9	14.0	9.3
	G252L	77.9	18.1	10.1	5.8	121.9	21.8	80.1	21.4	14.7	8.1
	A256I	75.5	17.4	8.4	5.6	114.7	20.0	87.2	19.4	17.3	9.8
	V260I	70.7	18.6	8.4	5.6	115.9	19.7	78.6	19.3	13.1	7.3
Cavities III & IV	A164L/M208I	71.4	17.0	7.8	5.4	98.4	20.8	67.5	17.5	12.5	7.6
Cavities IV & V	A250L/A164L	74.4	16.4	7.0	6.7	132.5	21.9	55.1	16.1	14.5	8.5
Cavities III and V	A250L/M208I	62.8	22.8	20.0	13.3	100.6	22.7	89.1	18.6	12.4	6.6

Table 3.2 Statistics of the volume fluctuation of the cavities during MD simulation. All numbers including the average volumes (<Vol>) and the standard deviations (σ_{vol}) are in $\text{\AA}^3$. The numbers in bold correspond to the <Vol> and σ_{vol} of the cavities targeted by the cavity-filling mutations.

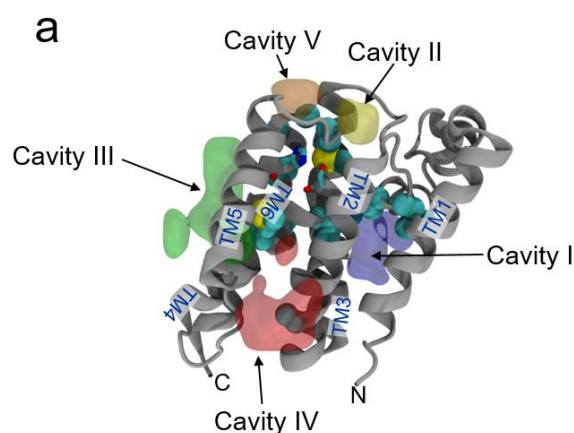


Figure 3.4 Analysis of cavity volume and packing on the snapshot structures from MD simulation. (a) Display of the major cavities in a structural snapshot from MD simulation of WT GlpG. Each cavity shows the void volume, which is presented at least over 10% of simulation time. **(b)** The difference occluded surface packing (OSP) values between WT and each variant ($\Delta\text{OSP}[X] = \text{OSP}_{\text{variant}}[X] - \text{OSP}_{\text{WT}}[X]$, X denotes the residue number). A positive value indicates improved packing around the residue. The mutated residue and the residues surrounding the cavity targeted for mutation are marked with blue and orange, respectively.

Figure 3.4 (cont'd)

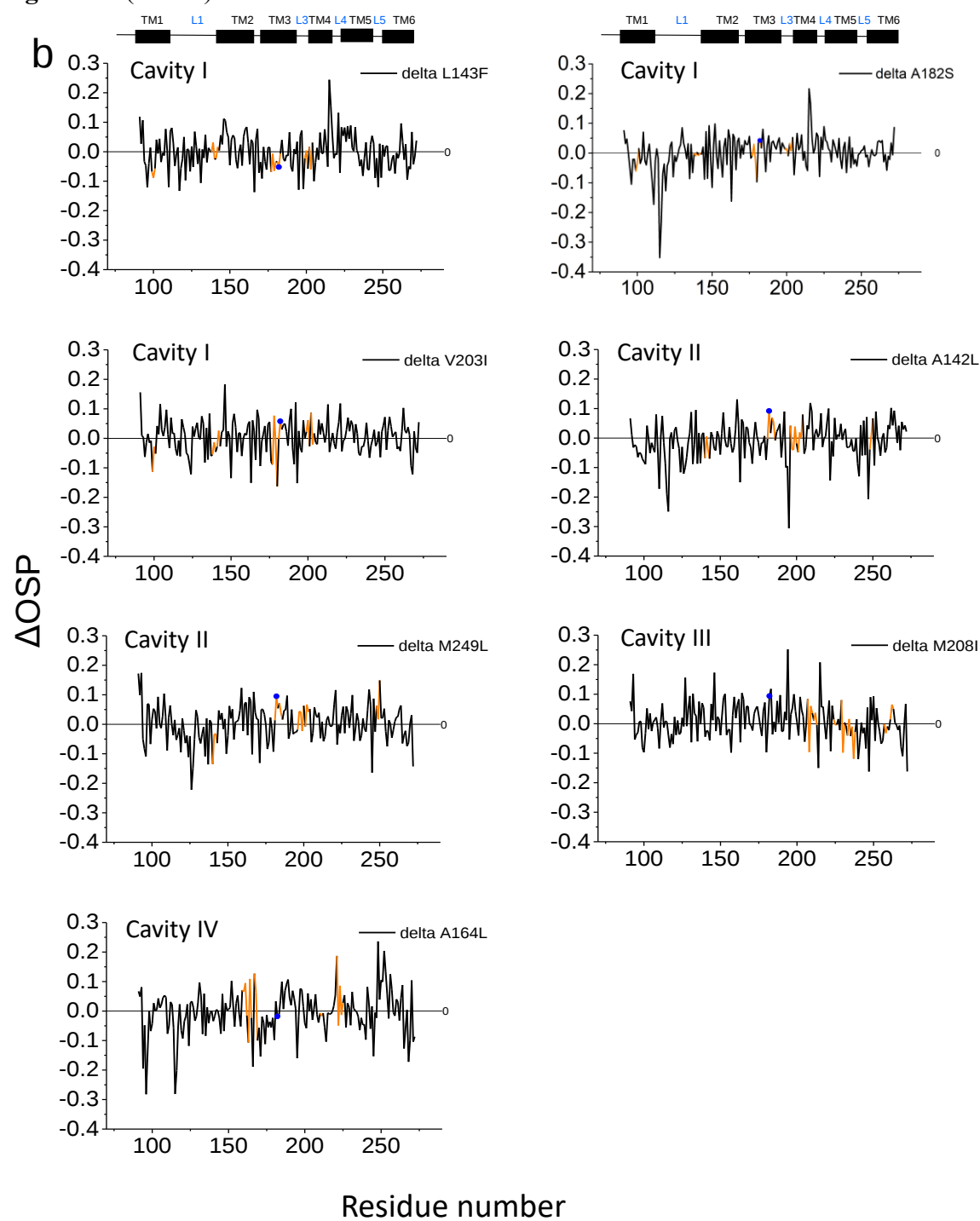
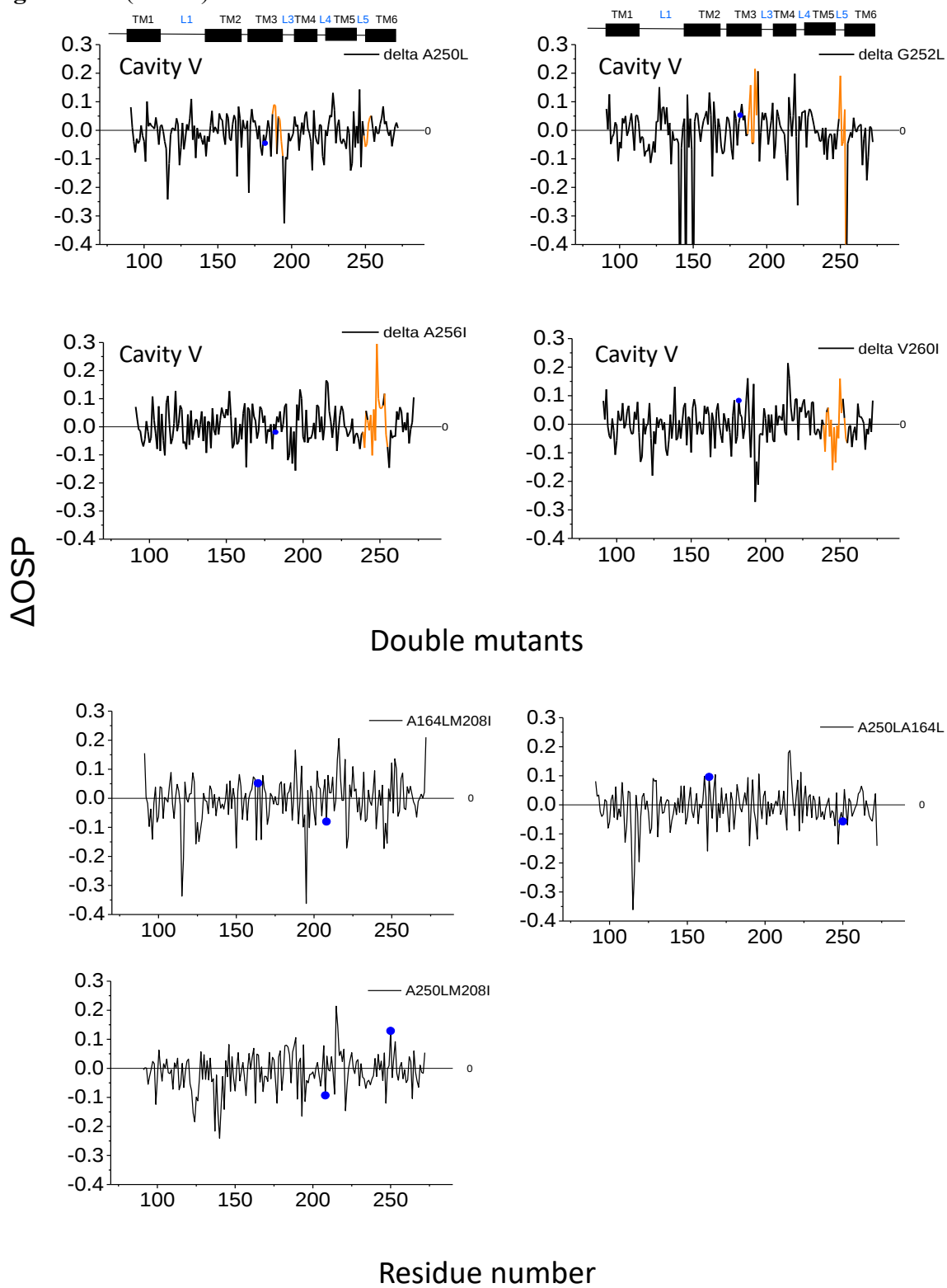


Figure 3.4b (cont'd)



The effect of cavity-filling mutants on stability and activity of GlpG

Next, we investigated the impacts of each cavity-filling mutation on the stability and activity of GlpG in dodecylmaltoside (DDM) micelles, which are widely used for the folding and functional studies of GlpG. To measure the stability, we employed the steric trapping strategy, which couples unfolding of a doubly-biotinylated protein to competitive binding of bulky monovalent streptavidin (mSA). Steric trapping is advantageous over conventional chemical denaturation methods because the thermodynamic stability (ΔG°_U) of proteins can be directly measured under native lipid and solvent conditions. The mutants were made in the background of a double-biotin variant 172/267C-biotin₂, in which the biotin pair approximately covers the C-terminal half of GlpG. The proteolytic activity of GlpG was measured using the second TM segment of lactose permease of *E. coli* (LYTM2) as a model substrate. For details of the steric trapping method for stability and activity, please refer to Chapter 2.

Interestingly, the impacts of cavity-filling mutations on stability and activity exhibited a unique pattern depending on the targeted cavity. First, we found several mutations (L143F, A182S and V203L; the increases in the side chain volume, $\Delta\Delta V_{sc} = 4.1\text{--}28.9 \text{ \AA}^3$) that did not significantly change either stability or activity (**Figure 3.5a**). All these mutations were mapped onto the same cavity (Cavity I: molecular volume, $V_{ms} = 55.6 \text{ \AA}^3$) surrounded by the residues in TM1, L1, TM3, L3 and TM4 (**Figure 3.2**). This cavity deeply penetrates into the core of GlpG and is open towards the bilayer center. Notably, in 14 out of 26 available structures of *E. coli* GlpG (www.rcsb.org), this cavity is occluded by single or multiple detergent or lipid tails. Therefore, by the small-to-large mutations, it is likely that the van der Waals interactions between hydrocarbon tails and cavity will be replaced by the protein packing interactions. However, the improvement of the local protein packing by the mutations did not lead to the net stabilization of GlpG.

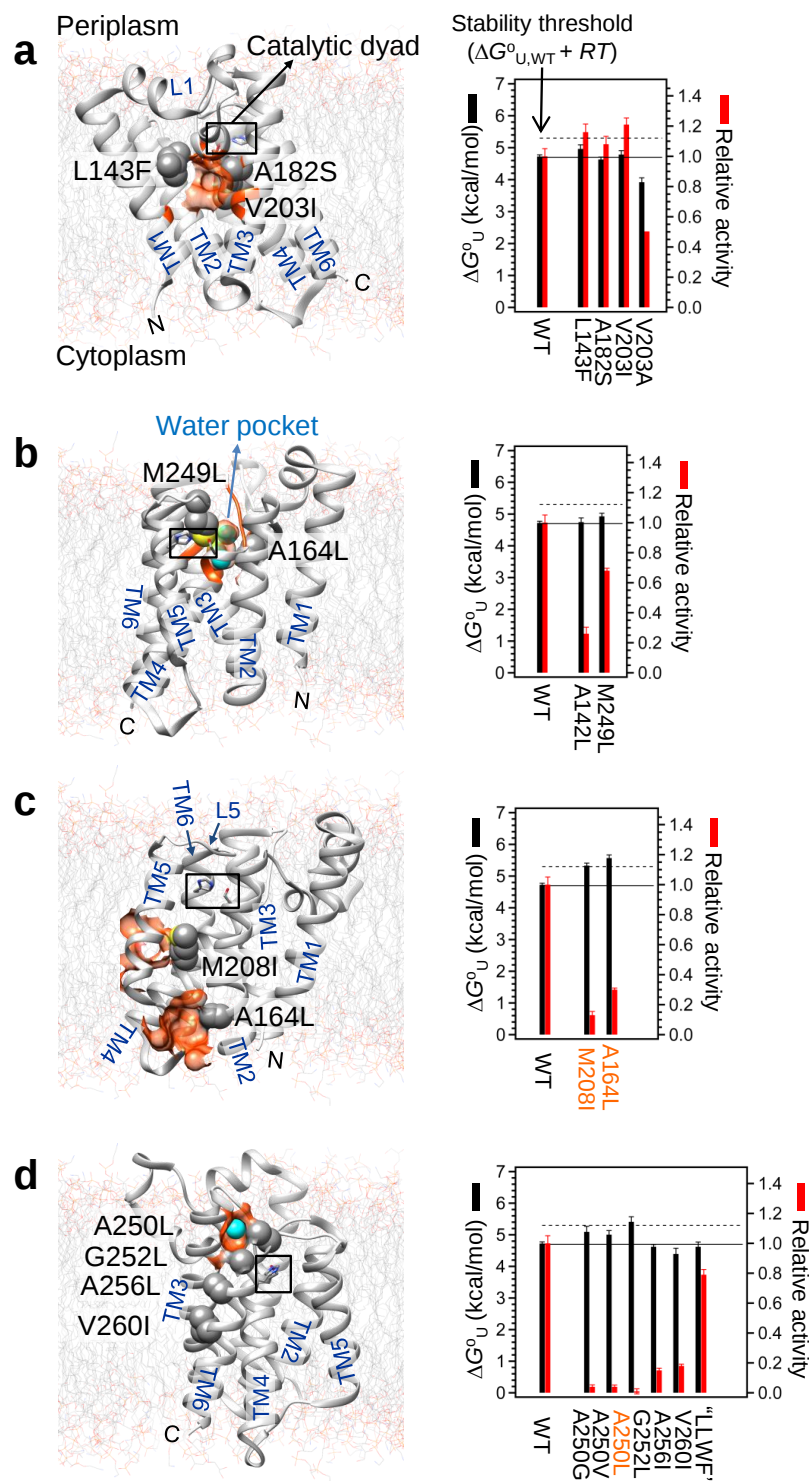


Figure 3.5 Impacts of single cavity-filling mutations on the stability and activity of GlpG (a-d) (Left) Locations of the targeted cavity and mutation sites. (Right) The stability (ΔG°_U) and activity of cavity-filled variants. “The stability threshold” denotes the stability level above which is defined as significant stabilization. This threshold was set to $\Delta G^{\circ}_{U,WT} + RT$, where $\Delta G^{\circ}_{U,WT}$ and RT designates the stability of WT and thermal fluctuation energy (0.59 kcal/mol), respectively.

Next, we targeted the water-filled cavity near the active site (Cavity II, $V_{ms} = 83.9 \text{ \AA}^3$) (**Figure 3.2**). This cavity contains three crystallographic water molecules which form five hydrogen bonds with the protein²⁶, and commonly exists in all three rhomboid structures. This cavity is thought to serve as a water-retention site, providing water molecules for the proteolytic active site²⁶. Large perturbation of the cavity by A142L ($\Delta V_{sc} = 55\text{--}75 \text{ \AA}^3$) substantially reduced the activity by $\sim 70\%$, while mild perturbation of the same cavity by M249L ($\Delta V_{sc} = -6.6\text{--}3.3 \text{ \AA}^3$) led to the moderate reduction of activity (**Figure 3.5b**). Despite the activity reduction, both mutations retained the stability. Therefore, improving the protein packing at the expense of perturbing the native water-protein interaction did not induce net protein stabilization. Taken together, this result confirms that the water-filled cavity indeed plays an important role in the proteolytic function of GlpG. Also, the formation of the water-filled cavity seems to be neutral to the protein stability because of favorable water-protein hydrogen bonding interactions as well as van der Waals interactions between them.

The crystal structures of GlpG indicate that, although its backbone fold is overall rigid, the segment L5-TM5 exhibit considerable plasticity. For example, in a few structures, TM5 is largely unfolded^{23, 27}, while a variation of its tilt angle relative to the bilayer normal have been observed²⁷⁻²⁹. Intriguingly, several deletion mutations at the TM2-TM5 interface dramatically enhance the activity^{18, 30}. Thus, TM5 has been assigned as a “gate” for the entry of TM substrates³⁰. Also, the loop L5 connected to TM5 displays “closed cap”, “open cap” or disordered conformations whose flexibility has been suggested to be a mechanism to allow the substrate access to the active site³¹⁻³³.

We found large cavities (Cavity III and IV) involving the gating helix TM5, which may form a structural basis of the flexibility of this region (**Figure 3.2**). Surprisingly, mutations targeting these

cavities significantly enhanced the stability but substantially reduced the activity (**Figure 3.5c**). M208I mutation, which modified the large free space (Cavity III, $V_{ms} = 141.7 \text{ \AA}^3$) between TM4 and TM5 stabilized GlpG by $\Delta\Delta G^0_U = 0.6 \pm 0.2 \text{ kcal/mol}$ ($>RT$, thermal fluctuation energy), but almost completely inactivated the protein. This inactivation is surprising because the mutated site is not only distant from the catalytic dyad but also not interfering with the predicted substrate entry site. A164L mutation on the cavity between TM2 and TM5 (Cavity IV, $V_{ms} = 64.5 \text{ \AA}^3$) induced the largest stabilization ($\Delta\Delta G^0_U = 1.0 \pm 0.2 \text{ kcal/mol}$) among tested mutants with the $\sim 70\%$ reduction in activity.

We also found a cavity located at the periplasmic end of the TM3-TM6 interface (Cavity V, $V_{ms} = 47.3 \text{ \AA}^3$) (**Figure 3.2**). TM6 harbors His254 which constitutes the catalytic dyad with Ser201. In the structural and mechanistic study of GlpG using the mechanism-based inhibitor diisopropyl fluorophosphonate (DFP), the Ha group observed an outward pivoting motion of TM6 involving a rotation of His254 towards a groove between TM5 and TM 6 upon binding of the inhibitor. Therefore, the flexibility of TM6 is important to allow His254 to carry out this critical catalytic step. In an effort to modify the pivoting motion of TM6, we step wisely increased the side chain volume at the residue 250 (A250V and A250L, $\Delta V_{sc} = 35\text{--}50$ and $55\text{--}80 \text{ \AA}^3$, respectively) on Cavity 5. Interestingly, the mutations increased the stability by 0.4 ± 0.2 and $0.7 \pm 0.2 \text{ kcal/mol}$, respectively, whereas completely inactivating (**Figure 3.5d**). The possible reason is that the bulky side chains blocks the pivoting motion of TM6, preventing His254 from rotating into the catalytically competent position ideal for carrying out the hydrolysis reaction. While for mutation G252L ($\Delta V_{sc} = 75\text{--}102 \text{ \AA}^3$) regain some activity compared to A250 mutants. Interestingly, the structure complexed with another mechanism-based inhibitor, isocoumarin exhibits an inward movement of TM6, opposite to the effect of DFP³³. Those conformational changes indicate the

involvement of TM6 for the catalytic cycle of GlpG. Our result combined with the previous structural studies suggests that TM3-TM6 helical interface harboring Cavity V serve as a “hinge” in this movement. To further test this suggestion, we successively introduced a bulky side chain at the lipid-exposed face on TM6 farther down from the hinge region (A256I and V260I). Indeed, these mutations gradually restored the activity without significant changes in the stability. Therefore, our strategy using cavity-filling mutation can be powerful to delineate the functionally important movement of proteins.

The interaction free energy between three stabilizing mutants

We identified three single mutations (A164L, M208I and A250L) that significantly stabilized GlpG ($\Delta\Delta G^\circ_U > RT$) (**Figure 3.6a**). All of these mutations were mapped onto the more flexible region of the protein and inactivating. This result implies that these cavity-filling mutations “locked” GlpG into the inactive conformations by inhibiting the functionally important “gating” (A164L, M208I) or “hinge” (A250L) movements. Are the stabilization effects of these mutations additive? To answer this question, we generated three pairwise double mutants and the triple mutant. Although all multiple mutants exhibited enhanced stability relative to WT, the stabilization effects of individual single mutations were not additive (**Figure 3.6b**). Thermodynamic cycle analysis to measure the interaction free energy ($\Delta\Delta G_{\text{Inter}}$) between two mutated sites yielded slight positive cooperative between A164L and M208I ($\Delta\Delta G_{\text{Inter}} = -0.7$ kcal/mol) and moderate negative cooperativity ($\Delta\Delta G_{\text{Inter}} = 1.2\text{--}1.6$ kcal/mol) between A164L and A250L and between M208I and A250L (**Figure 3.6c**). Interestingly, the weak positively cooperative interaction occurred between spatially close Cavity III and Cavity IV, both of which are located in the regions around the gating helix TM5. In contrast, the stronger negative cooperativity was observed between the TM3-TM6 hinge region (Cavity V) and the gating helix (Cavity III and Cavity IV). This result implies that

the flexible C-terminal region of GlpG (L5, TM5 and TM6) is a heterogeneous ensemble of multiple subsets of conformations, and stabilizing one subset by mutation may stabilize another or some of these subsets can be mutually exclusive. Based on this analysis, we suggest that the “gating” motions involving TM5 and the “catalytic” motions involving TM6 may not occur at the same time during the proteolytic cycle of GlpG.

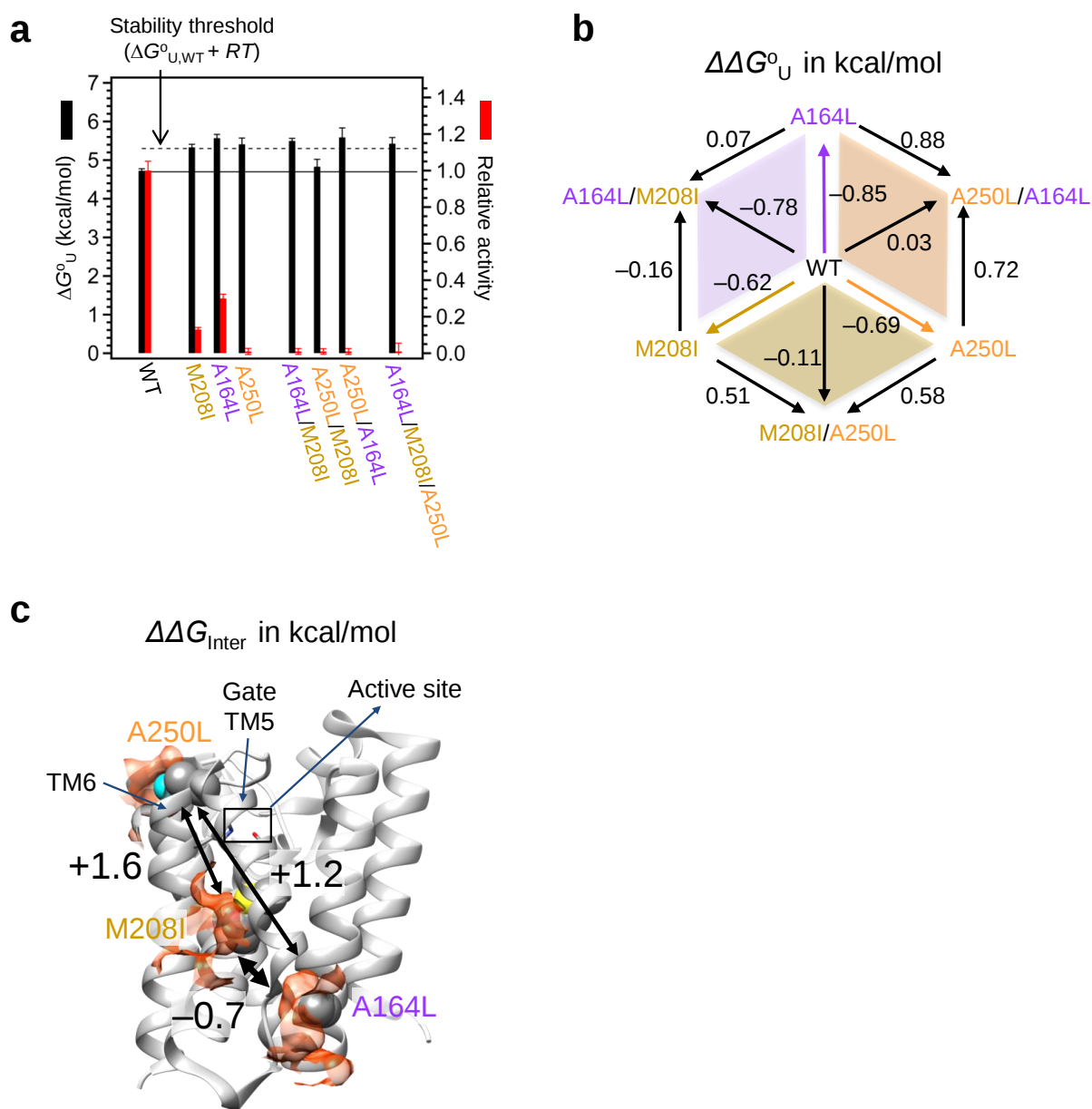


Figure 3.6 Additivity and cooperativity of stabilizing cavity-filling mutations. (a) The impacts of double and triple mutations combining individual stabilizing mutations on the stability and activity of GlpG. **(b)** Thermodynamic cycles describing the stability changes associated with the

stabilizing single and corresponding double mutations. The interaction energies ($\Delta\Delta G_{\text{Inter}}$) representing the degree of energetic coupling between a specific residue pair is calculated using this plot. **(c)** Cooperativity between the sites of the stabilizing mutations. The positive $\Delta\Delta G_{\text{Inter}}$ indicates the negatively cooperative interaction; where as the negative $\Delta\Delta G_{\text{Inter}}$ represents the positively cooperative interaction.

Cooperativity interactions of the stabilizing mutants

Lastly, we asked the question whether the stabilizing interactions induced by the cavity-filling mutations would be effective only in the flexible C-terminal region or propagated throughout the protein. We answered this question by further measuring GlpG stability using the biotin pair located at the N-terminal region (95/172_N-biotin₂)³⁴. By design, steric trapping captures transient opening of a specific biotin pair, thus enabling measurement of the local stability of the region encompassing the biotin pair³⁴. Interestingly, the same mutations (A164L, M208I and A250L) that significantly stabilized the C-terminal region of GlpG ($\Delta\Delta G^{\circ}_{\text{U}} = -0.6 \sim -1.1$ kcal/mol) did not lead to a similar degree of stabilization of the N-terminal region ($\Delta\Delta G^{\circ}_{\text{U}} = -0.3 \sim +0.1$ kcal/mol). Therefore, the effects of the stabilizing mutations seem to be largely limited to the more flexible C-terminal region.

Discussion

In this study, we elucidated how packing defects in the membrane proteins modulate the stability and function of an intramembrane protease GlpG. Improving the packing by engineering cavities did not generally lead to protein stabilization. Interestingly, stabilizing mutations were all mapped onto the regions with conformational flexibility and of functional importance. By modifying the cavity sizes, we were able to trap GlpG into multiple inactive conformations and delineate functionally critical movements and their coupling during the catalytic mechanisms. The cavities in membrane proteins appears to be randomly distributed and highly dynamic, thus identifying

cavities of functional may be challenging. In membrane proteins, the existence of cavities seems to be critical to functional movements but may compromise the stability. Our integrated computational and experimental tools using cavity-filling mutations may serve as a strategy to elucidate the structure-function-stability relationship of membrane proteins.

Materials and Method

Homology Modeling and Molecular Dynamic (MD) Simulation of Human Rhomboid Protease RHBDL2.

The sequence alignment of human RHBDL2 and *E. coli* GlpG and the predicted regions of transmembrane (TM) helices were obtained from Lemberg and Freeman (**Figure 3.1b**)³⁵. The alignment and specified location of TM helices are passed to the Rosetta software suite (Rosetta3, build 2016.32.58837) for the comparative modeling of membrane proteins with multiple templates². Three structures (PDB IDs: 2CI8, 2XOV, and 3B45) were used as templates. The fragment library for the target was generated with the online server Robetta³⁶. 10 models were constructed and the three with the lowest energy were selected for relaxation.

The modeled structure with the lowest energy after relaxation was chosen as the starting point of the subsequent molecular dynamics (MD) simulation for further relaxation. The structure was aligned to an POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine, C_{16:0}C_{18:1c9}PC) explicit membrane (65Å x 65Å) constructed using the Membrane Builder module in VMD (Visual Molecular Dynamics, version 1.9.2)³⁷. Then the system was solvated on VMD. A sequence of scripts were executed to combine the components of the system, the protein, the water molecules, and the lipids. The collisions were resolved by first deleting lipids that overlap with the protein, and then deleting the water molecules that overlapped with the protein or the lipids. Using the MD software NAMD (Nanoscale Molecular Dynamics, build CVS-2016-06-26 for Linux-x86_64-

multicore-CUDA)³⁸ and the NAMD protocol for membrane protein simulation³⁹, the system underwent a 10,000 step energy minimization, a 0.5 ns simulation with the protein fixed to melt the membrane, another 0.5 ns simulation with the protein restrained, and finally a 4 ns simulation with everything released. The force field used in CHARMM27⁴⁰. The purpose of this simulation was by no means to explore the mechanism of the protein but to improve the stability of the modeled structure and to confirm the validation of the structure. The last frame of the MD simulation was used for structural analysis.

Identification of Common Cavities among Three Rhomboid proteases

To analyze the cavity properties among the rhomboid family, we superimposed the modeled structure of human RHBDL2 (Hs_RHBDL2) and the experimentally determined structures of the two bacterial rhomboid proteases, GlpG of *E. coli* (PDB: 3B45) and GlpG of *Haemophilus influenzae* (PDB: 2NR9). 3B45 and 2NR9 were structurally aligned to Hs_RHBDL2 using the Matchmaker tool in the UCSF Chimera5 software. The RMSD for 88 matched C α pairs between 3B45 and Hs_RHBDL2 was 1.171 Å, and the RMSD of 67 C α pairs between 2NR9 and Hs_RHBDL2 was 1.056 Å. The superimposed as well as the separate structures were submitted to the CASTp (Computed Atlas of the Surface Topography of proteins) server²². The cavities in each protein, the sequence of each protein, the result of structure superposition are shown in **Figure 3.1a-b**. To identify the cavities that are located in spatially common regions in three rhomboid structures, we performed the cavity analysis directly on the superimposed structure using CASTp. The identified cavities would correspond to the free volumes common to all structures, or stem from the increased surface roughness upon superposition. The fictitious cavities in the latter were filtered out by mapping them back onto the structure of *E. coli* GlpG. If the heavy atoms surrounding a certain cavity in *E. coli* GlpG shared at least one atom with the atoms from *E. coli*

GlpG that participated in the formation of the corresponding cavity in the superimposed structure, we defined the cavity surrounded by the shared atoms as “common”. This procedure was repeated for the cavities in *H. influenza* GlpG and Hs_RHBDL2. Out of the total 24 cavities in *E. coli* GlpG, we identified 13 overlapping cavities as “common”. A quantitative analysis of cavity conservation among the proteins was also performed.

Double mutant cycle analysis

To measure the pairwise interaction energies between cavity-filled mutation sites, double-mutant cycle analysis was employed⁴¹. A double-mutant cycle involves wild type protein (WT), two single mutants and the corresponding double mutant. If the change in thermodynamic stability (ΔG^0_U) upon the double mutation ($\Delta\Delta G^0_{U,XY-X'Y'} + \Delta\Delta G^0_{U,X'Y-X'Y'}$) differs from the sum of the changes due to the single mutations ($\Delta\Delta G^0_{U,XY-XY'} + \Delta\Delta G^0_{U,XY-X'Y}$), the two residues in WT are coupled and the magnitude of the difference (interaction energy: $\Delta\Delta G^0_{\text{Inter}}$) is related to the strength of interaction between them. X and Y denote wild type residues of interest and X' and Y' designate the substituted residues for X and Y, respectively.

$$\begin{aligned}\Delta\Delta G^0_{\text{Inter}} &= - [(\Delta\Delta G^0_{U,XY-XY'} + \Delta\Delta G^0_{U,XY-X'Y}) - (\Delta\Delta G^0_{U,XY-X'Y'} + \Delta\Delta G^0_{U,X'Y-X'Y'})] \\ &= - [(\Delta\Delta G^0_{U,XY-XY'} + \Delta\Delta G^0_{U,XY-X'Y}) - (\Delta\Delta G^0_{U,XY-XY'} + \Delta\Delta G^0_{U,XY'-X'Y'})] \quad (4)\end{aligned}$$

REFERENCES

REFERENCES

1. Eriksson, A. E.; Baase, W. A.; Wozniak, J. A.; Matthews, B. W., A cavity-containing mutant of T4 lysozyme is stabilized by buried benzene. *Nature* **1992**, 355, 371-3.
2. Richards, F. M., Areas, volumes, packing and protein structure. *Annual review of biophysics and bioengineering* **1977**, 6, 151-76.
3. Lim, W. A.; Sauer, R. T., Alternative packing arrangements in the hydrophobic core of lambda repressor. *Nature* **1989**, 339, 31-6.
4. Wen, J.; Chen, X.; Bowie, J. U., Exploring the allowed sequence space of a membrane protein. *Nat Struct Biol* **1996**, 3, 141-8.
5. Liang, J.; Dill, K. A., Are proteins well-packed? *Biophys J* **2001**, 81, 751-66.
6. Eriksson, A. E.; Baase, W. A.; Zhang, X. J.; Heinz, D. W.; Blaber, M.; Baldwin, E. P.; Matthews, B. W., Response of a Protein-Structure to Cavity-Creating Mutations and Its Relation to the Hydrophobic Effect. *Science* **1992**, 255, 178-183.
7. Joh, N. H.; Oberai, A.; Yang, D.; Whitelegge, J. P.; Bowie, J. U., Similar energetic contributions of packing in the core of membrane and water-soluble proteins. *Journal of the American Chemical Society* **2009**, 131, 10846-7.
8. Kellis, J. T., Jr.; Nyberg, K.; Fersht, A. R., Energetics of complementary side-chain packing in a protein hydrophobic core. *Biochemistry* **1989**, 28, 4914-22.
9. Dill, K. A., Dominant forces in protein folding. *Biochemistry* **1990**, 29, 7133-55.
10. Willis, M. A.; Bishop, B.; Regan, L.; Brunger, A. T., Dramatic structural and thermodynamic consequences of repacking a protein's hydrophobic core. *Structure* **2000**, 8, 1319-28.
11. Tseng, Y. Y.; Liang, J., Estimation of amino acid residue substitution rates at local spatial regions and application in protein function inference: a Bayesian Monte Carlo approach. *Mol Biol Evol* **2006**, 23, 421-36.
12. Fernandez, A., Packing defects functionalize soluble proteins. *FEBS Lett* **2015**, 589, 967-73.
13. Kadirvelraj, R.; Sennett, N. C.; Polizzi, S. J.; Weitzel, S.; Wood, Z. A., Role of packing defects in the evolution of allostery and induced fit in human UDP-glucose dehydrogenase. *Biochemistry* **2011**, 50, 5780-9.
14. He, Y.; Liu, S.; Jing, W.; Lu, H.; Cai, D.; Chin, D. J.; Debnath, A. K.; Kirchhoff, F.; Jiang, S., Conserved residue Lys574 in the cavity of HIV-1 Gp41 coiled-coil domain is critical for six-helix bundle stability and virus entry. *J Biol Chem* **2007**, 282, 25631-9.
15. Kunji, E. R.; Robinson, A. J., The conserved substrate binding site of mitochondrial carriers. *Biochim Biophys Acta* **2006**, 1757, 1237-48.
16. Popot, J. L.; Engelman, D. M., Membrane protein folding and oligomerization: the two-stage model. *Biochemistry* **1990**, 29, 4031-7.

17. Fleming, K. G.; Engelman, D. M., Specificity in transmembrane helix-helix interactions can define a hierarchy of stability for sequence variants. *Proc Natl Acad Sci U S A* **2001**, *98*, 14340-4.
18. Baker, R. P.; Urban, S., Architectural and thermodynamic principles underlying intramembrane protease function. *Nature chemical biology* **2012**, *8*, 759-68.
19. Rother, K.; Hildebrand, P. W.; Goede, A.; Gruening, B.; Preissner, R., Voronoia: analyzing packing in protein structures. *Nucleic Acids Res* **2009**, *37*, D393-5.
20. Fleming, P. J.; Richards, F. M., Protein packing: dependence on protein size, secondary structure and amino acid composition. *J Mol Biol* **2000**, *299*, 487-98.
21. Hildebrand, P. W.; Rother, K.; Goede, A.; Preissner, R.; Frommel, C., Molecular packing and packing defects in helical membrane proteins. *Biophys J* **2005**, *88*, 1970-7.
22. Dundas, J.; Ouyang, Z.; Tseng, J.; Binkowski, A.; Turpaz, Y.; Liang, J., CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues. *Nucleic Acids Res* **2006**, *34*, W116-8.
23. Lemieux, M. J.; Fischer, S. J.; Cherney, M. M.; Bateman, K. S.; James, M. N., The crystal structure of the rhomboid peptidase from *Haemophilus influenzae* provides insight into intramembrane proteolysis. *Proc Natl Acad Sci U S A* **2007**, *104*, 750-4.
24. Wang, Y.; Maegawa, S.; Akiyama, Y.; Ha, Y., The role of L1 loop in the mechanism of rhomboid intramembrane protease GlpG. *J Mol Biol* **2007**, *374*, 1104-13.
25. Counterman, A. E.; Clemmer, D. E., Volumes of individual amino acid residues in gas-phase peptide ions. *Journal of the American Chemical Society* **1999**, *121*, 4031-4039.
26. Zhou, Y.; Moin, S. M.; Urban, S.; Zhang, Y., An internal water-retention site in the rhomboid intramembrane protease GlpG ensures catalytic efficiency. *Structure* **2012**, *20*, 1255-63.
27. Wu, Z.; Yan, N.; Feng, L.; Oberstein, A.; Yan, H.; Baker, R. P.; Gu, L.; Jeffrey, P. D.; Urban, S.; Shi, Y., Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry. *Nat Struct Mol Biol* **2006**, *13*, 1084-91.
28. Vinothkumar, K. R.; Strisovsky, K.; Andreeva, A.; Christova, Y.; Verhelst, S.; Freeman, M., The structural basis for catalysis and substrate specificity of a rhomboid protease. *EMBO J* **2010**, *29*, 3797-809.
29. Zoll, S.; Stanchev, S.; Began, J.; Skerle, J.; Lepsik, M.; Peclinovska, L.; Majer, P.; Strisovsky, K., Substrate binding and specificity of rhomboid intramembrane protease revealed by substrate-peptide complex structures. *EMBO J* **2014**, *33*, 2408-21.
30. Baker, R. P.; Young, K.; Feng, L.; Shi, Y.; Urban, S., Enzymatic analysis of a rhomboid intramembrane protease implicates transmembrane helix 5 as the lateral substrate gate. *Proc Natl Acad Sci U S A* **2007**, *104*, 8257-62.
31. Wang, Y.; Ha, Y., Open-cap conformation of intramembrane protease GlpG. *Proc Natl Acad Sci U S A* **2007**, *104*, 2098-102.
32. Wang, Y.; Zhang, Y.; Ha, Y., Crystal structure of a rhomboid family intramembrane protease. *Nature* **2006**, *444*, 179-80.

33. Xue, Y.; Ha, Y., Catalytic mechanism of rhomboid protease GlpG probed by 3,4-dichloroisocoumarin and diisopropyl fluorophosphonate. *J Biol Chem* **2012**, *287*, 3099-107.
34. Guo, R.; Gaffney, K.; Yang, Z.; Kim, M.; Sungsuwan, S.; Huang, X.; Hubbell, W. L.; Hong, H., Steric trapping reveals a cooperativity network in the intramembrane protease GlpG. *Nature chemical biology* **2016**, *12*, 353-360.
35. Lemberg, M. K. & Freeman, M. Functional and evolutionary implications of enhanced genomic analysis of rhomboid intramembrane proteases. *Genome Res* **2007**, *17*, 1634-1646.
36. Kim, D. E., D. Chivian, and D. Baker. Protein Structure Prediction and Analysis Using the Robetta Server. *Nucleic Acids Res* **2004**, *32*, W526-31.
37. Humphrey, W., Dalke, A. and Schulten, K., VMD: visual molecular dynamics. *J Mol Graph* **1996**, *14*, 33-8, 27-8
38. Phillips, J. C., Braun, R., Wang, W., Gumbart, J., Tajkhorshid, E., Villa, E., Chipot, C., Skeel, R. D., Kale, L. and Schulten, K., Scalable molecular dynamics with NAMD. *J Comput Chem* **2005**, *26*, 1781-802.
39. Aksimentiev, A., Sotomayor, M. and Wells, D., *Membrane Proteins Tutorial*. University of Illinois at Urbana Champaign **2012**.
40. Brooks, B. R., Brooks, C. L., 3rd, Mackerell, A. D., Jr., Nilsson, L., Petrella, R. J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflisch, A., Caves, L., Cui, Q., Dinner, A. R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuczera, K., Lazaridis, T., Ma, J., Ovchinnikov, V., Paci, E., Pastor, R. W., Post, C. B., Pu, J. Z., Schaefer, M., Tidor, B., Venable, R. M., Woodcock, H. L., Wu, X., Yang, W., York, D. M. and Karplus, M., CHARMM: the biomolecular simulation program. *J Comput Chem* **2009**, *30*, 1545-614.
41. Horovitz, A., Double-mutant cycles: A powerful tool for analyzing protein structure and function. *Folding & Design* **1996**, *1*, R121-R126.

Chapter 4 Is the lipid bilayer a good solvent for the denatured state of membrane proteins?

Kristen Gaffney, Ruiqiong Guo, Michael Bridges, Wayne Hubbell, and Heedeok Hong

This chapter will be submitted as an article later.

I worked with Kristen Gaffney for the research of this chapter for which we will be listed as co-1st authors after the publication. Kristen performed the experiments in **Figure 4.3e**, **Figure 4.4** and part of **Figure 4.2c**. I would like to acknowledge Dr. Michael Bridges for the extensive DEER measurements. I would also like to thank Professor Wayne Hubbell for advice on the manuscript.

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 inter-spin 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 broader 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 inter-helical loops and protection of transmembrane helices. Our distance data agree well with the “ Θ ” solvent scaling behavior based on the polymer model, suggesting a delicate balance between protein-protein and protein-lipid interactions in maintaining the denatured state in the bilayer. Our work provides an insight into the role of protein-lipid interactions at the early folding stage and a guideline in defining thermodynamic stability of membrane proteins in cell membranes.

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 translocation¹⁻⁴. 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^{5,6}. 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 water⁷⁻⁹. In contrast, the denatured state is poorly understood for membrane proteins which account for 25–30% of all genes in most genomes¹⁰. So far, the denatured states of helical membrane proteins have been mainly studied using chaotic agents including anionic detergent SDS and polar organic solutes, urea and GdnHCl in micellar solution¹¹⁻¹⁶. 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^{12,16}.

The folding of helical membrane proteins can be divided into two thermodynamically distinct stages¹⁷: 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 inter-helical loops before folding into the native state (*i.e.*, the denatured state ensemble, DSE). Nonetheless, the current approaches using chaotic 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.

In this study, we successfully reconstituted 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. Our results demonstrate that the DSE of GlpG is expanded in the lipid bilayers, and highly heterogeneous and dynamic. By applying the solvent-scaling models from polymer theory, we show that the degree of expansion fits well with the “ Θ -solvent model”, suggesting that, the DSE's of helical membrane proteins are reasonably well accommodated by the lipid bilayers with balanced protein-protein and protein-lipid interactions.

Results

Reconstitution of the on-pathway DSE in the 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^{18,19}. We overcame this difficulty by employing a steric trapping^{20,21}, which couples unfolding of a doubly-biotinylated protein to competitive binding of bulky monovalent streptavidin (mSA) (**Figure 4.1a**). 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 (see Chapter 2 for details)²¹. 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-iodoacetamide)²¹. 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 interspin distances between biotinylated sites using DEER at the same time (The different usages of these two probes are described below). DEER is adequate for measuring the dimension of the denatured states because both long-range distance (15–60 Å) and distribution can be obtained²².

The sterically trapped DSEs of double-biotin variants of GlpG were first prepared in dodecylmaltoside (DDM) micelles upon addition of excess wild-type mSA (mSA-WT) that tightly binds to biotin labels ($K_{d, \text{biotin}} \approx 10^{-14}$ M; $k_{\text{off}, \text{biotin}} \approx \text{weeks}$) (**Figure 4.1a**)²³⁻²⁵. Next, the DSE's were reconstituted in the 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 = 4:1:1.8; lipid-to-detergent molar ratio, $q=2.8$; disk diameter ≈ 30 nm²⁶) 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.

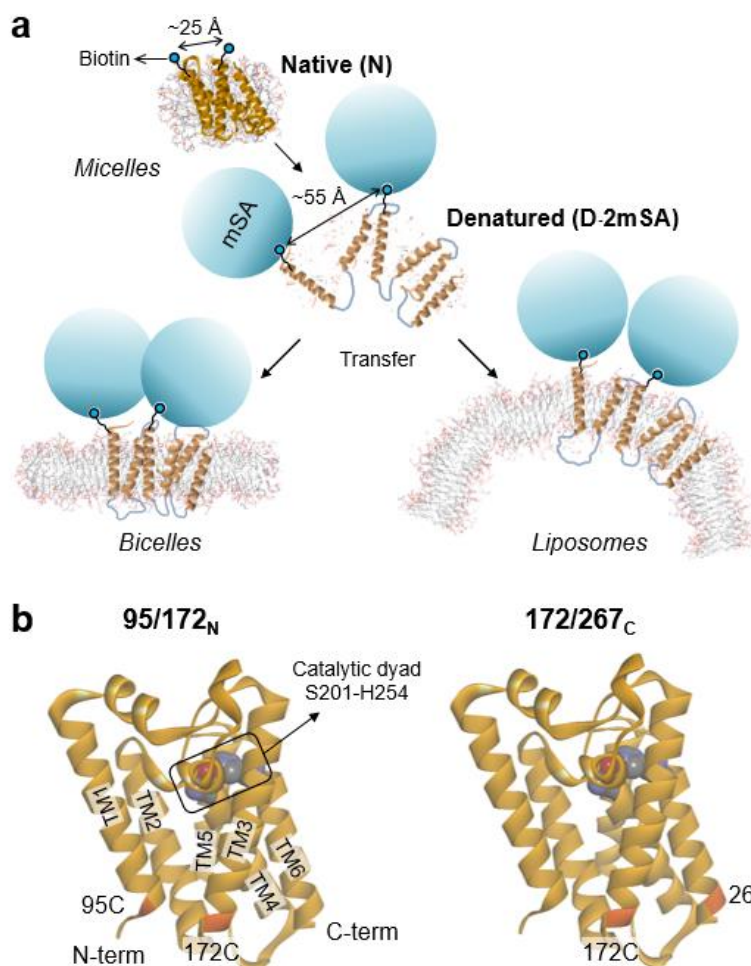


Figure 4.1 Steric trapping strategy to reconstitute denatured GlpG (D-2mSA) in the lipid bilayers. (a) Doubly-biotinylated GlpG was first denatured using a steric trapping in DDM micelles. For reconstitution in bicelles, denatured GlpG was directly injected into preformed DMPC/DMPG/CHAPS (molar ratio=4:1:1.8) 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 (see **Materials and Methods**). (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 (see details in Chapter 2 **Figure 2.1**).

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-BtnPyr₂ and 172/267_C-BtnPyr₂) and the bicelles containing the quencher (dabcyl)-labeled lipid (DOPE-dabcyl) (**Figure 4.2a**). Pyrene fluorescence from the DSE's of both double biotin variants was substantially quenched after injection to the bicelles close to the levels of full incorporation, indicating partition of the

DSE's in the bilayered region. Incorporation of the DSE's to *E. coli* liposomes was tested using a liposome floatation assay (**Figures 4.2b** and **4.3a**). 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 was completely resistant to sodium carbonate extraction, indicating membrane integration (**Figure 4.3b**).

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 (**Figure 4.2c**). In this assay (**Figures 4.3c** and **4.3d** for detailed description), we used GlpG labeled with BtnRG-thiopyridine (95/172_N-BtnRG₂ and 172/267_C-BtnRG₂), 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 released 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.

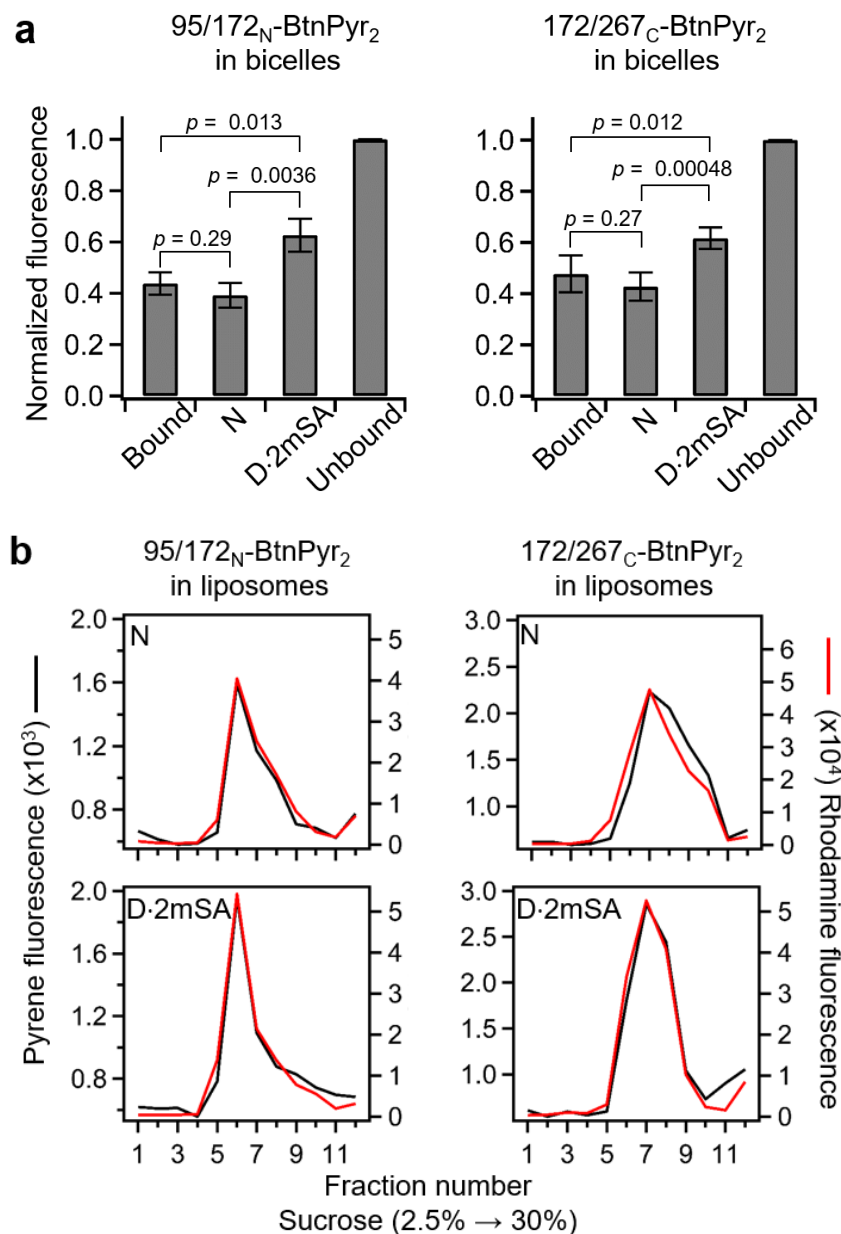
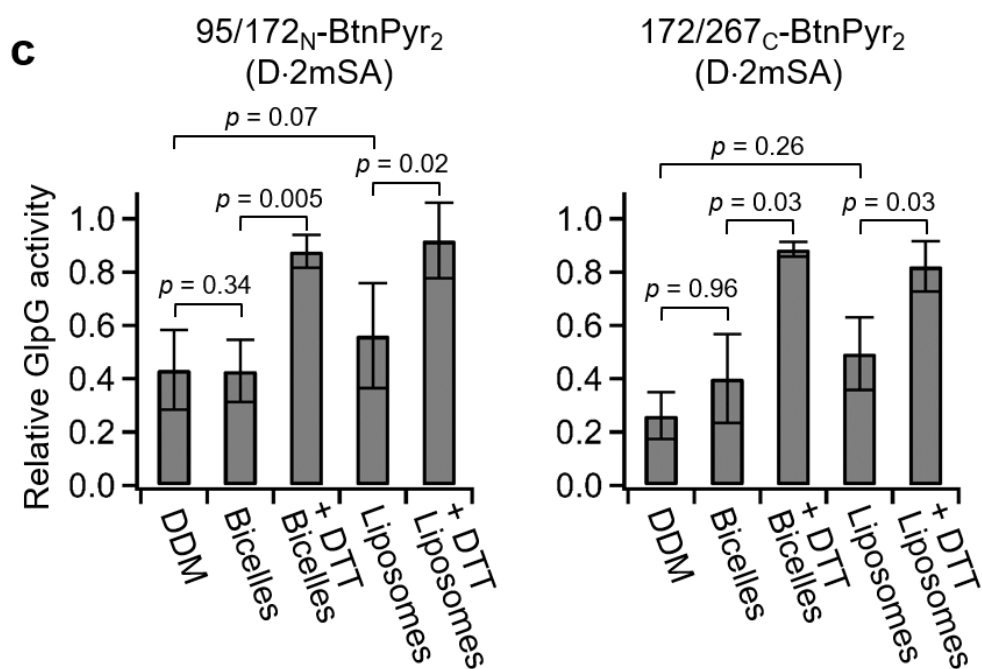


Figure 4.2 Reconstitution of denatured GlpG in the native lipid and solvent environments. **(a)** Fluorescence quenching assay to measure bicelle-association of native (N) and denatured (D·2mSA) GlpG. Binding of pyrene-labeled GlpG (double biotin variants, 95/172_N-BtnPyr₂ and 172/267_C-BtnPyr₂) to dabcyI (quencher)-labeled bicelles induced quenching of pyrene fluorescence. Pyrene-labeled mSA, which is 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-BtnPyr₂ and 172/267_C-BtnPyr₂) co-floated with rhodamine-labeled liposomes (see also **Figures 4.3a**). **(c)** The proteolytic activity of denatured GlpG (95/172_N-BtnRG₂ and 172/267_C-BtnRG₂) in micelles, bicelles and liposomes to test the maintaining 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. In (a) and (c), error bars denote $\pm$ SEM. ($n = 3$). P values were obtained using Student's t -test.

Figure 4.2 (cont'd)



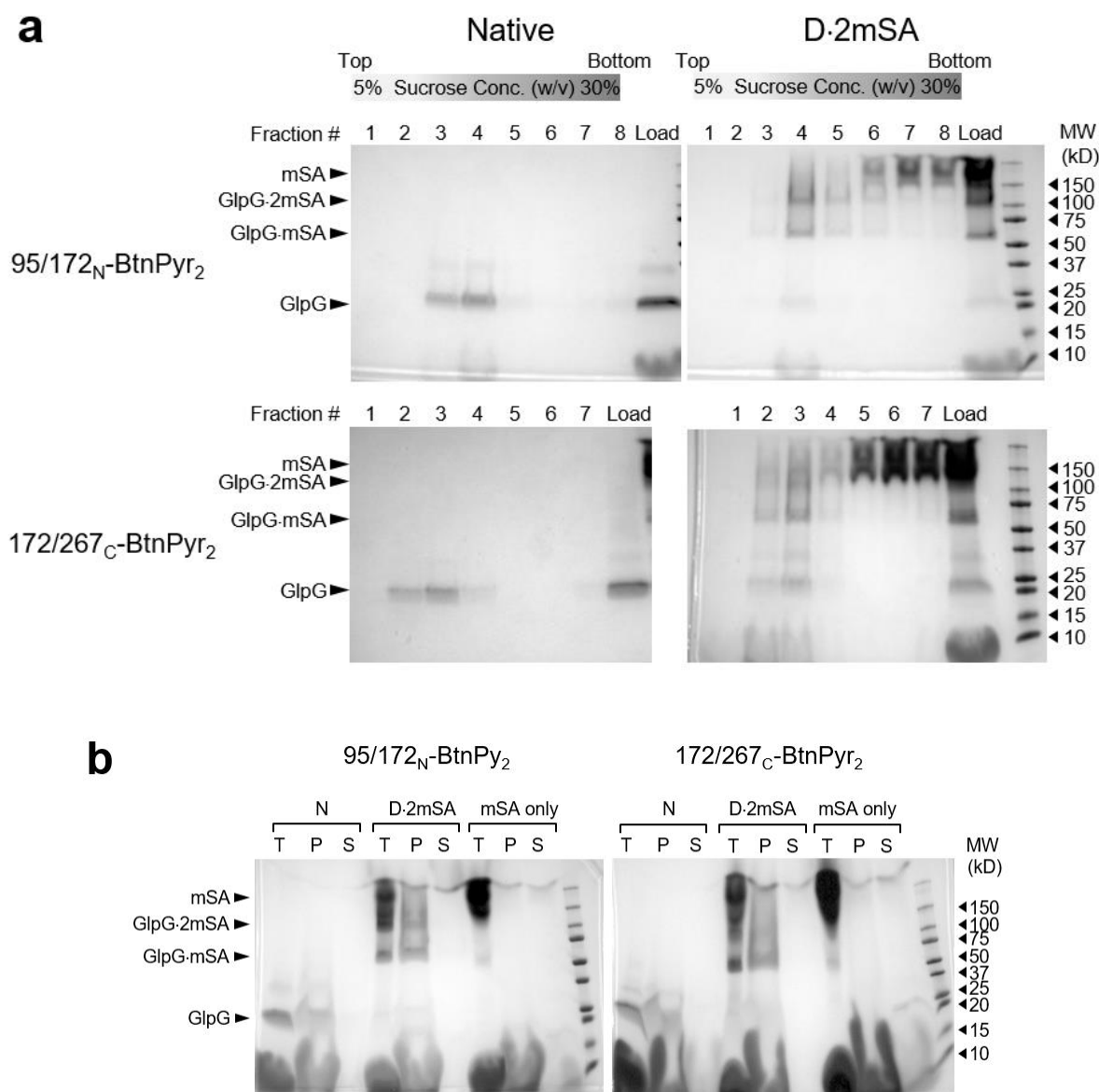
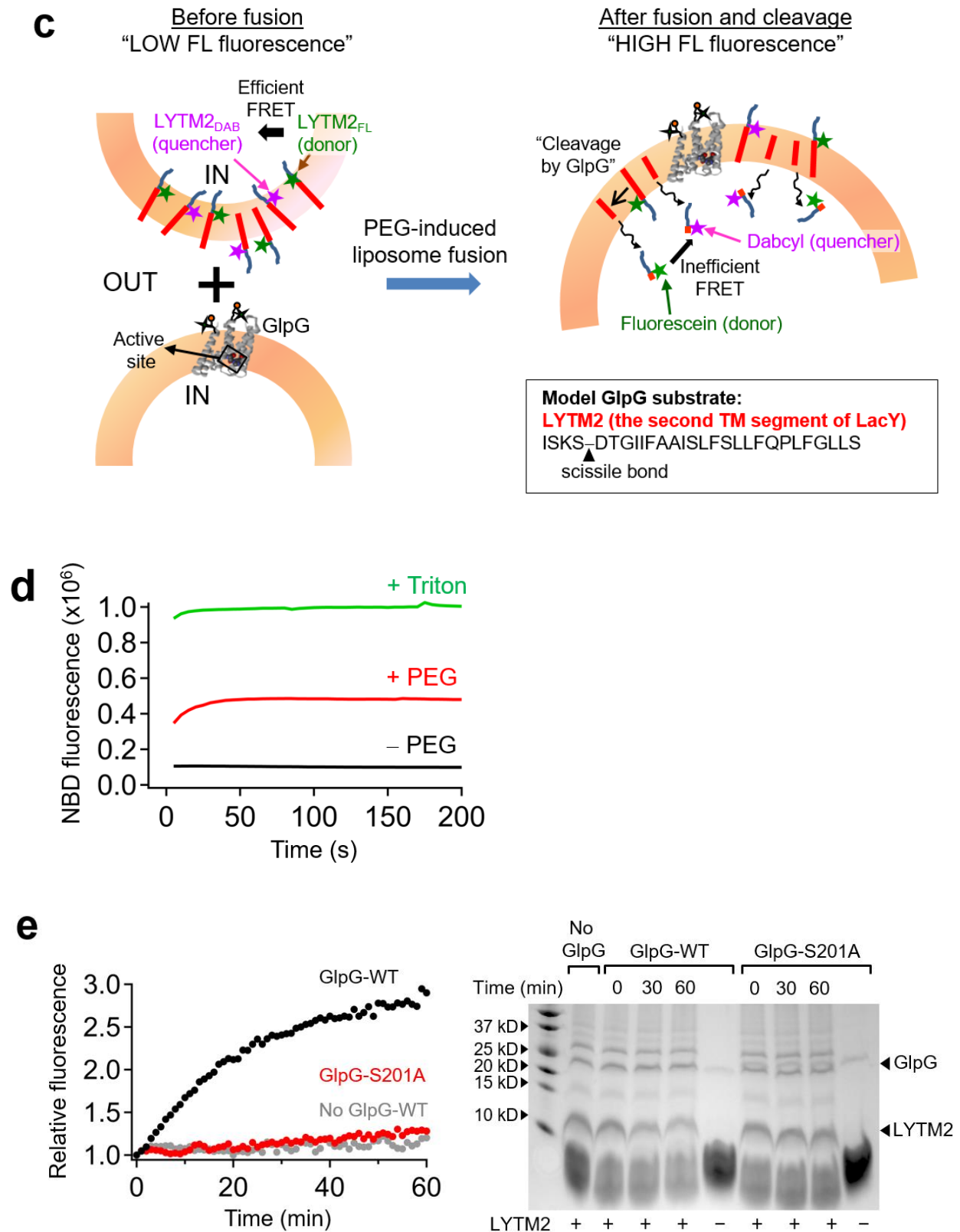


Figure 4.3 Reconstitution of the denatured states in the lipid bilayer environments. (a) Liposome floatation assay for (*left*) native or (*right*) sterically trapped denatured GlpG (D·2mSA) reconstituted in liposomes. Sucrose concentration (w/v) increased from 5% (top layer, Fraction 1) to 30% (Fraction 8, bottom layer). For the native state of 172/267_N-BtnPyr₂, 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. **(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. See **Materials and Methods** for detailed procedures. **(c)** The principle of the activity 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 nonfluorescent quencher dabcyI (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} 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-ethanolamine and the liposomes containing unlabeled wild-type GlpG and LacYTM2) at 37°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. **(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, de-quenching of FL fluorescence is indicative of cleavage of LacYTM2 by GlpG. See **Materials and Methods** for detailed description.

Figure 4.3 (cont'd)



The global flexibility of the DSE measured by proteolysis is higher in the bilayers

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 (**Figure 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 segments²⁷. Time-dependent proteolysis was measured for the DSE's trapped at two different biotin pairs (95/172_N-BtnRG₂ and 172/267_C-BtnRG₂) using SDS-PAGE (**Figure 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 digested, we expected that ~50% of GlpG would be fragmentized, which was the case.

Combined with the activity data (**Figure 4.2c**), this result illustrates that double binding of mSA induced an increase in conformational flexibility, demonstrating protein denaturation by steric trapping. In micelles, the DSE has trapped at the different biotin pairs displayed clearly different proteolysis patterns: the DSE trapped at the N-terminal half (95/172_N-BtnRG₂) were proteolyzed yielding only smaller fragments (< 8 kDa), whereas the DSE trapped at the C-terminal half (172/267_C-BtnRG₂) yielded three larger fragments (17, 13 and 11 kDa) (**Figure 4.4, top right**). Previously, we have shown that, in micelles, the state trapped at the N-terminal biotin pairs in 95/172_N-BtnRG₂ is globally denatured, while the state trapped at the C-terminal biotin pair in 172/267_C-BtnRG₂, is partially denatured²⁸. The proteolysis to multiple larger fragments observed for 172/267_C-BtnRG₂ suggests a partially denatured state with heterogeneous conformations with varied degrees of compactness, supporting our previous finding. In bicelles, ProK induced

maximal proteolysis (*i.e.*, proteolysis to only smaller fragments with <8kDa) for the denatured state trapped at 95/172_N-BtnRG₂, while yielding one larger fragment (~19 kDa) and maximally proteolyzed fragments for 172/267_C-BtnRG₂. In liposomes, the DSE's were maximally proteolyzed regardless of the location of the biotin pair.

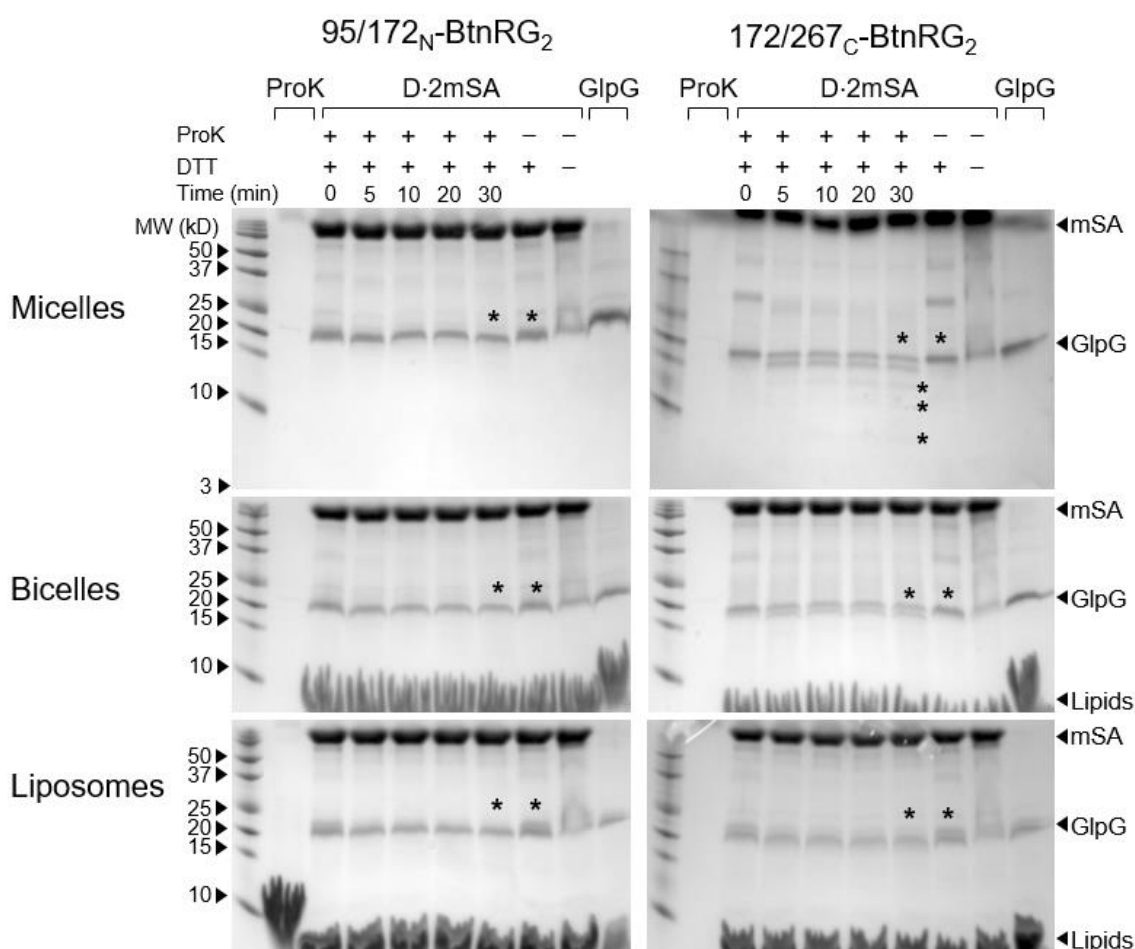


Figure 4.4 Limited proteolysis of denatured GlpG (D-2mSA, 95/172_N-BtnRG₂ and 172/267_C-BtnRG₂) by proteinase K (ProK) in (top) DDM micelles, (middle) DMPC:DMPG:CHAPS bicelles, and (bottom) *E. coli* liposomes. After termination of proteolysis reactions, DTT was added to release bound mSA from GlpG. 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 bioinylation reactions of double cysteine variants are not complete. Single- labeled and unlabeled GlpG are not subject to steric trapping and thus not denatured. These species remain folded and are protected from ProK.

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 labels (95/172_N-BtnRG₂ or 172/267_C-BtnRG₂) were measured for the native and sterically trapped DSEs in bicelles and liposomes using DEER. For native GlpG in micelles, BtnRG reports a slightly longer inter-spin distance by 2–4 Å than widely used spin label R1²⁸. We have shown that upon denaturation by steric trapping, the median inter-spin distance (d_{Med}) increased from 28 Å to 49 Å for 95/172_N-BtnRG₂ and from 26 Å to 51 Å for 172/267_C-BtnRG₂ (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 inter-spin distances for the DSE's without a dominant distance component (**Figure 4.5**), 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 inter-spin distances (d_{Prob}) in the bilayer environments were overall similar to those in micelles (d_{Prob} = 27–28 Å for 95/172_N-BtnRG₂ and d_{Prob} = 24–30 Å for 172/267_C-BtnRG₂, **Figure 4.5** and **Table 4.1**). 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-BtnRG₂ and from 30 Å to 47 Å for 172/267_C-BtnRG₂, *i.e.*, the d_{Prob} 's increased by 1.3 and 1.6 folds in the DSE's relative to the native state. Nonetheless, relative to micelles, bicelles did not induce a large expansion of the DSE's.

		95/172 _N -BtnRG ₂		172/267 _C -BtnRG ₂	
		<i>Native</i>	<i>Denatured</i>	<i>Native</i>	<i>Denatured</i>
Micelles (DDM)	$\langle d \rangle^a$	32	43	25	48
	σ^b	5	12	7	5
	$\langle d^2 \rangle^c$	1,112	2,469	653	2,412
	d_{Med}^d	28	49	26	51
	d_{Prob}^e	27	54	28	52
Bicelles (DMPC/DMPG/CHAPS)	$\langle d \rangle$	33	40	33	46
	σ	7	11	5	8
	$\langle d^2 \rangle$	1,227	1,753	1,207	2,319
	d_{Med}	29	39	31	45
	d_{Prob}	28	41	30	42
Liposomes (<i>E. coli</i> phospholipids)	$\langle d \rangle$	32	42	31	49
	σ	5	11	7	8
	$\langle d^2 \rangle$	1,111	1,976	1,132	2,562
	d_{Med}	29	41	26	53
	d_{Prob}	27	56	24	54

^aMean distance

^bStandard deviation

^cRoot-mean-square distance

^dMedian distance

^eThe most probable distance

Table 4.1 Statistical parameters of the interspin distance distributions in the native and sterically trapped denatured states of GlpG in micelles, bicelles and liposomes. These values were calculated from the distance distributions obtained by Tikhonov regulation fitting of DEER data.

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-BtnRG₂ and from 24 Å to 52 Å for 172/267_C-BtnRG₂. 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, leading to an overestimation of inter-spin distances. To test this possibility, we further increased L/P up to 12,000 or co-incorporated an inactive variant of unlabeled GlpG at a various molar excess relative to spin-labeled GlpG. Under all tested conditions, the overall inter-spin distances in the DSE's did not significantly change, demonstrating that the observed distance distributions mainly originated from intra-molecular dipolar coupling (**Figure 4.6a, 4.6c** and **Table 4.2**). We also questioned if bulky mSA molecules which were doubly bound to GlpG to trap the DSE's would distort the inter-spin distance distributions because of steric repulsion. To test this, we first obtained a DSE of the variant 172/267_C-BtnRG₂ using SDS and measured the distance distribution with and without bound mSA utilizing the fact that the biotin-streptavidin interaction is resistant to SDS. The statistical parameters of the distance distributions (**Figure 4.6b** and **Table 4.3**) were very similar regardless of the presence of bound mSA. This result validates the observed distance distributions for the sterically trapped DSE's closely reflect the true dimension and conformational heterogeneity of the intrinsic DSE's of the protein.

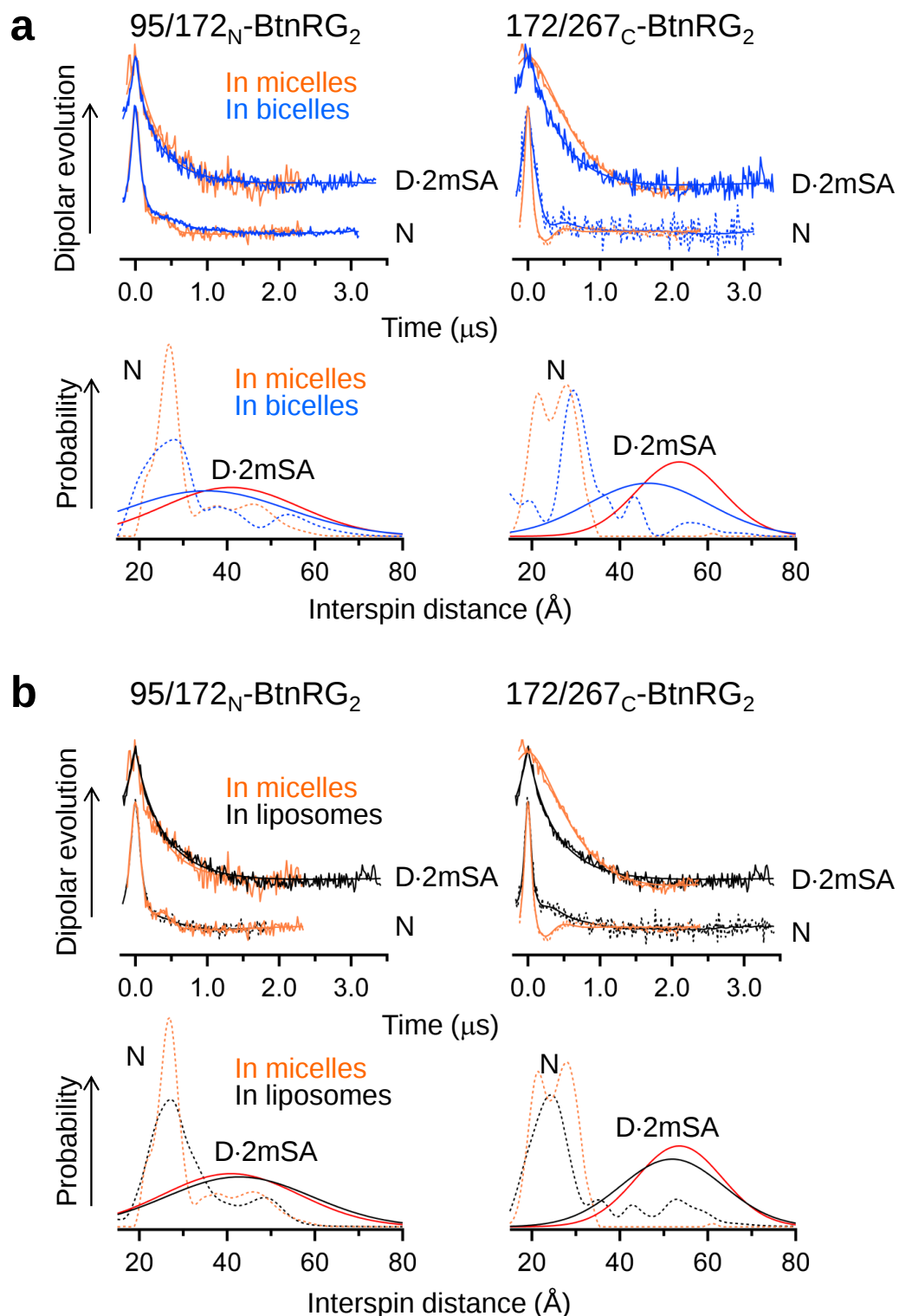


Figure 4.5 Distance distributions in the denatured states of GlpG measured by DEER. (a–b) (Top) Background-subtracted dipolar evolution data and their fits and (Bottom) inter-spin distance distributions for native (N) and sterically trapped denatured (D·2mSA) states of GlpG (95/172_N-BtnRG₂ and 172/267_C-BtnRG₂). The fitting was performed under the assumption that the

probabilities of inter-spin 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.

Molar excess of unlabeled GlpG ^a	95/172 _N -BtnRG ₂			172/267 _C -BtnRG ₂		
	x 0	x 3	x 6	x 0	x 3	x 6
$\langle d \rangle$	47	44	42	44	50	49
σ	9	11	11	11	8	8
$\langle d^2 \rangle$	2436	2127	1976	2134	2733	2562
d_{Med}	53	42	41	45	54	53
d_{Prob}	57	55	56	59	56	54

^a Inactive variant S201A

Table 4.2 Statistical parameters of the inter-spin distance distributions of the sterically trapped denatured states at an increasing molar excess of unlabeled native GlpG in *E. coli* liposomes.

	172/267 _C -BtnRG ₂	
	SDS	SDS + mSA
$\langle d \rangle$	48	47
σ	9	12
$\langle d^2 \rangle$	2538	2482
d_{Med}	52	52
d_{Prob}	54	53

Table 4.3 Statistical parameters of the interspin distance distributions of the SDS-induced denatured states in the presence and absence of bound mSA.

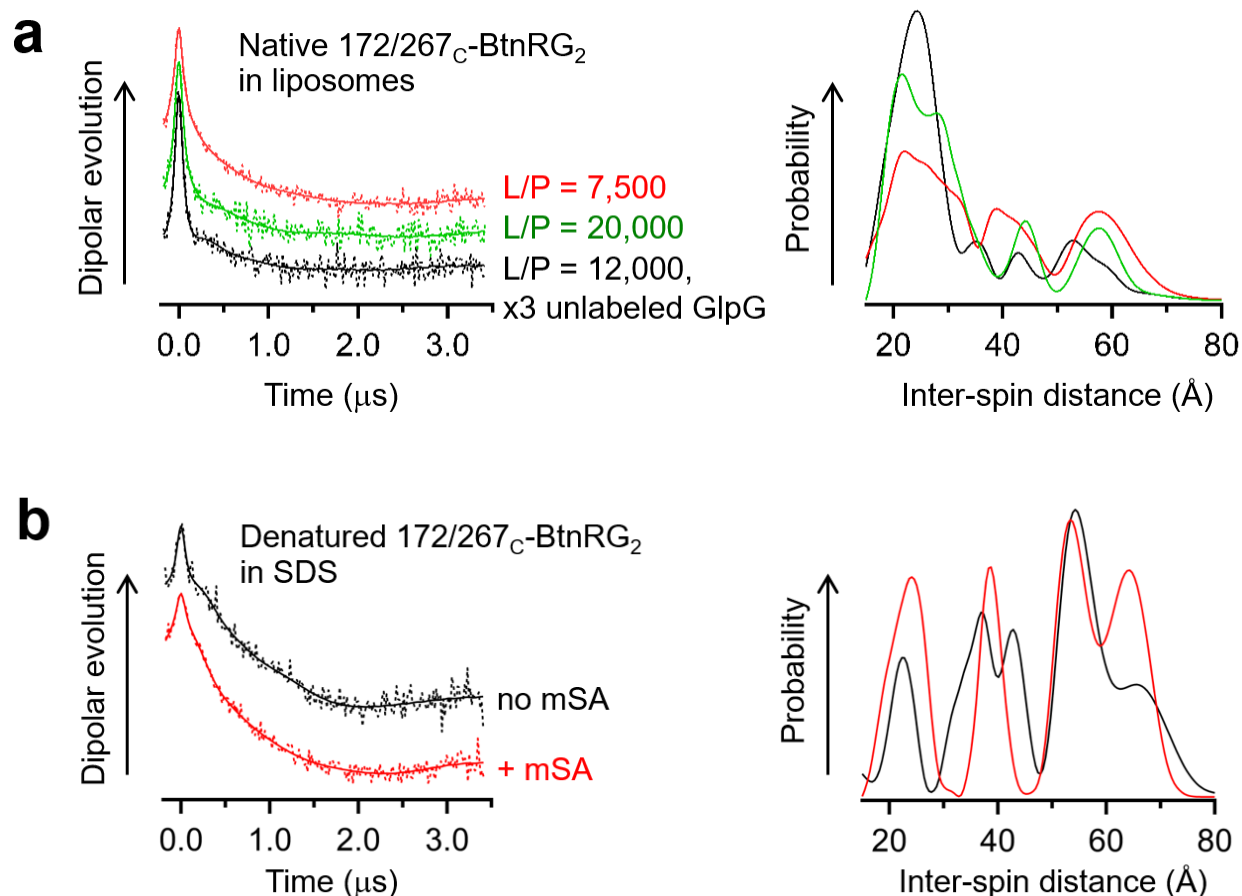
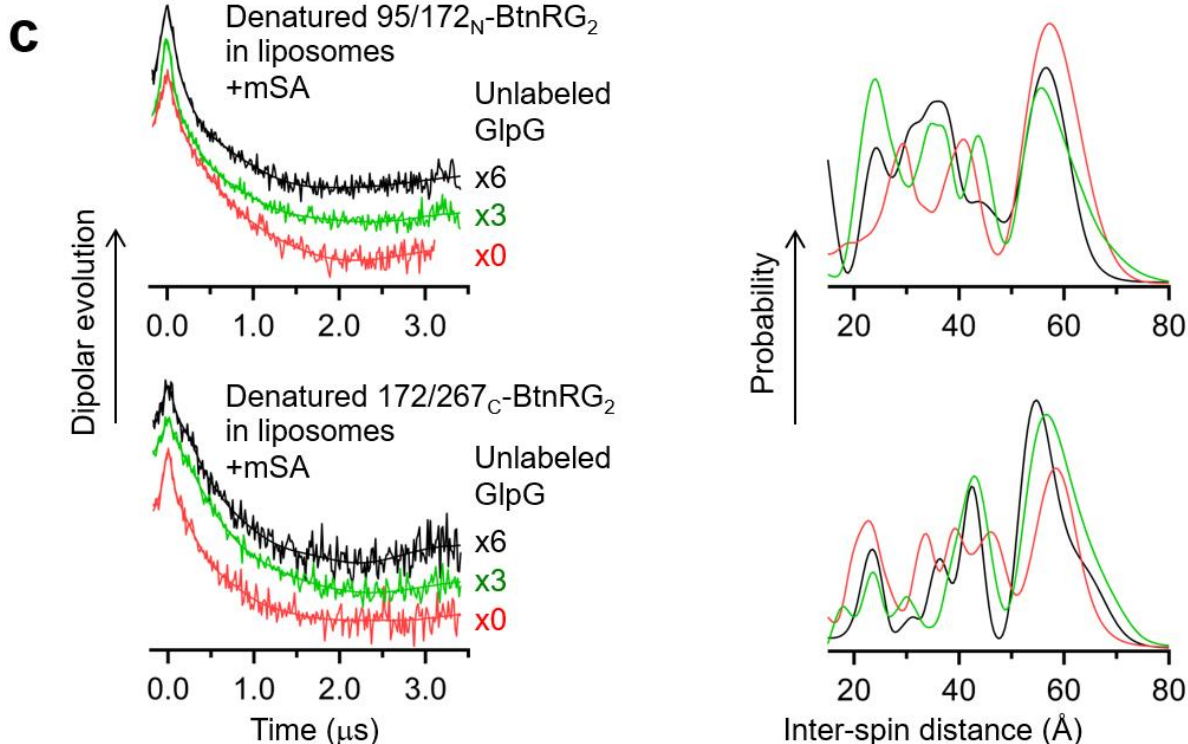


Figure 4.6 Effects of sample reconstitutions on DEER measurements. (a) To test if two bound mSA molecules in the denatured state affect the compactness of the denatured state ensemble, the inter-spin distances were measured for the SDS-induced denatured state of GlpG 172/267_C-BtNRG₂ and the same denatured state bound with mSA molecules. The result indicates that once GlpG is denatured, bound mSA molecules did not significantly change the overall degree of expansion of the denatured state. **(b)** Optimization of the sample condition to measure the intramolecular spin-spin distances for the native state of GlpG in liposomes. For the native state of 172/267_C-BtNRG₂ variant in liposomes at the protein-to-lipid molar ratio (L/P) of 7,500, we observed a large contribution of the long distance components (~40 Å and ~60 Å) that were not observed in micelles and bicelles. We suspected these components originated from the unwanted intermolecular dipolar interaction. To suppress the possible inter-molecular contribution, we increased L/P to 20,000 or incorporated a 3-times molar excess of unlabeled GlpG (inactive variant S201A) relative to spin-labeled GlpG. Indeed, both attempts substantially reduced the occurrence of the longer distance components, which proved the existence of intermolecular contribution and led to a successful optimization of the sample condition. In **Figure 4b**, the DEER result with increased L/P and incorporated unlabeled GlpG was added. **(c)** To test if we truly measure the intra-molecular inter-spin distances in the denatured states in liposomes, we incorporated unlabeled native GlpG at an increasing molar excess (x0, x3 and x6) relative to spin-labeled GlpG. The unlabeled protein inclusions can suppress unwanted inter-molecular dipolar coupling by reducing the collision frequency between spin-labeled GlpG. The inter-spin distance distributions did not significantly change in the presence of unlabeled proteins, validating our result.

Figure 4.6 (cont'd)



The lipid bilayers exhibit “ Θ -solvent” behavior for the denatured state of GlpG

Finally, we quantitatively evaluated the ability of the amphiphilic environments tested in this study for solubilizing the DSE's based on the distance information obtained from DEER. From the polymer theory, the solvents in which a given type of long chain homopolymers are dissolved can be classified into three types, “good”, “theta (Θ)” and “poor”, depending on the relative strengths between intra-chain and chain-solvent interactions^{29,30}. 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 intra-chain 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, R_G) as a function of the number of monomeric units in a polymer chain (*i.e.*, number of amino acids in a polypeptide chain). In case of a polypeptide chain in three-dimensional space, R_G is described using the following equation³¹:

$$R_G = R_o N_{AA}^\nu \quad (1)$$

, where $R_o = 1.98 \text{ \AA}$, 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”²⁹. Alternatively, when an end-to-end distance between a residue pair in a polypeptide chain is measured, equation (1) can be modified into the following equation^{31,32}:

$$\left(\langle R^2 \rangle \right)^{1/2} = (6)^{1/2} R_o N_{AA}^\nu \quad (2)$$

, where $(\langle R^2 \rangle)^{1/2}$ 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 ($D = \sim 30 \text{ \AA}$). 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)³³:

$$\left(\langle R^2 \rangle\right)^{1/2} = (6)^{1/2} \left(R_o^5 / D\right)^{1/4} N_{AA}^\nu \quad (3)$$

Under a good solvent condition, $\nu = 0.75$. Interestingly, the equation for a “ Θ solvent” condition under the quasi-two dimensional constraints collapses into the same equation as equation (2) under the three-dimensional condition with the same characteristic exponent (*i.e.*, $\nu = 0.5$)^{29,33}.

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 (**Figure 4.7**). According to the two stage model and previous experimental results^{11,17}, 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 states³¹. Intriguingly, our experimental inter-spin RMSD values determined by DEER for the DSE’s in micelles, bicelles and liposomes fell into the range close to the predicted values for the 2D or 3D “ Θ -solvent” model (**Figure 4.7**).

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 bilayer³⁴. In their study, although the denatured states retains 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 (**Figure 4.7**). Because the fraction of the native contacts in

bilayer simulation is similar to that in vacuum simulation, their inter-residue distances in bilayer simulation may represent 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.

Discussion

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 study that can provide more accurate reference distance information under each solvent condition. For example, MD simulation could be performed in an explicit bilayer mimicking *E. coli* membranes, and the lipid solvation strength for the denatured state could be changed to modulate the extent the native contacts for modeling different solvent qualities.

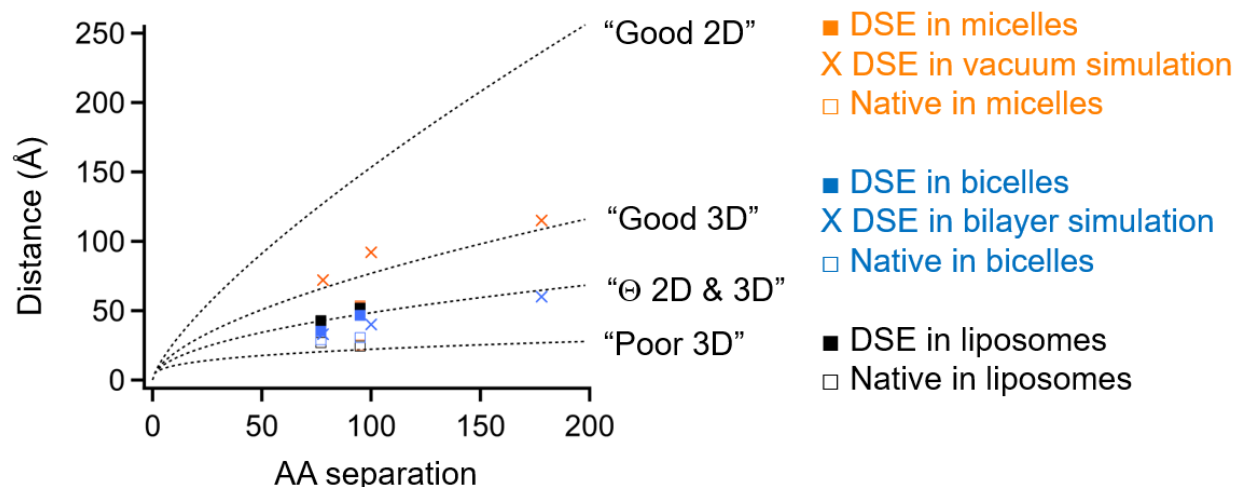


Figure 4.7 The values of the intrachain RMSDs as a function of residue separation obtained from DEER. Those values corresponds to the most probable distance from a single-Gaussian fit of the time-dependent dipolar evolution data. The dashed lines indicate the predicted RMSDs from the solvent scaling 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, $(\langle R^2 \rangle)^{1/2} = (6)^{1/2} R_0 N_{AA}^n$, where $(\langle R^2 \rangle)^{1/2}$: root-mean-square distance (RMSD); $R_0 = 1.98 \text{ Å}$, 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), $(\langle R^2 \rangle)^{1/2} = (6)^{1/2} (R_0^5/D)^{1/4} N_{AA}^n$, where D : the height of the slit in which the polymer is confined, $D = 30 \text{ Å}$ (the hydrophobic thickness of a bilayer). For the polymers in quasi-2D space under “ 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.

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 and exhibit global flexibility. 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 still necessary because certain intrinsically unstable membrane proteins would be still subject to misfolding and aggregation in the membranes crowded with other membrane proteins. Also, environmental stresses such as heat or oxidation would increase the risk of misfolding. Overall, our study provides fundamental insights into the physical properties of the cell membranes as a medium for the folding of membrane proteins.

Materials and Methods

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 = 4:1) 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 °C. Bicelle stocks were kept at -20 °C prior to use.

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-BtnRG₂ and 24 hours for 172/267_C-BtnRG₂. Native and denatured GlpG were directly injected into preformed bicelles to the final concentrations of 5 μM

GlpG, 25 μ M 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.

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 μ M 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 hr 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 ($\epsilon_{\text{molar}} = 43,000 \text{ M}\cdot\text{cm}^{-1}$) 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] = 4:1) 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 μ M 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 μ M. 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 1–2 hr. In the first removal step, the samples were incubated at 4 °C for 2 hours and then moved to room temperature in the subsequent removal steps. The resulting proteoliposomes were extruded again using 0.2 μ M 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 dabcyI-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 dabcyI 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-dabcyI at the final concentrations of 3% bicelles and the final pyrene concentration of 5 μ M as measured by UV-Vis absorbance at 346 nm ($\epsilon_{\text{Molar}} = 43,000 \text{ M}^{-1}\text{cm}^{-1}$). 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 to the bicelles.

Preparation of empty *E. coli* liposomes

Dried *E. coli* lipid (Avanti Polar Lipids) film was hydrated with 20 mM Na₂HPO₄ (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).

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

25 µM GlpG variant 172/267_C-BtnPyr₂ or 172/267_C-BtnRG₂ 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 hr at room temperature. In the first removal step, the samples were incubated at 4 °C for 2 hours 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-BtnPyr₂ or 95/172_N-BtnRG₂ 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 μ M, and incubated at room temperature overnight. Then a 5 times molar excess WT-mSA was added and incubated for 1 hr 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 hr. Then GlpG samples in detergent was mixed with empty liposome. The following steps were the same as those for 172/267_C variants described above.

Flotation assay of liposome samples

Pyrene labeled GlpG (95/172_N-BtnPyr₂ or 172/267_C-BtnPyr₂) 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 μ L) was mixed well with 60% (w/v) sucrose in 20 mM HEPES (pH 7.5) (50 μ L). 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 μ L of sucrose solution at a lower concentration (20%, 10%, 5% and 2.5%). The tube was centrifuged at 35,000 rpm at 4 °C for 2 hr 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 μ L 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 560nm and 345nm 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 μ L sample was taken out from each fraction and solubilized with 2% β -OG and SDS sample loading buffer.

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 μ L of each sample was incubated with 500 μ L of pre-chilled 0.1 M Na_2CO_3 buffer (pH 11.0) for 30 min on ice. Then the mixture was ultra-centrifuged at 4 °C 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 °C 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 for SDS-PAGE. In the gel, 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 μ L) that had not been treated with sodium carbonate, was solubilized with β -OG followed by the addition of SDS sample buffer.

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-LYTM2 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-LYTM2 alone. For activity measurement

in bicelles, both SN-LYTM2 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 μ M 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 μ M LYTM2_{DAB} and 25 μ M LYTM2_{FL} were added while vortexing, following an incubation at room temperature for 30 minutes. For detergent removal, three portions of Bio-Beads (Bio-Rad) were added (200 mg/mL for each) stepwise. In each step, the mixture was gently stirred for 1–2 hr at room temperature. The resulting proteoliposomes were extruded using 0.2 μ M pore size membrane.

For activity assay, the proteoliposomes containing LYTM2 (10 μ L) were mixed with the proteoliposomes (5 μ L) containing 5 μ M GlpG and 18.5 μ L of buffer (20 mM Na₂HPO₄, 40 mM NaCl, pH 7.5). Fusion of proteoliposomes was initiated by addition of 16.5 μ L 36% PEG₃₃₅₀, 365 mM NaCl. Time-dependent change of fluorescein fluorescence was monitored at 37 °C 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 LYTM2 mixed with the liposomes without GlpG.

Liposome fusion assay induced by PEG

This assay was for obtaining the time scale of mixing between the enzyme GlpG and the substrate LYTM2, which forms a basis for our GlpG activity assay in liposomes. We employed a FRET-

based lipid mixing assay³. To prepare the proteoliposomes containing the substrate, Cys-less LYTM2 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 μ M 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 μ M 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 μ L) and unlabeled liposomes (9.45 μ L) 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) PEG₃₃₅₀. Total volume was 1.4 mL in a Hellma fluorescence 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 °C.

As a negative control that represents no fusion, no PEG was added. As a positive control for a homogeneously mixing state, 12 μ L of 100% Triton X-100 was added to a final concentration of 0.08% (w/v) to solubilize the liposomes.

Proteinase K digestion

5 μ M GlpG (95/172_N-BtnRG₂ or 172/267_C-BtnRG₂) in the absence and presence of 25 μ M mSA was prepared in 10 mM DDM, 10 mM DMPC/DMPG/CHAPS bicelles and 10 mM *E. coli* liposomes, as described above. 2 mM CaCl₂ was added to enhance the stability of proteinase K (Sigma). Proteolysis was initiated by addition of 0.14 μ g/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.

Sample preparation for DEER

To obtain the sterically-trapped denatured state in DDM micelles, 120 μ L of GlpG variants 95/172_N-BtnRG₂ or 172/267_C-BtnRG₂ (25 μ M) was incubated with a 5 times molar excess of mSA-WT in 40 mM DDM, 20 mM Na₂HPO₄ (pH 7.5), 40 mM NaCl at room temperature for three days (95/172_N-BtnRG₂) or overnight (172/267_C-BtnRG₂). Then the samples were concentrated to about ~50 μ L using Amicon Ultra 0.5 mL (MWCO = 10k Da, 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-BtnRG₂ and 172/267_C-BtnRG₂ (5 μ M GlpG without or with 25 μ M 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 Na₂HPO₄ (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-BtnRG₂ and 172/267_C-BtnRG₂ 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 Cysless 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 hours. The resulting pellets were resuspended in 20 mM Na₂HPO₄ (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 °C.

REFERENCES

REFERENCES

- 1 Dill, K. A. and Shortle, D. Denatured states of proteins. *Annu Rev Biochem* **1991**, 60, 795-825.
- 2 Saibil, H. Chaperone machines for protein folding, unfolding and disaggregation. *Nat Rev Mol Cell Biol* **2013**, 14, 630-642.
- 3 Matouschek, A. Protein unfolding--an important process in vivo? *Curr Opin Struct Biol* **2003**, 13, 98-109.
- 4 Sauer, R. T. and Baker, T. A. AAA+ proteases: ATP-fueled machines of protein destruction. *Annu Rev Biochem* **2011**, 80, 587-612.
- 5 Tanford, C. Protein denaturation. *Adv Protein Chem* **1968**, 23, 121-282.
- 6 Dill, K. A. Theory for the folding and stability of globular proteins. *Biochemistry* **1985**, 24, 1501-1509.
- 7 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. Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. *Science* **2017**, 358, 238-241.
- 8 Meng, W., Luan, B., Lyle, N., Pappu, R. V., and Raleigh, D. P. 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* **2013**, 52, 2662-2671.
- 9 Aznauryan, M., Delgado, L., Soranno, A., Nettels, D., Huang, J. R., Labhardt, A. M., Grzesiek, S., and Schuler, B. Comprehensive structural and dynamical view of an unfolded protein from the combination of single-molecule FRET, NMR, and SAXS. *Proc Natl Acad Sci U S A* **2016**, 113, E5389-5398.
- 10 Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E. L. L. Predicting transmembrane protein topology with a hidden Markov model: Application to complete genomes. *Journal of Molecular Biology* **2001**, 305, 567-580.
- 11 Krishnamani, V., Hegde, B. G., Langen, R., and Lanyi, J. K. Secondary and tertiary structure of bacteriorhodopsin in the SDS denatured state. *Biochemistry* **2012**, 51, 1051-1060.
- 12 Dutta, A., Kim, T. Y., Moeller, M., Wu, J., Alexiev, U., and Klein-Seetharaman, J. Characterization of membrane protein non-native states. 2. The SDS-unfolded states of rhodopsin. *Biochemistry* **2010**, 49, 6329-6340.
- 13 Dutta, A., Tirupula, K. C., Alexiev, U., and Klein-Seetharaman, J. Characterization of membrane protein non-native states. 1. Extent of unfolding and aggregation of rhodopsin in the presence of chemical denaturants. *Biochemistry* **2010**, 49, 6317-6328.

- 14 Dockter, C., Volkov, A., Bauer, C., Polyhach, Y., Joly-Lopez, Z., Jeschke, G., and Paulsen, H. Refolding of the integral membrane protein light-harvesting complex II monitored by pulse EPR. *Proc Natl Acad Sci U S A* **2009**, *106*, 18485-18490.
- 15 Joh, N. H., Min, A., Faham, S., Whitelegge, J. P., Yang, D., Woods, V. L., and Bowie, J. U. Modest stabilization by most hydrogen-bonded side-chain interactions in membrane proteins. *Nature* **2008**, *453*, 1266-1270.
- 16 Jacso, T., Bardiaux, B., Broecker, J., Fiedler, S., Baerwinkel, T., Mainz, A., Fink, U., Vargas, C., Oshkinat, H., Keller, S., and Reif, B. The mechanism of denaturation and the unfolded state of the alpha-helical membrane-associated protein Mistic. *J Am Chem Soc* **2013**, *135*, 18884-18891.
- 17 Popot, J. L. and Engelman, D. M. Membrane protein folding and oligomerization: the two-stage model. *Biochemistry* **1990**, *29*, 4031-4037.
- 18 Burton, R. E., Huang, G. S., Daugherty, M. A., Calderone, T. L., and Oas, T. G. The energy landscape of a fast-folding protein mapped by Ala-->Gly substitutions. *Nat Struct Biol* **1997**, *4*, 305-310.
- 19 Gillespie, J. R. and Shortle, D. Characterization of long-range structure in the denatured state of staphylococcal nuclease. I. Paramagnetic relaxation enhancement by nitroxide spin labels. *J Mol Biol* **1997**, *268*, 158-169.
- 20 Blois, T. M., Hong, H., Kim, T. H., and Bowie, J. U. Protein unfolding with a steric trap. *J Am Chem Soc* **2009**, *131*, 13914-13915.
- 21 Guo, R.; Gaffney, K.; Yang, Z.; Kim, M.; Sungsuwan, S.; Huang, X.; Hubbell, W. L.; Hong, H. Steric trapping reveals a cooperativity network in the intramembrane protease GlpG *Nat. Chem. Biol.* **2016**, *12*, 353–360.
- 22 Jeschke, G. DEER distance measurements on proteins. *Annu Rev Phys Chem* **2012**, *63*, 419-446.
- 23 Jefferson, R. E., Blois, T. M., and Bowie, J. U. Membrane proteins can have high kinetic stability. *J Am Chem Soc* **2013**, *135*, 15183-15190.
- 24 Howarth, M., Chinnapen, D. J., Gerrow, K., Dorrestein, P. C., Grandy, M. R., Kelleher, N. L., El-Husseini, A., and Ting, A. Y. A monovalent streptavidin with a single femtomolar biotin binding site. *Nat Methods* **2006**, *3*, 267-273.
- 25 Srisa-Art, M., Dyson, E. C., deMello, A. J., and Edel, J. B. Monitoring of real-time streptavidin-biotin binding kinetics using droplet microfluidics. *Anal Chem* **2008**, *80*, 7063-7067.
- 26 Glover, K. J., Whiles, J. A., Wu, G., Yu, N., Deems, R., Struppe, J. O., Stark, R. E., Komives, E. A., and Vold, R. R. Structural evaluation of phospholipid bicelles for solution-state studies of membrane-associated biomolecules. *Biophys J* **2001**, *81*, 2163-2171.
- 27 Wu, C. C., MacCoss, M. J., Howell, K. E., and Yates, J. R., 3rd. A method for the comprehensive proteomic analysis of membrane proteins. *Nat Biotechnol* **2003**, *21*, 532-538.

- 28 Guo, R., Gaffney, K., Yang, Z., Kim, M., Sungsuwan, S., Huang, X., Hubbell, W. L., and Hong, H. Steric trapping reveals a cooperativity network in the intramembrane protease GlpG. *Nat Chem Biol* **2016**, 12, 353-360.
- 29 Flory, P. J. *Principles of Polymer Chemistry*. pp. 399-431 Cornell University Press, **1953**.
- 30 Chan, H. S. and Dill, K. A. Polymer Principles in Protein-Structure and Stability. *Annual Review of Biophysics and Biophysical Chemistry* **1991**, 20, 447-490.
- 31 Fitzkee, N. C. and Rose, G. D. Reassessing random-coil statistics in unfolded proteins. *Proc Natl Acad Sci U S A* **2004**, 101, 12497-12502.
- 32 Fehr, N., Garcia-Rubio, I., Jeschke, G., and Paulsen, H. Early folding events during light harvesting complex II assembly in vitro monitored by pulsed electron paramagnetic resonance. *Biochimica Et Biophysica Acta-Bioenergetics* **2016**, 1857, 695-704.
- 33 Daoud, M. and Degennes, P. G. Statistics of Macromolecular Solutions Trapped in Small Pores. *Journal De Physique* **1977**, 38, 85-93.
- 34 Schafer, N. P., Truong, H. H., Otzen, D. E., Lindorff-Larsen, K., and Wolynes, P. G. Topological constraints and modular structure in the folding and functional motions of GlpG, an intramembrane protease. *Proceedings of the National Academy of Sciences of the United States of America* **2016**, 113, 2098-2103.

Chapter 5 Concluding remarks

In this dissertation research, I successfully developed a series of steric trapping-based methods for their general application to membrane proteins by synthesizing novel biotin probes processing fluorophores or spin labels. Our advanced steric trap methods does not rely on functional assays specific to the protein of interest so that they can be applied to other α -helical membrane proteins. With the fluorescence-based assays, we can precisely determine the global and local conformational stability, which enables the investigation of several key elements of membrane protein folding such as thermodynamic stability and cooperativity. Also, the development of a high-throughput assay for measuring the proteolytic activity of GlpG will potentially serve as an accurate and efficient method for other membrane-bound proteases.

By applying those methods to GlpG in micelles, I elucidated a detailed asymmetrical energy landscape, subglobal unfolding of the region encompassing the active site, and a network of cooperative and localized interactions to maintain the stability. Steric trap enables measurements of the local stability by placing a biotin pair to a specific region. This capability would open up the new possibilities for a more detailed study on the protein folding process under native lipid and solvent conditions.

Using steric trapping, the denatured state of membrane proteins can be obtained in a large quantity without disruption of the native protein-lipid interactions. Therefore, for the first time, I was able to study the conformation of the denatured state was in the native lipid bilayer environments, which is crucial to define thermodynamic stability and folding mechanisms of membrane proteins. By using the novel spin-labeled biotin derivative conjugated to GlpG, I measured the inter-spin distance between the two biotinylated sites in the sterically trapped denatured state by DEER spectroscopy in the native lipid bilayer environments. It was demonstrated that the denature state in lipid bilayers is a large expanded and dynamic conformational ensemble despite the quasi-two

dimensional physical constraints of the lipid bilayer. By comparing this result to the polymer models, I suggest that the lipid bilayer is reasonably good at solubilizing the denatured states of membrane proteins and this feature of bilayer may help membrane proteins to prevent the formation of collapsed misfolded states and to fold depending on the specific intra- or inter-molecular interactions. This finding implies the important role of lipid bilayers for the membrane protein folding, which further demonstrates the advantages of the non-disrupting steric trap methods.

Packing interaction is one of the critical driving force in the second-stage of membrane protein folding. However, it has been found that the protein interior is not optimized for tight packing. The packing defects, including pockets and voids, are prevalent in the protein interior. It has been speculated that packing defects may be required for ligand binding, transport or conformational changes that are necessary for function. Steric trap method does not depend on the activity readout of the target protein so that it can be used for stability measurement for inactive mutants, enabling the study of stability-function relationship directly under native conditions. By carefully designing the cavity-filling mutations and testing their impacts on the stability and activity of GlpG, we suggest that the packing defects are required for the functionally important movement of the structural elements in GlpG. This study provides an example of applying steric trapping method into the study of molecular driving forces in membrane protein folding.

Overall, in this dissertation research, I not only developed a handful of powerful tools for studying membrane protein folding, but also discovered some crucial elements related to the folding, stability, conformation and function of a rhomboid protease GlpG. Further efforts are being made to obtain the folding energy landscape of GlpG in the lipid bilayer environment and the detailed map of the cooperativity network in micelles and bilayers. Steric trapping could be applied to

explore the role of bilayer environment and the role of other driving forces in membrane protein folding. What's more, the whole set of studies is promising to be transferred to other α -helical membrane proteins.

Conclusion

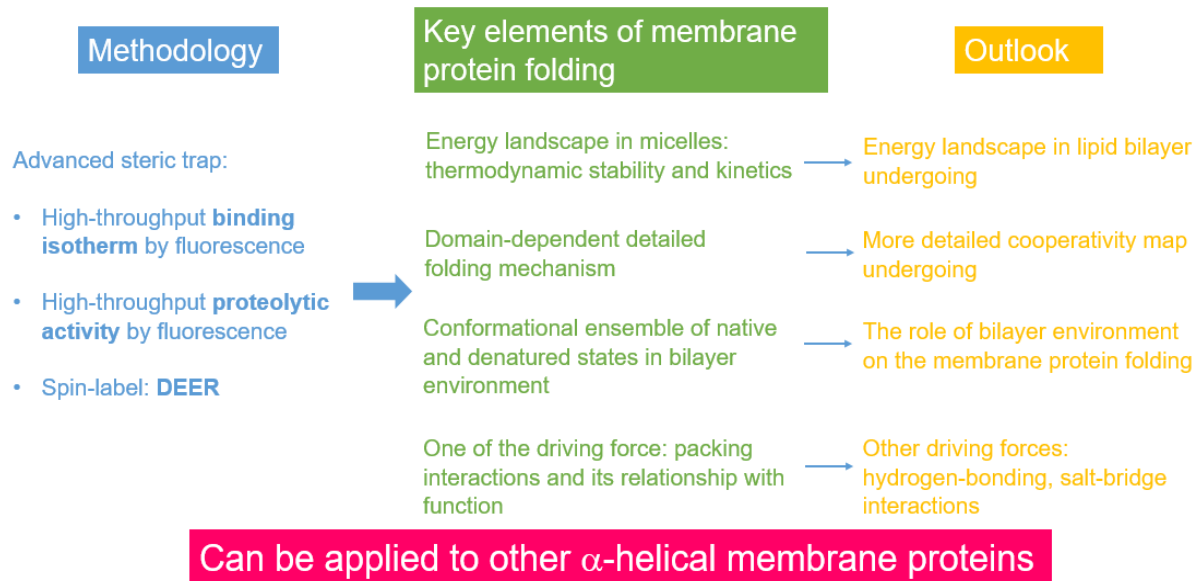


Figure 5.1 Conclusion and outlook of the dissertation research.