

FOLDING, STABILITY, AND DEGRADATION OF MEMBRANE PROTEIN IN THE
BILAYER

By

Fathima Shaima Muhammed Nazaar

A DISSERTATION

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

Chemistry-Doctor of Philosophy

2022

ABSTRACT

FOLDING, STABILITY, AND DEGRADATION OF MEMBRANE PROTEIN IN THE BILAYER

By

Fathima Shaima Muhammed Nazaar

Protein quality control involves the regulation of functional protein concentration at an optimal level in cells. To achieve this cellular need, a variety of biomolecular phenomena including protein synthesis, protein folding, chaperone action, and protein turnover are coordinated and balanced. While many studies on protein quality control focus on water-soluble proteins, it is not well understood how the quality control of membrane proteins is maintained. However, this question has been challenging to address due to difficulties in establishing tractable model systems in the lipid bilayer environment. This dissertation aims to answer two specific problems in membrane biology: 1) How does the lipid bilayer influence the folding and cooperativity of membrane proteins? 2) How do the intrinsic folding properties of membrane proteins influence their susceptibility to degradation? Using the intramembrane protease GlpG as a model, I find that, compared to micelles, the lipid bilayer enhances the stability of the protein by facilitating residue burial in the protein interior and strengthening the cooperative interaction network. Also, I find that conformational stability is not a major determinant of degradation rates of membrane proteins, and rather, the hydrophobicity of transmembrane segments or the conformational distribution of denatured state ensembles impact more. This finding suggests that the rate-limiting step of FtsH-mediated degradation of membrane proteins is not substrate denaturation but the dislocation of the hydrophobic transmembrane segments from the membrane to water. My studies will contribute to the fundamental understanding of the lipid bilayer as a solvent mediating folding, function, and quality control of membrane proteins.

Copyright by
FATHIMA SHAIMA MUHAMMED NAZAAR
2022

*Dedicated to my beloved parents
Nazaar and Kashiban*

ACKNOWLEDGMENTS

First, I would like to express my deepest gratitude to my Ph.D. mentor Dr. Heedeok Hong for his guidance, inspiration, and support during the research and thesis work. I am fortunate to get the opportunity to work with you. In the past six years, you helped me with research, from designing small experiments to developing advanced experiments that can explore new avenues in the future. You shaped me into a person who can strongly face any question on and off the lab and find solutions under any circumstances. These six years would not have been easy without your support. Under your guidance, I found my passion and developed the courage to pursue it, which means much more than I can express in my words here.

I am grateful to my graduate committee members Dr. David Weliky, Dr. Danial Jones, and Dr. James Geiger. Thank you for your support, guidance, and valuable suggestions for my graduate work.

I want to thank my lab members, past and present, whom I have enjoyed all these years and exchanging ideas, expertise, and support. Without you, my lab life would have been less enjoyable. I would like to thank Dr. Miyeon Kim for being thoughtful and warm and always making me happy with the cutest gifts. I would also like to thank Dr. Kristen Gaffeny, Dr. Ruiqiong Guo, and Dr. Yiqing Yang for being wonderful lab mates. You guys are so welcoming and gave me the support and guidance to understand the research better. I would also like to thank Mihiravi Gunasekara, Jiaqi Yao, Manoj Rana, Zhen Li, and Saba Kanwal for being a friend in need even outside the lab. I value all your friendship and valuable time, and friendly discussions beyond science. I thank Seung-Hyo Rhee, the former undergraduate student who worked with me for chapter 2.

I would like to thank Dr. Seung-Gu Kang, collaborator for Chapter 2, who performed the extensive MD simulations. I extend my thanks to all the members of the Department of Chemistry, including faculty members, staff, and the entire Michigan State University community, for giving me the opportunity and the environment to pursue my highest-level degree.

I express my sincere gratitude to parents, for installing the importance of education and for their understanding, support, encouragement, and unconditional love. I thank my siblings Imara, Raza, and Najeela for their love, support, and motivation. Most importantly, I would like to thank my amazing and supportive husband, Ishan, for being the pillar of my life.

Finally, I would like to thank my teachers, friends, and colleagues for their invaluable support throughout my carrier.

TABLE OF CONTENTS

LIST OF TABLES	x
LIST OF FIGURES	xi
CHAPTER 1 Introduction to membrane protein folding, stability, and degradation	1
1.1 Protein quality control in cellular health.....	2
1.2 Membrane protein structure, function, and biogenesis	4
1.2.1 Membrane protein folding problem: The two-stage model.....	7
1.2.2 Emerging roles of lipids in membrane protein structure, stability, and function.....	10
1.3 Methods to study membrane protein folding and thermodynamic stability	13
1.4 Rhomboid proteases as a model to study membrane protein folding	15
1.4.1 iRhoms: Catalytically inactive rhomboids homologous	17
1.4.2 Folding studies of <i>E. coli</i> GlpG: A rhomboid model substrate	18
1.5 Membrane protein degradation in cells.....	21
1.5.1 AAA+ proteases as degradation machines in bacteria	22
1.5.2 How are substrates recognized?	23
1.6 FtsH: Universally conserved and only membrane-bound protease in <i>E.coli</i>	24
1.6.1 Domain arrangement of FtsH	25
1.6.2 Three-dimensional structure of FtsH.....	25
1.6.3 Previous studies on FtsH-mediated protein degradation	27
1.7 Project description	28
REFERENCES	31
CHAPTER 2 Lipid bilayer strengthens the cooperative network of the rhomboid intramembrane protease GlpG.....	40
2.1 Summery	41
2.2 Introduction.....	42
2.3 Materials and methods	45
2.3.1 Expression and purification of GlpG.....	45
2.3.2 Labelling of GlpG and determining labeling efficiency by gel-shift assay	45
2.3.3 Expression, purification, refolding, and labeling of mSA.....	46
2.3.4 Expression and purification of GlpG substrate SN-LacYTM2	48
2.3.5 Preparation of native and sterically denatured GlpG in micelles	48
2.3.6 Activity-binding equilibrium to obtain reversibility	49
2.3.7 Fitting of second binding phase to obtain the thermodynamic stability of GlpG	49
2.3.8 Construction of binding isotherms to determine the thermodynamic stability of GlpG.....	50
2.3.9 Mapping the cooperative interactions in bilayers and micelles	51
2.3.10 Measuring the incorporation of native and denatured GlpG into bilayer phase of bicelles by fluorescence quenching assay	52

2.3.11 Measuring the intrinsic biotin affinity of mSA variants in bicelles	53
2.3.12 Proteinase K digestion to prove denaturation of GlpG by steric trapping	56
2.3.13 Probing the side chain contribution to GlpG stability by mutation.....	56
2.3.14 Molecular dynamics simulations of GlpG wild type in bilayers and micelles...	57
2.3.15 Examining the equilibrium of protein and amphiphiles	58
2.4 Results	59
2.4.1 Establishing the reversible folding system of GlpG in bicelles	59
2.4.2 Lipids promote the residue burial in the folding	63
2.4.3 Lipids strengthen the residue interaction network of GlpG	68
2.4.4 Lipids weakly solvate GlpG compared to micelles.....	71
2.5 Discussion	77
2.5.1 Enhanced stability and cooperativity can be explained by inefficient lipid solvation	77
2.5.2 Role of lipid solvation in membrane protein folding	79
2.5.3 Implications in function and disease mechanisms involving membrane proteins	79
2.6 Concluding remarks and outlook	81
APPENDICES	82
REFERENCES	100

CHAPTER 3 Dissecting folding-degradation relationship of membrane proteins in bilayer.....	107
3.1 Introduction.....	108
3.2 Material and methods.....	111
3.2.1 FtsH expression and purification.....	111
3.2.2 Expression and purification of GlpG for <i>in vitro</i> degradation assay.....	112
3.2.3 NBD-labelling of GlpG variants	113
3.2.4 Measuring the degradation rate of GlpG variants by NBD fluorescence.....	113
3.2.5 Expression, purification, and labeling of variants of GlpG TM to obtain thermodynamic stability using steric trapping	115
3.2.6 Measuring the intrinsic denaturation rate of GlpG by ProK digestion.....	116
3.2.7 Measuring the proteolytic activity of GlpG in DMPC/DMPG/CHAPS Bicelles	117
3.2.8 Measuring GlpG degradation <i>in vivo</i>	117
3.3 Results	118
3.3.1 The effect of mutations on conformational stability and activity of GlpG	118
3.3.2 Degradation rates of the membrane protein GlpG are independent of the conformational stability	125
3.3.3 The hydrophobicity of TM segments and interfacial loop residues contributed to the degradation rate	129
3.3.4 Conserved prolines in iRhoms are highly destabilizing, but its effect is compensated by a combination of inactivating mutations and evolved to be degraded similarly to active rhomboids	133
3.4 Discussion	138
APPENDICES	143

REFERENCES	149
CHAPTER 4 Concluding remarks and outlook	154
REFERENCES	159

LIST OF TABLES

Table 2.1.	The change in thermodynamic stability ($\Delta\Delta G^{\circ}_{WT-Mut}$), activity (relative to wild type) and residue burial (f_{ASA} : fraction of buried side-chain area) of GlpG variants in DMPC/CHAPS bicelles.....	83
Table 2.2.	The change in thermodynamic stability ($\Delta\Delta G^{\circ}_{WT-Mut}$), activity (relative to wild type) and residue burial (f_{ASA} : fraction of buried side-chain area) of GlpG variants in DDM micelles.....	84
Table 3.1.	The change in thermodynamic stability ($\Delta\Delta G^{\circ}_{WT-Mut}$), activity (relative to wild type), and residue burial (f_{ASA} : fraction of buried side-chain area) of GlpG variants.....	123
Table 3.2.	Kinetic parameters for GlpG degradation by FtsH.....	145
Table 3.3.	Kinetic parameters of GlpG degradation by FtsH with iRhom mutations...	146

LIST OF FIGURES

Figure 1.1.	Protein homeostasis network in cells.....	4
Figure 1.2.	The two-stage model of membrane protein folding.....	7
Figure 1.3.	Schematic illustration of proteolysis mediated by AAA+ proteases.....	23
Figure 1.4.	The domain arrangement and structure of FtsH.....	26
Figure 2.1.	Establishment of the reversible folding system of GlpG in a lipid bilayer environment	60
Figure 2.2.	Biotin labeling of double-cysteine variants of GlpG.....	61
Figure 2.3.	Reversibility of the folding and mSA binding of GlpG and their coupling induced by steric trapping in bicelles.....	65
Figure 2.4.	Comparison of GlpG stability measured at N- and C-subdomains in micelles and bicelles	67
Figure 2.5.	Comparison of mutational impacts on the stability in micelles and bicelles depending on the degree of the residues targeted for mutation for N-subdomain.....	67
Figure 2.6.	Comparison of mutational impacts on the stability in micelles and bicelles depending on the degree of the residues targeted for mutation for C-subdomain.....	68
Figure 2.7.	Lipid bilayer strengthens the cooperative network in GlpG.....	71
Figure 2.8.	All-atom MD simulation of GlpG WT in a DMPC lipid bilayer and DDM micelles.....	75
Figure 2.9.	Dissociation of lipid and detergent molecules at the protein surface or from them selves in the bulk phase measured by the lipid-contact autocorrelation function on time.....	76
Figure 2.10.	Biotin labeling of double-cysteine variants of GlpG.....	85
Figure 2.11.	Steric trapping induced GlpG denaturation monitored by proteinase K (ProK) digestion.....	86
Figure 2.12.	Activity assay to measure GlpG denaturation induced by steric trapping...	87

Figure 2.13.	Determination of the binding affinities ($K_{d, \text{biotin}}$) of mSA variants to biotin labels (BtnPyr) on GlpG using Förster resonance energy transfer in bicelles.....	88
Figure 2.14.	Determining the dissociation time constant ($\tau_{\text{off, pseudo}}$) of mSA variant from sterically denatured GlpG in bicelles.....	89
Figure 2.15.	Comparison of the intrinsic binding affinities ($K_{d, \text{biotin}}$) of mSA variants to biotin labels (BtnPyr) in micelles and bilayers.....	90
Figure 2.16.	Incorporation of native and sterically denatured GlpG into preformed bicelles.....	91
Figure 2.17.	Binding isotherms between double-biotin variants of GlpG and monovalent streptavidin (mSA) to determine the thermodynamic stability of WT and variants using steric trapping in DMPC/CHAPS bicelles.....	92
Figure 2.18.	Mapping of mutation-induced stability changes onto GlpG structure.....	94
Figure 2.19.	Schematic description of the cooperativity profiling method.....	95
Figure 2.20.	The features of cooperativity profiles in lipid bilayers still preserve those in micelles but to a less extent.....	96
Figure 2.21.	The effects of the location of biotin labels and hydrophobic environment on the proteolytic activity GlpG and variants.....	97
Figure 2.22.	Mapping of mutation-induced activity changes of GlpG onto the structure.....	98
Figure 2.23.	Modeling of the micellar systems for MD simulation.....	99
Figure 3.1.	Steric trapping scheme.....	119
Figure 3.2.	Mutational impact on the stability and activity of GlpG.....	122
Figure 3.3.	Spontaneous denaturation monitored by ProK digestion.....	125
Figure 3.4.	Targeted mutations to study thermodynamic stability and spontaneous denaturation rate in bicelles.....	126
Figure 3.5.	Fluorescence-based degradation assay to monitor membrane protein degradation.....	127
Figure 3.6.	The correlation plots of degradation rate by FtsH ₆ vs conformational stability of GlpG.....	129

Figure 3.7.	Effect of increased hydrophobicity on degradation.....	131
Figure 3.8.	Effect of increased hydrophilicity on degradation.....	132
Figure 3.9.	Multi-sequence alignment of metazoan iRhoms to <i>E. coli</i> GlpG.....	135
Figure 3.10.	Thermodynamic stabilities of GlpG mutants measured by steric trapping	136
Figure 3.11.	Correlation plots for iRhom sequence modifications.....	137
Figure 3.12.	Proteolytic activity of GlpG variants.....	144
Figure 3.13.	Spontaneous denaturation of GlpG monitored by ProK digestion.....	147
Figure 3.14.	The correlation between the conformational stability of soluble substrate Titin-I27 and the degradation rate by the water-soluble AAA ⁺ protease, ClpXP ₆	148

CHAPTER 1

Introduction to membrane protein folding, stability, and degradation

1.1. Protein quality control in cellular health

Proteins are versatile and structurally complex biological macromolecules. They are synthesized on ribosomes as linear chains of amino acids and then fold into a well-defined three-dimensional structure to function. The native conformations of most water-soluble proteins are known to be optimized at a free energy minimum structure as first proposed by Anfinsen.¹ The stability of water-soluble proteins is marginal, being of only $\Delta G^{\circ} = -5$ to -10 kcal/mol in their physiological environment.² The small free energy difference is the outcome of large unfavorable contributions from the conformational entropy and large favorable contributions from the enthalpic internal energy (e.g., backbone and side-chain H-bonding interactions, charge-charge interactions, and Van der Waals interaction) and entropic hydrophobic effect.^{2,3}

Protein folding occurs in a highly crowded cellular environment with cytosolic protein concentrations of 300-400 g l⁻¹.⁴ Resultant excluded volume effects strongly increase the non-native and structurally flexible proteins to aggregate.⁵ Thus, due to the marginal stability, conformational stability, and macromolecule crowding, protein folding is error-prone, and protein quality control mechanisms have evolved to ensure efficient folding and prevent aggregation. Maintaining proteome homeostasis or ‘proteostasis’ is essential for cellular and organismal health.⁶ Proteostasis involves complex, interconnected pathways that influence protein synthesis, folding, trafficking, disaggregation, and degradation. Major components of the proteostasis network are molecular chaperones and proteases, which assist in protein folding and degrade proteins with abnormal conformation, respectively.^{7,8}

Molecular chaperones are evolved to assist the folding in several ways. The chaperons that participate in de novo folding recognize generic structural features of nonnative proteins, primarily exposed hydrophobic amino acid residues.⁹ Then, they bind to the hydrophobic residues and

promote folding by the kinetic partition of nonnative states.¹⁰ Although the native conformation of a given protein is encoded by its amino acid sequence, in cells, many proteins require assistance from molecular chaperones to fold efficiently in a biologically relevant time scale.¹ In addition to their role in de novo folding, chaperones are also involved in protein refolding, disaggregation, trafficking, and oligomeric assembly. There are two types of chaperones with different modes of action: 1) foldases (or ATPases)¹¹, which are ATP-dependent and accelerate the transition of non-native conformations towards native conformations, and 2) holdases¹², which do not use ATP and protect the client protein from aggregation. In addition to chaperones, cellular degradation machinery also carefully chooses and degrades misfolded, aggregated, and functionally no longer needed proteins to prevent toxic aggregation.¹³ Two major pathways carry this in eukaryotes: ubiquitin-proteasome and autophagy-lysosome. In bacteria, this is predominantly carried out by AAA+ proteases.¹⁴ Therefore, the balance between folding and degradation is crucial in the quality control of proteins.

Understanding how cells maintain proteostasis in membranes is vital because 20-30% of genes encode membrane proteins¹⁵ and membrane proteins carry out numerous essential functions such as material exchange, cellular signaling, synthesis of metabolic energy, catalysis, and maintenance of ionic balance. Also, imbalances in membrane protein folding and degradation resulting from protein homeostasis deficiencies have been linked to aging and human diseases.¹⁶ The imbalances include: 1) excessive degradation of proteins that can lead to “loss of function” diseases such as cystic fibrosis (excessive degradation of cystic fibrosis transmembrane regulator)¹⁷, Charcot-Marie-Tooth’s diseases (excessive degradation of peripheral myelin protein 22)¹⁸, and 2) accumulation of misfolded and aggregated proteins that can lead to “gain of function” diseases such as Alzheimer’s disease (aggregation of A β peptides derived from amyloid precursor protein

or aggregation of phosphorylated tau proteins)¹⁹, Parkinson's disease (aggregation of α -synuclein fibrils).¹⁹ Therefore, understanding the detailed molecular mechanism of membrane protein quality control will aid in understanding the molecular determinants of disease mechanisms and finding cures to the diseases.

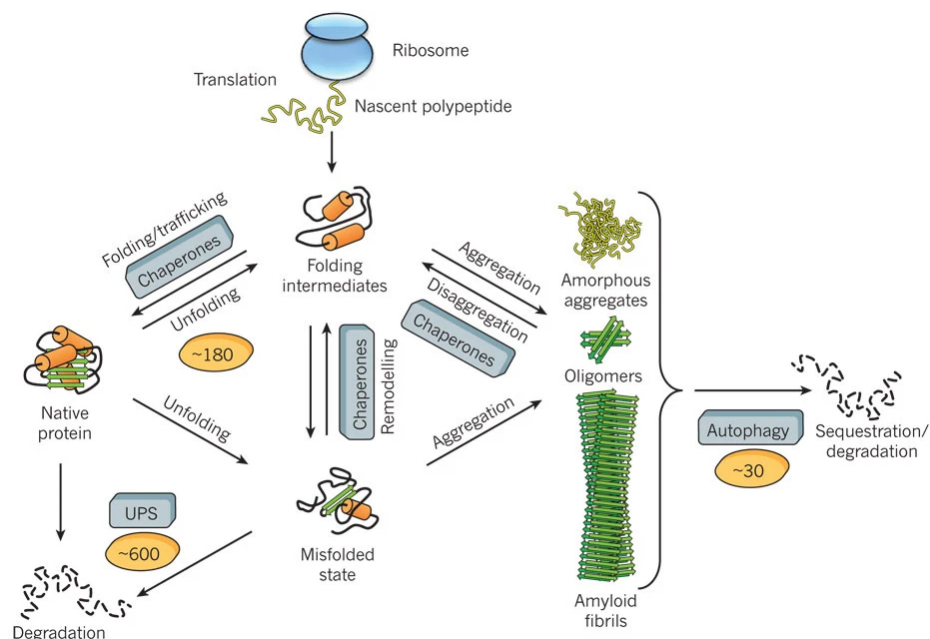


Figure 1.1. Protein homeostasis network in cells. The number in the yellow oval represents the number of cellular components involved in the process.⁵

1.2. Membrane protein structure, function, and biogenesis

The basic unit of the cell membranes are phospholipids which are organized in two monolayers with the polar head group exposed to water and the acyl chains buried in a central hydrophobic core. Cell membranes also contain glycolipids and sterols, which regulate the membrane's fluidity, cell signaling, and cell-cell communications.²⁰ As described from X-ray diffraction studies of the hydrated lipid bilayers, the total membrane thickness is $\sim 55\text{-}60 \text{ \AA}$, with a hydrocarbon core that occupies a total of $30 \text{ \AA}$, and each interfacial region accounts for $10\text{-}15 \text{ \AA}$.²¹ The overall bilayer structure is maintained through a balanced, complex lateral pressure profile.^{22,23} This encompasses

a positive pressure at lipid head groups and hydrocarbon chains due to the electrostatic charges, steric repulsions, and chain collisions. The pressure at the water-bilayer interface is typically negative due to the cohesive hydrophobic effect.

In contrast to water-soluble proteins, membrane proteins evolved to fold and function in the lipid bilayer, physically anisotropic and chemically heterogeneous. Membrane proteins (MP) constitute ~30% of the total surface area of the cell membranes.²⁴ Membrane proteins are classified into two classes, α -helical and β -barrel, depending on the secondary structure of the lipid-embedded region.²⁵ α -helical membrane proteins are dominant in all types of the cellular membranes except for the outer membranes of gram-negative bacteria, mitochondria, and chloroplasts where the β -barrel type prevails. Structural studies reveal that membrane proteins share several key structural features^{25,26}: 1) The lipid-contacting surfaces are dominantly composed of nonpolar residues, but the average polarity of the protein interior varies in the two types of membrane proteins. For α -helical, the protein interior is almost nonpolar and packed tightly as soluble proteins while the interior is largely polar for β -barrel; 2) The average length of traversing secondary structure elements is 15–25 amino acids for α -helical and 10–12 amino acids for β -barrel, and they expand over the hydrophobic thickness of 25–30 Å in the bilayer; 3) Amino acid composition of membrane proteins depends on the membrane depth. For example, Arg and Lys residues are more abundant in the cytoplasmic loops than in the periplasmic or extracellular loops of the proteins,²⁷ known as the positive inside rule. The positive charges are known to stabilize the topology of membrane proteins. Polar aromatic residues such as Tyr and Trp are enriched in the water-membrane interfacial regions. Nonpolar residues including Val, Leu, Ile, and Phe are most probable in the center of the bilayer core.

My dissertation research focuses on the folding and degradation of α -helical membrane proteins. The biogenesis process of helical membrane proteins is highly conserved in the kingdoms of life and achieved via a complex translocon-mediated membrane-insertion process discovered by Blobel & Dobberstein.²⁸ Nascent polypeptide chain of an integral membrane protein is recognized by hydrophobic stretches (10–20 amino acids) or cleavable N-terminal signal sequence by the signal recognition particle (SRP) on the ribosome and binding of SRP to the signal sequence slows the translation. This translational complex is targeted to the endoplasmic reticulum (ER) in eukaryotes or the cytoplasmic membranes in bacteria and archaea via the SRP receptor (SR). The nascent polypeptide chain in the SRP-SR complex is then transferred to the membrane-integrated protein conduction channel, called translocon. The SRP-SR complex disassembles, and translational pause is relieved. These events are coordinated by GTP hydrolysis on the ER membrane.²⁹ During translation, the translocon integrates the polypeptide segments into the membrane or across the membrane based on their hydrophobicity. It is well accepted that the membrane integration of the polypeptide segments is mediated by their thermodynamic partitioning between the translocon and the lipidic environment. However, it is unclear how hydrophobic stretches of amino acids are positioned in the translocon (i.e., within the pore of the translocon vs. outside of the translocon) and where they adopt a helical conformation. Interestingly, a hydropathy analysis of helical membrane proteins in *E. coli* predicts that about half of the transmembrane (TM) segments have a low tendency to insert into the membrane ($\Delta G_{\text{app, insertion}} > 0$).³⁰

1.2.1. Membrane protein folding problem: The two-stage model

The two-stage model for helical membrane protein folding has been proposed by Popot and Engelman based on the structure and folding studies of bacteriorhodopsin.³¹ The folding of helical membrane proteins can be divided into two thermodynamically distinct stages. In stage 1, stable TM helices are formed upon insertion of hydrophobic polypeptide segments into the membrane. In stage 2, the TM helices laterally interact to form a tertiary fold with a functional native structure. This model was modified to include a third stage in 2003, which involves the incorporation of prosthetic groups, folding of the loop region, and oligomerization.³² This model was further dissected into a four-step model involving partitioning into the water/lipid interfaces, folding in the interfaces, insertion into the bilayer, and association of TM helices within the lipid bilayer.²⁵ Regardless of the detailed mechanism of membrane insertion, these models assume the preformation of stable TM helices as a requirement for association with one another to achieve the native structure.

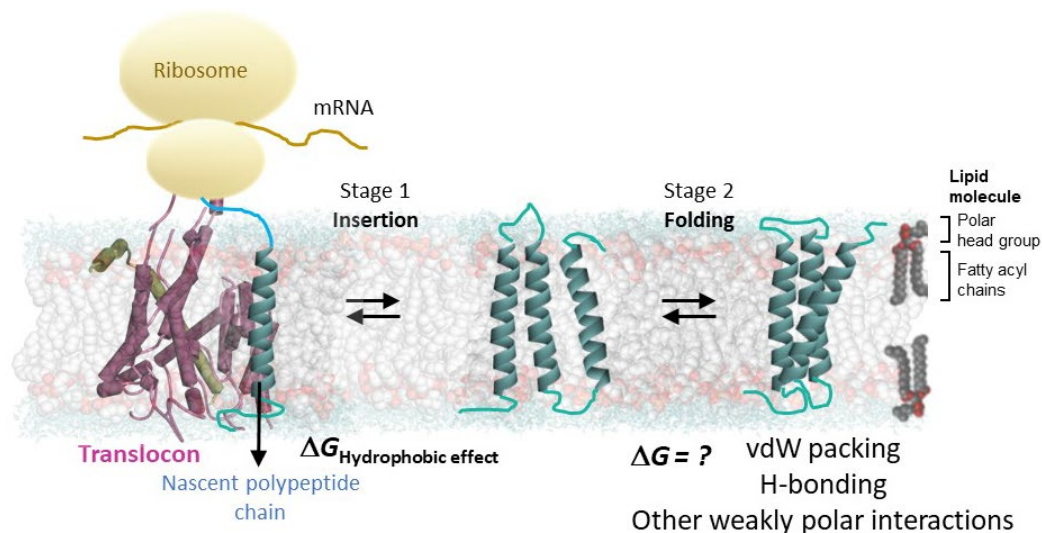


Figure 1.2. Two stage model of membrane protein folding. In the first stage, hydrophobic polypeptide segments are inserted into the membrane bilayer to form stable TM helices. In the second stage, individual TM helices assemble to form native structure.²⁶

Various studies have been carried out to understand the driving forces of membrane folding. Stage 1 is known to be driven by the hydrophobic effect caused by the burial of nonpolar residues and the favorable formation of the backbone hydrogen bonds in the nonpolar bilayer core. Over the past decades, to understand the effect of hydrophobicity in stage 1, various hydrophobicity scales have been derived using experimental and computational methods. The hydrophobic effect has been empirically related to the side-chain solvent-accessible surface area (ASA) through the energy termed the nonpolar solvation parameter. This relationship is quantified as the free energy (cal mol^{-1}) gain per area ($\text{\AA}^2$) of the nonpolar surface excluded from the water and buried in a nonpolar solvent. Toward the effort to experimentally obtain an accurate transfer free energy value for each amino acid residue, White and coworkers developed a pentapeptide (Ace-WLxLL: x can be any of the 20 naturally occurring amino acids) based host-guest system to derive a hydrophobicity scale for each residue. To measure the partition between water and the hydrophobic bilayer core, they used octanol as a hydrophobic medium. To measure the partition between water and the water-membrane interfacial regions, they employed 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC) vesicles.^{33,34}

Hessa and von Heijne developed a biological hydrophobicity scale for helical membrane proteins based on an insertion assay using a test segment (H segment) that is inserted between the two N terminal TM segments of a leader peptidase (Lep) and the water-soluble P2 domain of an alkaline phosphatase.³⁵ Two acceptor sites for N-linked glycosylation flank the H segment. If the H-segment is hydrophobic enough and inserted into the membrane to form a TM topology, the P2 domain resides in the cytoplasm, and only one site upstream of the H-segment is glycosylated. The construct is challenged against an in vitro translation system that contains the SRP targeting machinery and the ER-derived endosomes with a translocon. If the H-segment is hydrophilic, it is

translocated across the ER membrane, and both sites reside in the ER lumen and are glycosylated. The number of glycosylated sites induces the size difference between inserted and translocated H-segments, and the fraction of insertion can be quantified on SDS-PAGE. By properly designing the sequence of the H-segment, the ability of each amino acid residue to induce membrane insertion, as well as the position-dependence of the ability along with the membrane depth, can be quantified in the free energy scale.

Later, Moon and Fleming developed a hydrophobicity scale based on the guanidine hydrochloride (GdnHCl)-induced equilibrium folding of the outer membrane protein phospholipase A (OmpLA) from the aqueous phase to large unilamellar vesicles composed of 1,2-dilauroyl-sn-glycero-3-phosphocholine (DLPC).³⁶ This scale has the advantage of using OmpLA, which has a defined structure in the membrane and undergoes spontaneous insertion into the membranes from a water-soluble unfolded state such that the folding free energy can be determined using GdnHCl denaturation. In this study, the amino acid substitution of the wild-type (WT) alanine residue in the center of the bilayer core is made for one of the rest 19 amino acid residues, and the difference free energy change ($\Delta\Delta G^{\circ}_{N-U, WT-Mut}$) between WT and a mutant is measured. For a given substituting residue in the mutant, $\Delta\Delta G^{\circ}_{N-U, WT-Mut}$ represents the transfer of free energy of the residue from water to the center of the membrane relative to Ala (the wild-type residue at the membrane center). By changing the position of the substitution along a membrane-spanning β -strand, the membrane-depth dependence of the free energy change can be measured for a given residue.

Other than experimental scales, computational and statistical scales have also been developed to predict the hydrophobicity of individual amino acid residues. Liang's group reported a computational scale using the prediction of the folding free energy of the outer membrane β -

barrel proteins (OMPs) by combining an empirical energy function with a reduced state-space model.³⁷ This method yields the depth-dependent free energy transfer for 20 amino acids. Moreover, the regions necessary for protein function and structural anomalies can be predicted by analyzing the context-dependence of transfer free energies at specific positions in OmpLA.

The molecular forces that drive Stage 2 contribute to attaining the native tertiary fold and remain elusive. This is mainly due to the inherent difficulties in achieving reversible folding of helical membrane proteins in the bilayer environment. For water-soluble proteins burial of non-polar amino acids in the protein core is mainly driven hydrophobic effect. For membrane proteins, the free energy gained from the hydrophobic effect is primarily consumed during the insertion step, and they cannot experience hydrophobic forces as the bilayer lacks water. Therefore, it is possible that other molecular forces such as hydrogen bonding interactions, van der Waals packing interactions, weak polar interactions, and salt bridge interactions could play an essential role in the association of helices. However, it is unclear how these individual forces are balanced to stabilize membrane proteins in a bilayer environment.

1.2.2. Emerging roles of lipids in membrane protein structure, stability, and function

The host lipid bilayers are fascinating two-dimensional microenvironments whose composition can regulate membrane protein function. This regulation may depend on specific interactions between proteins and individual molecules in the bilayer and non-specific interactions between proteins and the bilayer behaving as a physical entity with collective material properties (e.g., thickness, intrinsic monolayer curvature, or elastic moduli).²³ Biological membranes consist of a variety of lipid types. The predominant lipid species in the cytoplasmic and subcellular membranes of eukaryotic cells are phosphatidylcholine

(PC:41-50%), phosphatidylethanolamine (PE: 17-38 mol %), phosphatidylserine (PS: 1-6 mol %), sphingomyelin (SM: 2-20 mol %), and cholesterol (~1 mol %).³⁸ In contrast, the inner membrane of gram-negative bacteria contains PE (71.4 mol %) predominantly with a smaller amount of phosphatidylglycerol (PG: 23.4 mol %) and cardiolipin (5.3 mol %).³⁹ The lipid composition profile (e.g., lipid head group size and charge, acyl chain length, etc.) is thought to be optimized for membrane protein insertion, folding, stability, and function.^{40,41,42} However, it is largely unknown how the complex lipid environment modulates the membrane protein folding and stability at the molecular level.

High-resolution crystal structures have provided insights into how lipids interact with membrane proteins. One example is the detergent extracted crystal structure of formate dehydrogenase-N (Fdh-N). Fdh-N is crystalized as a physiological trimer in which cardiolipin molecules mediate the quaternary contacts between individual subunits.⁴³ Another example is the alternative complex III (ACIII), from the bacterium, *Flavobacterium johnsoniae*, crystallized with a detergent-free approach using styrene-maleic acid copolymer (SMA).⁴⁴ 11 phospholipid (PL) molecules have been resolved bound to the protein, and the electron densities of PLs are assigned to two key regions: The first region is between two of the ACIII subunits, suggesting the role of PLs in the stability of the protein. The second region is close to the menaquinol entry site in the ActB subunit suggesting the role of PLs in the protein's function. Bilayer curvature stress and lateral lipid-packing pressure have been shown to increase membrane protein stability. Bacteriorhodopsin is more resistant to irreversible thermal denaturation in DMPC/DOPC vesicles than in DMPC vesicles, presumably due to the higher curvature stress in the DMPC/DOPC vesicles.⁴⁵ The thermodynamic stability of LeuT increases as the PE content in PC liposomes increases.⁴⁰ Also, the addition of PE to PC vesicles stabilizes

the glycophorin A TM dimer, whereas adding lysoPC destabilizes it (thus lowering the lateral chain pressure).⁴⁶ In addition to protein stability, the curvature stress and lateral pressure can affect folding rates and intermediate formation in membrane protein folding. When the EmrE transporter is refolded or reconstituted into DOPC/DOPE or DOPG/DOPE vesicles, the refolding rate increases as the percentage of PE increases. However, the functional protein recovered decreases.⁴⁷ This is consistent with the hypothesis that an increase in lateral pressure inhibits insertion but facilitates the packing of TM helices or oligomerization. Also, refolding experiments of bacteriorhodopsin support the same idea. When bacteriorhodopsin folds into DMPC/DHPC bicelles, the rate constant for a rate-limiting folding step decreases as the DMPC fraction increases. It is also shown that the rate of formation and the population of folding intermediates of bacteriorhodopsin are directly affected by the lateral pressure when the folding in more stressed DPOPC bilayers is compared to that in DMPC/DHPC micelles.⁴⁸

Another emerging idea regarding the contribution of lipids and bilayers to membrane protein stability is the entropically driven “lipophobic effect,” which is analogous to the hydrophobic effect in soluble proteins.⁴⁹ The lipophobic effect explains the association of TM helices to reduce the lipid-exposed surface area, releasing the ordered solvating lipids to the more dynamic bulk lipids. This is likely to increase the entropy of the membrane and explains why membrane protein favors oligomeric states. To investigate the thermodynamic stability of naturally occurring bacteriorhodopsin lattice, the residues at the protein interface have been mutated to small amino acids, either Gly or Ala. Although most mutations destabilize the lattice as predicted, the I45A mutant located at the interface between interacting B and D helices of neighboring BR monomers stabilizes it; authors hypothesize this is due to the increased lipid entropy.⁵⁰ In another study,⁵¹ the entropic contribution to the solvation was up to -4 kcal/mol when the free energy of

glycophorin A helix dimerization was measured in SDS. During the dimerization, $\sim 400 \text{ \AA}^2$ lipid-exposed surface areas are buried. The favorable entropy of binding can be attributed to the release of SDS molecules from the dimerization interface to the bulk of SDS molecules.

1.3. Methods to study membrane protein folding and thermodynamic stability

Several methods have been developed to study the reversible folding of membrane proteins. Chemical denaturation using the strong anionic detergent sodium dodecyl sulfate (SDS) is a successful tool for studying the reversible folding of helical membrane proteins. This method has been applied to the folding studies of diacylglycerol kinase (DGK),⁵² bacteriorhodopsin,⁵³ and GlpG⁵⁴. A detergent-solubilized native protein is denatured by increasing the mole fraction of SDS ($X_{\text{SDS}} = [\text{SDS}] / ([\text{SDS}] + [\text{other mild detergents}] + [\text{protein}])$). Reversible refolding is achieved by increasing the mole fraction of non-denaturing mild detergents. Folding reactions can be monitored by cofactor binding, tryptophan fluorescence, and protease resistance.^{30,52} The equilibrium denaturation curves are typically fitted to a two-state model involving only the native and denatured states. The free energy of denaturation (i.e., $\Delta G^{\circ}_{\text{N-D}}$, the free energy difference between the native and denatured states) under native conditions is obtained by linearly extrapolating the $\Delta G^{\circ}_{\text{N-D}}$ values in the transition region to zero SDS mole fraction.⁵⁵ SDS denaturation has provided insights into the driving forces and transition states in membrane protein folding.^{45,53} However, the mechanism of SDS denaturation, the conformation of the denatured state, and the validity of the linearity between $\Delta G^{\circ}_{\text{N-D}}$ and SDS mole fraction are still elusive. A study with a series of model TM helical segments solubilized in SDS suggests that there may not be a large conformational change that corresponds to a true unfolding transition. Also, membrane proteins with relatively short, stable extra membranous loops may still preserve the tertiary structure in SDS.⁵⁶ Moreover,

the studies using steric trapping have revealed a nonlinearity of $\Delta G^{\circ}_{\text{N-D}}$ at lower SDS mole fractions, implying a complex interaction between membrane proteins and the mixed micelles composed of SDS and nondenaturing detergents.^{57,58}

Another method involves single-molecule force spectroscopy using magnetic tweezers. The Bowie's and Yoon's labs pioneered applying this method for studying membrane protein folding within the lipid environment:⁵⁹. A single membrane protein molecule is covalently linked to two DNA handles at the N and C-termini. One DNA handle with a biotin label is anchored to PEG-coated solid support via biotin-avidin binding, and the other DNA handle is tethered to a magnetic bead. The change in bead height is then measured as a function of the force applied, allowing the protein to unfold and refold in the bilayer environment. After multiple cycles of pulling and relaxing cycles, force-extension curves are constructed.

The steric trap is an ensemble method to study the thermodynamic and kinetic folding of both water-soluble and membrane proteins by coupling the unfolding of a biotin-tagged protein to the binding of bulky monovalent streptavidin (mSA, 52 kDa). This method has been applied to studying the association of glycophorin A TM dimer in detergent micelles and lipid bilayers.⁶⁰ It has also been applied to study the dissociation kinetics of DGK timer in detergent micelles and the thermodynamic stability of bacteriorhodopsin in bicelles.^{61,62} Steric trap has two requirements: 1) Two biotin tags that are spatially close in the native three-dimensional structure and distant in the primary sequence; 2) A method to monitor denaturation of the target protein or binding of mSA. The first mSA can bind to one of the biotin tags when the protein is in the folded state, and the second mSA binding is hindered by the steric clash between bulky mSA molecules. When the protein becomes transiently denatured, the second binding of mSA binds to the unoccupied biotin label and traps the target protein in the denatured state. Therefore, the first mSA binds with an

intrinsic affinity of biotin to mSA, and the second mSA binding coupled to denaturation is attenuated depending on the thermodynamic stability of the target protein. Refolding can be tested by inducing dissociation of bound mSA molecules upon the addition of an excess concentration of free biotin that competes with the biotin labels for mSA. Therefore, steric trapping can reversibly control the folding and denaturation reactions in the native bilayer and aqueous environments without using perturbants such as SDS and pulling force. The steric trap has broader application to various types of protein, including nonfunctional and misfolded, due to the development of new biotin probes.⁶³ The versatile biotin probes included three essential features: 1). A biotin group to bind to mSA, 2) a thiol-reactive group for conjugation to engineered cysteine residues on a target protein, and 3) a fluorescent or paramagnetic reporter group to sense the mSA binding or protein unfolding.

1.4. Rhomboid proteases as a model to study membrane protein folding

Rhomboid proteases are intramembrane serine proteases that hydrolyze a peptide bond near the lipid bilayer. The rhomboid family proteins are found in all branches of life, and their functions fall into four general categories: 1) activate growth factor signaling by liberating the membrane-anchored inactive form of growth factors via the cleavage of the peptide bond between the growth factor and membrane anchor.⁶⁴ For example, during *Drosophila* development, Rhomboid-1 in the Golgi apparatus cleaves off the epidermal growth factor (EGF) Spitz, from the membrane-anchor after being transported from the endoplasmic reticulum. Free Spitz is then secreted to the extracellular space and triggers EGF signaling in the target cells; 2) Mitochondrial homeostasis.⁶⁴ A rhomboid protease PARL in *Saccharomyces cerevisiae* cleaves PINK1 in the inner mitochondrial membrane to reduce Parkin recruitment to mitochondria. Without PARL cleavage,

PINK1 accumulates in mitochondria and fails to adequately recruited to damaged mitochondria; 3) Parasite invasion.⁶⁵ Malaria parasite-encoded rhomboids cleave parasite transmembrane adhesins to disassemble the junction between parasite and host at the end of the invasion; 4) Quorum sensing.⁶⁶ The rhomboid AarA of *Providencia stuartii* activates TatA by removing a small amino-terminal extension. This allows TatA to assemble into the Twin-arginine translocation machinery required for exporting protein (presumably quorum-sensing signal).⁶⁷ It is expected that there are many other uncharacterized rhomboids, and their biological functions are still to be discovered.

GlpG, the rhomboid protease of *E. coli*, was the first intramembrane protease whose crystal structure has been solved, and its structure and the catalytic mechanism are best characterized among rhomboids. Later, the crystal structure of GlpG from *Haemophilus influenzae* have been solved.^{68,69} The structure revealed the Ser/His catalytic dyad located ~ 10 Å below the bilayer plane, presenting the structural basis of intramembrane proteolysis. Towards the extracellular side, an aqueous cavity is found and known to provide water molecules essential for catalysis.⁶⁹ The catalytic core domain of GlpG is composed of six TM helices which are compactly packed with an asymmetric shape. Multiple sequence alignment suggests that rhomboids' core structures are highly conserved and may share a common catalytic mechanism.⁷⁰ Structural and mutational studies have identified four essential motifs that are thought to constitute the catalytic active site:⁷¹

- 1) The HxxxN motif in TM2 (His150 and Asn154 in *E.coli* GlpG). The His and Asn residues are known to stabilize the Ser201 oxyanion hole that is formed during the catalytic cycle; 2) the GxSG motif in TM4 (Ser201); 3) (A/G)H motif in TM6 (His254). Ser201 in TM4 and His254 in TM6 form the catalytic dyad; 4) Two G(A)xxxG(A) motifs in TM4 and TM6. In the crystal structures, these motifs allow close packing of the two TM helices harboring the catalytic dyad. The Ser/His

catalytic dyad is unique in the rhomboid protease family since conventional serine proteases possess a Ser/His/Asp catalytic triad.

Detailed mechanisms of proteolysis and substrate recognition mediated by rhomboid proteases are still not fully understood. Rhomboid proteases are known to cleave single-spanning membrane proteins. As a substrate, TM helices need to adopt a helical conformation to satisfy the hydrogen-bonding requirements of polar peptide groups, minimizing the energetically unfavorable exposure to the hydrophobic lipid bilayers. However, such secondary structural elements are generally poor substrates and need to be destabilized to become susceptible to proteolysis.⁷² Consistent with this, the natural substrates for *Drosophila* Rhomboid-1 (e.g., Spitz) possess helix-destabilizing residues, particularly the glycine–alanine motif, to facilitate local helix unfolding and proteolysis.⁷³ Interestingly, diverse rhomboids from bacteria, archaea, or mammals can cleave the substrates of *Drosophila* Rhomboid-1, which suggest that these enzymes recognize common conformational features in substrates.⁷⁴

1.4.1. iRhoms: Catalytically inactive rhomboids homologous.

Besides proteolytic activity, rhomboids have evolved to fulfill non-regulatory roles in metazoans. These inactive rhomboids are called iRhoms, classified as highly conserved rhomboid-like proteins.⁷⁵ In *Drosophila*, iRhom-1 regulates epidermal growth factor receptor signaling (e.g., EGFR pathways) in ER by inducing the degradation of EGF ligands. When EGF is overexpressed in cells, it binds to the iRhom-1 on the ER membrane which targets EGF to the endoplasmic reticulum-assisted degradation (ERAD) machinery for degradation.^{76,77} In mice, iRhom-2 works as a trafficking chaperone for membrane-bound protease ADAM17/TACE, the primary activator of inflammatory signaling induced by tumor necrosis factor.⁷⁸ In both cases, the iRhoms' function

seems to be mediated by binding to the single TM anchor of its client protein. Nonetheless, the mechanism of action is largely unknown for a majority of iRhoms.

Structurally, iRhoms have an extended cytoplasmic amino terminus. Most strikingly, they contain a highly conserved cysteine-rich luminal loop between the first two TM helices, named the iRhom homology domain. The role of the iRhom homology domain is unknown, but its conservation implies its functional significance. Multiple sequence alignment of iRhoms shows they lack essential catalytic residues (His and Ser, which form the catalytic dyad) for proteolysis. Some iRhoms are missing the serine residue, and others are missing the histidine residue. In some cases, they miss both of the residues. A common anomaly in all iRhoms is the proline residue adjacent to the position of the catalytic serine in the primary sequence.⁷⁵ This observation suggests that proline has been acquired before the loss of the catalytic dyad residues, disrupting the active site structure and removing the selective pressure to maintain either Ser or His.

1.4.2. Folding studies of *E.coli* GlpG: A rhomboid model substrate

The folding and stability of GlpG have been carried out by several groups using different methods in various hydrophobic environments. Urban and Baker studied the architectural principles that may support the intramembrane proteolytic function.⁷⁹ They carried out thermal and SDS denaturation studies of GlpG in DDM micelles over 150 mutants. They found four ‘key stone’ regions that are essential to the conformational integrity of GlpG and two main regions that are critical to the function: 1) Strong helix-helix packing between the asymmetric “glycine zipper,” GxxxGxxxA on TM6 and GxxxAxxG on TM4; 2) The opposite face of TM4, together with the intruded L3 loop, forms a second critical packing area mediated by larger nonpolar residues; 3) The hydrogen-bonded network between the TMs near the cytosolic side of GlpG; 4) The hydrogen-

bonded network that stabilizes the hairpin conformation of the L1 loop that lies parallel on the membrane plane.

Otzen group carried out an ϕ -value analysis of GlpG over 69 residue sites in DDM/SDS mixed micelles to study the folding mechanism.⁸⁰ ϕ -value analysis is developed to map the structure of transition state in protein folding by obtaining the ratio of the change in activation free energy of folding to the change in free energy of folding upon mutation of a specific site. $\phi = 0$ indicates that the site of mutation is unfolded and $\phi = 1$ indicates that the site of mutation is folded in the transition state. The chevron plot displaying the unfolding and refolding rates as a function of denaturant concentration has a V shape, showing that GlpG folds through a two-stage process without populating stable intermediates. The larger positive values cluster at the cytosolic side of TM1 and TM2, which is assigned as the folding nucleus. Interestingly, negative ϕ values are found in the loops 1-3, and these unusual values have been interpreted as the “back-tracking,” that is, this region undergoes conformational rearrangements to correctly position the folding core assisting the folding of the rest of the domain. TM helices 3–6 yield near-zero ϕ values, indicating that this part of the protein has denatured state-level structure in the transition state. This observation is reasonable as TM4 and TM6 harbor the active site, and TM5 is thought to have some flexibility to allow the substrates to bind to the active site.

Min and coworkers have carried out single-molecule pulling studies with magnetic tweezers for GlpG in the bilayer environment provided by DMPC/CHAPSO bicelles.⁸¹ GlpG unfolds cooperatively over a wide range of applied forces (13–31 pN), and the kinetic barrier for unfolding from the folded state is high, so the folded state has a long lifetime ($t_{1/2} \sim 3.5$ h). Cooperativity and a high kinetic barrier of unfolding would limit the existence of incorrect partially folded structures, which would be more prone to aggregation. Nonetheless, ‘force-jump’

experiments where force is increased rapidly and maintained at the constant value identify intermediates during unfolding. The two transient intermediates have been observed, and dwell time relative to the dwell time of the unfolded state is small (τ_1 and $\tau_2 < 2\%$ of τ_u), thus, supporting the largely cooperative unfolding of GlpG. Directionality of the mechanical unfolding of GlpG is important in the folding mechanism. Mutational studies for destabilizing a local structure suggest that unfolding starts at C-terminal and propagates to N-terminal as proposed by ϕ -value analysis.

The Hong lab has studied thermodynamic stability, folding cooperativity, and compactness of the denatured state of GlpG in DDM micelles using steric trapping.⁸² Using the novel thiol-reactive biotin tags with a spectroscopic reporter group (fluorophore or spin-label) and moving the position of the biotin pairs for probing the stability of the specific region, they have obtained the local and global stabilities of GlpG and elucidated two subdomains with distinct folding properties (rigid, stable N-subdomain and flexible, unstable C-subdomain). Also, using the “cooperativity profiling” method, which quantifies the degree of propagation of mutation-induced local structural perturbation, they have mapped the cooperative, localized, and over propagated side-chain interaction network in micelles. The cooperativity map reveals that cooperative interactions are clustered in multiple distinct regions, the central packing core in the protein interior, the TM4/TM6 interface harboring the catalytic dyad, and the residues near the water-retention site are critical for proteolytic function. They have further investigated the conformation of the denatured state ensemble (DSE) using limited proteolysis, mass spectrometry, and double electron-electron resonance spectroscopy (DEER) in DDM micelles DMPC/DMPG/CHAPS bicelles, and E.coli liposomes.⁸³ The DSE was identified as highly dynamic TM segments that involve membrane-topology changes and transient unfolding. Interestingly, in the lipid bilayer environment, it has been shown that the DSE retains a substantial degree of compactness relative to the fully expanded

model, implying that the lipid bilayer does not serve as a good solvent for membrane protein folding.

1.5. Membrane protein degradation in cells.

In eukaryotic cells, regulated degradation is mediated by the ubiquitin-proteasome system, which functions in the cytoplasm and nucleus. Ubiquitin, a small globular protein, is conjugated to a substrate protein and serves as a molecular tag for subsequent degradation.⁸⁴ Ubiquitin molecules are often added to the target protein, forming a linear or branched chain. At least four linked ubiquitin molecules are needed to bind to the 26S proteasome. Once the target protein binds to the proteasome, it is unfolded and translocated, driven by ATP hydrolysis on the AAA+ ring, and degraded in a central protease chamber. Ubiquitin is not degraded, is released from the proteasomes before translocation, and recycled.

Lysosomes are also key players in protein degradation. The lysosome is a membrane-enclosed organelle that contains a variety of hydrolytic enzymes.⁸⁵ Protein degradation in the lysosome occurs via a completely different mechanism than the ubiquitin-proteasome system. Proteins from the Golgi apparatus or plasma membrane destined for degradation are first collected in the vesicles and transported to the endosomes. The membrane fusion between the vesicles and endosomes incorporates the protein into the endosomes. Vesicles containing the proteins bud from the interior of the endosome, creating a multivesicular body (MVB) with many internal vesicles. The outer membrane of MVBs fuses with the lysosome membrane, exposing the internal vesicles to lysosomal hydrolytic enzymes for degradation.

In autophagy, double-membrane vesicles termed autophagosomes engulf membrane proteins, organelles, cytosolic proteins, protein aggregates, and even invasive pathogens and

transport these cargos to lysosomes.⁸⁶ There, the outer membrane of the autophagosome fuse with the lysosomal membrane, and the inner vesicle and its cargo are degraded—the degraded contents from the autophagic body are transported back to the cytosol for reuse.

1.5.1. AAA+ proteases as degradation machines in bacteria.

Energy-dependent proteolysis mediated by AAA+ proteases (ATPases associated with various cellular activities) is a major degradation pathway in bacteria. *E. coli* cells possess five AAA+ proteases: ClpXP, ClpAP, HslUV, Lon, and FtsH. All AAA+ proteases function as a large protein complex having two major functional and structural modules, a hexameric AAA+ ring ATPase and a compartmental protease. In FtsH and Lon, the AAA+ and protease domains are encoded in a single polypeptide chain and form functional hexamers. In ClpXP, ClpAP, and HslUV, the AAA+ and protease domains are encoded in separate polypeptide chains. Each module (ClpX, ClpA, and HslU: AAA+ ATPases; ClpP and HslV: compartmental proteases) is independently assembled, and then specific AAA+ and protease modules are associated with forming a functional AAA+ protease unit.⁸⁷

The recognition of proper substrates is mediated by a stretch of flexible peptide segments called a degradation marker or degron that binds to the pore residues in the AAA+ ring. Upon substrate binding, the conformational changes in the AAA+ ring powered by ATP binding and hydrolysis generate power-stroke motions and create pulling force against the bound substrate. Pulling the substrate occurs along the axial pore, and repeated cycles of ATP hydrolysis induce unfolding and translocation of the substrate into the protease chamber for degradation.⁸⁷

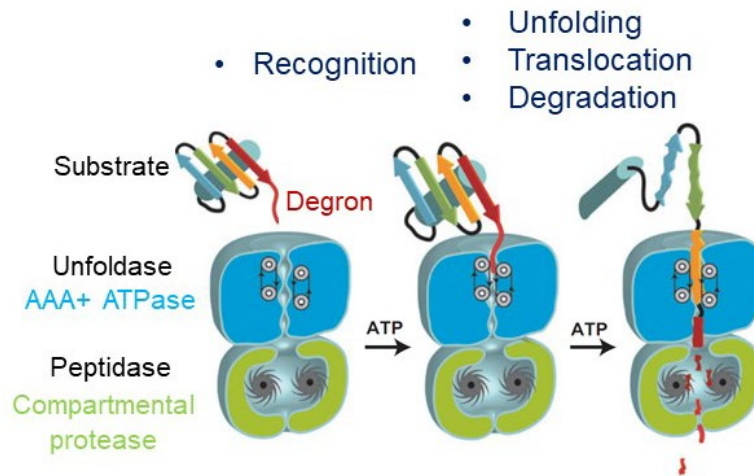


Figure 1.3. Schematic illustration of proteolysis mediated by AAA+ proteases. The AAA+ ring ATPase recognizes a degron. Cycles of ATP hydrolysis power substrate unfolding and translocation into compartmental protease in which substrate is proteolyzed.⁸³

1.5.2. How are substrates recognized?

Degradation is an irreversible chemical process. Hence, substrates must be chosen carefully. Each AAA+ protease has preferred amino acid sequences of degradation markers, but the sequence preference is broad and often overlaps between different AAA+ proteases. In some cases, specialized adapter or delivery proteins modulate the substrate selectivity by increasing the substrate binding affinity.⁸⁷ Degradation markers on substrates become accessible to AAA+ ATPases by unfolding, subunit dissociation, or cleavage by another protease. For example, the LexA repressor, which regulates the pleiotropic response (the SOS response) to DNA damage, undergoes autocleavage to produce an N-terminal and C-terminal fragment that can be recognized degraded by ClpXP. Without cleavage, intact LexA is not degraded by ClpXP.

Degradation markers for AAA+ proteases are typically a flexible, hydrophobic peptide stretch with a length of >20 amino acids. One example is the ssrA-tag located at the C-terminal end of the target protein. The sequence (-AANDENYALAA) comprises nonpolar amino acids

targeted to ClpXP, ClpAP, FtsH, and Lon. LpxC is a substrate of FtsH. LpxC possesses an unstructured C-terminal tail whose sequence resembles the ssrA tag (-LAFKAPSAVLA).⁸⁸ Best characterized markers for Lon protease include a cluster of hydrophobic residues with at least two aromatic residues, preferentially phenylalanine and tryptophan (e.g., -WEFAWFP).⁸⁹

1.6. FtsH: Universally conserved and only membrane-bound protease in *E.coli*.

Among AAA+ proteases, FtsH is only membrane-bound. Also, FtsH is the only growth-essential AAA+ protease in *E. coli* cells. FtsH-family proteases are widely conserved in the inner membranes of eubacteria and mitochondria and the thylakoid membranes of chloroplasts. FtsH malfunction causes severe phenotypes such as cell division defects, growth arrest, and envelope stresses in bacteria. In chloroplasts, FtsH is responsible for the turnover of photodamaged D1 core protein in the photosystem II reaction center (PSII). In humans, mutations on the mitochondrial homolog paraplegin are known to cause an autosomal recessive form of hereditary spastic paraplegia. FtsH is crucial in the quality control of many soluble and membrane proteins. Soluble substrates include the heat shock transcription factor σ^{32} , LpxC, an enzyme involved in the lipid metabolism, SsrA-tagged proteins, and the activator λ -CII. Known membrane protein substrates of FtsH include the uncomplexed SecY subunit part of the translocon, the F_0 component of ATP synthase, and the seven-membrane spanning protein YccA with unknown function and virulence protein MgtC. Recognition of membrane proteins requires exposed N or C terminal tails with 10–20 residues. Other than the length requirement, no specific sequence motif that preferentially binds to FtsH has been identified.

1.6.1. Domain arrangement of FtsH.

The primary sequence of *E. coli* FtsH comprises ~650 amino acid residues with a molecular weight of 70.7 kDa. FtsH contains four domains: the N-terminal TM domain with an intervening periplasmic globular domain and the C-terminal ATPase and protease domains. A 15–20 residue-long glycine-rich linker connects the membrane-anchored part to the AAA domain. The ~ 250-residue ATPase domain comprises an N-terminal alpha-beta-alpha fold and a small C-terminal helical bundle commonly referred to as large and small subdomains. The AAA+ domain contains the walker A and walker B motifs and the second region of homology (SRH) motif characteristic of the AAA+ family. Walker A and B motifs are involved in ATP binding and hydrolysis, respectively. The aromatic and nonpolar FVG motif from each AAA+ subunit is aligned to form an entry pore in the hexameric AAA+ ring, which is known to recognize and bind the degradation marker on a substrate.

The C-terminal part of the polypeptide chain bears the ‘zincin’ HExxH motif (His-Glu-xx-His, where x denotes any residue) characteristic of Zn^{2+} -dependent metalloproteases. The two histidine residues in the HExxH motif are coordinated to the Zn^{2+} ion, and the glutamate residue serves as a catalytic base.⁹⁰

1.6.2. Three-dimensional structure of FtsH.

The first structure of full-length FtsH was a yeast homolog Yta12 (or m-AAA in which m denotes the AAA+ domain located in the matrix of a mitochondrion) in detergents and solved at 12 Å resolution using cryo-electron microscopy (EM).⁹¹ Cryo-EM analysis shows the intact m-AAA+ protease in a hexameric assembly with a height of 137 Å and diameter of 130 Å. The structure is divided into a funnel-shaped upper-density corresponding to the N-terminal TM and

intermembrane-space domains and a more rigid lower body consisting of the ATPase and protease domains. Moreover, six 25 Å lateral openings in the protease domain may function as exit gates, allowing the cleaved peptides to leave the protease. The 13 Å gap between the ATPase domain and the membrane is considered a limiting factor for substrate entry. The narrow gap is large enough to accommodate only unfolded but not folded substrates. However, a recent structural and mutagenesis study suggests that the gap can be enlarged via tilting the linker region. Therefore, the 13 Å gap may not be the critical limiting factor that determines the size of substrate proteins.⁹²

The cryo-EM structure for the water-soluble catalytic domains of yeast homolog YME1 provides the first atomic-level picture (at 3.4 Å resolution) with the bound ATP and substrate.⁹³ These FtsH structure indicates that the ATP-binding induces a “spiral staircase” mechanism in substrate translocation.^{93, 84} The structure shows three distinct nucleotide-bound states within a hexamer: four ATP-bound, one ADP-bound, and one empty site. These states are correlated to the three modes of interaction with the substrate involving the highly conserved tyrosine residues in the pore loop. However, these structures do not explain the transition between an ADP-bound state in resting conditions and an ATP-bound state during the substrate loading and translocation.

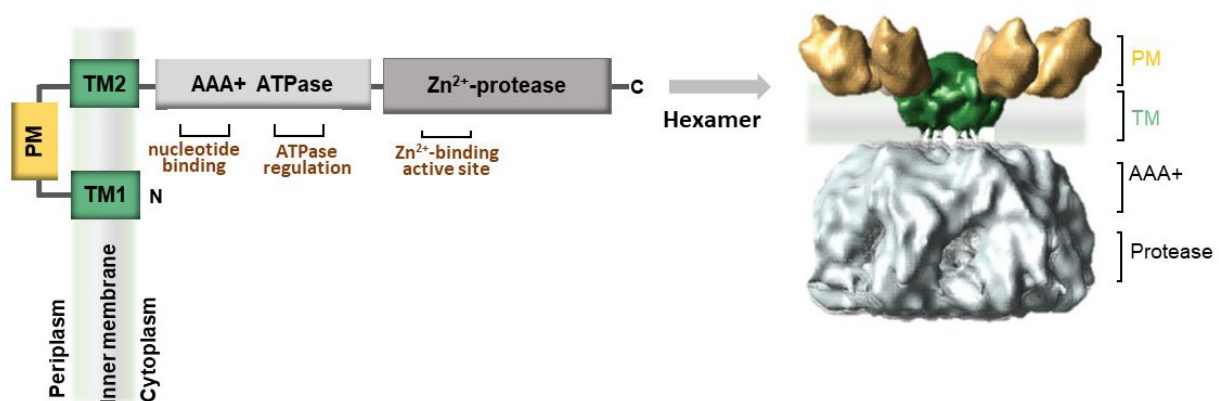


Figure 1.4. Domain arrangement and structure of FtsH.^{83,87}

1.6.3. Previous studies on FtsH-mediated protein degradation.

Studies of FtsH-mediated protein degradation have been mainly carried out for water-soluble substrates. It has been reported that FtsH can degrade the heat-shock transcription factor sigma-32 (σ^{32}) with a half-life of 18 min at 42 °C, which is 2-fold faster than at 37 °C in NP-40 detergents.⁹⁴ Among various tested nucleotides as energy sources, only cytidine triphosphate (CTP) can partially substitute for ATP. In contrast, other nucleotides such as guanosine triphosphate (GTP), uridine triphosphate (UTP), and non-hydrolyzable ATP analogs such as adenosine 5'-(α,β -methylene) triphosphate and adenosine 5'-(β,γ -methylene) triphosphate cannot.⁹⁴ Another study shows that FtsH-dependent degradation of both membrane-bound (YccA substrate: $t_{1/2} \sim 10$ -13 min; F_{oa} substrate: $t_{1/2} \sim 2.5$ min) and soluble substrates are retarded when the cells are treated with carbonyl cyanide-3-chlorophenylhydrazone or 2,4-dinitrophenol couplers, which dissipates the proton motive force.⁹⁵

ClpXP and ClpAP degrade ssrA-tagged substrates regardless of their thermodynamic stabilities. In contrast, FtsH only degrades unstable or metastable substrates. For example, when a ssrA-tag is placed at the C-terminus of the green-fluorescent protein (GFP-ssrA, $t_{1/2}$ of spontaneous unfolding = ~ 20 years) with very high kinetic stability, the resulting protein is degraded efficiently by ClpXP and ClpAP in vitro ($t_{1/2} = 1.55$ min),⁹⁶ but FtsH cannot degrade it. Similarly, two other substrates, barnase and dihydrofolate reductase (DHFR), have been tested for the role of protein stability in degradation. FtsH cannot degrade DHFR-ssrA but can degrade DHFR-CP Asp87, a circularly permuted mutant of DHFR with a faster unfolding rate than wild type. Also, barnase-ssrA is resistant to degradation by FtsH, while the faster unfolding mutant barnase-(I25A)-ssrA is degraded by FtsH. Moreover, faster unfolding Arc variants are degraded according to their thermodynamic stability. These studies suggest that FtsH lacks a robust unfoldase activity, but this

weak unfoldase activity confers FtsH the substrate selectivity preferentially targeting unstable or faster unfolding proteins.⁹⁷ However, the Hong lab has recently demonstrated that FtsH can degrade stable membrane protein GlpG while overcoming the dual energetic burden of substrate unfolding and dislocation with the ATP cost (1.7-2.5 ATP/residue) comparable to other robust AAA+ proteases (0.2-6.6 ATP/residue).⁹⁸

1.7. Project description

In general, protein quality control involves the regulation of functional protein concentrations at an optimal level in the cells. To achieve this, the synthesis, folding, trafficking, disaggregation, and degradation of proteins need to be coordinated. Since proteins are dynamic molecules, constant surveillance by chaperons and clearance by proteases operate to balance the fluxes towards folding and degradation. Most studies on protein quality control have targeted water-soluble proteins. However, it is not well understood how the cellular level of membrane proteins is regulated. Hence, quantitative studies on membrane protein folding and degradation can provide an insight into the critical molecular determinants of the balance for membrane protein homeostasis. However, this remains challenging since the tractable model systems to study folding and degradation in the native bilayer environment have been lacking.

On the “folding” side, the molecular forces that drive the association of TM helices to form the native structure in the bilayer are not fully understood yet. Membrane protein folding and stability studies have been conducted in membrane mimetic environments, primarily in detergent micelles. However, the lipid bilayer is the native environment for many membrane proteins. Therefore, it is essential to establish tools to study the unique role of the lipid bilayer as a solvent mediating protein folding and stability.

I employed GlpG from *E.coli* as a model membrane protein in my research. The folding properties of GlpG have been studied using various methods in different hydrophobic environments, including DDM micelles, bicelles, and lipid vesicles.^{59,63,80,99} The reported ΔG°_{N-D} values widely vary from 3.8-8.2 kcal/mol. The discrepancies are expected to stem from the distinct properties of mimetic membrane environments (SDS vs. bilayer or DDM vs. bilayer), the validity of the model-dependent extrapolation, and the conformation of the denatured state. In Chapter 2, I applied steric trapping to quantitatively study the thermodynamics of membrane protein folding without perturbants in the bilayer environment to resolve the discrepancies. Moreover, it has not been investigated how the bilayer environment would impact the residue interaction network for mediating the stability and cooperativity of membrane proteins. This is important to understand the functions of many types of membrane proteins such as receptors, channels, and transporters where the signal from one region needs to be propagated across the bilayer.

On the “degradation” side, the widely accepted model for FtsH-mediated degradation is that ATP-hydrolysis on the AAA+ domain cannot drive substrate unfolding but is only used to translocate the substrate to the protease domain for degradation. Therefore, the key determinant to inducing the degradation is the intrinsic kinetic stability of the substrate. However, this model is derived from the studies with the water-soluble substrate. Since membrane proteins and water-soluble proteins have distinct folding principles, a new conceptual framework is needed to understand detailed molecular mechanisms of membrane protein degradation. Thus, in Chapter 3, I dissect the membrane degradation into two thermodynamically unfavorable stages: 1) Substrate denaturation involving the disruption of tertiary interactions within the membrane; 2) Substrate translocation involving the dislocation of hydrophobic TM helices from the membrane to the aqueous active site in the protease domain. I hypothesized that the molecular forces shifting the

equilibrium in each step would impact the degradation rate. To test this hypothesis, it is essential to choose a membrane protein substrate with controllable folding properties in the native bilayer environment. Therefore, I employ GlpG, whose folding has been intensively studied by many research groups, including the Hong lab. Various intrinsic folding properties of GlpG (e.g., thermodynamic stability, denaturation rate, and hydrophobicity) have been modified by mutation and characterized. Next, the resulting variants with the known folding properties are challenged to FtsH, and their susceptibility to degradation is determined by measuring the degradation rates. The folding property that most impacts the degradation rate will be identified. Finally, based on the information, the rate-determining step of degradation will be defined. This lay an essential step toward understanding the molecular mechanism of FtsH-mediated degradation of membrane proteins.

REFERENCES

REFERENCES

1. Anfinsen, C. B. (1973) Principles that govern the folding of protein chains. *Science*. 181, 223–230
2. Pace, C. N. (2009) Energetics of protein hydrogen bonds. *Nat. Struct. Mol. Biol.* 16, 681–682.
3. Dill, K. A. (1990) Dominant Forces in Protein Folding. *Biochemistry*. 29, 7133–7155.
4. Brown, G. C. (1991) Total cell protein concentration as an evolutionary constraint on the metabolic control distribution in cells. *J. Theor. Biol.* 153, 195–203.
5. Ellis, R. J., and Minton, A. P. (2006) Protein aggregation in crowded environments. *Biol. Chem.* 387, 485–495.
6. Douglas, P. M., and Dillin, A. (2010) Protein homeostasis and aging in neurodegeneration. *J. Cell Biol.* 190, 719–729.
7. Klaips, C. L., Jayaraj, G. G., and Hartl, F. U. (2018) Pathways of cellular proteostasis in aging and disease. *J. Cell Biol.* 217, 51–63.
8. Hartl, F. U., Bracher, A., and Hayer-Hartl, M. (2011) Molecular chaperones in protein folding and proteostasis. *Nature* 475, 324–332.
9. Rüdiger, S., Buchberger, A., and Bukau, B. (1997) Interaction of Hsp70 chaperones with substrates. *Nat. Struct. Biol.* 4, 342–349.
10. Kim, Y. E., Hipp, M. S., Bracher, A., Hayer-Hartl, M., and Ulrich Hartl, F. (2013) Molecular chaperone functions in protein folding and proteostasis. *Annu. Rev. Biochem.* 82, 323–355.
11. Santra, M., Farrell, D. W., and Dill, K. A. (2017) Bacterial proteostasis balances energy and chaperone utilization efficiently. *Proc. Natl. Acad. Sci. U. S. A.* 114, E2654–E2661.
12. Haslbeck, M., Weinkauff, S., and Buchner, J. (2019) Small heat shock proteins: Simplicity meets complexity. *J. Biol. Chem.* 294, 2121–2132.
13. Balchin, D., Hayer-Hartl, M., and Hartl, F. U. (2016) In vivo aspects of protein folding and quality control. *Science*. 353, aac4354.1–aac4354.12
14. Mogk, A., Huber, D., and Bukau, B. (2011) Integrating protein homeostasis strategies in prokaryotes. *Cold Spring Harb. Perspect. Biol.* 3, 1–19.

15. Almén, M. S., Nordström, K. J. V., Fredriksson, R., and Schiöth, H. B. (2009) Mapping the human membrane proteome: A majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biol.* 7, 50.
16. Koga, H., Kaushik, S., and Cuervo, A. M. (2011) Protein homeostasis and aging: The importance of exquisite quality control. *Ageing Res. Rev.* 10, 205–215.
17. Arora, K., Moon, C., Zhang, W., Yarlagadda, S., Penmatsa, H., Ren, A., Sinha, C., and Naren, A. P. (2014) Stabilizing rescued surface-localized Δ f508 CFTR by potentiation of its interaction with Na⁺/H⁺ exchanger regulatory factor 1. *Biochemistry* 53, 4169–4179.
18. Schleich, J. P., Narayan, M., Alford, C., Mittendorf, K. F., Carter, B. D., Li, J., and Sanders, C. R. (2015) Conformational Stability and Pathogenic Misfolding of the Integral Membrane Protein PMP22. *J. Am. Chem. Soc.* 137, 8758–8768.
19. Ainslie, A., Huiting, W., Barazzuol, L., and Bergink, S. (2021) Genome instability and loss of protein homeostasis: Converging paths to neurodegeneration? *Open Biol.* 11. 200296 1–15
20. Brown, D. A., and Rose, J. K. (1992) Sorting of GPI-anchored proteins to glycolipid-enriched membrane subdomains during transport to the apical cell surface. *Cell* 68, 533–544.
21. Kučerka, N., Liu, Y., Chu, N., Petrache, H. I., Tristram-Nagle, S., and Nagle, J. F. (2005) Structure of fully hydrated fluid phase DMPC and DLPC lipid bilayers using x-ray scattering from oriented multilamellar arrays and from unilamellar vesicles. *Biophys. J.* 88, 2626–2637.
22. Gullingsrud, J., and Schulten, K. (2004) Lipid Bilayer pressure profiles and mechanosensitive channel gating. *Biophys. J.* 86, 3496–3509.
23. Corin, K., and Bowie, J. U. (2020) How bilayer properties influence membrane protein folding. *Protein Sci.* 29, 2348–2362.
24. Lindén, M., Sens, P., and Phillips, R. (2012) Entropic Tension in Crowded Membranes. *PLoS Comput Biol* 8, 1002431.
25. White, S. H., and Wimley, W. C. (1999) Membrane protein folding and stability: Physical principles. *Annu. Rev. Biophys. Biomol. Struct.* 28, 319–365.
26. Rees, D. C., DeAntonio, L., and Eisenberg, D. (1989) Hydrophobic organization of membrane proteins. *Science.* 245, 510–513.
27. Elazar, A., Weinstein, J. J., Prilusky, J., and Fleishman, S. J. (2016) Interplay between hydrophobicity and the positive-inside rule in determining membrane-protein topology. *Proc. Natl. Acad. Sci. U. S. A.* 113, 10340–10345.

28. Blobel, G., and Dobberstein, B. (1975) Transfer of proteins across membranes: I. Presence of proteolytically processed and unprocessed nascent immunoglobulin light chains on membrane-bound ribosomes of murine myeloma. *J. Cell Biol.* 67, 835–851.
29. Shao, S., and Hegde, R. S. (2011) Membrane protein insertion at the endoplasmic reticulum. *Annu. Rev. Cell Dev. Biol.* 27, 25–56.
30. Whitley, P., and Mingarro, I. (2014) Stitching proteins into membranes, not sew simple. *Biol. Chem.* 395, 1417–1424.
31. Folding, M. P., and Model, T. (1990) in *Biochemistry Membrane Protein Folding*. *Biochemistry* 29, 4031–4037.
32. Matthews, E. E., Lee, A. S., Curran, A. R., Senes, A., Dixon, A. M., Engelman, D. M., Reshetnyak, Y. K., Dupuy, A. D., Chin, C.-N., Popot, J.-L., Chen, Y., and Lehnert, U. (2003) Membrane protein folding: beyond the two stage model. *FEBS Lett.* 555, 122–125.
33. Wimley, W. C., Creamer, T. P., and White, S. H. (1996) Solvation energies of amino acid side chains and backbone in a family of host - Guest pentapeptides. *Biochemistry* 35, 5109–5124.
34. Wimley, W. C., and White, S. H. (1996) At Membrane Interfaces. *Nat. Structural Biol.* 3, 842–848.
35. Hessa, T., Kim, H., Bihlmaier, K., Lundin, C., Boekel, J., Andersson, H., Nilsson, I. M., White, S. H., and Von Heijne, G. (2005) Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* 433, 377–381.
36. Moon, C. P., and Fleming, K. G. (2011) Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. *Proc. Natl. Acad. Sci. U. S. A.* 108, 10174–10177.
37. Lin, M., Gessmann, D., Naveed, H., and Liang, J. (2016) Outer Membrane Protein Folding and Topology from a Computational Transfer Free Energy Scale. *J. Am. Chem. Soc.* 138, 2592–2601.
38. Yang, Y., Lee, M., and Fairn, G. D. (2018) Phospholipid subcellular localization and dynamics. *J. Biol. Chem.* 293, 6230–6240.
39. Ilgü, H., Jeckelmann, J. M., Gachet, M. S., Boggavarapu, R., Ucurum, Z., Gertsch, J., and Fotiadis, D. (2014) Variation of the detergent-binding capacity and phospholipid content of membrane proteins when purified in different detergents. *Biophys. J.* 106, 1660–1670.
40. Sanders, M. R., Findlay, H. E., and Booth, P. J. (2018) Lipid bilayer composition modulates the unfolding free energy of a knotted α -helical membrane protein. *Proc. Natl. Acad. Sci. U. S. A.* 115, E1709–E1808.

41. Gupta, K., Donlan, J. A. C., Hopper, J. T. S., Uzdaviny, P., Landreh, M., Struwe, W. B., Drew, D., Baldwin, A. J., Stansfeld, P. J., and Robinson, C. V. (2017) The role of interfacial lipids in stabilizing membrane protein oligomers. *Nature* 541, 421–424.
42. McDowell, M. A., Heimes, M., Fiorentino, F., Mehmood, S., Farkas, Á., Coy-Vergara, J., Wu, D., Bolla, J. R., Schmid, V., Heinze, R., Wild, K., Flemming, D., Pfeffer, S., Schwappach, B., Robinson, C. V., and Sinning, I. (2020) Structural Basis of Tail-Anchored Membrane Protein Biogenesis by the GET Insertase Complex. *Mol. Cell* 80, 72–86.
43. Jormakka, M., Törnroth, S., Byrne, B., and Iwata, S. (2002) Molecular basis of proton motive force generation: Structure of formate dehydrogenase-N. *Science* (80-.). 295, 1863–1868.
44. Sun, C., Benlekbir, S., Venkatakrishnan, P., Wang, Y., Hong, S., Hosler, J., Tajkhorshid, E., Rubinstein, J. L., and Gennis, R. B. (2018) Structure of the alternative complex III in a supercomplex with cytochrome oxidase. *Nature* 557, 123–126.
45. Allen, S. J., Curran, A. R., Templer, R. H., Meijberg, W., and Booth, P. J. (2004) Controlling the folding efficiency of an integral membrane protein. *J. Mol. Biol.* 342, 1293–1304.
46. Hong, H., and Bowie, J. U. (2011) Dramatic destabilization of transmembrane helix interactions by features of natural membrane environments. *J. Am. Chem. Soc.* 133, 11389–11398.
47. Miller, D., Charalambous, K., Rotem, D., Schuldiner, S., Curnow, P., and Booth, P. J. (2009) In vitro Unfolding and Refolding of the Small Multidrug Transporter EmrE. *J. Mol. Biol.* 393, 815–832.
48. Allen, S. J., Curran, A. R., Templer, R. H., Meijberg, W., and Booth, P. J. (2004) Folding kinetics of an α helical membrane protein in phospholipid bilayer vesicles. *J. Mol. Biol.* 342, 1279–1291.
49. Duneau, J. P., and Sturgis, J. N. (2013) Lateral organization of biological membranes role of long-range interactions. *Eur. Biophys. J.* 42, 843–850.
50. Isenbarger, T. A., and Krebs, M. P. (2001) Thermodynamic stability of the bacteriorhodopsin lattice as measured by lipid dilution. *Biochemistry* 40, 11923–11931.
51. Fisher, L. E., Engelman, D. M., and Sturgis, J. N. (1999) Detergents modulate dimerization, but not helicity, of the glycophorin A transmembrane domain. *J. Mol. Biol.* 293, 639–651.
52. Lau, F. W., and Bowie, J. U. (1997) A method for assessing the stability of a membrane protein. *Biochemistry* 36, 5884–5892.
53. Booth, P. J., Flitsch, S. L., Stern, L. J., Greenhalgh, D. A., Kim, P., and Khorana, H. G. (1995) Intermediates in the folding of the membrane protein bacteriorhodopsin. *Struc.biol.* 2, 139–143.

54. Baker, R. P., and Urban, S. (2012) Architectural and thermodynamic principles underlying intramembrane protease function. *Nat. Chem. Biol.* 8, 759–768.
55. Joh, N. H. J., Min, A., Faham, S., Whitelegge, J. P., Yang, D., Woods, V. L., and Bowie, J. U. (2008) Modest stabilization by most hydrogen-bonded side-chain interactions in membrane proteins. *Nature* 453, 1266–1270.
56. Tulumello, D. V., and Deber, C. M. (2009) SDS micelles as a membrane-mimetic environment for transmembrane segments. *Biochemistry* 48, 12096–12103.
57. Guo, R., Gaffney, K., Yang, Z., Kim, M., Sungsuwan, S., Huang, X., Hubbell, W. L., and Hong, H. (2016) Steric trapping reveals a cooperativity network in the intramembrane protease GlpG. *Nat. Chem. Biol.* 12, 353–60.
58. Otzen, D. E., Pedersen, J. N., Somavarapu, A. K., Clement, A., Ji, M., Petersen, E. H., Pedersen, J. S., Urban, S., and Schafer, N. P. (2021) Cys-labeling kinetics of membrane protein GlpG: a role for specific SDS binding and micelle changes? *Biophys. J.* 120, 4115–4128.
59. Min, D., Jefferson, R. E., Bowie, J. U., and Yoon, T. Y. (2015) Mapping the energy landscape for second-stage folding of a single membrane protein. *Nat. Chem. Biol.* 11, 981–987.
60. Hong, H., Blois, T. M., Cao, Z., and Bowie, J. U. (2010) Method to measure strong protein-protein interactions in lipid bilayers using a steric trap. *Proc. Natl. Acad. Sci. U. S. A.*
61. Jefferson, R. E., Blois, T. M., and Bowie, J. U. (2013) Membrane proteins can have high kinetic stability. *J. Am. Chem. Soc.* 135, 15183–15190.
62. Chang, Y. C., and Bowie, J. U. (2014) Measuring membrane protein stability under native conditions. *Proc. Natl. Acad. Sci. U. S. A.* 111, 219–224.
63. Guo, R., Gaffney, K., Yang, Z., Kim, M., Sungsuwan, S., Huang, X., Hubbell, W. L., and Hong, H. (2016) Steric trapping reveals a cooperativity network in the intramembrane protease GlpG. *Nat. Chem. Biol.* 12, 353–60.
64. Urban, S., Lee, J. R., and Freeman, M. (2001) Drosophila Rhomboid-1 Defines a Family of Putative Intra membrane Serine Proteases. *Cell*. 107, 1–10.
65. Srinivasan, P., Coppens, I., and Jacobs-Lorena, M. (2009) Distinct roles of Plasmodium Rhomboid 1 in parasite development and malaria pathogenesis. *PLoS Pathog.* 5.
66. Rather, P. N., Ding, X., Baca-DeLancey, R. R., and Siddiqui, S. (1999) Providencia stuartii genes activated by cell-to-cell signaling and identification of a gene required for production or activity of an extracellular factor. *J. Bacteriol.* 181, 7185–7191.
67. Urban, S., and Dickey, S. W. (2011) The rhomboid protease family: A decade of progress on function and mechanism. *Genome Biol.* 12. 1–10.

68. Lemieux, M. J., Fischer, S. J., Cherney, M. M., Bateman, K. S., and James, M. N. G. (2007) The crystal structure of the rhomboid peptidase from *Haemophilus influenzae* provides insight into intramembrane proteolysis. *Proc. Natl. Acad. Sci. U. S. A.* 104, 750–754.
69. Wang, Y., Zhang, Y., and Ha, Y. (2006) Crystal structure of rhomboid family intramembrane protease. *Nature*. 444, 179–183.
70. Koonin, E. V., Makarova, K. S., Rogozin, I. B., Davidovic, L., Letellier, M. C., and Pellegrini, L. (2003) The rhomboids: a nearly ubiquitous family of intramembrane serine proteases that probably evolved by multiple ancient horizontal gene transfers. *Genome Biol.* 4, 1–12.
71. Ha, Y., Akiyama, Y., and Xue, Y. (2013) Structure and mechanism of rhomboid protease. *J. Biol. Chem.*
72. Hubbard, S. J. (1998) The structural aspects of limited proteolysis of native proteins. *Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol.* 1382, 191–206.
73. Urban, S., and Freeman, M. (2003) Substrate specificity of Rhomboid intramembrane proteases is governed by helix-breaking residues in the substrate transmembrane domain. *Mol. Cell* 11, 1425–1434.
74. Urban, S., Schlieper, D., and Freeman, M. (2002) Conservation of intramembrane proteolytic activity and substrate specificity in prokaryotic and eukaryotic rhomboids. *Curr. Biol.* 12, 1507–1512.
75. Adrain, C., and Freeman, M. (2012) New lives for old: Evolution of pseudoenzyme function illustrated by iRhoms. *Nat. Rev. Mol. Cell Biol.* 13, 489–498.
76. Zettl, M., Adrain, C., Strisovsky, K., Lastun, V., and Freeman, M. (2011) Rhomboid family pseudoproteases use the ER quality control machinery to regulate intercellular signaling. *Cell* 145, 79–91.
77. Cao, S. S., and Kaufman, R. J. (2011) iRhoms: ERADicating the Messenger in Growth Control Signaling. *Dev. Cell* 20, 414–416.
78. Adrain, C., Zettl, M., Christova, Y., Taylor, N., and Freeman, M. (2012) Tumor necrosis factor signaling requires iRhom2 to promote trafficking and activation of TACE. *Science*. 335, 225–228.
79. Baker, R. P., and Urban, S. (2012) Architectural and thermodynamic principles underlying intramembrane protease function. *Nat. Chem. Biol.* 8(9), 759–768.
80. Paslawski, W., Lillelund, O. K., Kristensen, J. V., Schafer, N. P., Baker, R. P., Urban, S., and Otzen, D. E. (2015) Cooperative folding of a polytopic α -helical membrane protein involves a compact N-terminal nucleus and nonnative loops. *Proc. Natl. Acad. Sci. U. S. A.* 112, 7978–7983.

81. Min, D., Jefferson, R. E., Bowie, J. U., Yoon, T., Korea, S., Korea, S., and Angeles, L. (2016) Single Membrane Protein *11*, 981–987.
82. Guo, R., Gaffney, K., Yang, Z., Kim, M., Sungsuwan, S., Huang, X., Hubbell, W. L., and Hong, H. (2016) Steric trapping reveals a cooperativity network in the intramembrane protease {GlpG}. *Nat. Chem. Biol.* *12*, 353–360.
83. Gaffney, K., Guo, R., Bridges, M., MuhammedNazaar, S., Chen, D., Kim, M., Yang, Z., Schimiller, A., Faruak, A., Peng, X., Jones, D., Kim, K., Sun, L., Hubbel, W., Sosnick, T., Hong, H., (2021) Lipid bilayer induces contraction of the denatured state ensemble of a helical-bundle membrane protein. *PNAS.* *119*, 1–12.
84. Avci, D., and Lemberg, M. K. (2015) Clipping or Extracting: Two Ways to Membrane Protein Degradation. *Trends Cell Biol.* *25*, 611–622.
85. Foot, N., Henshall, T., and Kumar, S. (2017) Ubiquitination and the regulation of membrane proteins. *Physiol. Rev.* *97*, 253–281.
86. He, C., and Klionsky, D. J. (2009) Regulation mechanisms and signaling pathways of autophagy. *Annu. Rev. Genet.* *43*, 67–93.
87. Sauer, R. T., and Baker, T. A. (2011) AAA+ Proteases: ATP-fueled machines of protein destruction. *Annu. Rev. Biochem.* *80*, 587–612.
88. Narberhaus, F., Obrist, M., Führer, F., and Langklotz, S. (2009) Degradation of cytoplasmic substrates by FtsH, a membrane-anchored protease with many talents. *Res. Microbiol.* *160*, 652–659.
89. Gur, E., and Sauer, R. T. (2008) Recognition of misfolded proteins by Lon, a AAA+ protease. *Genes Dev.* *22*, 2267–2277.
90. Bieniossek, C., Schalch, T., Bumann, M., Meister, M., Meier, R., and Baumann, U. (2006) The molecular architecture of the metalloprotease FtsH. *Proc. Natl. Acad. Sci. U. S. A.* *103*, 3066–3071.
91. Lee, S., Augustin, S., Tatsuta, T., Gerdes, F., Langer, T., and Tsai, F. T. F. (2011) Electron cryomicroscopy structure of a membrane-anchored mitochondrial AAA protease. *J. Biol. Chem.* *286*, 4404–4411.
92. Carvalho, V., Prabudiansyah, I., Kovacik, L., Chami, M., Kieffer, R., Van Der Valk, R., De Lange, N., Engel, A., and Aubin-Tam, M. E. (2020) The cytoplasmic domain of the AAA+ protease FtsH is tilted with respect to the membrane to facilitate substrate entry. *J. Biol. Chem.* *296*, 100029.

93. Puchades, C., Rampello, A. J., Shin, M., Giuliano, C. J., Wiseman, R. L., Glynn, S. E., and Lander, G. C. (2017) Structure of the mitochondrial inner membrane AAA+ protease YME1 gives insight into substrate processing. *Science*. 358, 1-10.
94. Tomoyasu, T., Gamer, J., Bukau, B., Kanemori, M., Mori, H., Rutman, a J., Oppenheim, a B., Yura, T., Yamanaka, K., and Niki, H. (1995) Escherichia coli FtsH is a membrane-bound, ATP-dependent protease which degrades the heat-shock transcription factor sigma 32. *EMBO J.* 14, 2551–2560.
95. Akiyama, Y. (2002) Proton-motive force stimulates the proteolytic activity of FtsH, a membrane-bound ATP-dependent protease in Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* 99, 8066–8071.
96. Kim, Y. I., Burton, R. E., Burton, B. M., Sauer, R. T., and Baker, T. A. (2000) Dynamics of substrate denaturation and translocation by the ClpXP degradation machine. *Mol. Cell* 5, 639–648.
97. Herman, C., Prakash, S., Lu, C. Z., Matouschek, A., and Gross, C. A. (2003) Lack of a robust unfoldase activity confers a unique level of substrate specificity to the universal AAA protease FtsH. *Mol. Cell* 11, 659–669.
98. Yang, Y., Guo, R., Gaffney, K., Kim, M., Muhammednazaar, S., Tian, W., Wang, B., Liang, J., and Hong, H. (2018) Folding-Degradation Relationship of a Membrane Protein Mediated by the Universally Conserved ATP-Dependent Protease FtsH. *J. Am. Chem. Soc.* 140, 4656–4665.
99. Min, D., Jefferson, R. E., Bowie, J. U., and Yoon, T.-Y. (2015) Mapping the energy landscape for second-stage folding of a single membrane protein. *Nat. Chem. Biol.* 11, 981-987.

CHAPTER 2

Lipid Bilayer Strengthens the Cooperative Network of the rhomboid intramembrane protease GlpG

Shaima Muhammednazaar[†], Ruiqiong Guo^{†,‡}, Seung-Hyo Rhee^{†,§},

Seung-gu Kang^{*,‡} and Heedeok Hong^{*,†,‡}

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

[‡]Computational Biology Center, IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598, USA

2.1. Summary

The lipid bilayer provides a solvent-like environment for membrane proteins mediating their folding and function in cells. Here, we investigate how the native lipid environment stabilizes a helical membrane protein GlpG and engages the protein's residue interaction network compared to nonnative detergent micelles. We find that the bilayer stabilizes GlpG by facilitating the residue burial in the protein interior relative to micelles. Strikingly, while the cooperative residue interactions are clustered in multiple distinct regions in micelles, they span the entire packed regions of GlpG in the bilayer. Molecular dynamics simulation predicts that lipids bind weaker and exchange faster at the protein surface than detergents, suggesting that the lipid-induced stability and cooperativity enhancement may stem from the intraprotein interactions that outcompete the weak protein-lipid interactions. This result reveals a unique role of lipids in stabilizing membrane proteins and allowing efficient propagation of local perturbation within a protein, which may benefit function.

2.2. Introduction

The solvent environment plays a fundamental role in shaping the folding energy landscape and function of proteins.^{1,2} For example, conformation of water-soluble proteins is determined by the energetic balance between intraprotein, protein–water, and water–water interactions. In this balance, the hydrophobic effect, which refers to the unfavorable ordering of water molecules around nonpolar residues, provides a critical driving force for folding by inducing the collapse of nonpolar residues in the protein interior and the release of the solvating water into the bulk aqueous phase. Involving the collective formation and dismantling of the water hydrogen-bond (H-bond) network, the hydrophobic effect evokes a free energy barrier between the native and denatured states, inducing cooperativity in folding. Prevalent in protein folding and function, cooperativity links the behaviors of distant sites within a protein.³⁻¹¹

Unlike water-soluble proteins, membrane proteins fold and function in a lipid bilayer which provides a quasi-two-dimensional, hydrophobic solvent environment. The folding of helical membrane proteins, which are dominant in all types of the cell membranes except for the outer membranes of bacteria, mitochondria and chloroplasts, can be divided into two-stages:¹² In Stage I, nonpolar polypeptide segments insert into the bilayer to form transmembrane (TM) helices. In Stage II, TM helices associate into a compact native structure. Recent studies suggest that the folding is more complex than expected from the model. Prior to association into the native structure, individual TM helices can flip across the membrane, unfold at the membrane surface, or partially associate with one another.¹³⁻¹⁷ Nonetheless, the two-stage model provides a useful thermodynamic framework for dissecting driving forces in membrane protein folding.

The formation of TM helices (Stage I) is majorly driven by the hydrophobic effect that induces the burial of nonpolar side chains as well as the energetic penalty of unraveling backbone H-bonded partners in the nonpolar bilayer core.^{18,19} To drive Stage II, attractive interactions between TM helices should overcome likewise favorable interactions between individual helices and solvating lipids.¹² Thermodynamic stabilities of helical-bundle membrane proteins measured in various hydrophobic environments (*i.e.*, the free energy change in Stage II, $\Delta G^{\circ}_{\text{N-D}}$) fall into the typical stability range of globular proteins in water (−5 to −10 kcal/mol).^{11,20} Within the lipid bilayer lacking water, interhelical van der Waals (vdW) packing and polar interactions would be crucial for stabilizing membrane proteins.²¹⁻²⁸ However, the contributions of vdW packing and side-chain H-bonds to the stability are known to be comparable between membrane and globular proteins,^{21,26} which may not fully compensate the lack of the hydrophobic effect in the bilayer. In contrast, although not quantified yet, the confinement of the polypeptide chains in the quasi-2D bilayer and their ordering into helices are expected to reduce the backbone entropic cost in Stage

II.²⁹ Notably, a recent NMR study shows that membrane proteins have more dynamic side-chain motions in the protein interior than globular proteins, suggesting a relatively small side-chain entropic cost in the same stage.³⁰ Finally, the physical properties of the lipid bilayer such as (*e.g.*, the lateral lipid packing pressure in the bilayer and the lipid deformation energy induced by either the hydrophobic mismatch between the protein and the bilayer or the exposed polar residues at the lipid–protein interface) are known to impact TM helix–helix interactions.³¹⁻³⁴

Although these studies suggest that membrane proteins fold through a delicate balance between various types of molecular forces, the fundamental role of lipid solvation in shaping the folding energy landscape and cooperativity of membrane proteins remains elusive. Do the lipid solvation compete or facilitate the association of TM helices, or serve as an inert hydrophobic

medium? Do the lipids modulate the cooperativity of membrane proteins as water does for water-soluble proteins? Cooperativity is essential for function of various types of membrane proteins including receptors, transporters, and enzymes, allowing propagation of chemical or physical stimuli on one side of the protein across the bilayer (~ 50 Å thick) to induce functional changes on the other. Here, we address these questions using experiment and molecular dynamics (MD) simulation. We employed the helical-bundle membrane protein GlpG of *E. coli* as a model, a member of the near-universally conserved rhomboid intramembrane protease family.^{35,36} Rhomboids are serine proteases with a Ser/His catalytic dyad and plays a variety of regulatory roles in signaling, protein quality control, membrane remodeling, quorum sensing, and apoptosis *via* the cleavage of a peptide bond in a membrane-bound substrate near the membrane.^{35,37}

We hypothesize that, if the lipid bilayer is an inert solvent for membrane proteins, the information for stability and cooperativity is entirely encoded in the protein sequence and thus, the contribution of individual residue interactions and their interaction network to stability and cooperativity will not be affected by the chemical and physical nature of the hydrophobic medium. Accordingly, we compared the folding of GlpG in two distinct hydrophobic media that are widely used for membrane protein research: bicelles,³⁸ the disc-shaped lipid bilayer fragments edge-stabilized by detergents, and micelles, the globular aggregates of detergent molecules. We find that the lipid environment induces the favorable residue burial in the protein interior and tightening of the residue interaction network relative to micelles. Interestingly, MD simulation suggests that the lipids strengthen the intraprotein interaction by the weak solvation. This weak lipid solvation stems from the strong lateral lipid–lipid interactions that facilitate dissociation of the lipid molecules solvating the protein surface. These results suggest a novel principle of membrane protein stabilization and cooperativity mediated by the lipid bilayer in the cell membranes.

2.3. Materials and methods

2.3.1. Expression and purification of GlpG.

GlpG transmembrane domain possessing residues from 87 to 276 with an N terminal His tag was expressed in *E. coli* BL21(DE3)RP cells. Cells were harvested and resuspended in 30 mL of 50 mM Tris-HCl buffer (pH 8.0) containing 5 mM EDTA, 0.5 mM TCEP and 0.5 mM PMSF. Then, the resuspended cells were lysed 5 times using a pressure homogenizer (Avestin). The lysate was centrifuged at 5,000 rpm for 20 min in the FS-34 rotor using a Sorvall RC6+ centrifuge. After that, clear supernatant was centrifuged to obtain the total membrane fraction at 24,000 rpm for 2 h in the 45Ti rotor using ultracentrifuge (Beckman-Coulter). Membrane pellet was resuspended in 25 mL of 50 mM Tris-HCl buffer (pH 8.0) containing 200 mM NaCl, 1 mM TCEP and 0.25 mM PMSF using a tissue homogenizer (Fisher Scientific). The membrane resuspension was solubilized by the addition of 0.7% (w/v) DDM followed by ultracentrifugation at 18,000 rpm for 25 min. The clear supernatant was mixed with 3 mL of Ni-NTA resin (Qiagen: 50% w/v) and incubated at 4°C for 1 h. Eluted GlpG was desalted and concentrated using Amicon centrifugal filter unit (Millipore Sigma, 10 kDa MWCO) in 0.1% DDM, 50 mM Tris-HCl buffer (pH 8.0) and 200 mM NaCl.

2.3.2. Labeling of GlpG and determining labeling efficiency by gel-shift assay.

50 μ M of GlpG double cysteine variants purified in 0.1% DDM, 50 mM Tris-HCl and 200 mM NaCl (pH 8.0) was incubated with 10 molar excess of TCEP for 1 h at room temperature. 40 molar excess of BtnPyr-IA dissolved in DMSO (~10% v/v) was added dropwise to the mixture while vortexing. The labeling reaction proceeded overnight at room temperature. Excess free labels were removed by washing GlpG bound to Ni-NTA resin with 30 mM imidazole in 0.05% DDM, 50 mM Tris-HCl and 200 mM NaCl (pH 8.0) and eluted with 300 mM imidazole in 0.1%

DDM, 50 mM Tris-HCl and 200 mM NaCl (pH 8.0). Free labels and imidazole were further removed by dialysis against the buffer containing 0.01% DDM, 50 mM Tris-HCl, and 200 mM NaCl (pH 8.0).

Labeling efficiency of GlpG variants was determined by measuring pyrene fluorescence ($\epsilon_{346\text{nm}} = 42,000 \text{ M}^{-1}\text{cm}^{-1}$) and protein concentration by 660 nm assay (Bio-Rad). The pyrene-to-protein molar ratio ranged from 1.4 to 2.2. SDS-PAGE gel shift assay was carried out as follows: 10 μL of 5 μM GlpG was incubated with the SDS sample buffer for 30 min to denature GlpG. 10 mL of 25 μM WT-mSA was added and incubated another 30 min for binding to biotin-labeled GlpG. The SDS-PAGE was run at 100 V for 90 min on ice to prevent tetrameric mSA from dissociation. Labeling efficiency was calculated by measuring the band intensities of single-mSA bound GlpG and double-mSA bound GlpG and accounting for their molecular masses.

2.3.3. Expression, purification, refolding and labeling of mSA.

The procedures are described in the previous literature.³⁹ Streptavidin (active/inactive) was expressed in *E. coli* BL21(DE3)RP cells. Active subunits refer to wild type streptavidin or weaker affinity streptavidin variants (W79M, S45A, S27A, and E51S) with a C-terminal His₆ tag. Inactive streptavidin refers to the N23A/S27D/S45A triple mutant (inactive, $K_{d,\text{biotin}} = 1.2 \times 10^{-3} \text{ M}$) streptavidin without His₆ tag. 50 ml cultures grown overnight in the presence of 0.1 g/L ampicillin at 37°C were inoculated in 1 L terrific broth media with 0.1 g/L ampicillin at 37°C until OD_{600nm} reached 0.6. The final concentration of 0.5 mM IPTG was added for inducing protein expression and the culture was incubated overnight at 37°C. Harvested cells were resuspended in 40 mL of 50 mM Tris-HCl, 0.75 M sucrose, and 1 mg/ml hen egg lysozyme (pH 8.0). The resuspended cells were lysed 5 to 7 times using a pressure homogenizer (Avestin). Inclusion bodies were collected

as a pellet by centrifuging at 12,000 rpm for 15 min at 4 °C using a 45Ti rotor. The pellets were washed with 35 mL of 50 mM Tris-HCl (pH 8.0), 1.5 M NaCl, 0.5% TritonX-100 (Sigma) using a tissue homogenizer, and centrifuged at 12,000 rpm for 15 min at 4 °C. The detergent-washing procedure was repeated 3 to 4 times. The pellets were finally washed with 35 mL of 50 mM Tris-HCl (pH 8.0), 1.5 M NaCl without TritonX-100. The final pellet was solubilized in 8 mL (per L-culture) of 6 M guanidine hydrochloride (GdnHCl, pH 2.0). Aggregates were removed by centrifuging the sample at 24,000 rpm for 45 min at 4 °C using a 45Ti rotor. OD_{280nm} of the supernatant was measured using a nanodrop (Thermo Scientific).

For refolding of streptavidin, GdnHCl-solubilized active and inactive subunits of streptavidin were mixed at a molar ratio of 1:4. This mixture was added dropwise to the buffer containing 15% glycerol in 20 mM sodium phosphate, 200 mM NaCl, and 0.5 mM TCEP (pH 7.5) while vortexing vigorously on ice. Aggregates were removed by centrifuging at 6,000 rpm for 30 min at 4 °C in a Sorvell RC6+ high-speed centrifuge. The clear supernatant was mixed with Ni-NTA resin (Qiagen) and incubated at 4°C for 1 h. Collected Ni-NTA resin by centrifugation was washed with 10 mM imidazole, 0.5 mM TCEP, 20 mM sodium phosphate, 200 mM NaCl (pH 7.5). Monovalent Streptavidin (mSA) was eluted with 50 mM imidazole, 0.5 mM TCEP, 20 mM sodium phosphate, 200 mM NaCl (pH 7.5). To further purify mSA, the second affinity chromatography in Ni-NTA resin was applied. To remove imidazole, eluted mSA fraction was desalted and concentrated using Amicon centrifugal filter unit (Millipore Sigma, 10 kDa MWCO).

To label mSA with a quencher, the active subunit was mutated at Y83C near the biotin-binding pocket. mSA was incubated with 5 times molar excess TCEP for 1 h at room temperature. 15 times molar excess of dabcyI-maleimide (AnaSpec) solubilized in DMSO (10% v/v) was added dropwise while vortexing and incubated at 4°C overnight. Excess free labels were removed by

desalting in 20 mM HEPES (pH 7.5), 100 mM KCl and further removed by overnight dialysis against 20 mM HEPES (pH 7.5).

2.3.4. Expression and purification of GlpG substrate SN-LacYTM2.

To measure the activity of GlpG as a folding indicator, we used the specific cleavage of the transmembrane substrate, the second TM domain of the lactose permease of *E. coli* fused to staphylococcal nuclease (SN-LacYTM2). The DNA construct of this fused substrate ligated to the pET30a vector, and the construct design was as follows: SN domain is linked to the TM domain of lactose permease with a C-terminal His₆ tag, in which a TEV cleavage site (TEV_{cut}) is engineered in the middle of the linker (SN-TEV_{cut}-LacYTM2-His₆). In LacYTM2, the residue five residues upstream of the scissile bond (P5 position) was mutated to cysteine for labeling with thiol-reactive, environment-sensitive fluorophore iodoacetyl-7-nitrobenz-2-oxa-1,3-diazol (IA-NBD amide, Setareh Biotech). The construct was expressed in the BL21(DE3)RP *E. coli* strain. The protein was expressed, purified, and labeled using the protocol described previously.¹

2.3.5. Preparation of native and sterically denatured GlpG in micelles.

20 μ M of GlpG doubly labeled with BtnPyr was incubated with 2.4 molar excess of mSA_{Dab}-E51S at room temperature until the maximum denaturation was reached. The extend of denaturation was monitored every 24 h using GlpG activity as a folding indicator. For the C-terminal biotin pair (172_M267_C-BtnPyr₂), maximum denaturation was reached within 24 h. For the N-terminal biotin pair (95_N172_M-Btn-Pyr₂), 8 mM SDS was added in the beginning to facilitate the denaturation and incubated for 5 h with 2.4 times molar excess of mSA_{Dab}-E51S at room temperature. Native and denatured GlpG was directly injected into bicelles to initiate the folding and denaturation reactions where the final concentrations of GlpG (0.5 μ M) and mSA_{Dab}-E51S

(1.2 μ M) and DMPC/CHAPS bicelles (3%, w/v) in 20 mM HEPES (pH 7.5), 40 mM KCl and 0.5 mM DTT.

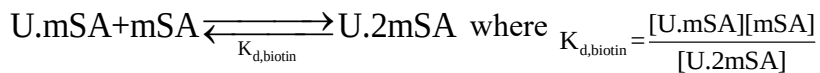
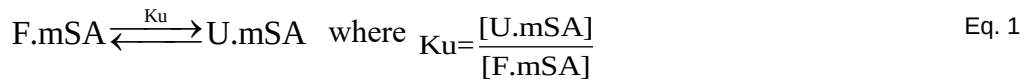
2.3.6. Activity-binding equilibrium to obtain reversibility.

Native and sterically denatured GlpG in DDM micelles for C-terminal and N-terminal biotin pair were titrated against mSA_{Dab}-E51S at varying concentrations at room temperature. Every 24 h, quenching of pyrene fluorescence was monitored for denaturation and refolding reactions at the excitation and emission wavelengths of 345 nm and 390 nm, respectively, using SpectroMax M5e until an equilibrium was reached. Data were averaged from three fluorescence readings.

Similarly, to test the folding equilibrium, GlpG activity was monitored using 20 times molar excess of NBD-labeled SN-LacYTM2 as a substrate. Fluorescence change with time indicates the cleavage of the substrate by GlpG. The activity was represented by the initial slope of fluorescence change versus time. Time-dependent changes of NBD fluorescence were monitored in 96-well plate using SpectraMax M5e plate reader with the excitation and emission wavelengths of 485 nm and 535 nm, respectively.

2.3.7. Fitting of the second binding phase to obtain the thermodynamic stability of GlpG.

To obtain the thermodynamic stability of GlpG, the attenuated second binding of mSA was fitted to the equation which was obtained by the following scheme:³⁹



The fitting equation for the second binding phase was:

$$F = \frac{1}{\left[1 + \left(K_{d,biotin} + \frac{K_{d,biotin}}{K_u} \right) \frac{1}{[mSA]} \right]} (F_{\infty} - F_0) + F_0 \quad \text{Eq. 2}$$

, 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 unhindered biotin affinity of mSA, and K_U is the equilibrium constant for the denaturation of GlpG. After obtaining the fitted K_U , the thermodynamic stability was calculated using the equation $\Delta G_{N-D}^0 = -RT \ln K_U$

2.3.8. Construction of binding isotherms to determine the thermodynamic stability of GlpG.

1 μ M of GlpG mutants labeled with BtnPyr was titrated against mSA labeled with thiol-reactive dabcyI (AnaSpec) at Y83C position of the active subunit at varying concentrations in 20 mM HEPES (pH 7.5), 40 mM KCl, 1 mM DTT. Depending on the stability of GlpG mutant, a series of mSA variants with a weaker biotin affinity (mSA_{DAB}-W79M, S45A, S27A, and E51S) were screened until an optimal second binding phase was obtained. The titrated samples were transferred to a 96 well plate (VWR), sealed with a polyolefin tape, and incubated at room temperature until the equilibrium was reached. The binding was monitored by quenching of GlpG conjugated pyrene fluorescence at 390 nm with an excitation wavelength of 345 nm using a SpectroMax M5e plate reader. Data were averaged from three fluorescence readings. The second attenuated binding phase was fitted to Eq.2.

2.3.9. Mapping the cooperative interactions in bilayers and micelles.

This method is developed to identify and characterize the side-chain interactions in DDM micelles.³⁹ First, specific side-chain interaction was perturbed by a single point mutation in the background of the N-terminal (95_N172_M -BtnPyr₂) or C-terminal biotin (172_M267_C-BtnPyr₂) pair set as ‘WT’. Then, the stability changes induced by the same point mutation was measured by steric trapping in both biotin-pair backgrounds. The change in the stability relative to WT is quantified as follows:

$$\Delta\Delta\Delta G_U^o = \left[\Delta G_{U,95/172_N-BtnPyr_2}^o(WT) - \Delta G_{U,95/172_N-BtnPyr_2}^o(mut) \right] - \left[\Delta G_{U,172/267_C-BtnPyr_2}^o(WT) - \Delta G_{U,172/267_C-BtnPyr_2}^o(mut) \right]$$

$$\Delta\Delta\Delta G_U^o = \left[\Delta\Delta G_{U,95/172_N-BtnPyr_2}^o(WT - mut) - \Delta\Delta G_{U,172/267_C-BtnPyr_2}^o(WT - mut) \right] \quad \text{Eq. 3}$$

If the mutation induces similar stability changes for both subdomains within the limit of thermal fluctuation energy, it indicates that the side-chain perturbation in one region is effectively propagated to the other, and we classify the mutated side chains are engaged cooperatively with the environment ($\Delta\Delta\Delta G_u^o \leq RT = 0.6$ kcal/mol where R : Gas constant and $T = 298$ K). If a mutation induces a larger stability change in the region containing the mutation, it indicates that the mutational perturbation is not effectively propagated to other regions, and we classify the perturbed side chains to be locally engaged with the environment ($\Delta\Delta\Delta G_u^o > RT$). If the mutation in one subdomain destabilizes the other subdomain not containing the mutation, we call that the perturbed interactions are over-propagated.

2.3.10. Measuring the incorporation of native and denatured GlpG into the bilayer phase of bicelles by fluorescence quenching assay.

To measure the incorporation of pyrene labeled GlpG into the bilayer phase FRET-based quenching assay was employed. For this, bicelles were made using dabcyl-labeled lipids and the degree of quenching of pyrene fluorescence by dabcyl was monitored. Native GlpG in dabcyl-labeled proteoliposomes were made as a positive control, which gives maximum quenching during incorporation. The procedure as follows: 7.5% (w/v) of dried DMPC lipids were mixed with dabcyl-POPE (Avanti polar lipids) at the molar ratio of 0.5:99.5 in chloroform in a glass tube and dried under stream of nitrogen. They were further dried under for 4 h. Dried lipids were resuspended in 500 mL of 20 mM HEPES (pH 7.5), 5% (w/v) β -octyl glucoside (Anatrace). GlpG in DDM was added to the resuspension and incubated in ice for 30 min. Bio beads (Bio-Rad) were added to remove the detergents in three steps. Each step 0.2 g/mL of bio bead was added and gently stirred for 6 to 12 h at room temperature. Resulting proteoliposomes were extruded using a 0.2 mm pore size membrane. The total phospholipid concentration was measured using an organic phosphate assay. Based on the lipid concentration, desired CHAPS was added to form bicelles of $q = 1.5$ (lipid-to-detergent molar ratio). Total protein concentration was measured using a 660 nm assay (Bio-Rad).

As a negative control (no quenching), water-soluble E51S-mSA labeled with pyrene was used. mSA labeling procedure as follows: E51S-mSA at Y83C position was used for site-specific cysteine labeling. 30 mM E51S-mSA in 20 mM sodium phosphate (pH 8.0), 200 mM NaCl was incubated with 5 molar excess of TCEP at room temperature for 1 h. 10 molar excess of NHS-Pyrene (Pierce) dissolved in 10% (v/v) DMSO was added to the E51S-mSA dropwise while vortexing. The mixture covered with aluminum foil was incubated at room temperature for 8 h.

Excess free labels were removed by binding the protein to Ni-NTA affinity column and washing with 10 mM imidazole, 20 mM sodium phosphate (pH 7.5), 200 mM NaCl and eluted with 50 mM imidazole, 20 mM sodium phosphate (pH 7.5), 200 mM NaCl. The eluted fraction was further dialyzed against 20 mM HEPES (pH=7.5) buffer at 4°C with two buffer exchanges. Pyrene labeled E51S-mSA was characterized by measuring pyrene fluorescence at 346 nm ($\epsilon_{346\text{nm}}$ 42,000 $\text{M}^{-1}\text{cm}^{-1}$). Protein concentration was determined by a 660 nm assay (Bio-Rad).

As an experimental sample, native and denatured GlpG in DDM was directly injected to dabcyl-labelled bicelles ($q = 1.5$) prepared by solubilizing the liposomes in CHAPS to the final GlpG conjugated-pyrene concentration of 1 mM, lipid concentration of 10 mM and DDM concentration of 5 mM in 20 mM HEPES buffer and 40 mM KCl. Also, for the positive and negative controls, the final measuring conditions were set to the experimental condition mentioned above. After mixing, the samples were incubated at room temperature for 24 h in a 96-well plate sealed with polyolefin tape. Pyrene quenching was measured at 390 nm and excited at 345 nm. The degree of quenching was determined by the ratio of experimental and control samples to pyrene fluorescence of GlpG in DDM.

2.3.11. Measuring the intrinsic biotin affinity of mSA variants in bicelles.

Biotin binding affinity of mSA variants at three cysteine sites (P95C, G172C, and V267C) was determined by FRET-based binding assay. The mSA variant with a weaker biotin-binding affinity, mSA_{DAB}-W79M (FRET acceptor), was titrated at 100 nM of GlpG singly labeled with BtnPyr (FRET donor) in 2% (w/v) DMPC/CHAPS bicelles, 40 mM KCl, 0.5 mM DTT, and 20 mM HEPES (pH 7.5). The titrated samples were transferred to a 96 well microplate sealed with

polyolefin tape and incubated at room temperature for 24 h. Changes in pyrene fluorescence were monitored at 390 nm exciting at 345 nm on a SpectroMax M5e plate reader. Data were averaged from three fluorescence readings. After 24 h of incubation, an excess of biotin to the final concentration of 2 mM was added and incubated more than 24 h to dissociate bound mSA from biotinylated GlpG and pyrene fluorescence was monitored. This data serves as a background signal. Background subtracted data were fitted to the following equation to obtain $K_{d, \text{biotin}}$ of mSA_{DAB}-W79M in bicelles.³⁹

$$F = A1 \times \frac{(P_T + [mSA] + K_{d, \text{biotin}}) - \sqrt{(P_T + [mSA] + K_{d, \text{biotin}})^2 - 4P_T[mSA]}}{2P_T} + A2 \quad \text{Eq. 4}$$

, where F is the measured fluorescence intensity, P_T is the total GlpG concentration, $[mSA]$ is the total mSA concentration, $K_{d, \text{biotin}}$ is the dissociation constant for the biotin-binding affinity of mSA_{DAB}, $A1$ is the net change in fluorescence and $A2$ is the fluorescence level without mSA_{DAB}. Fitted values include $A1$, $A2$, and $K_{d, \text{biotin}}$. The other values are fixed.

$K_{d, \text{biotin}}$'s of stronger biotin-binding variants (W79M, S45A, and S27A) were measured by FRET-based competitive binding assay. 1 μ M GlpG (G172C-BtnPyr) was preequilibrated with 2 to 5 times molar excess of the mSA variants labeled with DAB-MI for 3 h at room temperature. Under this condition, pyrene fluorescence was quenched. Next, the weaker biotin-affinity mSA variant without the label was titrated against the GlpG sample. Final concentrations were 3% DMPC/CHAPS ($q = 1.5$), 20 mM HEPES (pH 7.5), 40 mM KCl and 1 mM DTT. The titrated samples were transferred to a 96 well microplate, sealed with polyolefin tape, and incubated at room temperature for every 24 h. Changes in pyrene fluorescence were monitored at 390 nm exciting at 345 nm using a SpectroMax M5e plate reader. Data were averaged from three fluorescence readings. Under this condition, pyrene fluorescence increased due to the binding

competition between labeled and unlabeled mSA. Once equilibrium was reached (24 to 48 h), 2 mM biotin was added to replace bound mSA from GlpG and incubated from 7 to 24 h. Pyrene fluorescence was monitored until it reached an equilibrium. Once equilibrium was reached, the recorded fluorescence data served as a background signal. Background-subtracted data were fitted to Eq. 5.

This approach did not work for mSA_{DAB}-E51S because of the slow dissociation compared to the other mSA variants mentioned above. Therefore, the protocol was modified as follows: 1.5 μ M GlpG (G172C-BtnPyr) was first titrated against mSA-S27A without a label and pyrene fluorescence was monitored. After a stable read, this signal served as a background signal. Then, 2 μ M mSA_{DAB}-E51S was added, and pyrene quenching was monitored. After reaching the equilibrium (usually it takes 48 to 72 h), the recorded fluorescence data was subtracted from the background and net fluorescence change was fitted to the following equation.

$$F = A1 \times \frac{-\left[P_T + [mSA] + \frac{K_{unlabel}}{K_{dab}}(C_T - P_T)\right] + \sqrt{\left(P_T + [mSA] + \frac{K_{unlabel}}{K_{dab}}(C_T - P_T)\right)^2 + 4P_T[mSA]\frac{K_{unlabel}}{K_{dab}}}}{2P_T\frac{K_{unlabel}}{K_{dab}}} + A2 \quad \text{Eq. 5}$$

, where F is the measured fluorescence intensity, P_T is the total GlpG concentration, $[mSA]$ is the total mSA concentration, $K_{unlabel}$ is the dissociation constant for the biotin-binding affinity of mSA without dabcyyl label, K_{label} is the dissociation constant for mSA_{DAB}, $A1$ is the net change in fluorescence and $A2$ is the fluorescence level without mSA_{DAB}. Fitted values include $A1$, $A2$, and $K_{unlabel}$ and K_{label} . The other values were fixed.

2.3.12. Proteinase K digestion to prove denaturation of GlpG by steric trapping.

25 μ M GlpG (95_N172_M-BtnRG₂ and 172_M267_C-BtnRG₂) in the presence and absence of 125 μ M mSA-WT in 5 mM DDM, 20 mM HEPES (pH 7.5) and 40 mM KCl was incubated for 24 to 48 h at room temperature. After measuring the extent of denaturation using SN-LacYTM2 as a substrate, native and denatured samples were directly injected into 3% DMPC/CHAPS bicelles ($q = 1.5$) in 20 mM HEPES (pH 7.5) and 40 mM KCl. The final concentrations of GlpG and WT-mSA were 5 μ M and 25 μ M respectively. After incubating at room temperature for 24 h, Proteinase K was added to initiate the digestion reaction (the final concentration of 3.4 mg/mL). Samples were withdrawn for native GlpG at 0- and 30-min time points and for denatured GlpG, samples were taken at 0, 2, 10, 30 min. After the sampling, PMSF was added to the final concentration of 10 mM to quench the proteolytic reaction. Then, 10 mM DTT was added and incubated for 1 h at room temperature to remove the bound WT-mSA from biotinylated GlpG by cleaving the disulfide bond. Finally, 4% (w/v) SDS sample buffer was added, and SDS-PAGE gel was run at 100 V for 90 min on ice.

2.3.13. Probing the side chain contribution to GlpG stability by mutation.

Thirty seven-point mutations were made on GlpG in the background of N-terminal (95_N172_M-BtnRG₂) and C-terminal biotin (172_M267_C-BtnRG₂) pair. Residues were chosen based on the accessible surface area fraction (f_{ASA}). To calculate the f_{ASA} , the crystal structure of GlpG (PDB id: 2ic8) was provided as an input and the radius of water probe was set to 1.4 Å to the server, <http://www.scsb.utmb.edu/getarea/>. The mutations on completely buried residues ($f_{ASA} = 0$) were C104A, F139A, N154A, W158F, L161A, L174A, S181A, S201T, A206G, L207A, Y210F, A253V, G261A, and A265V. The mutations on partially buried residues ($f_{ASA} = 0.01$ to 0.20) were M100A, V119A, W125A, R137A, T140A, H150A, L155A, T178A, F197A, R214A, I223A,

L225A, M249A, A253V and D268A. The mutations on exposed residues ($f_{\text{ASA}} = 0.21$ to 1.0) were L123A, K132A, F133A, F135A, W136A, Y138F, F139A, L143A and N226A.

2.3.14. Molecular dynamics simulation of GlpG wild type in a bilayer and micelles.

The WT system was built with the x-ray crystal structure of *E. coli* GlpG³⁶ (PDB code: 2IC8; 2.1 Å resolution) embedded in a DMPC lipid bilayer (308 molecules) lying on the xy-plane, with the z-axis normal to the membrane. The lipid bilayer was constructed on the web-based CHARMM-GUI membrane builder.⁴⁰ Then, the composite system (*i.e.*, GlpG and the lipid bilayer) was immersed in the TIP3P water solvent⁴¹ followed by a charge neutralization and ionization with 150mM NaCl, which resulted in a final model of over 78,000 atoms in a box of 100×100×81 Å³. DDM micelles were built with 120 and 150 DDM molecules. Inter- and intramolecular potential energies were enumerated based on the CHARMM36 force field.⁴² The nonbonding van der Waals and short-range electrostatic interactions were treated with a typical cutoff distance of 12 Å, while the long-range electrostatic contributions were evaluated with the particle-mesh Ewald method.⁴³ All simulations were carried out using the NAMD2 software⁴⁴ massively parallelized on the GPU-accelerated IBM Power8 machine with a 2-fs time step in the semi-isotropic isobaric and isothermal (NPT) ensemble of 1 atm and 310 K, by which the membrane normal (z-axis) fluctuated independently from the lateral dimensions (xy-plane). The system was first subject to 10,000 steps of conjugate gradient energy minimization with restraints on lipids and GlpG to preserve their conformation and relative positions, followed by further pre-equilibration along 6 separated stages as the restraints gradually relaxed until no constraints applied. Then, we ran the MD of GlpG WT for ~150 ns. The final snapshot was employed as a starting structure for the 14 different variants as well as for WT. We obtained at least up to ~1.4 μs trajectories for WT and each variant.

2.3.15. Examining the equilibration of protein and amphiphiles.

To test the system equilibrium, we focused our observation on three parts: (1) protein, (2) lipids or detergents in the bulk, and (3) lipids or detergents in contact with protein. To test the equilibration of protein conformation, we monitored global structural fluctuations by calculating the heavy atom RMSD's during simulation until an enough number of conformations are obtained to reach equilibration (*i.e.*, ~500 ns). Next, to test the equilibration of lipid or detergent conformation in the bulk, we targeted lipid molecules in the bilayer without the protein, or detergent molecules in the micelles without the protein. Then, we calculated the *RMSD* (τ) as a function of the time lag τ by averaging over all available lipid molecules in the bulk as follows:

$$RMSD(\tau) = \frac{1}{N_L} \sum_{i=1}^{N_L} \langle RMSD_i(t, t + \tau) \rangle_t \quad \text{Eq. 6}$$

, where N_L is the number of the lipid or detergent molecules, and $\langle RMSD_i(t, t + \tau) \rangle_t$ is the heavy-atom root-mean square deviation between the i -th lipid conformations at the time t and $t + \tau$, averaged over the time t . Prior to RMSD calculation, the lipid, or detergent conformations in comparison at each t and $t + \tau$ were structurally aligned with each other by transrotating the heavy-atom conformations.

Lastly, for the interfacial lipid or detergent molecules in contact with protein, we focused on how fast the lipid or detergent dissociation would occur at the protein-lipid or protein-detergent interface by measuring the residence time using the autocorrelation function on time for the lipid heavy atom in contact with the protein as follows:

$$c(\tau) = \frac{1}{N_c} \sum_{i=1}^{N_c} \langle c_i(t, t + \tau) \rangle_t \quad \text{Eq. 7}$$

, where N_c is the number of contact events, and a single contact event is defined as a consecutive contact of a lipid or detergent with no non-contacting time gap longer than the lipid or detergent relaxation time measured above (*i.e.*, 20 ns). The autocorrelation function at the time τ of the i -th

contact event, $\langle c_i(t, t + \tau) \rangle_t$ is defined by $\langle q_i(t)q_i(t + \tau)/q_i^2(t) \rangle_t$, the normalized product of heavy atom contact numbers of a lipid or detergent in the i -th contact event, $q_i(t)$ and $q_i(t + \tau)$ at two-time moments (t and $t + \tau$), averaged over the time t . The retention time for lipid-lipid or detergent-detergent interactions were measured similarly for the lipid bilayer or micellar system without the protein.

2.4. Results

2.4.1. Establishing the reversible folding system of GlpG in bicelles.

To test our hypothesis, it is necessary to determine the thermodynamic stability ($\Delta G^{\circ}_{\text{N-D}}$) of GlpG in bicelles and micelles for the quantitative analysis of lipid effects. However, the reversible control of folding has been inherently difficult in a lipid bilayer environment. We have previously determined the $\Delta G^{\circ}_{\text{N-D}}$ of GlpG in dodecylmaltoside (DDM) micelles using the steric trapping strategy,³⁹ which couples spontaneous denaturation of a doubly biotinylated protein to the simultaneous binding of two bulky monovalent streptavidin (mSA) molecules (**Fig. 2.1**).^{39,45,46} By adjusting the biotin affinity of mSA by mutation on the biotin binding pocket, denaturation and mSA binding can be reversibly controlled. $\Delta G^{\circ}_{\text{N-D}}$ is determined by monitoring the attenuated second binding coupled to the denaturation without perturbing native protein–lipid or protein–water interactions. This method captures transient denaturation of the region encompassing a specific biotin pair on a protein, thus allowing for the measurement of local stability depending on the position of the biotin pair.^{39,47}

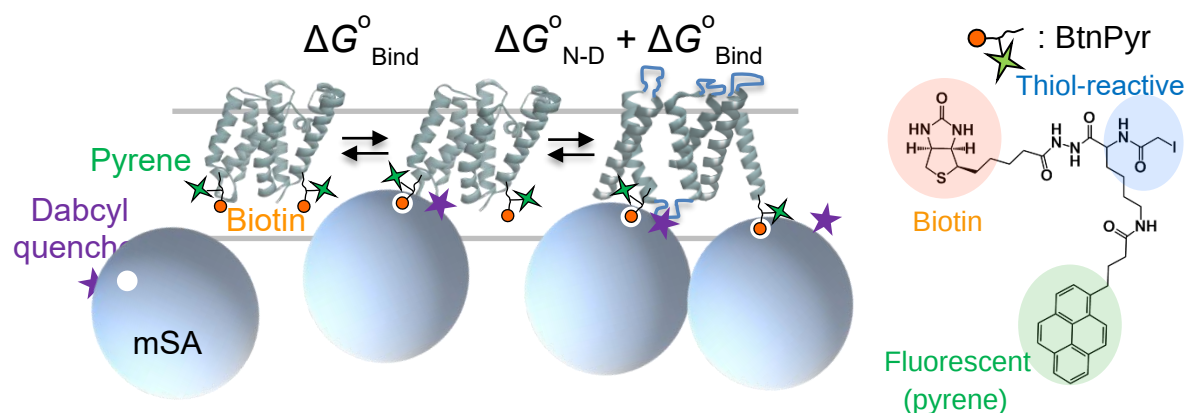


Figure 2.1. Establishment of the reversible folding system of GlpG in a lipid bilayer environment. The principle of steric trapping to determine the thermodynamic stability (ΔG°_{N-D}) of GlpG under native conditions. First, GlpG is labeled with biotin tags at two specific residues which are close in space in the native state but distant in the amino acid sequence. The first mSA (monovalent streptavidin) binds unhindered to either biotin label (ΔG°_{Bind}). The second mSA binds only when the tertiary contacts between biotinylated sites are transiently unraveled (ΔG°_{N-D}) because of the steric hindrance between bound mSA molecules. The coupling of mSA binding to denaturation attenuates the apparent binding affinity of the second mSA ($\Delta G^{\circ}_{Bind} + \Delta G^{\circ}_{N-D}$). The degree of attenuation is directly related to ΔG°_{N-D} , which is determined by fitting of the second binding phase to **equations (2–4) (Methods)**.

Here, we demonstrate the steric trapping scheme to determine ΔG°_{N-D} of GlpG in bicelles (**Figure 2.1**). Two double-cysteine variants of GlpG were generated at the residue pairs, 95_N172_M and 172_M267_C (N, M and C designate the N-terminal, middle and C-terminal helices, respectively, indicating the positions of cysteine mutation) to conjugate the thiol-reactive biotin derivative with fluorescent pyrene (BtnPyr) (**Figure 2.1. and Figure 2.10**).³⁹ The pyrenyl group on BtnPyr sensitizes binding of mSA labeled with the dabcyI quencher. The resulting double-biotin variants, 95_N172_M-BtnPyr₂ and 172_M267_C-BtnPyr₂ (**Figure 2.2 top**) are used to measure the stability approximately at the N- and C-terminal halves of GlpG, which are denoted as N- and C-subdomains, respectively.³⁹ We chose neutral DMPC:CHAPS bicelles (the lipid-to-detergent molar ratio = 1.5) as a bilayer medium. Under this condition, lipids are known to be largely segregated from detergents to form a stable bilayer phase,⁴⁸ and

denaturation of GlpG was reversibly coupled to mSA binding in the experimentally accessible timescale (~days).

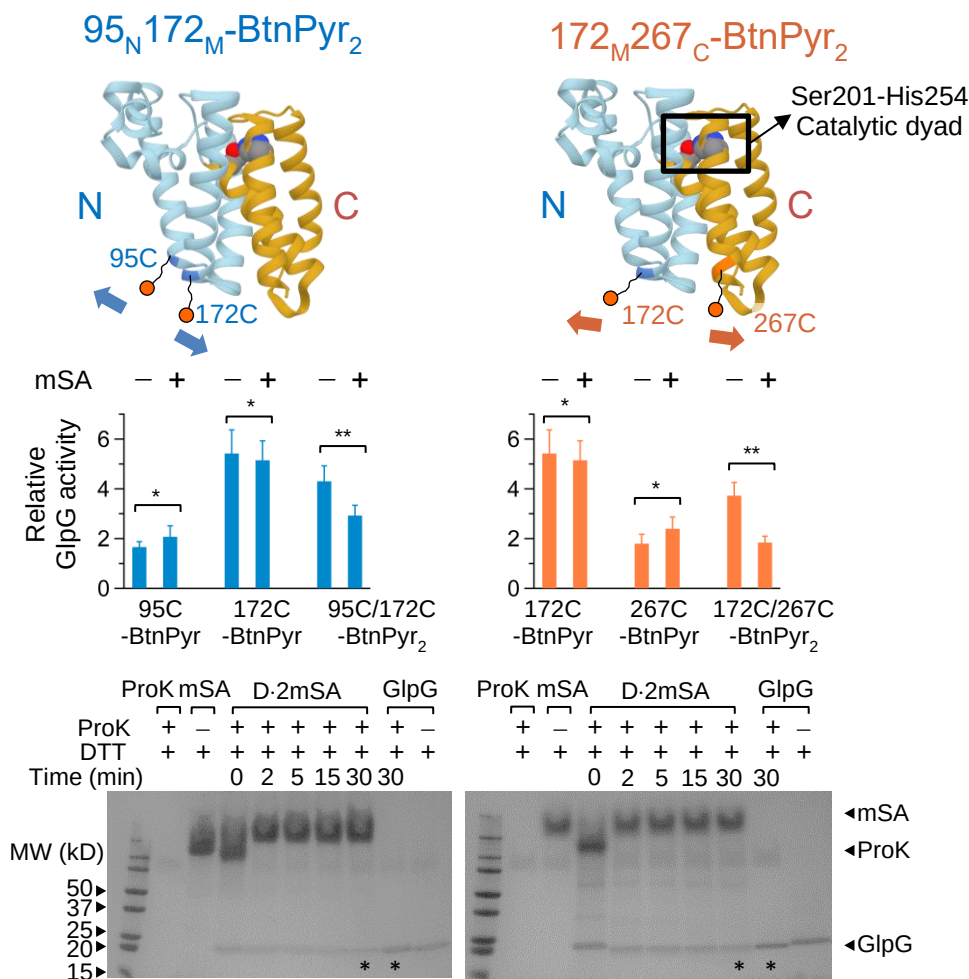


Figure 2.2. Biotin labeling of double-cysteine variants of GlpG. (Top) Double-cysteine variants of GlpG labeled with the thiol-reactive biotin derivative BtnPyr with a pyrene fluorophore or BtnRG with a spin label (Figures 2.10 and 2.11, respectively). (Middle) Proteolytic activities of single- and double-biotin variants of GlpG for the model substrate LYTM2 in the absence or presence of an excess concentration of mSA-WT (Figure 2.12). The results of the student *t*-test for pairwise comparisons showing the effect of mSA-WT on the activity are shown as asterisks (*: $p>0.05$; **: $p<0.05$). (Bottom) Comparing the susceptibilities of the native and sterically denatured states to proteolysis by proteinase K (ProK) (Figure 2.11).

The addition of mSA to the individual single-biotin variants (95_N -BtnPyr, 172_M -BtnPyr and 267_C -BtnPyr) did not affect the proteolytic activity of GlpG (Figure 2.2. middle) indicating

that the single binding of mSA to each biotinylated site did not perturb the native conformation of GlpG. In contrast, the addition of mSA to the samples containing the double-biotin variants decreased GlpG activity and induced their preferential digestion by Proteinase K (ProK) by 40–70% relative to the samples without mSA (**Figure 2.2 bottom**). The incomplete inactivation and proteolysis by ProK were correlated with the efficiency of double-biotin labeling of the double-cysteine variants (**Figures 2.11 and 2.12**). ProK proteolyzes an unfolded and water-exposed region on a protein with low sequence specificity.^{49,50} Thus, saturated binding of mSA to the double-biotin variants induced denaturation of GlpG increasing its conformational flexibility and water accessibility in bicelles.

Next, we tested the coupling between the mSA binding and GlpG denaturation, and their reversibility, which is necessary to determine $\Delta G^{\circ}_{\text{N-D}}$ (**Figure 2.3.a**). We first prepared folded or sterically denatured GlpG in micelles, which was then transferred to bicelles at an increasing concentration of mSA_{DAB-E51S} (the quencher-labeled mSA mutant with a reduced biotin binding affinity) (**Figures 2.13-2.15**). After incubation, the binding isotherms between GlpG and quencher-labeled mSA were constructed by monitoring the quenching of pyrene fluorescence. The degree of denaturation of GlpG was also monitored using GlpG activity as a folding indicator (**Figure 2.12**). For each double-biotin variant, the binding isotherm clearly display two binding phases, the tight first binding followed by the attenuated second binding. The attenuated second binding and denaturation phases agreed with each other regardless of the initial state (*i.e.*, folded vs sterically denatured) demonstrating the reversible control of GlpG folding by steric trapping. Transfer of folded and sterically denatured GlpG from micelles to bicelles was close to completion (**Figure 2.16**)⁵¹ indicating that the observed binding-denaturation coupling occurred in the bicellar phase.

Fitting of the attenuated second binding phases yielded $\Delta G^{\circ}_{\text{N-D,bicelles}} = -7.0 \pm 0.2$ kcal/mol for 95_N172_M-BtnPyr₂ (*i.e.*, the stability of N-subdomain) and -6.7 ± 0.2 kcal/mol for 172_M267_C-BtnPyr₂ (*i.e.*, the stability of C-subdomain) in bicelles (**Figure 2.3.b**). In micelles, we have shown that GlpG is organized into the more stable N-subdomain ($\Delta G^{\circ}_{\text{N-D,micelles}} = -5.8$ kcal/mol), the denaturation of which induces global denaturation, and the less stable C-subdomain ($\Delta G^{\circ}_{\text{N-D,micelles}} = -4.7$ kcal/mol), which harbors the catalytic dyad (Ser201–His254) and undergoes subglobal denaturation.³⁹ Relative to micelles, bicelles stabilized N- and C-subdomains by -1.2 ± 0.3 kcal/mol and -2.0 ± 0.3 kcal/mol, respectively. Thus, the lipid environment made the stability difference between the two subdomains insignificant (0.2 ± 0.3 kcal/mol) by preferentially stabilizing C-subdomain. This result raises a possibility that, while both bicelles and micelles support the native conformation of GlpG, the lipid environment provided by bicelles enhances the stability by strengthening intraprotein or protein–amphiphile interactions compared to micelles.

2.4.2. Lipids promotes the residue burial in the folding.

To test this possibility, we compared the contribution of individual residue interactions to the stability of GlpG in micelles and bicelles. Toward this goal, a total of 37 residues that cover many packed regions were targeted for mutation (mostly large-to-small mutations except for A253V, G261A and A265V at the TM4/TM6 interface) (**Figures 2.17-2.18** and **Table 2.1**), including 12 completely buried (the fraction of solvent-accessible surface area, $f_{\text{ASA}} = 0$), 15 moderately buried ($f_{\text{ASA}} = 0.01\text{--}0.3$), and 10 exposed residues ($f_{\text{ASA}} > 0.3$). Each mutation was made in the background of the double-biotin variant, 95_N172_M-BtnPyr₂ (the biotin pair at N-subdomain) or 172_M267_C-BtnPyr₂ (the biotin pair at C-subdomain) and the stability was measured for both double-biotin variants bearing the same mutation.

Overall, the mutation-induced stability changes in micelles ($\Delta\Delta G^{\circ}_{\text{WT-Mut,micelle}}$) and bicelles ($\Delta\Delta G^{\circ}_{\text{WT-Mut,bicelles}}$) displayed moderate linear correlation with the slope close to 1 ($m = 0.8 \pm 0.1$ and $m = 1.1 \pm 0.1$ when the stabilities were measured at N- and C-subdomains, respectively) (**Figure 2.4.**). This may imply that the contribution of individual residue interactions to the stability is similar in the two distinct hydrophobic environments.

Surprisingly, however, the mutational impacts on the stability showed a systematic difference in micelles and bicelles depending on the degree of burial of the residues targeted for mutation. In the $\Delta\Delta G^{\circ}_{\text{WT-Mut,bicelles}}$ vs $\Delta\Delta G^{\circ}_{\text{WT-Mut,micelle}}$ plots for the residues completely buried in the protein interior, the fitted slopes were significantly larger than one ($m = 1.2 \pm 0.2$ with $R = 0.93$ for N-subdomain and $m = 1.7 \pm 0.2$ with $R = 0.96$ for C-subdomain) (**Figure 2.5 and 2.6 left**), indicating that disruption of the native internal packing led to larger destabilization of GlpG in bicelles than in micelles. That is, the lipid environment induced more favorable burial of the native residues in the protein interior than detergents. This effect was more pronounced when the stability was measured at C-subdomain ($m = 1.7 \pm 0.2$) than at N-subdomain ($m = 1.2 \pm 0.2$). That is, in the lipid environment, C-subdomain is destabilized more than N-subdomain by an internal mutation regardless of the position of the mutated site (**Figure 2.5 and 2.6**). The smaller destabilization of C-subdomain in bicelles could be due to the difference in the degree of compactness of the denatured state conformation between the two subdomains. We have shown that, when denatured in micelles, bicelles and liposomes, N-subdomain is compact close to the level of chain collapse (*i.e.*, the poor solvent limit) while the TM helices in C-subdomain is expanded (*i.e.*, the good solvent limit).⁵¹ Thus, mutations in the protein interior will be less effective in destabilizing N-subdomain.

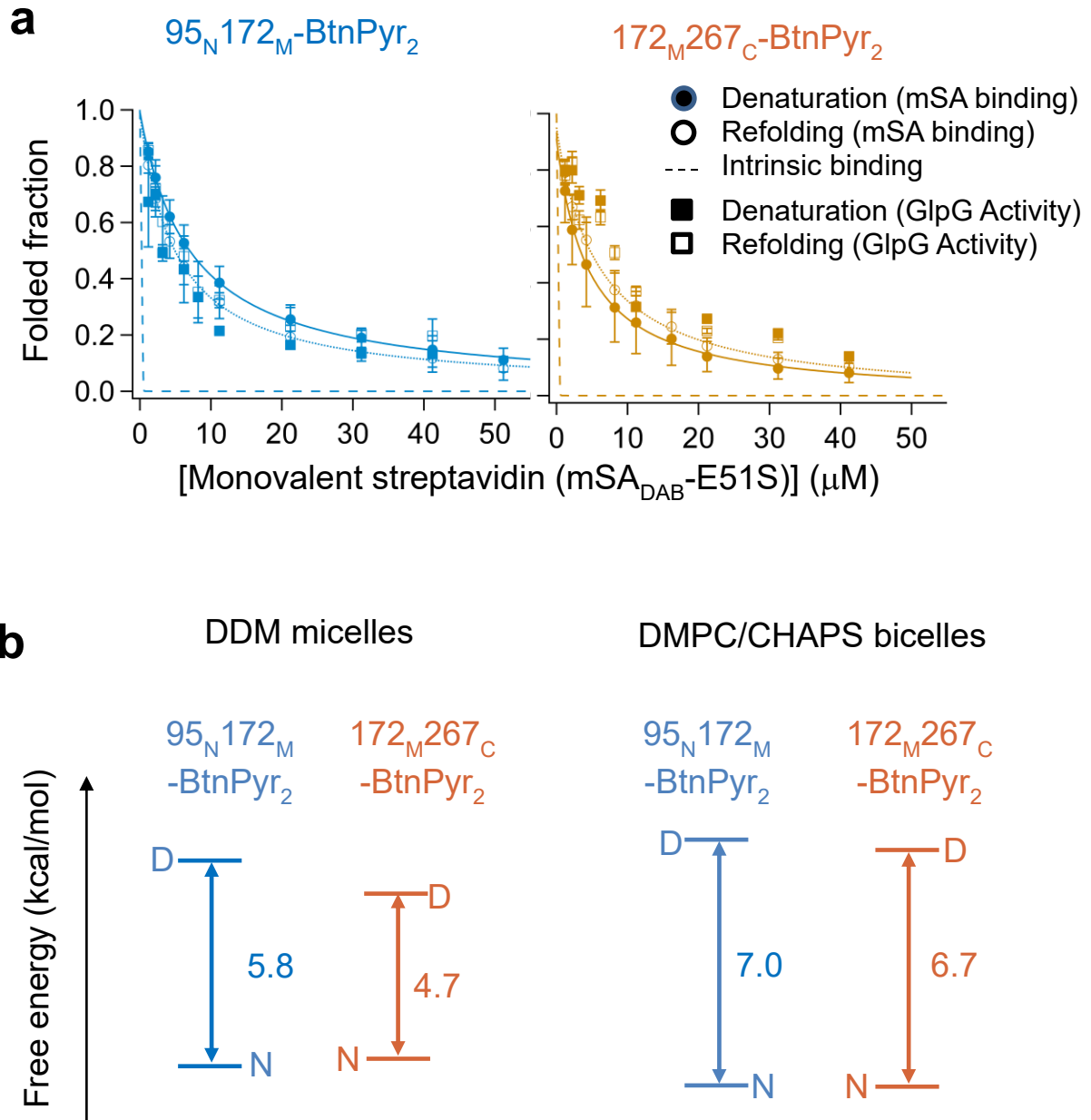


Figure 2.3. Reversibility of the folding and mSA binding of GlpG and their coupling induced by steric trapping in bicelles. (a) mSA binding was measured by quenching of pyrene fluorescence while denaturation by the proteolytic activity of GlpG for LYTM2. The data for mSA binding and denaturation were normalized to the amplitude of the second binding phase and the change in activity of GlpG, respectively. (b) The influence of the lipid bilayer on the stability of N- and C-subdomains of GlpG.

The slopes in the $\Delta\Delta G^{\circ}_{\text{WT-Mut,bicelles}}$ vs $\Delta\Delta G^{\circ}_{\text{WT-Mut,micelle}}$ plots decreased to $m = 0.9\text{--}1.0$ ($R = 0.72\text{--}0.82$) for the moderately buried residues (**Figure 2.5. middle**) and further to $m = 0.5\text{--}0.8$ ($R = 0.69\text{--}0.88$) for the exposed residues (**Figure 2.5. right**): As the degree of residue exposure increased to the hydrophobic environment, the lipid environment attenuated the mutational impact on the stability relative to detergents. This result may be explained by either of two scenarios: 1) lipids more strongly binds on the protein surface than detergents, and thus they can more favorably compensate the mutation-induced structural defects at the protein, or 2) lipids less tightly interact with the protein surface than detergents such that the surface mutations induce smaller impacts to the stability in the bilayer. Interestingly, our molecular dynamics (MD) simulation points to the latter (see below).

In summary, our results suggest that the contribution of individual residue interactions to membrane protein stability strongly depends on the hydrophobic solvent environment. The native buried residues contribute more to the stability of GlpG in the lipid environment than in micelles whereas the exposed residues contribute less.

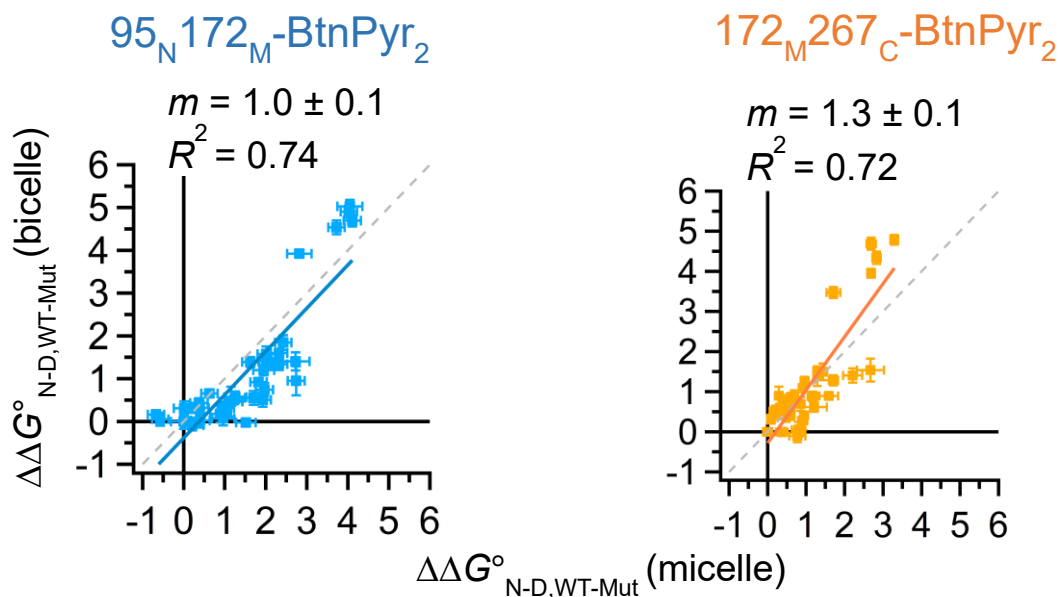


Figure 2.4. Comparison of GlpG stability measured at N- and C-subdomains in micelles and bilayers. Comparison of the mutational impacts on the stability ($\Delta\Delta G^{\circ}_{N-D,WT-Mut} = \Delta G^{\circ}_{N-D,WT} - \Delta G^{\circ}_{N-D,Mut}$) in micelles and bilayers. The dashed lines with the slope (m) = 1 are shown as a guide to indicate the equal stability in micelles and bilayers

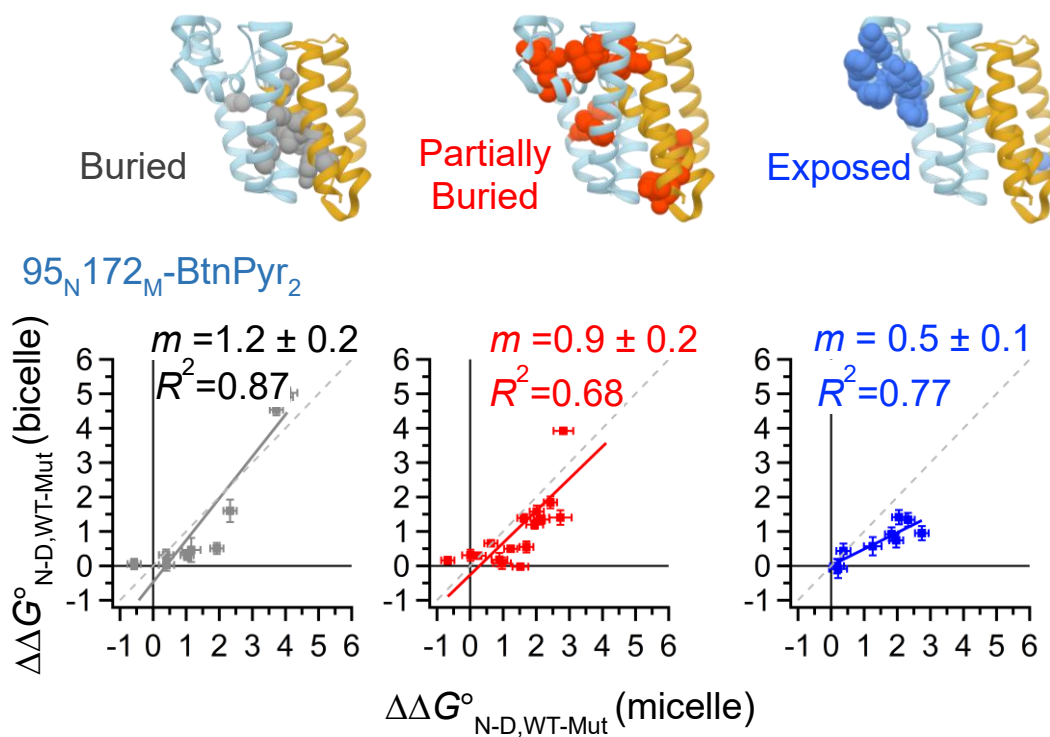


Figure 2.5. Comparison of the mutational impacts on the stability in micelles and bilayers depending on the degree of burial of the residues targeted for mutation for N-subdomain

(f_{ASA} , the fraction of the residual solvent accessible surface area). Errors in $\Delta G^{\circ}_{\text{N-D}}$ and $\Delta\Delta G^{\circ}_{\text{N-D,WT-Mut}}$ denote $\pm$ s. d. from fitting.

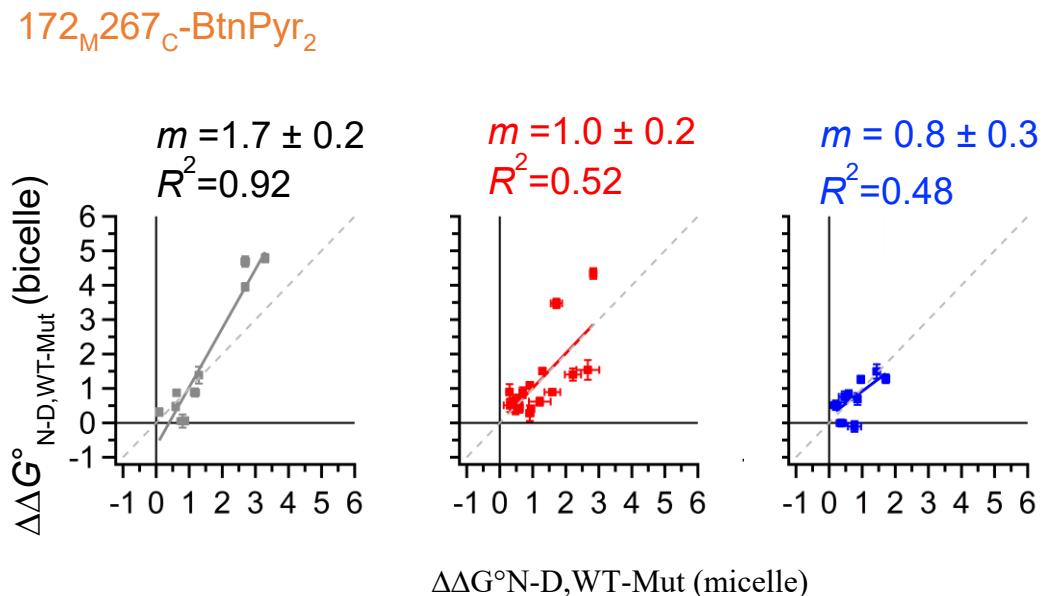


Figure 2.6. Comparison of the mutational impacts on the stability in micelles and bicelles depending on the degree of burial of the residues targeted for mutation for C-subdomain (f_{ASA} , the fraction of the residual solvent accessible surface area). Errors in $\Delta G^{\circ}_{\text{N-D}}$ and $\Delta\Delta G^{\circ}_{\text{N-D,WT-Mut}}$ denote $\pm$ s. d. from fitting.

2.4.3. Lipids strengthen the residue interaction network of GlpG.

These findings raise a question, whether these lipid-induced effects on GlpG stability are limited to the region where the stability was probed (*i.e.*, the position of the biotin pair or mutation) or globally impact the residue interaction network of the protein. To answer this question, we employed our “cooperativity profiling” method to analyze whether a given residue is engaged in cooperative or localized interactions with the surrounding.³⁹ This method quantifies the degree of spatial propagation of structural perturbation in a protein (**Figure 2.19**):³⁹ We first

make a point mutation (typically to alanine) to perturb specific residue interactions. Next, using steric trapping, the stability changes induced by the mutation are measured at the two biotin pairs located in different regions (*i.e.*, $\Delta\Delta G^0_{WT-Mut}^N$ and $\Delta\Delta G^0_{WT-Mut}^C$ measured with the biotin pairs at N- and C-subdomains, respectively). Finally, the differential effect of the mutation on the stability of each subdomain is calculated: $\Delta\Delta\Delta G = \Delta\Delta G^0_{WT-Mut}^N - \Delta\Delta G^0_{WT-Mut}^C$.

If $\Delta\Delta\Delta G$ for a given mutation is within the range of thermal energy ($|\Delta\Delta\Delta G| \leq RT$, R : gas constant and T : absolute temperature), the mutation similarly destabilizes the two subdomains, that is, the perturbation efficiently propagates throughout the protein, and the mutated site engages in “cooperative” interactions. When $|\Delta\Delta\Delta G| > RT$, two scenarios arise: If a mutation preferentially destabilizes the subdomain in which the mutation has been made, the perturbed interactions are “localized”. If a mutation on one subdomain induces larger destabilization of the other, the perturbed interactions are “overpropagated”. We applied four cut-off values, $\Delta\Delta\Delta G = -2RT, -RT, +RT$ and $+2RT$ to precisely resolve the degree of cooperativity of each residue interaction.³⁹

In micelles, we have previously obtained the cooperativity profiles of 20 residues in micelles.³⁹ In this study, we extended the analysis to 16 additional residues in micelles and newly identified the profiles of all 36 residues in bicelles. The resulting “cooperativity map” revealed cooperative, localized and overpropagated residue interactions in micelles (**Figure 2.7**). Notably, cooperative interactions are clustered in multiple distinct regions of the GlpG structure, including 1) The packing core near the bilayer center (Met100, Cys104, Leu174 and Thr178)³⁹; 2) The TM4/TM6 interface which is engaged in both cooperative and overpropagated interactions. This interface harbors three conserved residues in rhomboid proteases, the Ser201 (TM4)–His254 (TM6) catalytic dyad and Gly261 (TM6) in the middle of the glycine zipper motif;³⁹ 3) The residues

forming the water-conduction channel (Ser201, Met249, His150 and Asn154). This channel connects the water retention pocket to the catalytic dyad providing water molecules required for proteolysis.

Strikingly, the spatial pattern of the cooperativity profiles substantially changed in bicelles. Most of the localized and overpropagated residue interactions in micelles turned into cooperative interactions in bicelles. As a result, nearly all packed regions of the protein were engaged as a single cooperative unit in bicelles except for a few residues (Leu123, Leu223 and Gly261) in the extracellular or periplasmic interfacial regions. Interestingly, the reassignment of the cooperativity profiles in bicelles using the narrower cutoff values (*i.e.*, $\Delta\Delta\Delta G = -RT$, $-1/2RT$, $+1/2RT$ and $+RT$) yields a similar spatial distribution of cooperative interactions to that in micelles using the regular cutoff values (**Figure 2.20**), indicating that the cooperativity profiles in micelles are preserved in bicelles to a certain extent. Notably, the proteolytic activity of GlpG in bicelles increased by three times relative to that in micelles (**Figures 2.21-2.22**), which imply that the enhancement in cooperativity may lead to the enhancement in protein activity.

Our result indicates that the hydrophobic solvent affects how structural perturbation is propagated throughout the proteins. Compared to detergent micelles, the lipid environment strengthens the residue interaction network promoting the propagation of structural perturbation throughout the protein. Therefore, the observed lipid effects such as the increased and uniform local stabilities (**Figure 2.3.b**) and the favorable residue burial (**Figure 2.5 and 2.6**) stem from the increased cooperativity (*i.e.*, the strengthened residue interaction network) rather than the local effect.

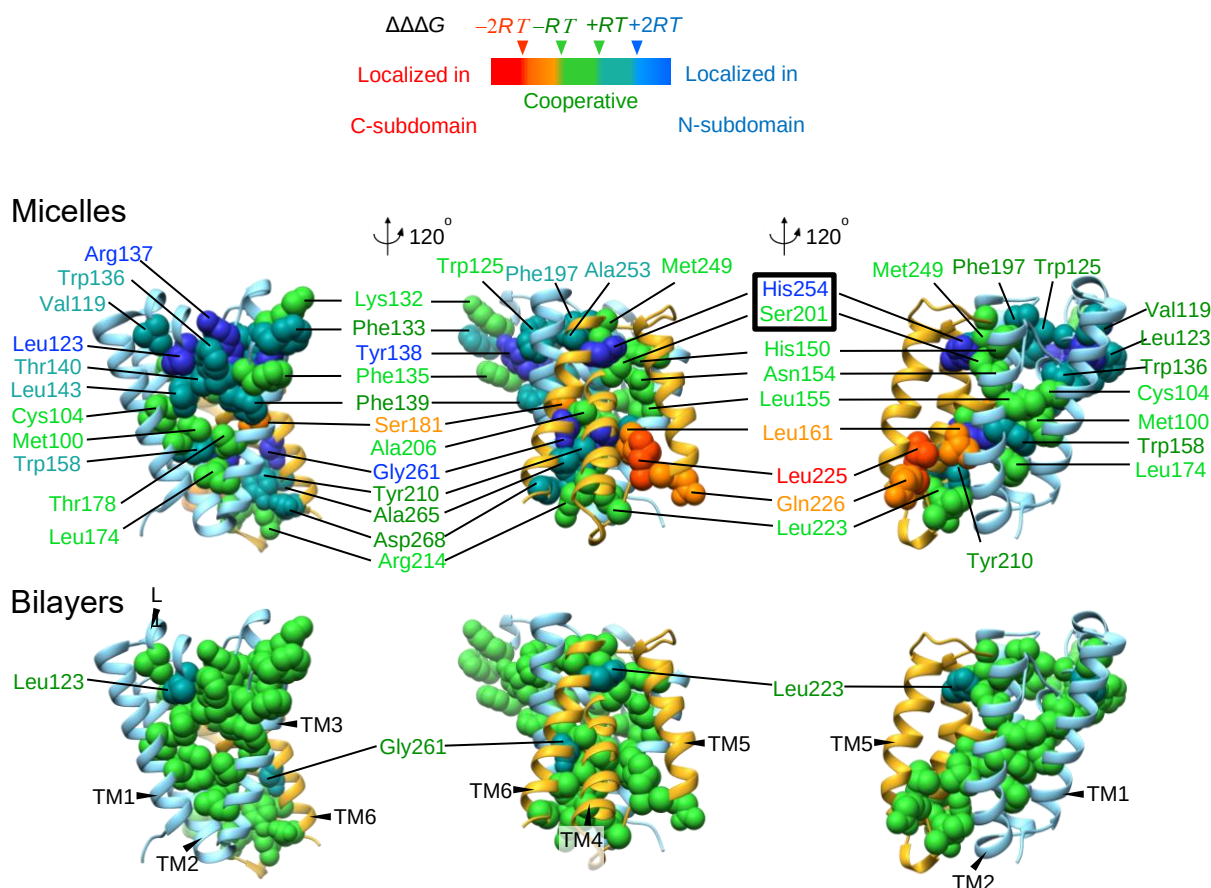


Figure 2.7. Lipid bilayer strengthens the cooperative network in GlpG. Cooperativity maps of GlpG in micelles and bicelles at a residue resolution. The cooperativity profiles of each residue are color-coded on the basis of the $\Delta\Delta\Delta G$ values, the “cooperative” (green, $|\Delta\Delta\Delta G| \leq RT=0.6$ kcal/mol), “moderately localized in N-subdomain” (tin, $2RT \geq \Delta\Delta\Delta G > RT$), “localized in N-subdomain” (blue, $\Delta\Delta\Delta G > 2RT$), “moderately localized in C-subdomain” (orange, $-RT > \Delta\Delta\Delta G \geq -2RT$), and “localized in C-subdomain” (red, $-2RT > \Delta\Delta\Delta G$). The cooperativity profiles of 20 residues in micelles was determined previously.

2.4.4. Lipids weakly solvate GlpG compared to micelles.

To track down the molecular origin of the environmental dependence of GlpG stability and cooperativity, we carried out all-atomic MD simulation of the GlpG–DMPC bilayer and GlpG–DDM micellar complexes up to 1.5 μ s and compared the solvation behaviors of the amphiphiles on the protein. The simulation was carried out by Dr. Seung-Gu Kang at IBM Watson Laboratory.

In general, studying protein stability requires an analysis of both native and denatured states. However, as the atomic-level modeling and simulation is challenging for the denatured states of membrane proteins, we infer the role of amphiphile solvation in GlpG stability from the simulation of only native GlpG.

In the initial modeling and simulation, we noticed that the protein–detergent interactions systemically depended on the detergent-to-protein molar ratio (Det/Pro). We finally chose the two ratios, 120 (or Det120) and 150 (or Det150) for detailed analysis. While Det120 better represented the experimental result regarding the micellar shape and aggregation number, increasing the Det/Pro ratio made the micelle more oblate-shaped and increased the detergent packing density (**Figure 2.23**).^{52,53} The bilayer and micelles without GlpG were separately simulated to evaluate the influence of the embedded protein on the amphiphile dynamics.

In both the bilayer and micelles, the backbone root-mean-square deviation (RMSD) of GlpG relative to the crystal structure was stabilized within ~300 ns at 1.5–2.5 Å (**Figure 2.8. top**). The RMSD (τ) of the lipid and detergent molecules not in contact with GlpG, which was measured as a function of the time lag τ ,⁴⁷ reached a plateau in ~10 ns at 2.1 Å for detergents and in ~30 ns at 3.5 Å for lipids (**Figure 2.8 bottom**). The conformation of lipids in the bilayer relaxed slower and fluctuated more at equilibrium than that of detergents in the micelle. Therefore, the conformation of both GlpG and the amphiphiles reached an equilibrium within a much shorter time than the total simulation time. The residue root-mean-square fluctuation (RMSF) indicates the rigid TM helices (~1 Å) and the relatively flexible interhelical loops (2–4 Å) (**Figure 2.8.b**). Taken together, the time- or residue-averaged dynamic features of GlpG were overall similar in the bilayer and micelles.

Next, to analyze the solvation dynamics of lipids and detergents on GlpG, we measured the average resident times (τ_R) for the lipid or detergent heavy atoms in contact with the protein ($\tau_{R,Pro-Lip}$ or $\tau_{R,Pro-Det}$) as well as with themselves ($\tau_{R,Lip-Lip}$ or $\tau_{R,Det-Det}$) by calculating the contact autocorrelation function on time after the systems reached an equilibrium (**Figure 2.9**).⁴⁷ τ_R was defined as the time when the correlation function decreased by $1/e$ from the start. We first compare the solvation dynamics of lipids to that of detergents at the more realistic Det/Pro ratio of 120 ($\tau_{R,Det-Det120}$ and $\tau_{R,Pro-Det120}$) and then describe how increasing the ratio affects the solvation of detergents (*i.e.*, $\tau_{R,Det-Det150}$ and $\tau_{R,Pro-Det150}$).

We expected that, because a lipid molecule has more hydrocarbon tails (double) than a detergent (single), lipids would have a higher probability to form vdW contacts, and thus bind tighter to the nonpolar surface of GlpG and themselves. During simulation, lipids indeed bound to themselves with a longer residence time than detergents at Det/Pro = 120 ($\tau_{R,Lip-Lip} = 38$ ns vs $\tau_{R,Det-Det120} = 24$ ns) (**Figure 2.9. left**). Contrary to our expectation, however, lipids resided on the protein surface with a shorter residence time than detergents ($\tau_{R,Pro-Lip} = 85$ ns vs $\tau_{R,Pro-Det120} = 93$ ns) probably because the lipids on the protein was strongly attracted by the bulk lipids facilitating dissociation. In this line, using the τ_R 's of protein–amphiphile vs amphiphile–amphiphile contacts, we calculated the solvation free energies of lipids and detergents on the protein, which were $\Delta G^o_{Solv,Pro-Lip} = -RT\ln(\tau_{R,Pro-Lip}/\tau_{R,Lip-Lip}) = -0.5$ kcal/atom and $\Delta G^o_{Solv,Pro-Det120} = -RT\ln(\tau_{R,Pro-Det}/\tau_{R,Det-Det120}) = -0.8$ kcal/atom (R : gas constant; T : absolute temperature). Therefore, while ΔG^o_{Solv} 's for both lipids and detergents were comparable to thermal energy (~ 0.6 kcal/atom), lipids solvated GlpG weaker than detergents.

Notably, the headgroup and hydrocarbon tail regions of lipid or detergent molecules differentially contributed to the overall protein–amphiphile interactions (**Figure 2.9. right**). While the headgroups and tails of lipids have similar residence times on GlpG ($\tau_{R, \text{Prot-LipHead}} = 64 \text{ ns}$ vs $\tau_{R, \text{Prot-LipTail}} = 60 \text{ ns}$), the tails of detergents resided on the protein much longer than the headgroups ($\tau_{R, \text{Prot-DetTail120}} = 88 \text{ ns}$ vs $\tau_{R, \text{Prot-DetHead120}} = 27 \text{ ns}$). That is, the solvation of GlpG by lipids is equally mediated by the headgroups and tails whereas that by detergents is mainly by the tails.

Interestingly, upon increase in the micellar volume by increasing the Det/Pro ratio from 120 to 150, detergent molecules associated with themselves tighter ($\tau_{R, \text{Det-Det120}} = 23 \text{ ns}$ vs $\tau_{R, \text{Det-Det150}} = 28 \text{ ns}$) but weaker with the protein ($\tau_{R, \text{Pro-Det120}} = 93 \text{ ns}$ vs $\tau_{R, \text{Pro-Det150}} = 80 \text{ ns}$) (**Figure 2.9 bottom**). As a result, the residence times in the GlpG–detergent complex became similar to those in the GlpG–lipid complex. This change could be explained by the rebalance of enthalpic and entropic contributions in solvating the protein: Enthalpically, the relatively close packing of detergent molecules in Det150 (the molecular area per detergent, $A_{\text{Det150}} = 72 \text{ \AA}^2$ vs $A_{\text{Det120}} = 77 \text{ \AA}^2$) will strengthen vdW interaction between themselves, facilitating their dissociation from GlpG. The dissociation will further be favored by the increase in mixing entropy of detergents in the larger micellar volume. This thermodynamic effect in the micellar systems (*i.e.*, enhanced dissociation of amphiphiles from the protein by strong attraction to bulk amphiphiles) may have the same physical origin as the inefficient solvation of GlpG by lipids.

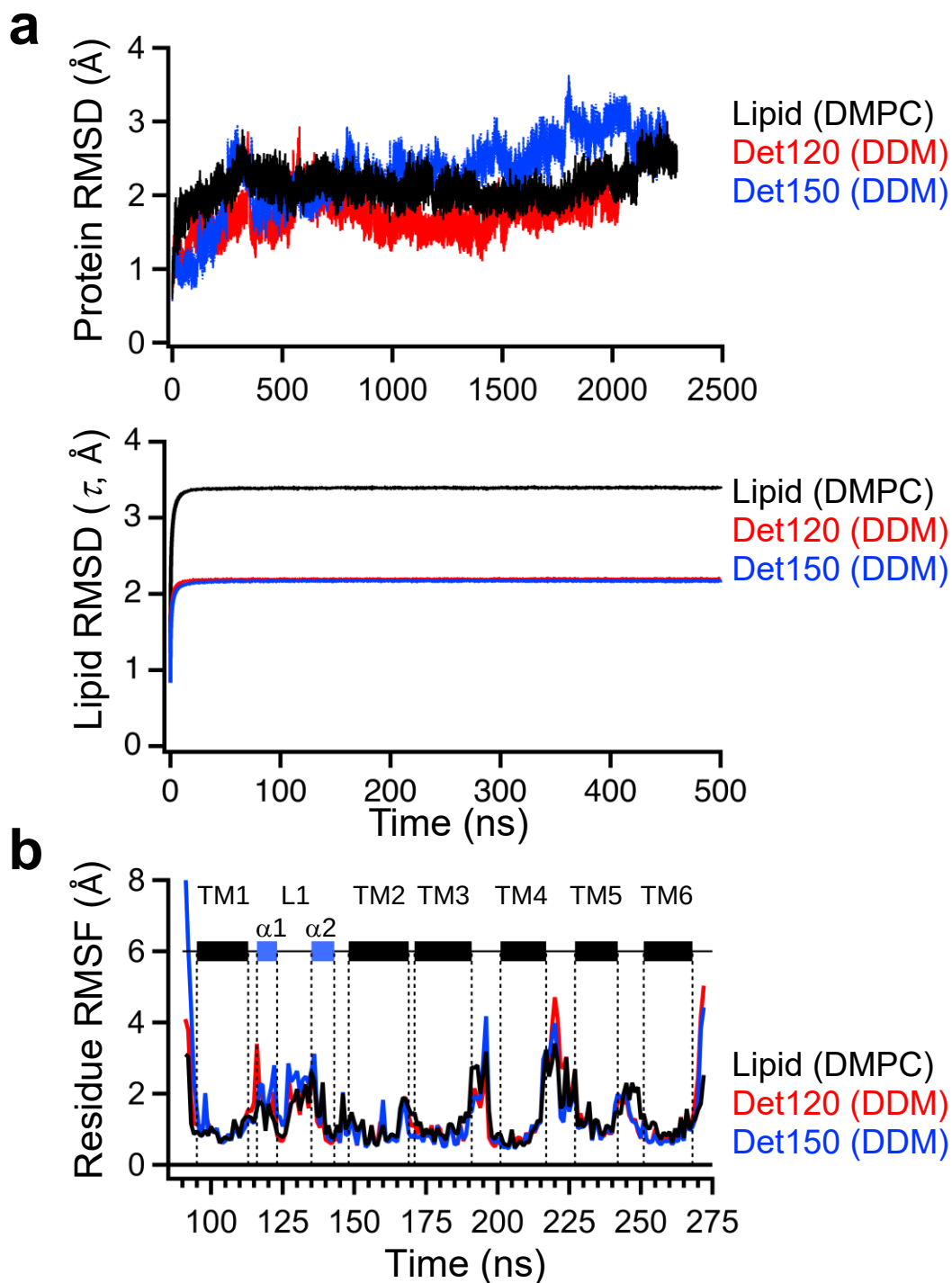


Figure 2.8. All-atom MD simulation of GlpG WT in a DMPC lipid bilayer and DDM micelles. **(a) Top:** The root-mean-square-displacement (RMSD) of the backbone heavy atoms (PDB code: 2IC8).³⁶ **Bottom:** Equilibration of the lipid and detergent conformation measured by the average RMSD (τ) as a function of the time lag τ . **(b)** The residue root-mean-square-fluctuation (RMSF) of GlpG WT.

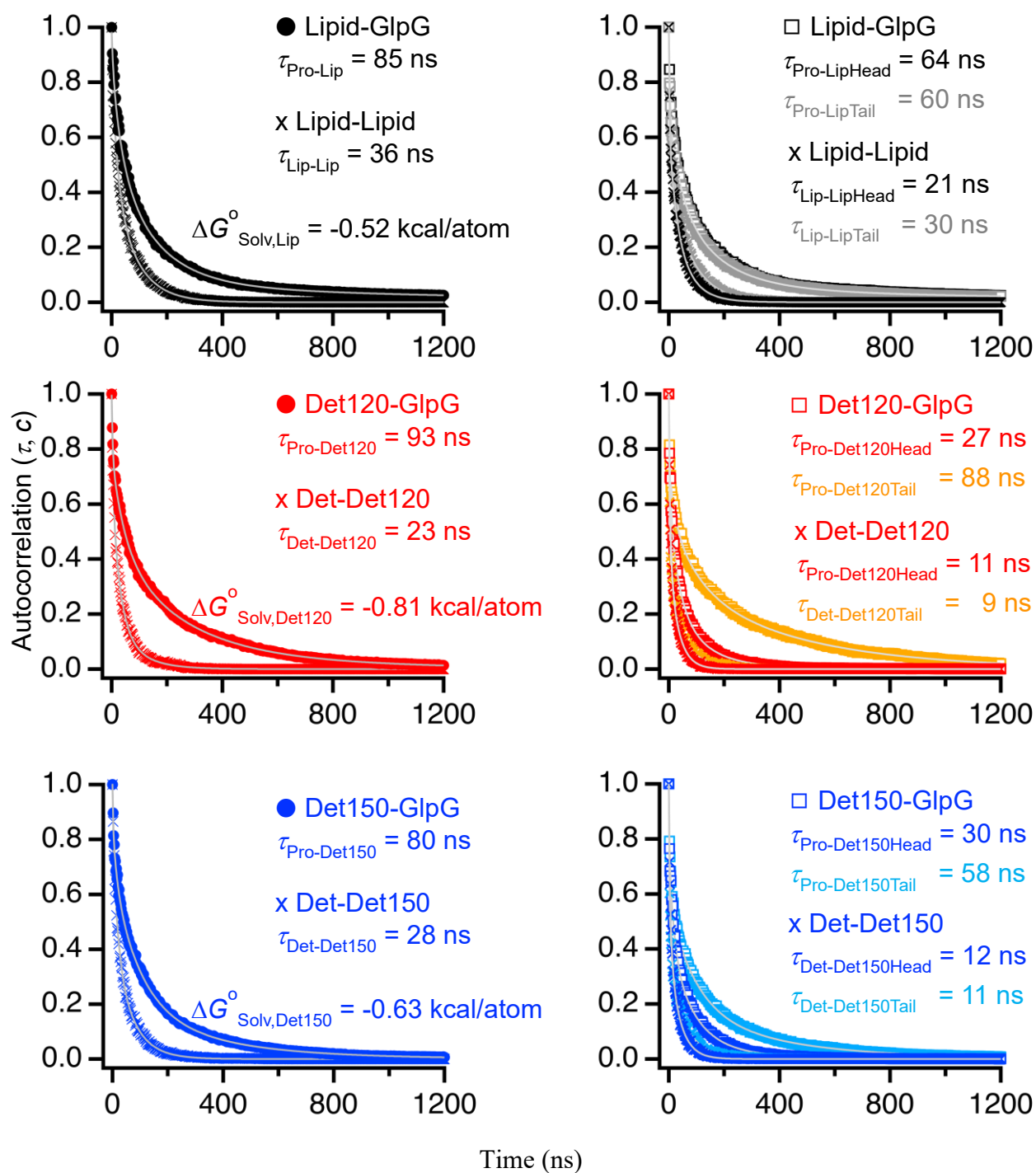


Figure 2.9. Dissociation of lipid and detergent molecules at the protein surface or from themselves in the bulk phase measured by the lipid-contact autocorrelation function on time. *Right:* Dissection of lipid and detergent interactions with the protein and themselves into the headgroup and tail regions. Retention time (τ) and the free energy of solvation ($\Delta G^{\circ}_{\text{Solv}}$) for each

case are shown. Det120 and Det140 denote the DDM-to-protein molar ratios of 120 and 140, respectively

2.5. Discussion

Although the native lipid bilayers in cell membranes have been regarded as an optimal environment for folding and function of membrane proteins, the chemical and physical basis of this perception remains elusive. Here, by comparing the folding of GlpG in nonnative detergent micellar vs native lipid bilayer environments, we discover that the solvation by lipids yields unique outcomes on the stability and cooperativity of the protein in comparison to that by detergents: 1) GlpG is stabilized by 1.5–2.0 kcal/mol (in $\Delta G^{\circ}_{\text{N-D}}$) and acquires nearly uniform local stabilities throughout the protein; 2) Destabilization effect by the mutation in the protein interior is amplified whereas that on the surface is attenuated; 3) The residues are engaged with one another in a highly cooperative manner to stabilize GlpG and thus, structural perturbation is widely propagated within the protein. These results commonly indicate that the lipid environment more tightly organize the protein interior than micelles, strengthening the protein's residue interaction network.

2.5.1. Enhanced stability and cooperativity can be explained by inefficient lipid solvation.

Our MD simulation suggests that the tight protein organization in the lipid environment may stem from the relatively inefficient lipid solvation on the protein. Stability of membrane proteins in an amphiphilic assembly (*e.g.*, the lipid bilayer or micelle) is defined as the free energy difference between the native and denatured states ($\Delta G^{\circ}_{\text{N-D}}$), which can be dissected into the free energy contributions from intraprotein, protein–amphiphile and amphiphile–amphiphile interactions ($\Delta G^{\circ}_{\text{N-D}} = \Delta G^{\circ}_{\text{N-D,intraprotein}} + \Delta G^{\circ}_{\text{N-D,protein-amphi}} + \Delta G^{\circ}_{\text{N-D,amphi-amphi}}$).²³

In the denatured state ensemble (DSE), we expect the favorable solvation of individual TM segments by amphiphiles, the perturbation of favorable amphiphile–amphiphile interactions by the

TM segments, and the protein conformations in which stable native intraprotein interactions have not yet been realized. Protein folding is a free energy-decreasing process ($\Delta G_{N-D}^0 < 0$). Thus, the combined energetic contributions will yield a high free energy level of the DSE relative to that of the native state. To reach a compact native structure, the energetic contributions will be rebalanced, involving the partial release of the amphiphiles that have favorably solvated the TM segments ($\Delta G_{N-D, \text{protein-amphi}}^0 > 0$), the restoration of the favorable amphiphile–amphiphile interactions ($\Delta G_{N-D, \text{amphiphile-amphi}}^0 < 0$), and the formation of the stable intraprotein protein interactions ($\Delta G_{N-D, \text{intraprotein}}^0 < 0$). In this scenario, the properties of amphiphiles will affect the overall protein stability (ΔG_{N-D}^0) by their ability to solvate the protein ($\Delta G_{N-D, \text{protein-amphi}}^0$) and strength to interact with themselves ($\Delta G_{N-D, \text{amphi-amphi}}^0$). If a given amphiphile type weakly solvates the protein and strongly associates with one another, it can favorably drive the folding by facilitating the release of solvating amphiphiles and the restoration of amphiphile–amphiphile interactions. This scenario is consistent with the trend of the weak solvation of the protein by lipids and the strong interactions between themselves compared to those for detergents,⁵⁴ explaining the increased GlpG stability and the tight organization of the intraprotein interactions in GlpG (**Fig. 2.9**).

2.5.2. Role of lipid solvation in membrane protein folding.

In Stage II of membrane protein folding, the association of the TM helices occurs when the TM helix–helix interactions outweigh the solvation of the individual helices by lipid molecules.^{12,23} Then, a question arises whether the solvating lipids compete or facilitate the folding. The lipid-induced enhancement in stability and cooperativity enhancement that we discovered in this study may indicate an active role of lipids that facilitates the folding. As discussed above, however, the physical origin of this enhancement is likely to be the weak lipid solvation that allows the intraprotein interactions to outcompete the solvation. In that sense, the lipid bilayer would rather be regarded as an inert hydrophobic medium such that the information for folding is majorly encoded in the protein sequence. For both lipids and detergents, the solvation free energies from our MD simulation are weakly favorable ($\Delta G^{\circ}_{\text{Solv}} = -0.5$ to -0.8 kcal/atom) and comparable to the thermal energy (~ 0.6 kcal/mol). This weak solvation may serve as a minimal safety threshold for preventing nonspecific collapse of the DSE. That is, any local or nonlocal intraprotein interactions in the DSE that can overcome thermal fluctuation will effectively drive folding to the native state.

2.5.3. Implications in function and disease mechanisms involving membrane proteins.

The cooperativity enhancement upon structural perturbation may serve as a double-daggered sword in membrane protein function and quality control. The cell membranes serve as a permeability barrier (~ 50 Å thick) that separates the intra-(sub)cellular space from the extra-(sub)cellular environment. Thus, many types of membrane proteins such as receptors, transporters and enzymes function by transmitting chemical or physical stimuli on one side of the membrane across the bilayer (~ 50 Å thick). This process often involves conformational changes in the entire

length scale of the proteins (*e.g.*, side-chain reorientation, helix rotation, domain rearrangement, and coordinated protein motions during catalysis) upon ligand or substrate binding, the change in membrane potential, or protein–protein interactions.⁵⁵⁻⁶⁰ Our finding that lipid solvation strengthens a membrane protein’s cooperative interaction network implies that the lipid bilayer is an adequate medium to facilitate such large-scale conformational changes in a cooperative manner.

On the other hand, our observation that the mutation of the buried residues of GlpG has a large detrimental impact on the stability in the lipid environment imply that the structural integrity of the membrane protein interior can be vulnerable to missense mutations in the cell membranes. The cellular folding of some human membrane proteins is highly inefficient as highlighted by the fact that only 20 to 30% of newly synthesized nascent polypeptide chains reach their final destination membranes.^{61,62} These findings hint at the delicate balance of various molecular interactions to maintain their stability and a subtle difference in the energy barriers towards the native *vs* misfolded state.⁶¹ Indeed, a variety of missense mutations are found at multiple residue sites in a given membrane protein and many of these mutations are the cause of a disease.⁶³ Many of these disease mutations are known to be detrimental to stability rather than directly disrupt the active site residues, increasing the susceptibility of mutated proteins to degradation by protein quality control machinery. Indeed, most of disease mutations that cause cystic fibrosis are mapped onto the structure of cystic fibrosis transmembrane regulator in the regions that are not directly involved in anion transport or nucleotide binding.⁶⁴ In line with our finding of the vulnerability of the membrane protein interior, a strong bias has been observed for finding disease-causing mutations in the TM region relative to the extramembraneous region. The bias is even stronger for the buried residues relative to the lipid-exposed residues.⁶⁵

2.6. Concluding remarks and outlooks

Here, we provide a few new fundamental insights into the role of a lipid bilayer as a solvent shaping the folding energy landscape and cooperativity of membrane proteins. We suggest that lipid solvation on the protein is favorable but comparable to the thermal fluctuation energy as predicted from MD simulation. This weak lipid solvation would allow the intraprotein interactions to outcompete the lipid solvation interactions and thus the information for folding is largely encoded in the amino acid sequence, not substantially affected by the lipids. Although we employ a simple single-component lipid bilayer system, the real cell membranes have a highly heterogeneous lipid composition that displays complex chemical and physical properties. Recent structural, thermodynamic, and mass spectroscopic studies indicate that some lipids are more actively involved in the conformation and stability of membrane proteins rather than in simple solvation, that is, bridging the tertiary or quaternary interactions, or inducing the lipid deformation by the hydrophobic mismatch between the bilayer and the protein.^{33,66-69} Our study will provide an additional contribution to the understanding of such multi-faceted protein–lipid interactions that are fundamental to folding, function and behavior of membrane proteins.

APPENDICES

Table 2.1. The change in thermodynamic stability ($\Delta\Delta G^{\circ}_{WT-Mut}$), activity (relative to wild type) and residue burial (f_{ASA} : fraction of buried side-chain area) of GlpG variants in DMPC/CHAPS bicelles

Secondary structure	Variant	f_{ASA}	N-subdomain (95 _N 172 _M)		C-subdomain (172 _M 267 _C)		$\Delta\Delta\Delta G^{\circ}$
			$\Delta\Delta G^{\circ}_{WT-Mut}$	Rel. Activity	$\Delta\Delta G^{\circ}_{WT-Mut}$	Rel. Activity	
TM1	M100A	0.14	1.4 ± 0.2	0.86 ± 0.09	1.4 ± 0.2	0.74 ± 0.08	0.0 ± 0.2
	C104A	0	0.5 ± 0.3	0.84 ± 0.09	0.1 ± 0.4	0.58 ± 0.06	0.4 ± 0.4
L1	V119A	0.05	1.6 ± 0.2	0.59 ± 0.07	1.1 ± 0.1	0.70 ± 0.07	0.5 ± 0.2
	L123A	0.23	1.4 ± 0.1	0.57 ± 0.06	0.7 ± 0.2	0.65 ± 0.07	0.7 ± 0.2
	W125A	0.09	3.9 ± 0.1	0.17 ± 0.02	3.5 ± 0.1	0.12 ± 0.01	0.5 ± 0.1
	K132A	0.56	-0.1 ± 0.1	0.79 ± 0.09	0.0 ± 0.1	0.72 ± 0.08	-0.1 ± 0.2
	F133A	0.80	0.6 ± 0.1	0.83 ± 0.10	0.8 ± 0.2	0.82 ± 0.09	-0.2 ± 0.2
	F135A	0.73	0.4 ± 0.1	1.02 ± 0.12	0.5 ± 0.1	0.75 ± 0.09	-0.1 ± 0.2
	W136A	0.46	1.0 ± 0.3	0.22 ± 0.02	1.3 ± 0.1	0.30 ± 0.03	-0.3 ± 0.3
	R137A	0.04	4.7 ± 0.1	0.01 ± 0.01	4.3 ± 0.2	0.01 ± 0.01	0.4 ± 0.2
	Y138F	0.25	0.9 ± 0.1	0.75 ± 0.08	0.8 ± 0.2	0.56 ± 0.07	0.1 ± 0.2
	F139A	0.38	0.7 ± 0.2	0.63 ± 0.07	1.3 ± 0.1	0.45 ± 0.05	0.5 ± 0.2
	T140A	0.20	1.4 ± 0.1	0.59 ± 0.07	0.9 ± 0.2	0.62 ± 0.07	0.5 ± 0.2
	L143A	0.25	1.3 ± 0.1	0.78 ± 0.09	1.5 ± 0.2	0.68 ± 0.08	-0.2 ± 0.2
TM2	H150A	0.01	0.3 ± 0.2	0.10 ± 0.01	0.5 ± 0.2	0.03 ± 0.00	-0.2 ± 0.3
	N154A	0	0.5 ± 0.1	0.12 ± 0.01	0.6 ± 0.1	0.04 ± 0.09	-0.1 ± 0.1
	L155A	0.15	1.4 ± 0.1	0.64 ± 0.07	0.9 ± 0.1	0.79 ± 0.08	0.5 ± 0.2
	W158F	0	0.3 ± 0.1	0.84 ± 0.10	0.3 ± 0.1	0.82 ± 0.09	0.0 ± 0.2
	L161A	0	1.2 ± 0.1	0.36 ± 0.04	1.5 ± 0.3	0.26 ± 0.03	-0.3 ± 0.3
TM3	L174A	0	4.5 ± 0.2	0.49 ± 0.05	4.8 ± 0.1	0.17 ± 0.02	-0.3 ± 0.2
	T178A	0.11	0.7 ± 0.1	0.77 ± 0.09	0.9 ± 0.2	0.66 ± 0.07	-0.2 ± 0.2
	S181A	0	0.1 ± 0.2	0.92 ± 0.11	0.5 ± 0.1	0.85 ± 0.09	-0.4 ± 0.2
	F197A	0.01	0.6 ± 0.2	0.02 ± 0.00	0.5 ± 0.2	0.00 ± 0.00	0.1 ± 0.2
TM4	S201T	0	0.0 ± 0.2	0.05 ± 0.01	0.0 ± 0.2	0.01 ± 0.00	0.0 ± 0.3
	A206G	0	0.3 ± 0.2	1.07 ± 0.12	0.9 ± 0.1	0.92 ± 0.10	-0.6 ± 0.2
	L207A	0	5.0 ± 0.2	0.12 ± 0.01	4.7 ± 0.1	0.16 ± 0.02	0.3 ± 0.2
	Y210F	0	0.5 ± 0.2	0.86 ± 0.10	0.9 ± 0.1	0.61 ± 0.07	-0.4 ± 0.2
	K214A	0.10	0.2 ± 0.2	0.70 ± 0.08	0.7 ± 0.2	0.52 ± 0.05	-0.5 ± 0.3
TM5	I223A	0.00	0.1 ± 0.2	0.56 ± 0.06	0.6 ± 0.2	0.30 ± 0.03	-0.6 ± 0.2
	L225A	0.03	0.2 ± 0.1	0.42 ± 0.05	0.4 ± 0.1	0.26 ± 0.03	-0.3 ± 0.1
	Q226A	0.57	0.0 ± 0.1	0.95 ± 0.11	-0.1 ± 0.2	0.67 ± 0.07	0.1 ± 0.2
TM6	M249A	0.01	0.3 ± 0.1	0.97 ± 0.11	0.4 ± 0.2	0.79 ± 0.09	-0.1 ± 0.2
	A253V	0	0.0 ± 0.1	0.03 ± 0.01	0.3 ± 0.2	0.03 ± 0.00	-0.3 ± 0.2
	H254A	0	0.7 ± 0.1	0.05 ± 0.01	-0.2 ± 0.2	0.01 ± 0.00	0.9 ± 0.2
	G261A	0	4.9 ± 0.1	0.05 ± 0.01	4.0 ± 0.1	-0.02 ± 0.01	1.0 ± 0.2
	A265V	0	1.6 ± 0.3	0.37 ± 0.04	1.4 ± 0.2	0.29 ± 0.03	0.2 ± 0.4
	D268A	0.15	1.8 ± 0.2	0.54 ± 0.06	1.5 ± 0.1	0.48 ± 0.05	0.3 ± 0.2

Table 2.2. The change in thermodynamic stability ($\Delta\Delta G^{\circ}_{WT-Mut}$), activity (relative to wild type) and residue burial (f_{ASA} : fraction of buried side-chain area) of GlpG variants in DDM micelles

Secondary structure	Variant	f_{ASA}	N-subdomain (95 _N 172 _M)		C-subdomain (172 _M 267 _C)		$\Delta\Delta G^{\circ}$
			$\Delta\Delta G^{\circ}_{WT-Mut}$	Rel. Activity	$\Delta\Delta G^{\circ}_{WT-Mut}$	Rel. Activity	
TM1	M100A	0.14	3.0 ± 0.3	0.55 ± 0.06	2.2 ± 0.3	0.64 ± 0.05	0.5 ± 0.4
	C104A	0	1.2 ± 0.3	0.69 ± 0.04	0.9 ± 0.1	0.70 ± 0.05	0.3 ± 0.3
L1	V119A	0.05	2.0 ± 0.2	0.57 ± 0.06	0.9 ± 0.1	0.76 ± 0.05	1.1 ± 0.2
	L123A	0.23	2.1 ± 0.2	0.55 ± 0.06	0.9 ± 0.1	0.68 ± 0.05	1.2 ± 0.2
	W125A	0.09	2.8 ± 0.3	0.04 ± 0.10	1.7 ± 0.2	0.00 ± 0.01	1.1 ± 0.3
	K132A	0.56	0.2 ± 0.3	0.71 ± 0.06	0.4 ± 0.1	0.52 ± 0.05	-0.2 ± 0.3
	F133A	0.80	1.3 ± 0.3	0.84 ± 0.06	0.5 ± 0.2	0.87 ± 0.05	0.8 ± 0.3
	F135A	0.73	0.4 ± 0.2	0.93 ± 0.06	0.2 ± 0.1	0.69 ± 0.05	0.2 ± 0.3
	W136A	0.46	2.7 ± 0.2	0.00 ± 0.02	1.7 ± 0.1	0.00 ± 0.03	1.0 ± 0.2
	R137A	0.04	4.1 ± 0.2	0.01 ± 0.01	2.8 ± 0.1	0.01 ± 0.01	1.3 ± 0.2
	Y138F	0.25	1.8 ± 0.2	0.95 ± 0.06	0.6 ± 0.1	0.93 ± 0.05	1.2 ± 0.2
	F139A	0.38	2.0 ± 0.2	0.47 ± 0.06	1.0 ± 0.1	0.47 ± 0.05	1.0 ± 0.2
	T140A	0.20	1.6 ± 0.2	0.85 ± 0.06	0.7 ± 0.1	0.60 ± 0.03	0.9 ± 0.2
	L143A	0.25	2.3 ± 0.2	0.76 ± 0.06	1.4 ± 0.1	0.65 ± 0.05	0.9 ± 0.2
TM2	H150A	0.01	0.0 ± 0.3	0.05 ± 0.08	0.3 ± 0.2	0.02 ± 0.13	-0.3 ± 0.3
	N154A	0	1.2 ± 0.2	0.01 ± 0.04	1.2 ± 0.3	0.01 ± 0.02	0.0 ± 0.4
	L155A	0.15	2.2 ± 0.2	0.75 ± 0.05	1.6 ± 0.2	0.60 ± 0.03	0.6 ± 0.3
	W158F	0	1.0 ± 0.2	0.92 ± 0.06	0.1 ± 0.1	0.85 ± 0.05	0.9 ± 0.2
	L161A	0	2.0 ± 0.3	0.16 ± 0.06	2.7 ± 0.3	0.10 ± 0.06	-0.7 ± 0.4
TM3	L174A	0	3.7 ± 0.2	0.35 ± 0.06	3.3 ± 0.1	0.07 ± 0.07	0.4 ± 0.2
	T178A	0.11	0.6 ± 0.2	0.77 ± 0.06	0.3 ± 0.1	0.66 ± 0.07	0.3 ± 0.2
	S181A	0	-0.6 ± 0.2	1.03 ± 0.06	0.6 ± 0.1	1.00 ± 0.05	-1.2 ± 0.2
	F197A	0.01	1.7 ± 0.2	0.01 ± 0.03	0.6 ± 0.1	0.00 ± 0.07	1.1 ± 0.2
TM4	S201T	0	0.4 ± 0.2	0.02 ± 0.01	0.8 ± 0.2	0.00 ± 0.03	-0.4 ± 0.3
	A206G	0	0.4 ± 0.2	0.09 ± 0.09	0.6 ± 0.1	0.09 ± 0.06	-0.2 ± 0.2
	L207A	0	4.1 ± 0.3	0.12 ± 0.01	2.7 ± 0.1	0.16 ± 0.02	1.4 ± 0.3
	Y210F	0	1.9 ± 0.2	0.50 ± 0.07	1.2 ± 0.1	0.66 ± 0.05	0.8 ± 0.2
	K214A	0.10	0.9 ± 0.2	0.41 ± 0.06	0.6 ± 0.1	0.43 ± 0.05	0.3 ± 0.3
TM5	I223A	0.00	1.0 ± 0.3	0.24 ± 0.06	0.5 ± 0.1	0.23 ± 0.11	0.5 ± 0.3
	L225A	0.03	-0.7 ± 0.2	0.27 ± 0.07	1.0 ± 0.1	0.10 ± 0.06	-1.6 ± 0.2
	Q226A	0.57	0.2 ± 0.2	0.82 ± 0.06	0.8 ± 0.2	0.51 ± 0.05	-0.6 ± 0.3
TM6	M249A	0.01	0.3 ± 0.2	0.59 ± 0.06	0.5 ± 0.2	0.85 ± 0.05	-0.2 ± 0.3
	A253V	0	1.5 ± 0.2	0.06 ± 0.01	0.9 ± 0.1	0.00 ± 0.06	0.6 ± 0.3
	H254A	0	1.5 ± 0.2	0.05 ± 0.01	-0.3 ± 0.1	0.01 ± 0.05	1.8 ± 0.3
	G261A	0	4.0 ± 0.2	0.05 ± 0.01	2.7 ± 0.1	-0.01 ± 0.06	1.3 ± 0.2
	A265V	0	2.3 ± 0.2	0.30 ± 0.06	1.3 ± 0.1	0.13 ± 0.05	1.0 ± 0.2
	D268A	0.15	2.4 ± 0.2	0.44 ± 0.07	1.3 ± 0.1	0.28 ± 0.05	1.1 ± 0.2

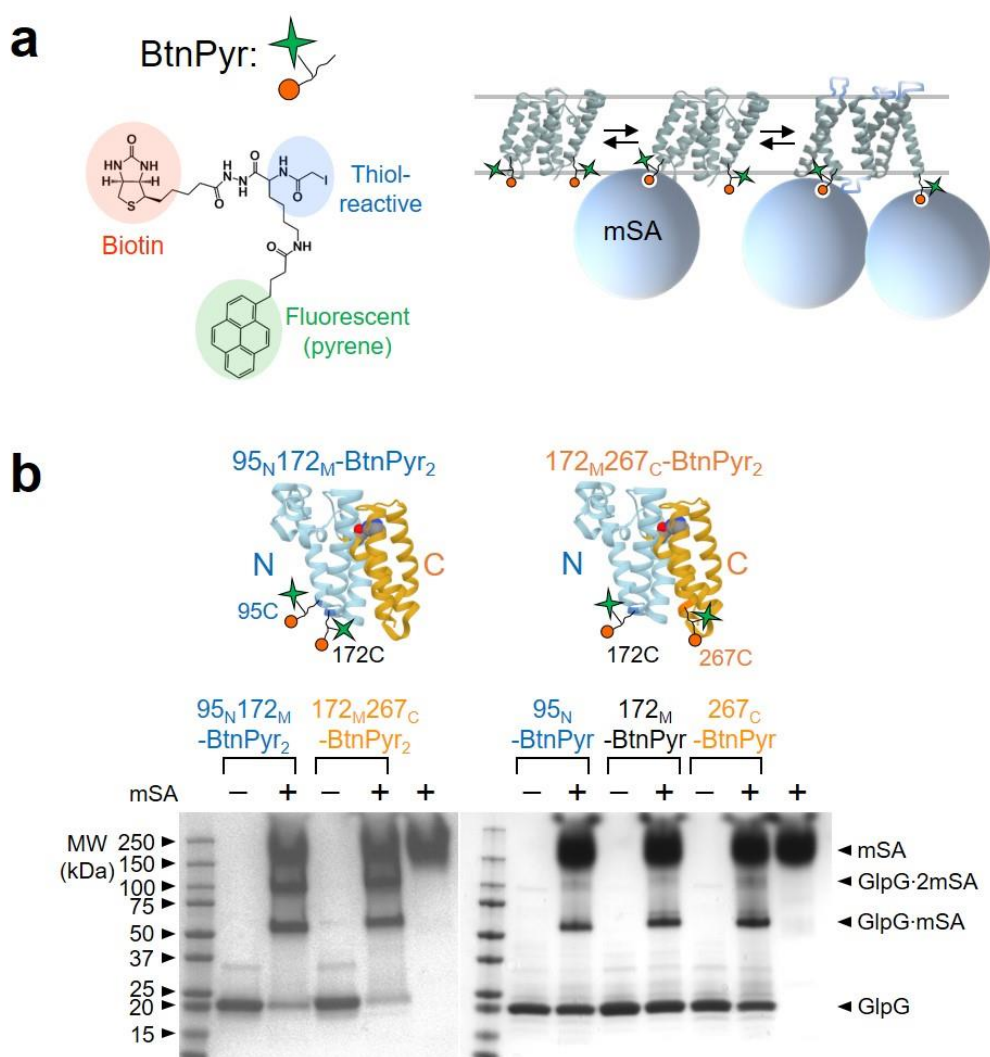


Figure 2.10. Biotin labeling of double-cysteine variants of GlpG.

(a) (Left) Structure of the thiol-reactive biotin derivative with a fluorescent pyrenyl group (BtnPyr). **(Right)** Addition of excess monovalent streptavidin (mSA) leads to the formation of double-biotin variants of GlpG doubly-bound with mSA.

(b) Identifying binding of mSA-WT with double-biotin variants of GlpG ($95_N172_M\text{-BtnPyr}_2$ or $172_M267_C\text{-BtnPyr}_2$) measured by SDS-PAGE without sample heating. Unbound (GlpG), single mSA-bound (GlpG·mSA) and double mSA-bound GlpG (GlpG·2mSA) are separated on the gel stained by Coomassie 6G. For $95_N172_M\text{-BtnPyr}_2$ and $172_M267_C\text{-BtnPyr}_2$, GlpG·2mSA is major with minor unbound and GlpG·mSA (labeling efficiency = $1.4\sim 1.6$ [BtnPyr]/[GlpG]). For single-biotin variants (95_N-BtnPyr , 172_M-BtnPyr and 267_C-BtnPyr), GlpG·mSA is major with minor unlabeled and GlpG·2mSA (over-labeled). mSA-WT (52 kDa) migrates as a larger species (>150 kDa) probably due to an abnormal number of SDS molecules bound to a tetrameric mSA molecule

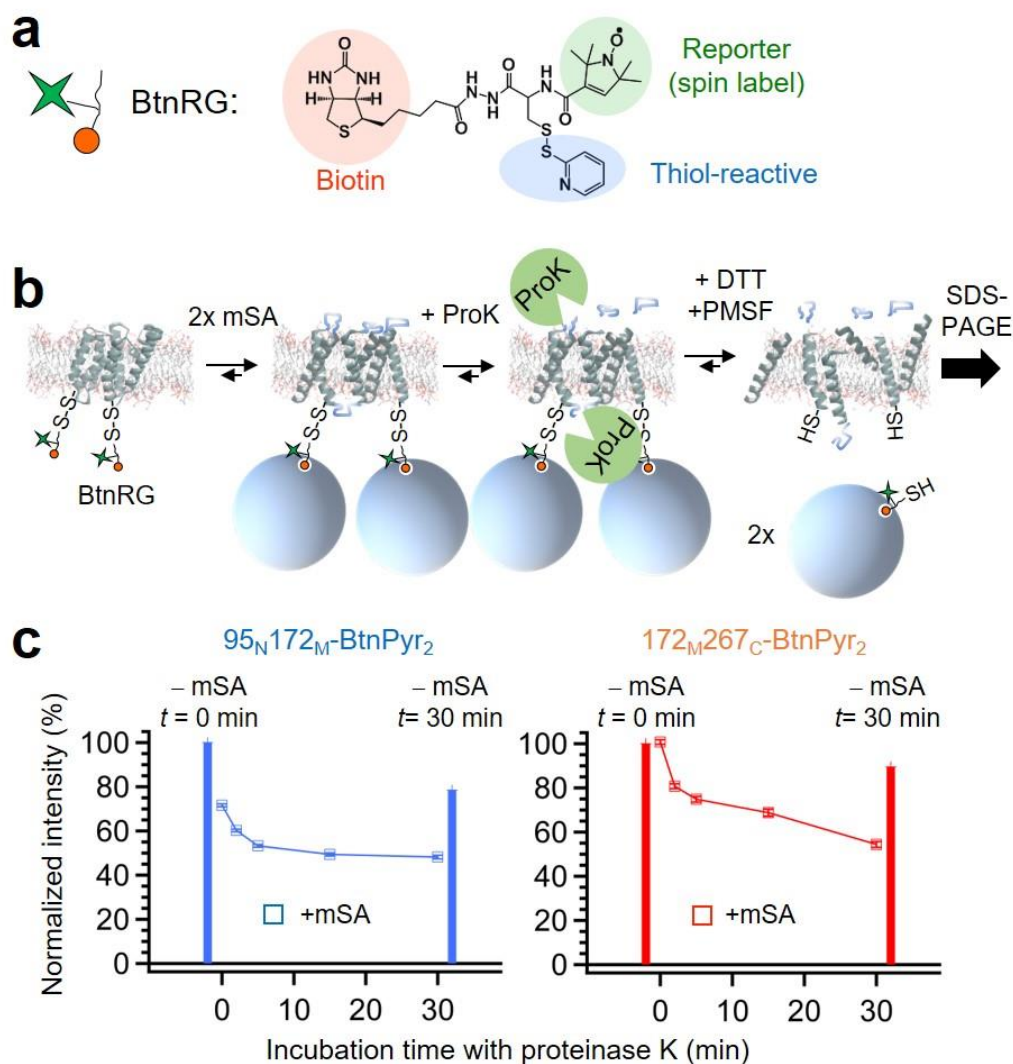


Figure 2.11. Steric trapping-induced GlpG denaturation monitored by proteinase K (ProK) digestion.

(a) Structure of reversible thiol-reactive biotin derivative (BtnRG-TP). Thiopyridine group reacts with the thiol group in a cysteine residue to form a disulfide linkage.

(b) Strategy to monitor sterically denatured GlpG by proteolysis. GlpG labeled with BtnRG is denatured by steric trapping. Denatured GlpG reacts with ProK for various incubation time. The reducing agent DTT is added to break the linkage between GlpG and the biotin label bound with mSA. The proteolysis products are analyzed by SDS-PAGE. **(c)** Proteolysis of sterically denatured GlpG as a function of incubation time with ProK. The band intensities on the SDS-PAGE gel which correspond to the molecular weight of GlpG were analyzed by the ImageJ program. As controls, the intensities of native GlpG in the absence of mSA at time 0 and 30 min are shown, respectively. The incomplete digestion in the presence of excess mSA is due to the incomplete double-biotin labeling of GlpG.

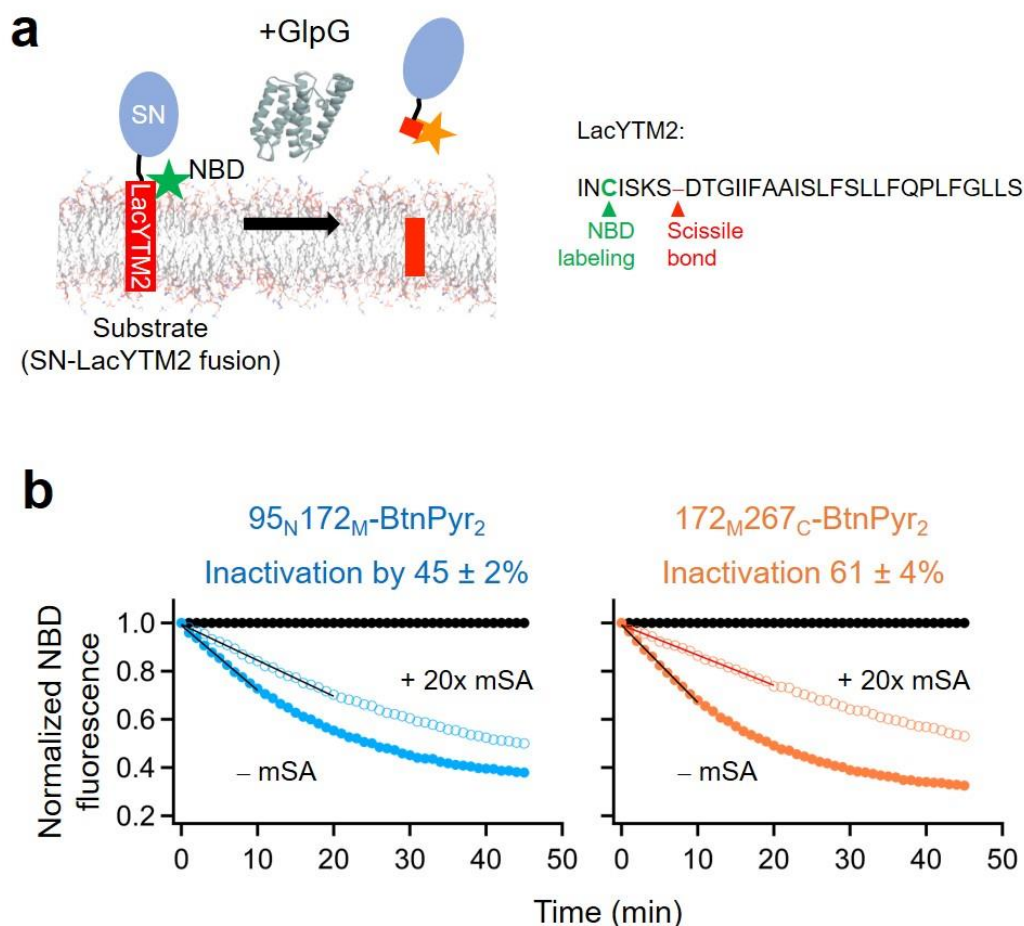


Figure 2.12. Activity assay to measure GlpG denaturation induced by steric trapping.

(a) (Left) The principle of measuring the sequence-specific proteolysis of the transmembrane (TM) substrate SN-LYTM2 by GlpG (SN: staphylococcal nuclease fusion; LYTM2: the second TM segment of *E. coli* lactose permease). The cleavage of LYTM2 induces the transfer of NBD labeled on the five residue upstream of the scissile bond from the hydrophobic bicellar phase to the aqueous phase **(Right)**. The transfer induces the decrease in NBD fluorescence, which enables the continuous detection of the cleavage reaction.

(b) Inactivation of double-biotin variants of GlpG (95_N172_M-BtnPyr₂ and 172_M267_C-BtnPyr₂) induced by steric trapping (+20xmSA). The activity is defined as the initial slope of time-dependent decrease of NBD fluorescence. The inactivation does not occur when mSA is added to individual single-biotin variants (95_N-BtnPyr, 172_M-BtnPyr and 267_C-BtnPyr).

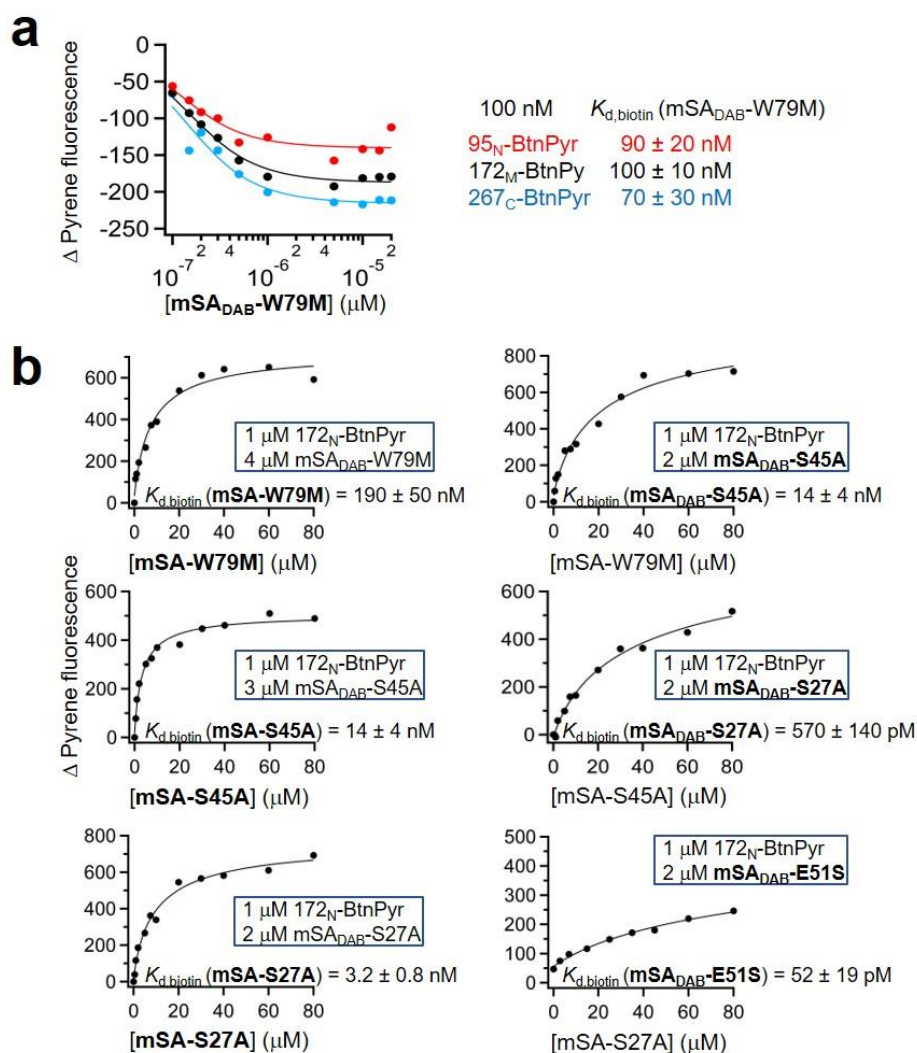


Figure 2.13. Determination of the binding affinities ($K_{d, \text{biotin}}$) of mSA variants to biotin labels (BtnPyr) on GlpG using Förster resonance energy transfer in bicelles.

(a) Binding isotherms between a weaker binding variant mSA-W79M labeled with a dabcyI quencher (mSA_{DAB}-W79M) and three single-biotin variants of GlpG. The data were fitted to Eq. 4 (Methods) to determine $K_{d, \text{biotin}}$. $K_{d, \text{biotin}}$'s of mSA to the three biotinylated sites are similar.

(b) Competition assay to determine $K_{d, \text{biotin}}$ between high-affinity mSA variants and a single-biotin variant label at the G172C site on GlpG. **Top left:** The complex of 172_N-BtnPyr and mSA_{DAB}-W79M (with known $K_{d, \text{biotin}}$) was incubated at an increasing concentration of mSA-W79M (no dabcyI label, unknown $K_{d, \text{biotin}}$). The replacement of mSA_{DAB}-W79M by mSA-W79M in the complex leads to an increase in pyrene fluorescence, which was fitted to Eq. 5 (Methods) to determine the $K_{d, \text{biotin}}$ of mSA-W79M. **Top right:** The $K_{d, \text{biotin}}$ of the next higher affinity variant mSA_{DAB}-S45A was determined using mSA-W79M with known $K_{d, \text{biotin}}$ obtained from the preceding plot. Using the same strategy, $K_{d, \text{biotin}}$'s of mSA_{DAB}-S27A and -E51S were determined consecutively.

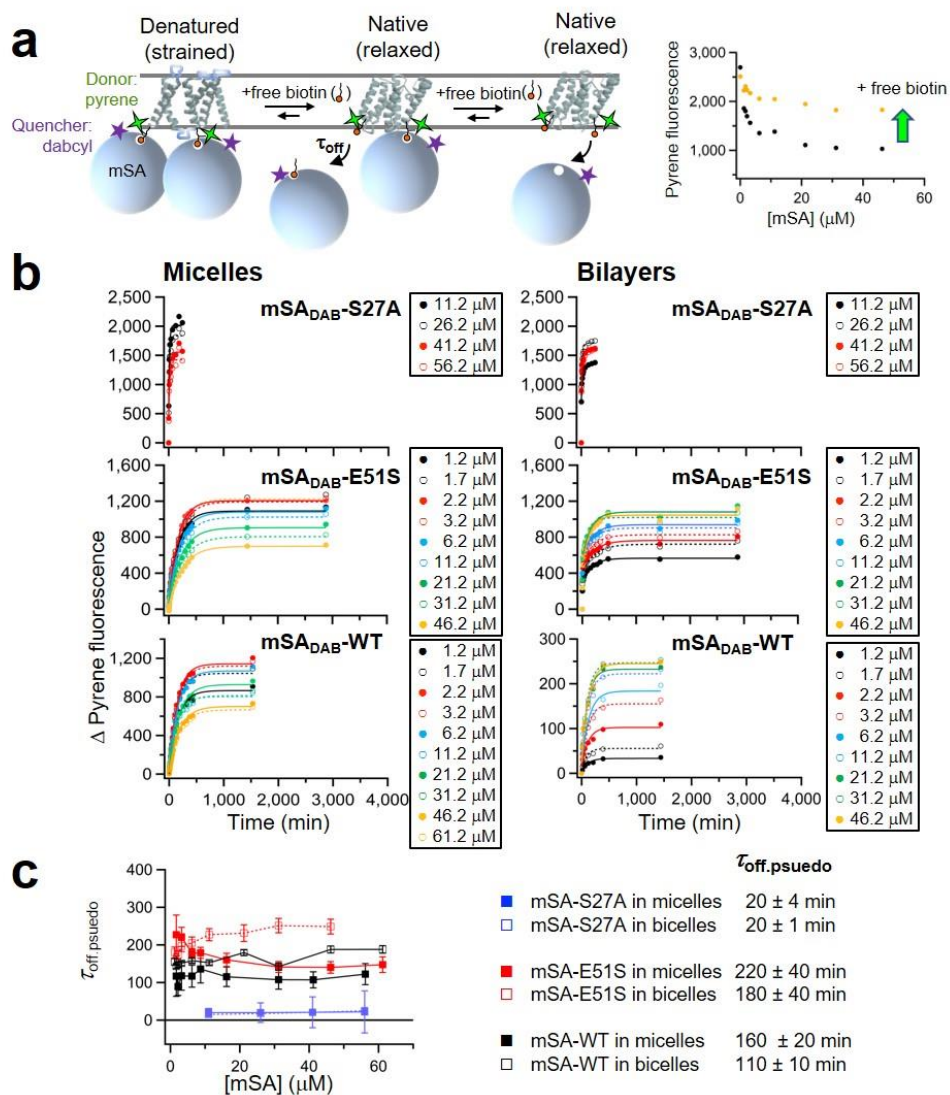


Figure 2.14. Determining the dissociation time constant ($\tau_{\text{off.psuedo}}$) of mSA variant from sterically denatured GlpG in bicelles.

(a) The principle of measuring τ_{off} . A double-biotin variant of GlpG, 172_M267_C-BtnPyr₂ (1 μM), is denatured at an increasing concentration of mSA_{DAB} variants. In the initial state, denatured GlpG is strained by the steric hindrance between bound mSA_{DAB} and pyrene fluorescence from the BtnPyr labels on GlpG are quenched by the quencher-labeled bound mSA_{DAB}. Upon addition of excess biotin (2 mM), bound mSA_{DAB} dissociates and pyrene fluorescence increases by dequenching.

(b) Dissociation kinetics of mSA_{DAB} variants bound to 172_M267_C-BtnPyr₂ upon addition of excess free biotin. The data was fitted to the single exponential function under the assumption of the pseudo-first order reaction.

(c) Pseudo-first order dissociation time constants ($\tau_{\text{off.psuedo}}$) of mSA_{DAB} variants are independent of the concentration of mSA_{DAB}.

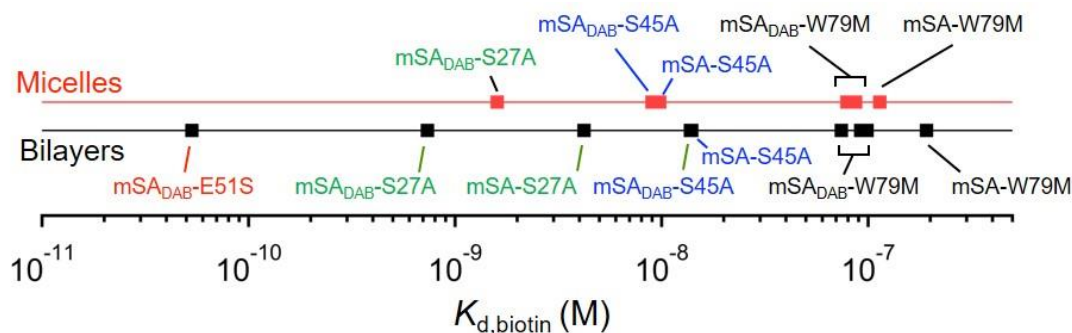


Figure 2.15. Comparison of the intrinsic binding affinities ($K_{d, \text{biotin}}$) of mSA variants to biotin labels (BtnPyr) in micelles and bilayers.

The $K_{d, \text{biotin}}$ values for mSA-W79M, mSA_{DAB}-W79M (dabcyl-labeled), mSA-S45A, mSA_{DAB}-S45A, mSA-S27A and mSA_{DAB}-S27A were determined previously in micelles (Guo et al. 2016 Nat Chem Biol 12, 353). Here, the $K_{d, \text{biotin}}$'s of mSA_{DAB}-W79M, mSA_{DAB}-W79M, mSA-S45A, mSA_{DAB}-S45A, mSA-S27A and mSA_{DAB}-S27A and mSA_{DAB}-E51S were determined in 2.0% (w/v) DMPC:CHAPS bicelles (molar ratio = 1.5:1). Overall, the maximal discrepancy between the $K_{d, \text{biotin}}$ values for the same variants in micelles and bilayers is less than two folds, which translates into the changes of ~0.4 kcal/mol in $\Delta G^{\circ}_{\text{N-D}}$.

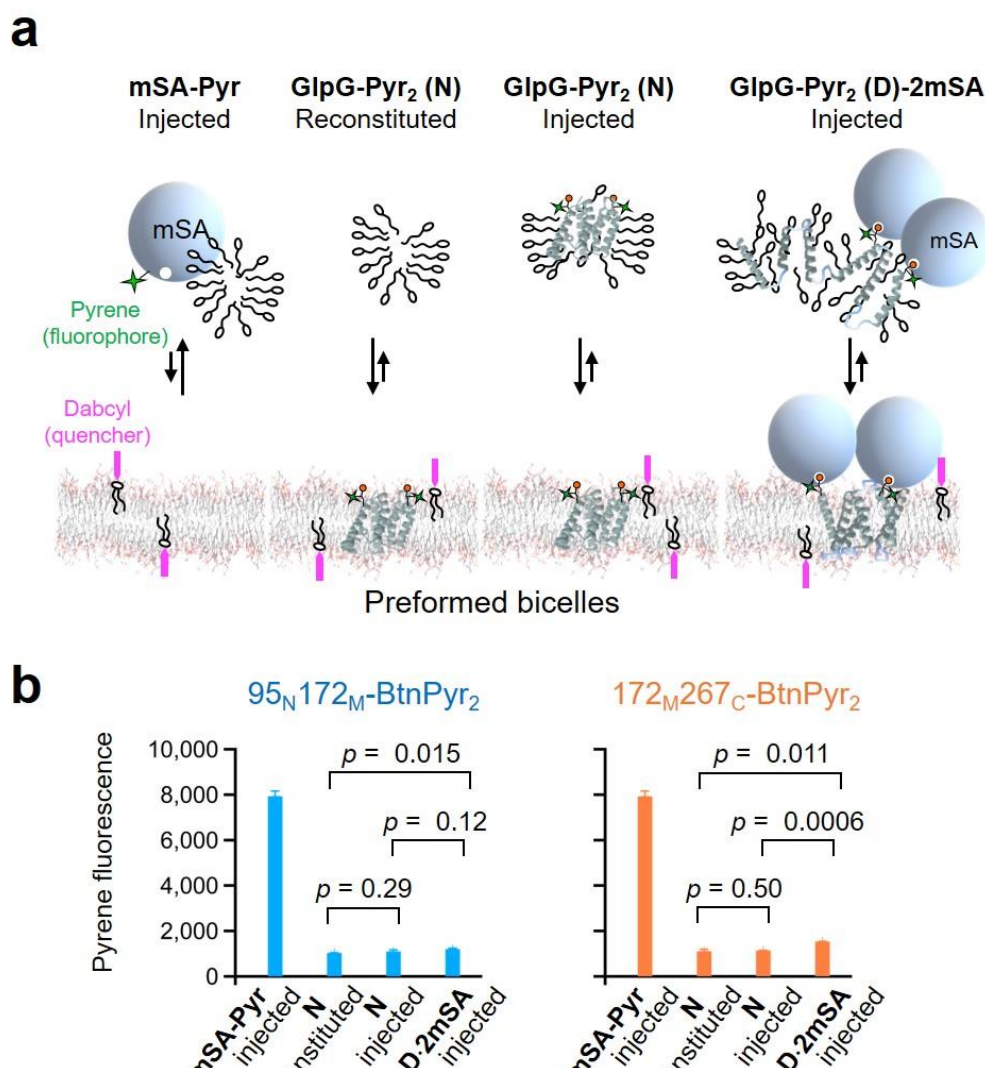


Figure 2.16. Incorporation of native and sterically denatured GlpG into preformed bicelles.

(a) Fluorescence quenching assay to measure the transfer of native (N: double-cysteine variants labeled with fluorescent BtnPyr, GlpG-BtnPyr₂, that is, 95_N172_M-BtnPyr₂ or 172_M267_C-BtnPyr₂) and sterically-denatured (D-2mSA) GlpG from the micellar to the bicellar phase by direct injection. The bicellar phase contains dabcyl (quencher)-labeled lipid (molar ratio, DMPC:dabcyl-DOPE = 1000:1). Pyrene-labeled mSA (mSA-Pyr), which is highly soluble in water, was used as a negative control (*i.e.*, no incorporation). Native GlpG, which was first reconstituted in DMPC/DMPG liposomes and then solubilized by CHAPS to form bicelles was used as a positive control (*i.e.*, incorporation).

(b) Binding of GlpG-BtnPyr₂ to bicelles induced quenching of pyrene fluorescence. Error bars denote $\pm$ SEM. ($n = 3$). P values were obtained using Student's t -test.

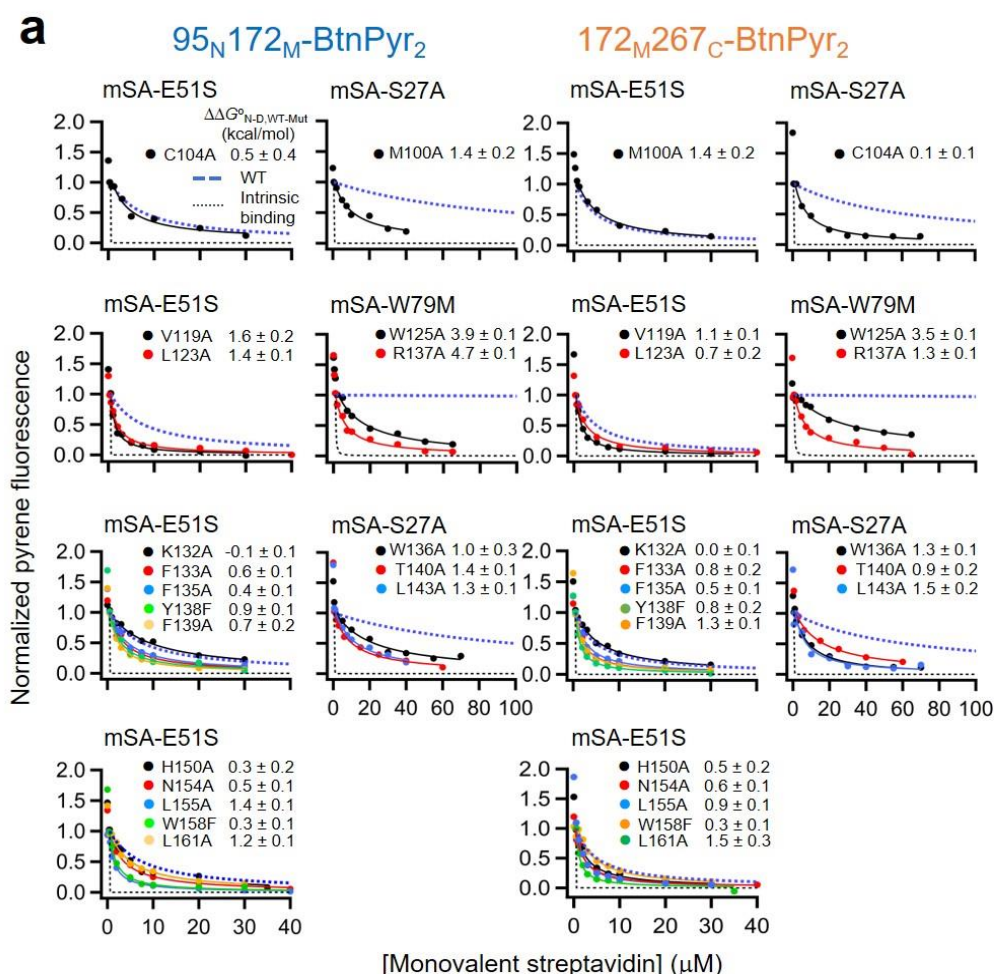
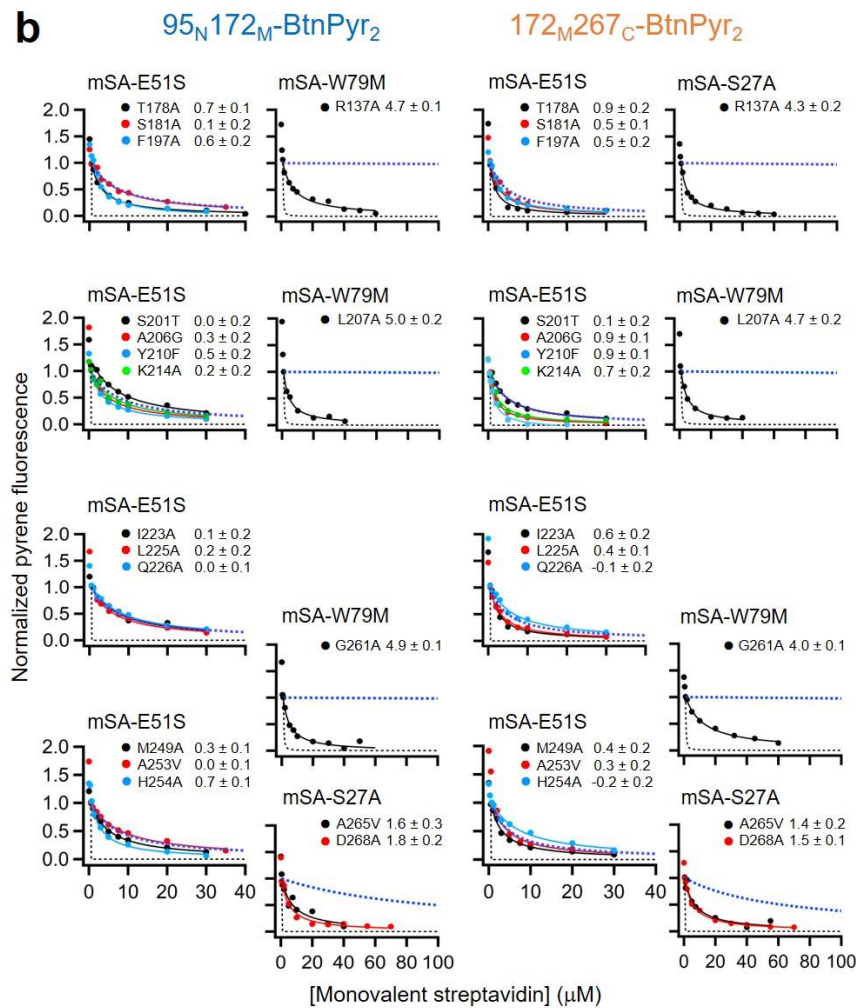


Figure 2.17. Binding isotherms between double-biotin variants of GlpG and monovalent streptavidin (mSA) to determine the thermodynamic stability of WT and variants using steric trapping in DMPC/CHAPS bicelles.

Binding was measured by quenching of pyrene fluorescence from BtnPyr labels, which was induced by mSA labeled with dabcyI quencher (mSA_{DAB}). Double-cysteine variants of GlpG were labeled with thiol-reactive fluorescent biotin derivative BtnPyr-IA at 95C/172C ($95_N172_M\text{-BtnPyr}_2$) or 172C/267C ($172_M267_C\text{-BtnPyr}_2$). When a mSA variant with a weaker biotin affinity (mSA_{DAB}-W79M, mSA_{DAB}-S27A or mSA-E51S) is used, the first mSA binds to either biotin label with intrinsic binding affinity (black dashed lines). Binding of the second mSA is attenuated depending on the stability of GlpG. $\Delta\Delta G^{\circ}_{N-D}$'s were obtained by fitting the attenuated second binding to Eq.'s 1–2 (Methods). In each plot, the fluorescence intensity was normalized to the intensity change of the second binding phase and the difference stability between WT and mutants ($\Delta\Delta G^{\circ}_{N-D, WT-Mut} = \Delta G^{\circ}_{N-D, WT} - \Delta G^{\circ}_{N-D, Mut}$) are shown. The more attenuated second binding indicates the higher stability (*i.e.*, larger ΔG°_{N-D}).

a) Binding isotherms for the variants with mutations on the segments TM1, L1 and TM2 of GlpG.



Continued-Figure 2.17. Binding isotherms between double-biotin variants of GlpG and monovalent streptavidin (mSA) to determine the thermodynamic stability of WT and variants using steric trapping in DMPC/CHAPS bicelles.

b: Binding isotherms for the variants with mutations on the segments TM3, TM4, TM5 and TM6 of GlpG.

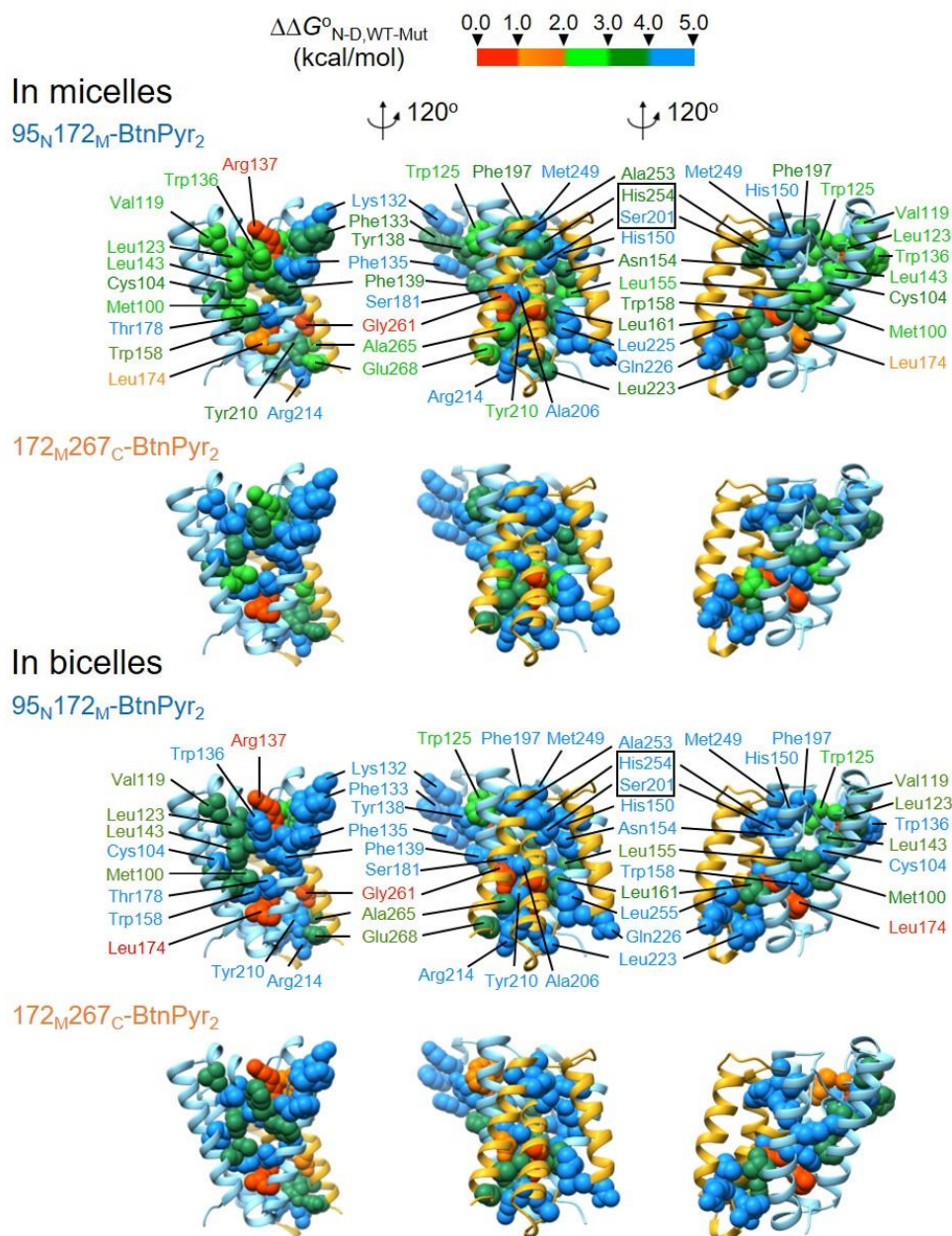


Figure 2.18. Mapping of mutation-induced stability changes onto GlpG structure.

The degrees of stability changes ($\Delta\Delta G^{\circ}_{N-D,WT-Mut}$) measured at the N and C subdomains (95_N172_M-BtnPyr₂ and 172_M267_C-BtnPyr₂, respectively) in micelles and bilayers were color-coded (**Top**) as a heat map.

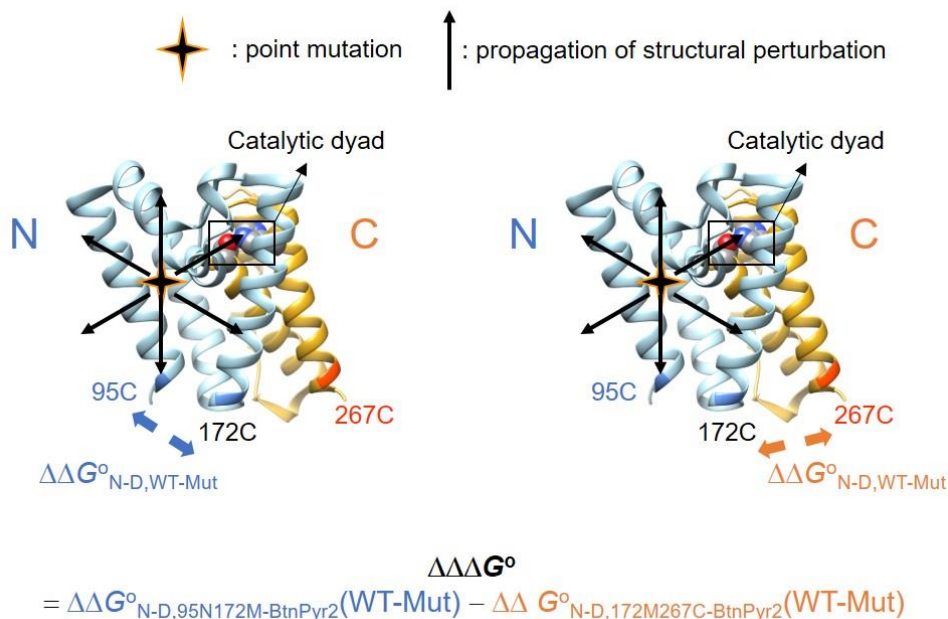


Figure 2.19. Schematic description of the cooperativity profiling method.

The method (Guo et al. 2016 *Nat Chem Biol* 12, 353) quantifies the degree of spatial propagation of the structural perturbation induced by a point-mutation using steric trapping. This is achieved by measuring the stability changes upon the mutation with the biotin pairs located in the two different regions. The differential effect of the same mutation on the stability of N- or C-subdomain ($\Delta\Delta\Delta G^\circ$) is defined as:

$$\begin{aligned}
 \Delta\Delta\Delta G^\circ &= [(\Delta G^\circ_{\text{N-D}, 95\text{N}172\text{M-BtnPyr}_2(\text{WT})} - \Delta G^\circ_{\text{N-D}, 95\text{N}172\text{M-BtnPyr}_2(\text{Mut})}) \\
 &\quad - (\Delta G^\circ_{\text{N-D}, 172\text{M}267\text{C-BtnPyr}_2(\text{WT})} - \Delta G^\circ_{\text{N-D}, 172\text{M}267\text{C-BtnPyr}_2(\text{Mut})})] \\
 &= \Delta\Delta G^\circ_{\text{N-D}, 95\text{N}172\text{M-BtnPyr}_2(\text{WT-Mut})} - \Delta\Delta G^\circ_{\text{N-D}, 172\text{M}267\text{C-BtnPyr}_2(\text{WT-Mut})}
 \end{aligned}$$

$\Delta\Delta G^\circ_{\text{N-D}, 95\text{N}172\text{M-BtnPyr}_2(\text{WT-Mut})}$ and $\Delta\Delta G^\circ_{\text{N-D}, 172\text{M}267\text{C-BtnPyr}_2(\text{WT-Mut})}$ designate the stability changes caused by the same mutation in the backgrounds of 95N172M-BtnPyr₂ and 172M267C-BtnPyr₂, respectively. Thus, $\Delta\Delta\Delta G^\circ$ represents the difference in the stability changes that are probed with two different biotin pairs upon the same mutation (*ibid*).

If a mutation causes a similar degree of destabilization for both double-biotin variants with a difference within thermal fluctuation energy ($|\Delta\Delta\Delta G^\circ| \leq RT = 0.6 \text{ kcal/mol}$; R : gas constant; $T = 298 \text{ K}$), the mutated site engages in a “cooperative” interaction. That is, the perturbation by the mutation similarly propagates to both subdomains. Among the cases where $|\Delta\Delta\Delta G^\circ| > RT$, if a mutation preferentially destabilizes the subdomain containing it, the perturbed interactions are “localized” within that subdomain. If mutation of a residue, which makes its side-chain contacts only with the subdomain containing it, preferentially destabilizes the other subdomain, we classified the perturbation as “over-propagated” (*ibid*).

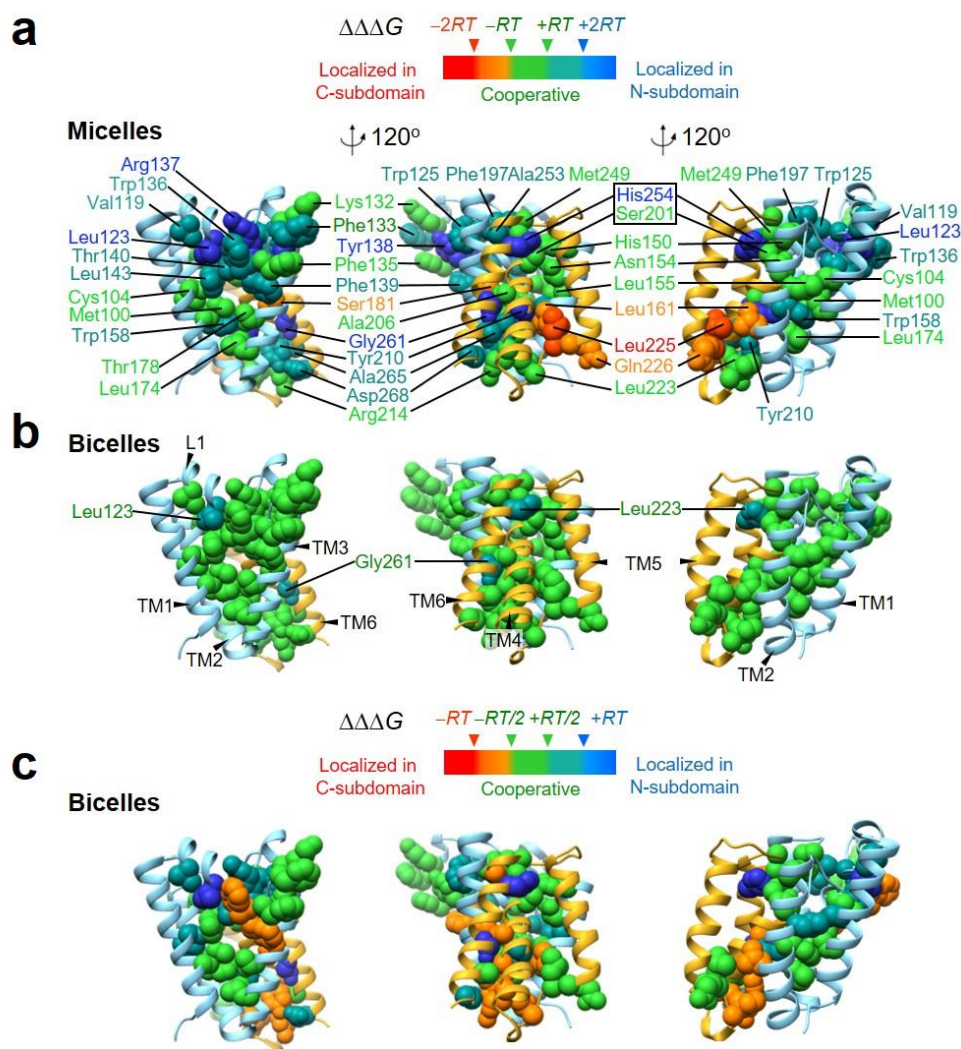


Figure 2.20. The features of cooperativity profiles in bicelles still preserve those in micelles, but to a less extent.

(a–b) Comparison of cooperativity profiles between micells (**Top**) and bilayers (**Bottom**) on the basis of the cut-off values, $-2RT$, $-RT$, $+RT$, and $+2RT$ (*i.e.*, the RT scale).

(c) Cooperativity profiles in bilayers on the basis of smaller cut-off values, $-RT$, $-1/2 RT$, $+1/2 RT$, and $+RT$ (*i.e.*, the $1/2 RT$ scale). Overall, except for several residues which display distinctively different profiles (Phe135, Phe136, Ala203, Ala206, Leu225, Gln226, and Arg214), the reconstructed profiles using the $1/2 RT$ scale in bicelles has an overall similarity to those using the RT scale in micelles. The preserved features include: the cooperative packing core (formed by TM1, TM2 and TM3), the cooperative cluster in the active site (Ser201, His150 and Asn154), the localized cluster in L1, and the overpropagated cluster at the TM4-TM6 interface

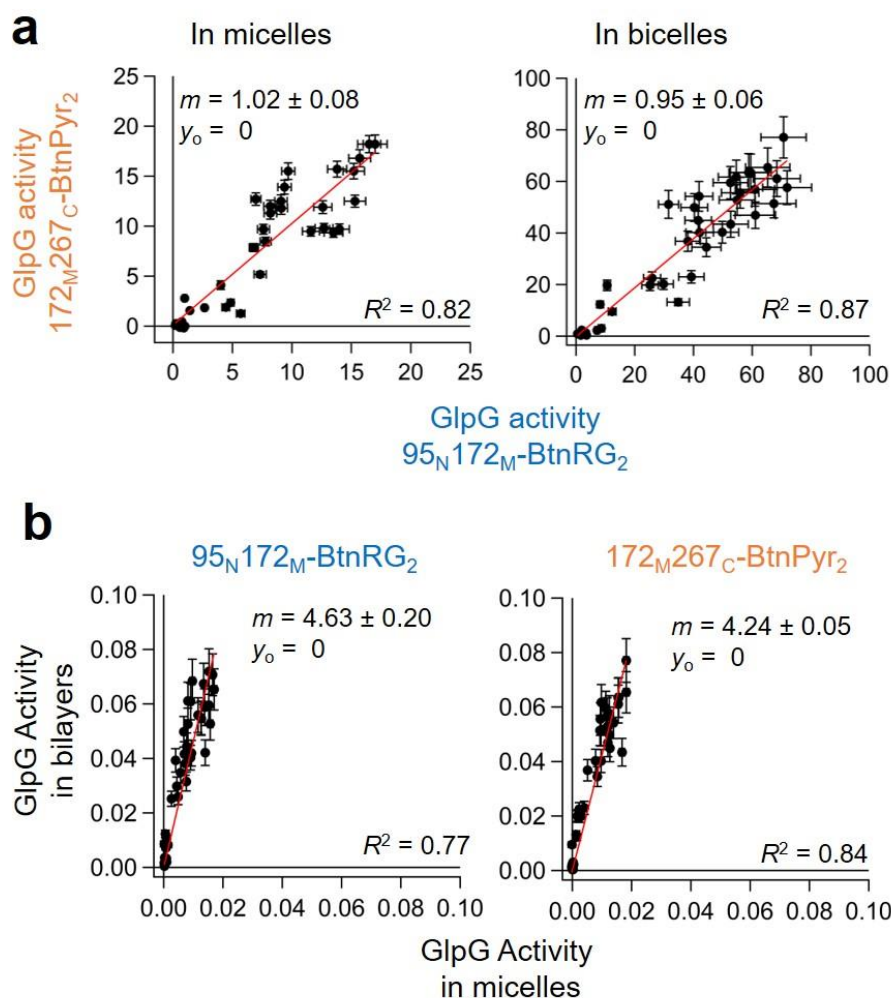


Figure 2.21. The effects of the location of biotin labels and hydrophobic environment on the proteolytic activity GlpG and variants.

(a) The effect of the location of biotin labels ($95_N 172_M$ -BtnPyr₂ vs $172_M 267_C$ -BtnPyr₂) on the activity of GlpG WT and variants ($1 \mu\text{M}$) for the TM substrate LYTM2 ($10 \mu\text{M}$) in micelles and bicelles. In both environments, the slopes representing the correlation between the activities in micelles and those in bicelles are close to the unity, indicating that the location of the biotin pair does not affect the mutational impacts on GlpG activity.

(b) The effect of the hydrophobic environment on GlpG activity. All activity values correspond to the fractional turnover rate (min^{-1}) of the substrate ($10 \mu\text{M}$) proteolyzed by GlpG ($1 \mu\text{M}$) in 5 mM DDM or 2% DMPC:CHPAS bicelles. Errors denote s.e.m. ($n = 3$).

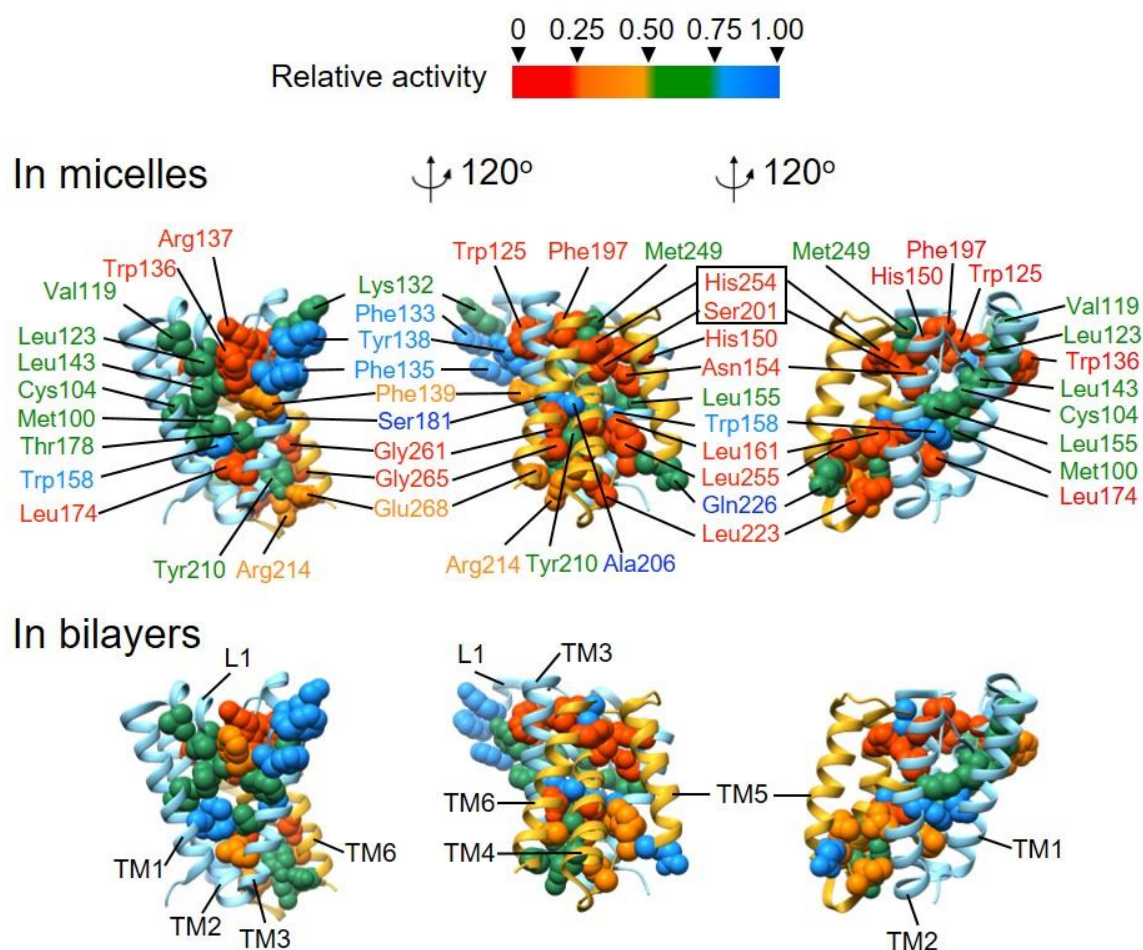
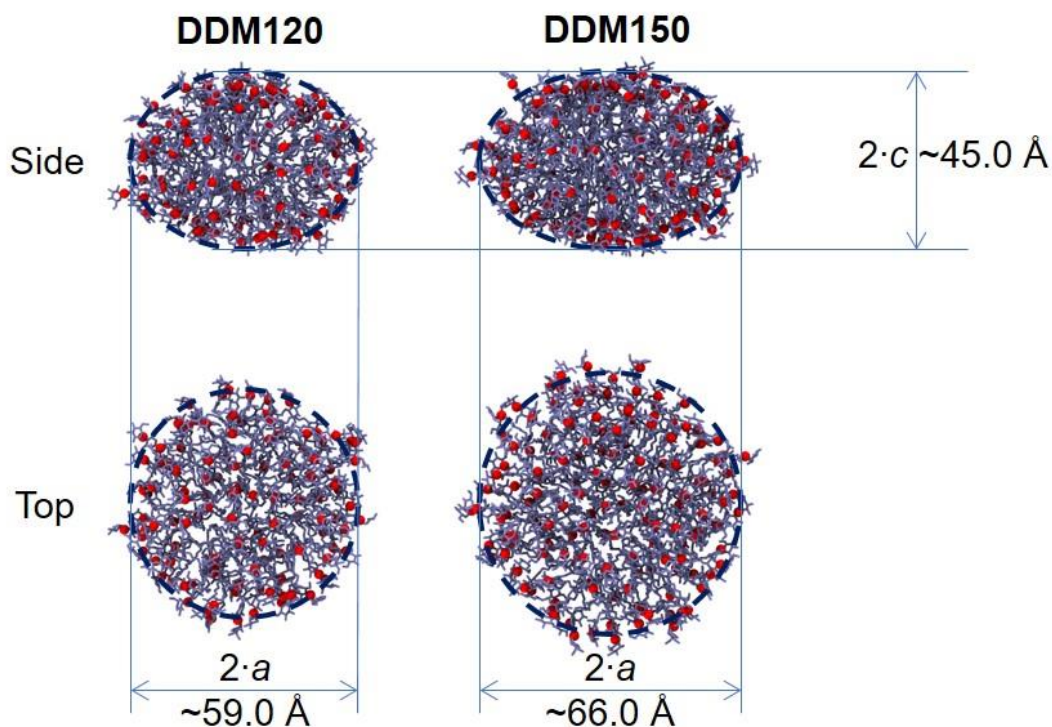


Figure 2.22. Mapping of the mutation-induced activity changes of GlpG onto the structure.

The activity of variants was presented as the proteolytic activity for the membrane-bound substrate LYTM2 (Extended Data Figure. 2.3) relative to that of WT. The relative activity shown represents an average of the relative activities of each mutant in the backgrounds of 95_N172_M-BtnPyr₂ and 172_M267_C-BtnPyr₂ (Extended Data Table 2.1), respectively, and color-coded (**Top**) in each heat map. Overall, the inactivating mutations are distributed near C-subdomain (the backbone in light orange) harboring the catalytic dyad (Ser201-His254) as well as Arg136 and Trp137 in the L1 loop (Wang & Ha 2007 *J Mol Biol* 374, 1104). Also, the mutational effect on activity is slightly more tolerant in bilayers.



Oblate structure	DDM120	DDM150
Radius, a	$29.5 \pm 0.8 \text{ \AA}$	$32.9 \pm 0.8 \text{ \AA}$
Radius, c	$22.6 \pm 1.1 \text{ \AA}$	$22.5 \pm 0.9 \text{ \AA}$
Oblateness	0.23 ± 0.06	0.31 ± 0.04
Area per DDM	$77.3 \pm 1.6 \text{ \AA}^2$	$72.1 \pm 1.7 \text{ \AA}^2$

Figure 2.23. Modelling of the micellar systems for MD simulation.

Two types of the micellar systems were modeled, one with 120 DDM molecules (DDM120) and the other with 150 DDM molecules (DDM150) that are comparable to the experimentally determined aggregation number of DDM (120 to 140). In both systems, the micelles are oblate-shaped with the axial dimension (the radius, c) maintained constant at $22.5 \text{ \AA}$. With the number of DDM molecules increased from 120 to 150, the equatorial dimension accordingly increased from $29.5 \text{ \AA}$ to $32.9 \text{ \AA}$. The area per DDM at the surface is larger in DDM120, which provides room for each DDM to relax fast relative to DDM150.

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* **21**, 167-195, doi:10.1002/prot.340210302 (1995).
2. Levy, Y. & Onuchic, J. N. Water mediation in protein folding and molecular recognition. *Annu Rev Biophys Biomol Struct* **35**, 389-415, doi:10.1146/annurev.biophys.35.040405.102134 (2006).
3. Hunter, C. A. & Anderson, H. L. What is cooperativity? *Angew Chem Int Ed Engl* **48**, 7488-7499, doi:10.1002/anie.200902490 (2009).
4. Hilser, V. J., Dowdy, D., Oas, T. G. & Freire, E. The structural distribution of cooperative interactions in proteins: analysis of the native state ensemble. *Proc Natl Acad Sci U S A* **95**, 9903-9908, doi:10.1073/pnas.95.17.9903 (1998).
5. Dill, K. A., Fiebig, K. M. & Chan, H. S. Cooperativity in protein-folding kinetics. *Proc Natl Acad Sci U S A* **90**, 1942-1946 (1993).
6. Bai, Y. & Englander, S. W. Future directions in folding: the multi-state nature of protein structure. *Proteins* **24**, 145-151, doi:10.1002/(SICI)1097-0134(199602)24:2<145::AID-PROT1>3.0.CO;2-I (1996).
7. Bedard, S., Mayne, L. C., Peterson, R. W., Wand, A. J. & Englander, S. W. The foldon substructure of staphylococcal nuclease. *Journal of molecular biology* **376**, 1142-1154, doi:10.1016/j.jmb.2007.12.020 (2008).
8. Chamberlain, A. K., Handel, T. M. & Marqusee, S. Detection of rare partially folded molecules in equilibrium with the native conformation of RNaseH. *Nature structural biology* **3**, 782-787 (1996).
9. Park, C. & Marqusee, S. Probing the high energy states in proteins by proteolysis. *Journal of molecular biology* **343**, 1467-1476, doi:10.1016/j.jmb.2004.08.085 (2004).
10. 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* **104**, 4347-4352, doi:10.1073/pnas.0607132104 (2007).
11. Dill, K. A. Dominant forces in protein folding. *Biochemistry* **29**, 7133-7155, doi:10.1021/bi00483a001 (1990).
12. Popot, J. L. & Engelman, D. M. Membrane protein folding and oligomerization: the two-stage model. *Biochemistry* **29**, 4031-4037 (1990).

13. Seurig, M., Ek, M., von Heijne, G. & Fluman, N. Dynamic membrane topology in an unassembled membrane protein. *Nat Chem Biol* **15**, 945-948, doi:10.1038/s41589-019-0356-9 (2019).
14. Van Lehn, R. C., Zhang, B. & Miller, T. F., 3rd. Regulation of multispinning membrane protein topology via post-translational annealing. *Elife* **4**, e08697, doi:10.7554/eLife.08697 (2015).
15. Woodall, N. B., Hadley, S., Yin, Y. & Bowie, J. U. Complete topology inversion can be part of normal membrane protein biogenesis. *Protein Sci* **26**, 824-833, doi:10.1002/pro.3131 (2017).
16. Vitrac, H., MacLean, D. M., Jayaraman, V., Bogdanov, M. & Dowhan, W. Dynamic membrane protein topological switching upon changes in phospholipid environment. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 13874-13879, doi:10.1073/pnas.1512994112 (2015).
17. Ojemalm, K., Halling, K. K., Nilsson, I. & von Heijne, G. Orientational preferences of neighboring helices can drive ER insertion of a marginally hydrophobic transmembrane helix. *Mol Cell* **45**, 529-540, doi:10.1016/j.molcel.2011.12.024 (2012).
18. Hessa, T. *et al.* Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature* **433**, 377-381, doi:10.1038/nature03216 (2005).
19. Cao, Z., Hutchison, J. M., Sanders, C. R. & Bowie, J. U. Backbone hydrogen bond strengths can vary widely in transmembrane helices. *J Am Chem Soc* **139**, 10742-10749, doi:10.1021/jacs.7b04819 (2017).
20. Sanders, M. R., Findlay, H. E. & Booth, P. J. Lipid bilayer composition modulates the unfolding free energy of a knotted alpha-helical membrane protein. *Proc Natl Acad Sci U S A* **115**, E1799-E1808, doi:10.1073/pnas.1714668115 (2018).
21. 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* **131**, 10846-10847, doi:10.1021/ja904711k (2009).
22. Mravic, M. *et al.* Packing of apolar side chains enables accurate design of highly stable membrane proteins. *Science* **363**, 1418-1423, doi:10.1126/science.aav7541 (2019).
23. 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* **98**, 14340-14344, doi:10.1073/pnas.251367498 (2001).
24. Chadda, R. *et al.* The dimerization equilibrium of a ClC Cl(-)/H(+) antiporter in lipid bilayers. *Elife* **5**, doi:10.7554/eLife.17438 (2016).

25. Bowie, J. U. Membrane protein folding: how important are hydrogen bonds? *Curr Opin Struct Biol* **21**, 42-49, doi:10.1016/j.sbi.2010.10.003 (2011).
26. Joh, N. H. *et al.* Modest stabilization by most hydrogen-bonded side-chain interactions in membrane proteins. *Nature* **453**, 1266-1270, doi:10.1038/nature06977 (2008).
27. Gratkowski, H., Lear, J. D. & DeGrado, W. F. Polar side chains drive the association of model transmembrane peptides. *Proc Natl Acad Sci U S A* **98**, 880-885, doi:10.1073/pnas.98.3.880 (2001).
28. Stanley, A. M., Chuawong, P., Hendrickson, T. L. & Fleming, K. G. Energetics of outer membrane phospholipase A (OMPLA) dimerization. *J Mol Biol* **358**, 120-131, doi:10.1016/j.jmb.2006.01.033 (2006).
29. Corin, K. & Bowie, J. U. How physical forces drive the process of helical membrane protein folding. *EMBO Rep* **23**, e53025, doi:10.15252/embr.202153025 (2022).
30. O'Brien, E. S. *et al.* Membrane Proteins Have Distinct Fast Internal Motion and Residual Conformational Entropy. *Angew Chem Int Ed Engl* **59**, 11108-11114, doi:10.1002/anie.202003527 (2020).
31. Hong, H. & Bowie, J. U. Dramatic destabilization of transmembrane helix interactions by features of natural membrane environments. *J Am Chem Soc* **133**, 11389-11398, doi:10.1021/ja204524c (2011).
32. Andersen, O. S. & Koeppe, R. E., 2nd. Bilayer thickness and membrane protein function: an energetic perspective. *Annu Rev Biophys Biomol Struct* **36**, 107-130, doi:10.1146/annurev.biophys.36.040306.132643 (2007).
33. Chadda, R. *et al.* Membrane transporter dimerization driven by differential lipid solvation energetics of dissociated and associated states. *Elife* **10**, e63288, doi:10.7554/eLife.63288 (2021).
34. Ballweg, S. *et al.* Regulation of lipid saturation without sensing membrane fluidity. *Nat Commun* **11**, 756, doi:10.1038/s41467-020-14528-1 (2020).
35. Freeman, M. The rhomboid-like superfamily: molecular mechanisms and biological roles. *Annu Rev Cell Dev Biol* **30**, 235-254, doi:10.1146/annurev-cellbio-100913-012944 (2014).
36. Wang, Y., Zhang, Y. & Ha, Y. Crystal structure of a rhomboid family intramembrane protease. *Nature* **444**, 179-180, doi:10.1038/nature05255 (2006).
37. Urban, S. & Dickey, S. W. The rhomboid protease family: a decade of progress on function and mechanism. *Genome Biol* **12**, 231, doi:10.1186/gb-2011-12-10-231 (2011).

38. Glover, K. J. *et al.* Structural evaluation of phospholipid bicelles for solution-state studies of membrane-associated biomolecules. *Biophys J* **81**, 2163-2171 (2001).
39. Guo, R. *et al.* Steric trapping reveals a cooperativity network in the intramembrane protease GlpG. *Nat Chem Biol* **12**, 353-360, doi:10.1038/nchembio.2048 (2016).
40. Jo, S., Kim, T., Iyer, V. G. & Im, W. CHARMM-GUI: a web-based graphical user interface for CHARMM. *J Comput Chem* **29**, 1859-1865, doi:10.1002/jcc.20945 (2008).
41. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. COMPARISON OF SIMPLE POTENTIAL FUNCTIONS FOR SIMULATING LIQUID WATER. *Journal of Chemical Physics* **79**, 926-935, doi:10.1063/1.445869 (1983).
42. Best, R. B. *et al.* Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J Chem Theory Comput* **8**, 3257-3273, doi:10.1021/ct300400x (2012).
43. Darden, T., York, D. & Pedersen, L. PARTICLE MESH EWALD - AN N.LOG(N) METHOD FOR EWALD SUMS IN LARGE SYSTEMS. *Journal of Chemical Physics* **98**, 10089-10092, doi:10.1063/1.464397 (1993).
44. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **26**, 1781-1802, doi:10.1002/jcc.20289 (2005).
45. Chang, Y. C. & Bowie, J. U. Measuring membrane protein stability under native conditions. *Proc Natl Acad Sci U S A* **111**, 219-224, doi:10.1073/pnas.1318576111 (2014).
46. 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* **107**, 19802-19807, doi:10.1073/pnas.1010348107 (2010).
47. Guo, R. *et al.* Structural cavities are critical to balancing stability and activity of a membrane-integral enzyme. *Proc Natl Acad Sci U S A* **117**, 22146-22156, doi:10.1073/pnas.1917770117 (2020).
48. Caldwell, T. A. *et al.* Low- q Bicelles Are Mixed Micelles. *J Phys Chem Lett* **9**, 4469-4473, doi:10.1021/acs.jpcllett.8b02079 (2018).
49. Junker, M. *et al.* Pertactin beta-helix folding mechanism suggests common themes for the secretion and folding of autotransporter proteins. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 4918-4923, doi:10.1073/pnas.0507923103 (2006).
50. Savas, J. N., Stein, B. D., Wu, C. C. & Yates, J. R., 3rd. Mass spectrometry accelerates membrane protein analysis. *Trends Biochem Sci* **36**, 388-396, doi:10.1016/j.tibs.2011.04.005 (2011).

51. Gaffney, K. A. *et al.* Lipid bilayer induces contraction of the denatured state ensemble of a helical-bundle membrane protein. *Proc Natl Acad Sci U S A* **119**, doi:10.1073/pnas.2109169119 (2022).
52. Lipfert, J., Columbus, L., Chu, V. B., Lesley, S. A. & Doniach, S. Size and shape of detergent micelles determined by small-angle X-ray scattering. *J Phys Chem B* **111**, 12427-12438, doi:10.1021/jp073016l (2007).
53. Oliver, R. C. *et al.* Dependence of micelle size and shape on detergent alkyl chain length and head group. *PLoS One* **8**, e62488, doi:10.1371/journal.pone.0062488 (2013).
54. Garavito, R. M. & Ferguson-Miller, S. Detergents as tools in membrane biochemistry. *J Biol Chem* **276**, 32403-32406, doi:10.1074/jbc.R100031200 (2001).
55. Drew, D. & Boudker, O. Shared Molecular Mechanisms of Membrane Transporters. *Annu Rev Biochem* **85**, 543-572, doi:10.1146/annurev-biochem-060815-014520 (2016).
56. Cho, S., Baker, R. P., Ji, M. & Urban, S. Ten catalytic snapshots of rhomboid intramembrane proteolysis from gate opening to peptide release. *Nat Struct Mol Biol* **26**, 910-918, doi:10.1038/s41594-019-0296-9 (2019).
57. Thomaston, J. L. *et al.* High-resolution structures of the M2 channel from influenza A virus reveal dynamic pathways for proton stabilization and transduction. *Proc Natl Acad Sci U S A* **112**, 14260-14265, doi:10.1073/pnas.1518493112 (2015).
58. Latorraca, N. R., Venkatakrishnan, A. J. & Dror, R. O. GPCR Dynamics: Structures in Motion. *Chem Rev* **117**, 139-155, doi:10.1021/acs.chemrev.6b00177 (2017).
59. Zhang, Z., Liu, F. & Chen, J. Conformational Changes of CFTR upon Phosphorylation and ATP Binding. *Cell* **170**, 483-491 e488, doi:10.1016/j.cell.2017.06.041 (2017).
60. Matthews, E. E. *et al.* Thrombopoietin receptor activation: transmembrane helix dimerization, rotation, and allosteric modulation. *FASEB J* **25**, 2234-2244, doi:10.1096/fj.10-178673 (2011).
61. Marinko, J. T. *et al.* Folding and Misfolding of Human Membrane Proteins in Health and Disease: From Single Molecules to Cellular Proteostasis. *Chem Rev* **119**, 5537-5606, doi:10.1021/acs.chemrev.8b00532 (2019).
62. Schleich, J. P. *et al.* Conformational Stability and Pathogenic Misfolding of the Integral Membrane Protein PMP22. *J Am Chem Soc* **137**, 8758-8768, doi:10.1021/jacs.5b03743 (2015).
63. Sanders, C. R. & Myers, J. K. Disease-related misassembly of membrane proteins. *Annu Rev Biophys Biomol Struct* **33**, 25-51, doi:10.1146/annurev.biophys.33.110502.140348 (2004).

64. Molinski, S. V. *et al.* Comprehensive mapping of cystic fibrosis mutations to CFTR protein identifies mutation clusters and molecular docking predicts corrector binding site. *Proteins* **86**, 833-843, doi:10.1002/prot.25496 (2018).
65. Oberai, A., Joh, N. H., Pettit, F. K. & Bowie, J. U. Structural imperatives impose diverse evolutionary constraints on helical membrane proteins. *Proc Natl Acad Sci U S A* **106**, 17747-17750, doi:10.1073/pnas.0906390106 (2009).
66. Allison, T. M. *et al.* Quantifying the stabilizing effects of protein-ligand interactions in the gas phase. *Nat Commun* **6**, 8551, doi:10.1038/ncomms9551 (2015).
67. Cong, X. *et al.* Determining membrane protein-lipid binding thermodynamics using native mass spectrometry. *J Am Chem Soc* **138**, 4346-4349, doi:10.1021/jacs.6b01771 (2016).
68. Laganowsky, A. *et al.* Membrane proteins bind lipids selectively to modulate their structure and function. *Nature* **510**, 172-175, doi:10.1038/nature13419 (2014).
69. Min, D. *et al.* Unfolding of a ClC chloride transporter retains memory of its evolutionary history. *Nat Chem Biol* **14**, 489-496, doi:10.1038/s41589-018-0025-4 (2018).

CHAPTER 3

Dissecting the folding-degradation relationship of membrane proteins in the bilayer

Shaima Muhammednazaar and Heedeok Hong

This chapter will be submitted as an article later.

3.1. Introduction

Protein degradation in cells serves as an essential tool to remove misfolded or damaged proteins and to control the activity of regulatory proteins by directly adjusting their concentration to cellular needs.¹ In bacteria, protein degradation, is majorly carried out by energy-dependent proteases, ClpXP, ClpAP, HslUV, Lon, and FtsH, which commonly consist of the two functional modules: AAA+, a ring-shaped hexameric motor ATPase, and a compartmentalized peptidase whose active sites are excluded from the cytosol.²

Proteolysis is by itself a thermodynamically downhill process, not requiring any external energy. However, the combination of the ATP hydrolytic and compartmentalized proteolytic activities enables highly regulated, and efficient protein degradation:³⁻⁵ AAA+ binds a water-exposed degradation marker (or degron) on a substrate. ATP hydrolysis on the AAA+ module induces power stroke motions of individual subunits and applies pulling force on the bound substrate, which leads to mechanical denaturation of the substrate. The same mechanical force induces translocation of the denatured substrate into the proteolytic chamber through the central pore in the AAA+ ring. This mode of action is also conserved by proteasomes in eukaryotic cells, which has an additional 19S regulatory particle plugged onto the 20S core particle with the AAA+ (β -ring) and protease (α -ring).⁶ Bacterial AAA+ proteases recognize a water-exposed, flexible peptide segment in the terminal or middle of the polypeptide chain, while the proteasomes use a ubiquitin chain conjugated to the substrate as a degron.³

Among bacterial AAA+ proteases, FtsH is unique in that it is only growth-essential in *E. coli* and only membrane-integrated with an N-terminal transmembrane (TM) domain. FtsH family proteins are widely conserved in the inner membranes of bacteria and mitochondria and the thylakoid membranes of chloroplasts⁷. In the respective membranes, FtsH and its orthologs serve

as key protein-quality inspectors, degrading misassembled and damaged membrane proteins as well as short-lived water-soluble transcription factors or enzymes.⁸⁻¹⁰ Notably, mutations in paraplegin and AFG3L2, which are FtsH orthologs in human mitochondria, are associated with neurological disorders such as spastic paraplegia and spinocerebellar ataxia.^{11,12}

Each AAA+ protease has a distinct ability to hydrolyze ATP and unfold a substrate. FtsH has been known as a weak ATPase and unfoldase.^{13,14} While robust AAA+ proteases such as ClpXP and ClpAP rapidly unfold and degrade thermodynamically or kinetically stable water-soluble proteins (*e.g.*, GFP),¹⁵ FtsH cannot degrade them.¹⁴⁻¹⁶ Also, FtsH degrades a misfolded variant of the membrane protein, diacylglycerol kinase in the *E. coli* cells but cannot degrade wild type or a thermostable variant.¹⁴ These studies suggest that FtsH only degrades the proteins that have been denatured, and ATP hydrolysis is used for translocation.¹⁴

Previously, we have successfully reconstituted membrane protein degradation by *E. coli* FtsH in a lipid bilayer environment using the intramembrane protease GlpG of *E. coli* as a model substrate.¹⁷ We have demonstrated that despite the suggestion that FtsH cannot actively denature proteins,^{13,14} FtsH possesses a substantial ability to actively denature the stable membrane protein GlpG, and the stability and hydrophobicity near the degradation marker affect the degradation rate. To degrade a membrane protein, FtsH should induce substrate denaturation and membrane dislocation to the protein's active site in the cytosol.¹⁷ Remarkably, FtsH overcomes this dual-energetic burden with the ATP cost (0.5–1.0 ATP hydrolysis/residue) comparable to that for water-soluble substrates by robust ClpAP/XP proteases (0.5–6.0 ATP hydrolysis/residue) by cooperatively coupling ~4 ATP hydrolysis events to degradation.^{17,18}

Nonetheless, the detailed molecular mechanism of FtsH-mediated membrane protein degradation is not well understood. Toward the goal, it is critical to determine the rate-limiting

step (*i.e.*, the step with the slowest rate) in the degradation. For membrane protein substrates, there are two thermodynamically unfavorable steps that resist degradation and thus can slow down the degradation rate: **1)** substrate denaturation involving disruption of tertiary interactions within the membrane (represented by the thermodynamic stability, $\Delta G^0_{\text{N-D}}$, and its kinetic barrier, $\Delta G^{0\ddagger}_{\text{N-D}}$); **2)** dislocation of the denatured hydrophobic TM segments from the membrane to the aqueous phase (represented by the transfer free energy, $\Delta G^0_{\text{transfer,bilayer-water}}$, and its kinetic barrier, $\Delta G^{0\ddagger}_{\text{transfer,bilayer-water}}$). This intrinsic energetics of membrane protein substrates is distinct from that of water-soluble substrates, which are dissected into unfolding in water and translocation of the unfolded chain into the similarly aqueous proteolytic chamber.¹⁷ In case of water-soluble substrates, their degradation rate by ClpXP strongly depends on the local stability and native secondary structure near the degradation marker, whereas the translocation rate is less affected by the detailed amino acid sequence.¹⁹⁻²¹ That is, the rate limiting step can be either the denaturation or translocation depending on the stability (*i.e.*, it is the denaturation when the substrate is stable or the translocation when the substrate is unstable).¹⁹

Here, we defined the rate limiting step in FtsH-mediated membrane protein degradation. We designed GlpG variants to generate each variable of $\Delta G^0_{\text{N-D}}$, $\Delta G^{0\ddagger}_{\text{N-D}}$ or $\Delta G^0_{\text{transfer,bilayer-water}}$ in a wide range, challenged the variants against FtsH in the lipid bilayer, and obtained the correlation between each variable and the activation energy of degradation ($\Delta G^{0\ddagger}_{\text{deg}}$). Strikingly, we find that FtsH degrades GlpG variants at a constant degradation rate with the negligible dependence on their intrinsic thermodynamic and kinetic stability. This is strong evidence that the rate-limiting step in membrane protein degradation is not the substrate denaturation but the membrane dislocation, which has an opposite trend to that in the degradation of stable globular water-soluble proteins.

3.2. Materials and Methods

3.2.1. FtsH Expression and purification

FtsH with a C-terminal His₆ tag was expressed in the *E.coli* C43 pLysS strain. FtsH was grown on an ampicillin (100 mg/ ml) plate at 37 °C. 50 mL of LB (Luria-Bertani) media (100 mg/ml ampicillin) was inoculated with a single colony and the cells were grown at 37 °C until the growth reached the stationary phase. The culture grown overnight was used to inoculate 1 L LB media (100 mg/ml ampicillin), and cells were grown at 37 °C until OD_{600nm} reached 1.2. Protein expression was induced with 1 mM isopropyl β-D-thiogalactopyranoside (IPTG, Gold Bio), followed by additional cultivation for 3 hours at 37 °C. Harvested cells were resuspended in 1/40 culture volume of resuspension buffer containing 25 mM Tris-HCl (pH 8.0), 0.1% (v/v) β-mercaptoethanol (BME), 15% (v/v) glycerol and 0.5 mM phenylmethylsulphonyl fluoride (PMSF). Cells were lysed five times using EmulsiFlex-C5 pressure homogenizer (Avestin). Cell debris was removed by centrifugation (Thermo-Fisher scientific, F21 rotor, 6,000 rpm, 30 min), and the total membrane fraction was obtained by ultracentrifugation (Beckman Coulter, Type 45 Ti rotor, 24,000 rpm, 2 h). The total membrane fraction was resuspended in 1/50 culture volume of base buffer containing 25 mM Tris-HCl (pH 8.0), 15% (v/v) glycerol, 200 mM KCl, and 2 % (w/v) Triton X-100. The pellets were removed by centrifugation at 12,000 rpm for 1 h. The resulting supernatant was incubated with 1 ml Ni²⁺-NTA resin at 4 °C for 1 h. After washing the resin with a 10-resin volume of wash buffer (25 mM Tris-HCl pH 8.0, 15% (v/v) Glycerol, 200 mM KCl, 0.1 % (w/v) Triton X-100, 0.1% (v/v) BME and 20 mM Imidazole), bound FtsH was eluted with a 10-resin volume of elution buffer (25 mM Tris-HCl pH 8.0, 15% (v/v) glycerol, 200 mM KCl, 0.1 % (w/v) Triton X-100, 0.1% (v/v) BME and 200 mM Imidazole). After removing excess imidazole on a desalting column, FtsH was concentrated to the final volume of 0.5-1.0 mL/

L-culture, frozen, and stored at -80 °C. Protein concentration was determined by a 660 nm assay, compatible with Triton X-100.

3.2.2. Expression and purification of GlpG for *in vitro* degradation assay.

MBP-TEV_{cleavage}-His₆-GlpGTM (residues 87-276)-108 (-SLLWS) was expressed in *E. coli* BL21(DE3)RP strain. Cysteine mutant G172C was generated by site-directed mutagenesis for thiol-reactive fluorophore (NBD) labeling. Cells were grown in kanamycin (50 mg/mL) plates overnight. A single colony was inoculated in a liquid culture (25 mL) and grown at 37 °C overnight. The large culture in LB media was initiated by adding the overnight-grown small culture and growing at 37 °C until OD_{600nm} reached 0.9 to 1.2. Protein expression was induced with 0.5 mM IPTG (Gold Bio), followed by additional cultivation for 18 h at 15 °C. The cells were harvested and resuspended in a 1/50-culture volume of 50 mM Tris-HCl (pH 8.0), 5 mM EDTA, 0.5 mM TCEP, and 0.5 mM PMSF. The resuspended cells were lysed five times using EmulsiFlex-C5 pressure homogenizer (Avestin). The cell debris was removed by centrifugation (Thermo fisher scientific, F21 rotor, 6000 rpm, 30 min). The supernatant was centrifuged to obtain the total membrane fraction (Beckman Coulter, Type 45 Ti rotor, 24,000 rpm, 2 h). the membrane pellets were resuspended in a 1/100 culture volume of 50 mM Tris-HCl (pH 8.0), 200 mM NaCl , 0.5 mM TCEP, 0.8% (w/v) n-dodecyl- β -D-maltoside (DDM, Anatrace) and 0.5 mM PMSF using tissue homogenizer. After removing aggregates by ultracentrifugation at 18,000 rpm for 25 min, TEV (tobacco etch virus) protease was added and incubated at room temperature for 3 h. The resulting mixture was incubated with 3 ml of Ni²⁺-NTA resin at 4 °C for 1 h. The cleaved His₆-GlpG-108 was bound to the Ni²⁺-NTA resin, and MBP and TEV proteas are eluted in the flow-through. Bound GlpG was washed with a 10-resin volume of wash buffer (50 mM Tris-HCl pH 8.0, 200 mM NaCl, 0.05% DDM, and 40 mM imidazole) and eluted with the 5-resin volume of elution

buffer (50 mM Tris-HCl pH 8.0, 200 mM NaCl, 0.1% DDM, and 400 mM imidazole). Excess imidazole was removed on a desalting column equilibrated with 50 mM Tris-HCl (pH 8.0), 200 mM NaCl, and 0.1% DDM. The purified protein was concentrated using an Amicon centrifugal filter unit (Millipore Sigma, 30 kDa MWCO). The protein concentration was measured using absorbance at 280 nm ($\epsilon_{280\text{nm}} = 69,940 \text{ M}\cdot\text{cm}^{-1}$).

3.2.3. NBD-labelling of GlpG variants.

~50 μM of purified variants of GlpG in 50 mM Tris-HCl (pH 8.0), 200 mM NaCl, and 0.1% DDM was incubated with 5 molar excess of TCEP at room temperature for 1 h. 10 molar excess of the thiol-reactive fluorophore IA-NBD amide dissolved in DMSO (~10 mg/ml) was added to the mixture while vortexing. The labeling reaction was further extended at 4 °C overnight. Excess-free labels were removed by running a desalting column 3 times equilibrated with 50 mM Tris-HCl (pH 8.0), 200 mM NaCl, and 0.1% DDM. The concentration of NBD fluorophore was determined by UV-VIS absorption ($\epsilon_{486\text{nm}} = 23,500 \text{ M}\cdot\text{cm}^{-1}$), and protein concentration was determined by a 660 nm protein assay (Bio-Rad). The typical labeling efficiency was 0.9 to 1.2 NBD per GlpG molecule.

3.2.4. Measuring the degradation rate of GlpG variants by NBD fluorescence.

The GlpG samples for the degradation assay were prepared as follows: GlpG variant labeled with NBD in 0.1% DDM was first reconstituted into liposomes composed of DMPC/DMPG (molar ratio = 4:1). The DMPC/DMPG mixture dissolved in chloroform was first dried in a stream of N_2 gas and vacuum desiccator and solubilized in 20 mM HEPES (pH 7.5), 100 mM KCl, 0.1% BME, and 3% β -OG (final). The stock of NBD-labeled GlpG (80 to 150 μM) was mixed with resuspended lipids (to 3% final concentration) and incubated on ice for 1 h. Then, Bio-

Beads (0.5 mg/ml-suspension, Bio-Rad) were added and incubated at room temperature for 4 h. Bio-Beads were changed every 4 to 6 h until a turbid solution with no foams was obtained. The resulting proteoliposomes were dialyzed against 20 mM HEPES (pH 7.5), 100 mM KCl, 0.1% BME, and 15% glycerol overnight and extruded through a 0.2 μ M pore-size polycarbonate membrane (Waters) to remove aggregates. The total phospholipid concentration was determined using an organic phosphate assay. Based on the lipid concentration, CHAPS was added to form the bicelles of $q = 2.8$. The final concentration of NBD-labeled GlpG was measured by a 660 nm protein assay (Bio-Rad).

Time-dependent GlpG degradation by FtsH was performed in the solution of 3% DMPC/DMPG/CHAPS bicelles (20 mM HEPES pH 7.5, 100 mM KCl, 0.1% BME, 400 μ M ZnCl₂, MgCl₂, and 15% glycerol). FtsH in Triton X-100 was directly injected into the bicelles to the final monomer concentration of 2 μ M and incubated on ice for 1 h. Degradation was obtained at varying concentrations of NBD-labeled GlpG in the presence and absence of ATP (2 mM ATP). 65 μ L of the final mixture was transferred to a 96-well plate (UV compatible, Greiner Bio-one). Quenching of NBD fluorescence by GlpG degradation over time (5 to 6 h) was measured at 545 nm with an excitation wavelength of 485 nm on a SpectraMax M5 plate reader. The net change in NBD fluorescence induced by GlpG degradation was obtained by obtaining the difference in time-dependent change of NBD fluorescence in the presence and absence of ATP at each GlpG concentration. GlpG degradation rate per FtsH hexamer per minute (v_{deg}) is defined as follows,

$$v_{deg} = \frac{\frac{\Delta F}{\min}}{\frac{\Delta F_{\infty}}{[GlpG]}} \times \frac{1}{[FtsH_6]} \quad (\text{Equation 1})$$

Where $\Delta F/\text{min}$ denotes the initial slope from the NBD signal decrease as a function of time, $\Delta F_{\infty}/[\text{GlpG}]$ denotes a fluorescence contribution per GlpG.

Degradation rates of GlpG as a function of mol fraction of GlpG were fitted to Michaelis- Menten equation:

$$v_{\text{deg}} = \frac{k_{\text{cat,deg}} X_{\text{GlpG}}}{K_{\text{m,deg}} + X_{\text{GlpG}}} \quad (\text{Equation 2})$$

Where $k_{\text{cat,deg}}$ is the maximal turnover number of FtsH hexamer per min, and $K_{\text{m,deg}}$ is the mol fraction of GlpG at which degradation rate reaches half maximum. X_{GlpG} is the mol fraction of GlpG denoted as $[\text{GlpG}]/([\text{DMPC}]+[\text{DMPG}]+[\text{CHAPS}]+[\text{FtsH}]+[\text{GlpG}])$.

3.2.5. Expression, purification, and labeling of variants of GlpG TM to obtain thermodynamic stability using steric trapping.

The TM domain of GlpG (residues 87-276) encoded by the pET15b vector was expressed in *E. coli* BL21(DE3)RP strain. Detailed purification procedures are described in Chapter 2. The construct with two cysteine residues at G172C and V267C served as a template for mutations. GlpG was labeled with the thiol-reactive biotin derivative possessing BtnPyr-IA as described in Chapter 2.

1 μM of GlpG (172_N267_M-BtnPyr) in 2% (w/v) DMPC/CHAPS bicelles ($q = 1.5$), 20 mM HEPES (pH 7.5), 1mM DTT and 40 mM KCl was titrated against mSA labeled with Y83C-DABMI (AnaSpec). The titrated samples were transferred to a 96-well plate, and binding was monitored by decreasing pyrene fluorescence at 390 nm with an excitation wavelength of 345 nm using a SpectraMax M5 plate reader. Averaged data of the second mSA binding phase was fitted to the following equation,

$$F = \frac{1}{1 + \left(K_{d,biotin} + \frac{K_{d,biotin}}{K_{N-D}} \right) \frac{1}{[mSA]}} (F_{\infty} - F_o) + F_o \quad (\text{Equation 3})$$

Where F is measured fluorescence intensity, F_o 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 biotin binding, and K_{N-D} is the equilibrium constant for denaturation of GlpG. Thermodynamic stability was obtained by steric trapping using the equation $\Delta G_{N-D}^0 = -RT \ln K_{N-D}$.

3.2.6. Measuring the intrinsic denaturation rate of GlpG by ProK digestion

5 μ M GlpG (172_M267_C-BtnPyr) was prepared in 2% (w/v) DMPC/DMPG/CHAPS bicelles in 20 mM HEPES (pH 7.5) and 100 mM KCl, and incubated on ice for 30 min. 2 mM CaCl₂ was added to enhance the stability of Proteinase K (Sigma). Proteolysis was initiated by adding 300 μ g/mL Proteinase K and samples were incubated at 37 °C. A 25 μ L aliquot of each sample was taken at a specific time, and the reaction was quenched by adding 10 mM PMSF. Proteolysis reactions were monitored by SDS-PAGE (4 to 20% gradient gels, Bio-Rad). The remaining fraction of GlpG after Proteinase K was quantified by measuring the band intensities of GlpG fractions using ImageJ software (<https://imagej.nih.gov/ij/download.html>). The band intensity at each time point was normalized to the control (GlpG without Proteinase K). The remaining fraction at each time point was fitted to an equation for first-order kinetics to obtain an intrinsic denaturation rate.

$$y = y_o + Ae^{-k_{denat} \cdot t} \quad (\text{Equation 4})$$

Where y is the GlpG fraction remaining after Proteinase K digestion; y_0 is the minimal GlpG fraction after Proteinase K digestion; A is the amplitude of the time-dependent change; k_{denat} is the intrinsic denaturation rate, and t is the time.

3.2.7. Measuring the proteolytic activity of GlpG in DMPC/DMPG/CHAPS bicelles.

1 μM GlpG (172_M267_C-BtnPyr) in 2% DMPC/DMPG/CHAPS bicelles, 20 mM HEPES (pH 7.5), and 100 mM KCl was incubated on ice for 30 min. The proteolytic reaction of GlpG was initiated by adding 10-molar excess of the substrate, NBD-labeled SN-LacYTM2, in 2% DMPC/DMPG/CHAPS bicelles at room temperature. The time-dependent decrease of NBD fluorescence, a measure of proteolytic activity, was monitored using a SpectraMax M5 plate reader at excitation and emission wavelengths of 485 nm and 535 nm, respectively. Fluorescence change was normalized to a control sample containing NBD-SN-LacYTM2 alone. The initial slope of the time-dependent change was taken as an activity.

3.2.8. Measuring GlpG degradation *in vivo*.

The *in vivo* construct was cloned to pBAD/HisA vector, including the transmembrane domain of GlpG (residues 87 to 276), an N-terminal FLAG epitope tag, and a C-terminal HA epitope tag, and a C-terminal 108 degradation marker for FtsH degradation. The strains of *E. coli* AR3289 (+*ftsH*) and AR3291 (-*ftsH*) were transformed with the plasmid encoding a GlpG variant plasmid. A single colony was inoculated in 7 mL of LB containing 100 $\mu\text{g}/\text{mL}$ ampicillin. AR3289 cells were grown at 37 °C, and AR3291 cells at 30 °C overnight. OD_{600nm} was measured for AR3289 (1.2 to 1.8) and AR3291 (0.5 to 0.9). Then, the expression of GlpG was induced by adding 0.05% (w/v) arabinose at 37 °C for 45 min. Protein synthesis was blocked by adding 300 $\mu\text{g}/\text{mL}$ spectinomycin (Sigma), immediately followed by the sample collection at 0 min. 500 μL aliquots

of culture were collected at different time points at 37 °C and flash-frozen with liquid nitrogen to monitor the degradation over time. For immunodetection, the cells were thawed and spun down at 13,000 rpm for 3 min using a bench-top centrifuge (Eppendorf, 5424R). Cells were resuspended in 150 µL of TE buffer containing 10 mM Tris-HCl (pH 8.0) and 1 mM EDTA, and then mixed with 150 µL of the protein sample buffer (final concentrations of 2% SDS (w/v), 0.1% (w/v) bromophenol blue, 10% (v/v) glycerol, 1% (v/v) BME and 50 mM Tris-HCl (pH 6.8)). Before loading to SDS-PAGE (4 to 20% gradient gel, Bio-Rad), samples were sonicated for 15 to 25 min and run for 18 min at 300 V. Western blotting analysis was performed against the N-terminal FLAG epitope, which can detect the degradation of the full-length transmembrane region of GlpG, which initiated from the C-terminus. GlpG was transferred to a Polyvinylidene difluoride (PVDF) membrane (Bio-Rad) at 100 V for 1 h. GlpG with the FLAG tag was detected using rabbit monoclonal anti-FLAG primary antibody (Cell Signaling Technology, 1:1,000 dilution) and anti-rabbit IgG-HRP secondary antibody (Cell Signaling Technology, 1:2,000 dilution). Chemiluminescent detection was performed using Clarity Western ECL substrate (Bio-Rad) and ChemiDoc Imager (Bio-Rad).

3.3 Results

3.3.1. The effect of mutations on conformational stability and activity of GlpG

To study the folding-degradation relationship of membrane proteins, it is essential to have a membrane protein substrate with well-characterized folding properties. Here we employed the intramembrane protease GlpG of *E. coli* as a model membrane substrate. GlpG is an advantageous model due to its reversible folding in various membrane-mimetic environments; thus, we can

obtain thermodynamic stability, and its proteolytic activity can provide a folding indicator. We first investigated the effects of various mutations on the conformational stability of GlpG. The thermodynamic stability of GlpG was obtained using steric trapping (**Figure 3.1**).²²⁻²⁵ Compared to other methods that typically use chemical denaturants or pulling force, steric trapping is advantageous because protein stability can be measured directly under native conditions (*i.e.*, the same condition as the degradation assay). Steric trap couples the unfavorable denaturation of a biotin-tagged protein to the binding of monovalent streptavidin (mSA). ΔG°_{N-D} can be obtained by fitting the model function (**Equation 3**) to the attenuated second binding phase of mSA.

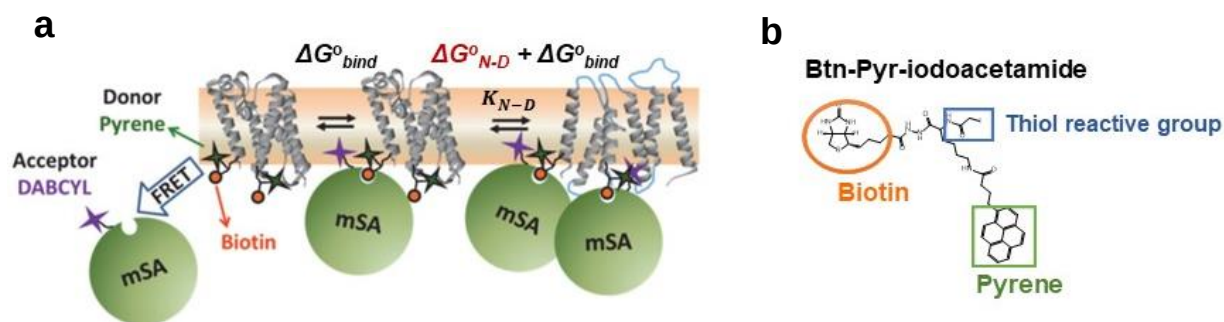


Figure 3.1. (a) Steric trapping scheme. Bulky mSA binds and captures the transiently denatured protein. Protein denaturation is coupled to binding of the second mSA, resulting in the attenuation of the apparent second mSA binding. ΔG°_{N-D} is obtained by fitting the second binding phase (Equation 3 in Materials and Methods). **(b) Biotin derivative (BtnPyr-IA) used for steric trapping.** The probe possesses a biotin (orange circle), pyrene (green square), and iodoacetamide (blue rectangle) motifs conjugated to a lysine template

We chose a double cysteine variant at positions G172 and V267 to label with thiol-reactive biotin derivative with fluorescent pyrene (BtnPyr). This biotin pair cover the approximate C-terminal half of GlpG (*i.e.*, C-subdomain). Steric trapping captures the transient separation of biotin pairs; thus, we can obtain the local stability of the region encompassing the biotin pair. Therefore, in this construct design, the local stability of the C-subdomain directly connected to the

degradation marker at the C-terminus (*i.e.*, 108 tag) is measured. The local stability of a protein near the degradation marker is known to be a strong determinant of the degradation rate.^{19,26} mSA was mutated at Y83 to cysteine to conjugate the dabcyI quencher. Binding of mSA to the biotin labels can be monitored by FRET between pyrene donor and dabcyI acceptor by FRET. We chose DMPC/CHAPS bicelles ([DMPC]/[CHAPS] =1.5) as a bilayer mimic, in which the reversible folding for measuring ΔG°_{N-D} has been established (**Chapter 2 Figure 2.3**). In parallel, we characterized the proteolytic activity of GlpG variants (**Table 3.1** and **Figure 3.12**) in the bicellar environment using the model substrate SN-LacYTM2 (Methods).²²

We classified GlpG mutants into four categories based on the stability, structural and functional contexts (**Figure 3.2.** and **Figure 3.12**). By this, the thermodynamic stability of GlpG was changed in a wide range of $\Delta\Delta G^{\circ}_{N-D,WT-Mut} = -0.1$ to 5.0 kcal/mol: First, catalytically critical residues include H145, H150, N154, S201, and H254. S201 and H254 form the catalytic dyad of GlpG. Single Ala or Thr mutations at these residues abolish the proteolytic activity of GlpG.^{27,28} ΔG°_{N-D} of the double-biotin variant without additional modifications (denoted as WT) was 6.7 ± 0.1 kcal/mol (Chapter 2). S201T showed similar stability as WT, while H254A was slightly stabilized $\Delta\Delta G^{\circ}_{N-D,WT-Mut} = 0.2 \pm 0.2$ kcal/mol (**Table 3.1**). H150A and N154A were mildly destabilizing with $\Delta\Delta G^{\circ}_{N-D,WT-Mut} = 0.5 \pm 0.2$ and 0.6 ± 0.1 kcal/mol (**Table 3.1**), respectively. H145A induced moderate destabilization with $\Delta\Delta G^{\circ}_{N-D,WT-Mut} = 2.6 \pm 0.2$ kcal/mol (**Table 3.1**).

The second category include substantially destabilized mutants ($\Delta\Delta G^{\circ}_{N-D,WT-Mut} \geq 3.5 \pm 0.1$ kcal/mol), substituting the residues that are critically involved in the packing interaction in the protein interior. Single Ala mutations on L174 and L207 were vastly destabilizing with $\Delta\Delta G^{\circ}_{N-D}$

of 4.8 ± 0.1 and 4.7 ± 0.2 kcal/mol, respectively (**Table 3.1**). They retain 10% and 14% activity respectively.

The mutations in the third category are both substantially destabilizing and functionally important.^{28,29} We generated three mutations, W125A, R137A, and G261A, which were destabilizing by $\Delta\Delta G^{\circ}_{N-D,WT-Mut} \geq 3.5 \pm 0.1$ kcal/mol with the residual proteolytic activity of $\leq 10\%$ relative to WT.

The target residues for mutation in the fourth category are involved in the favorable electrostatic interaction between the protein and membrane at the protein surface that can possibly affect the rate of the membrane dislocation step during degradation. We found four positively charged Arg and Lys residues that would be engaged in the attractive interaction with the negatively charged lipid headgroups and mutated each to Ala (R92A, K167A, K191A, and R217A). These residues were selected based on the proximity (≤ 6.5 Å) between the negatively charged phosphate head group and the centroid C atom of the arginine or lysine residues. The proteolytic active site of FtsH resides in the cytosol. Thus, we expect that the Ala mutation of the positively charged residue in the periplasmic side (*i.e.*, the *trans* side) may accelerate the degradation by the disruption of the attractive electrostatic protein-lipid interaction and reducing the energetic cost of translocating the positive charge across the bilayer. The neutralization of the positively charged residue in the cytosolic side (*i.e.*, the *cis* side) would also accelerate degradation by abolishing the attractive interaction with the membrane. K191A at the periplasmic water-membrane interface displayed minimal destabilization of the protein ($\Delta\Delta G^{\circ}_{N-D,WT-Mut} = 0.2 \pm 0.2$ kcal/mol) with only 20% activity reduction relative to WT while R92A, K167A, R217A, and R92A/K167A at the cytosolic interface, induced moderate destabilization of $\Delta\Delta G^{\circ}_{N-D,WT-Mut} = 0.9$ to 1.1 kcal/mol.

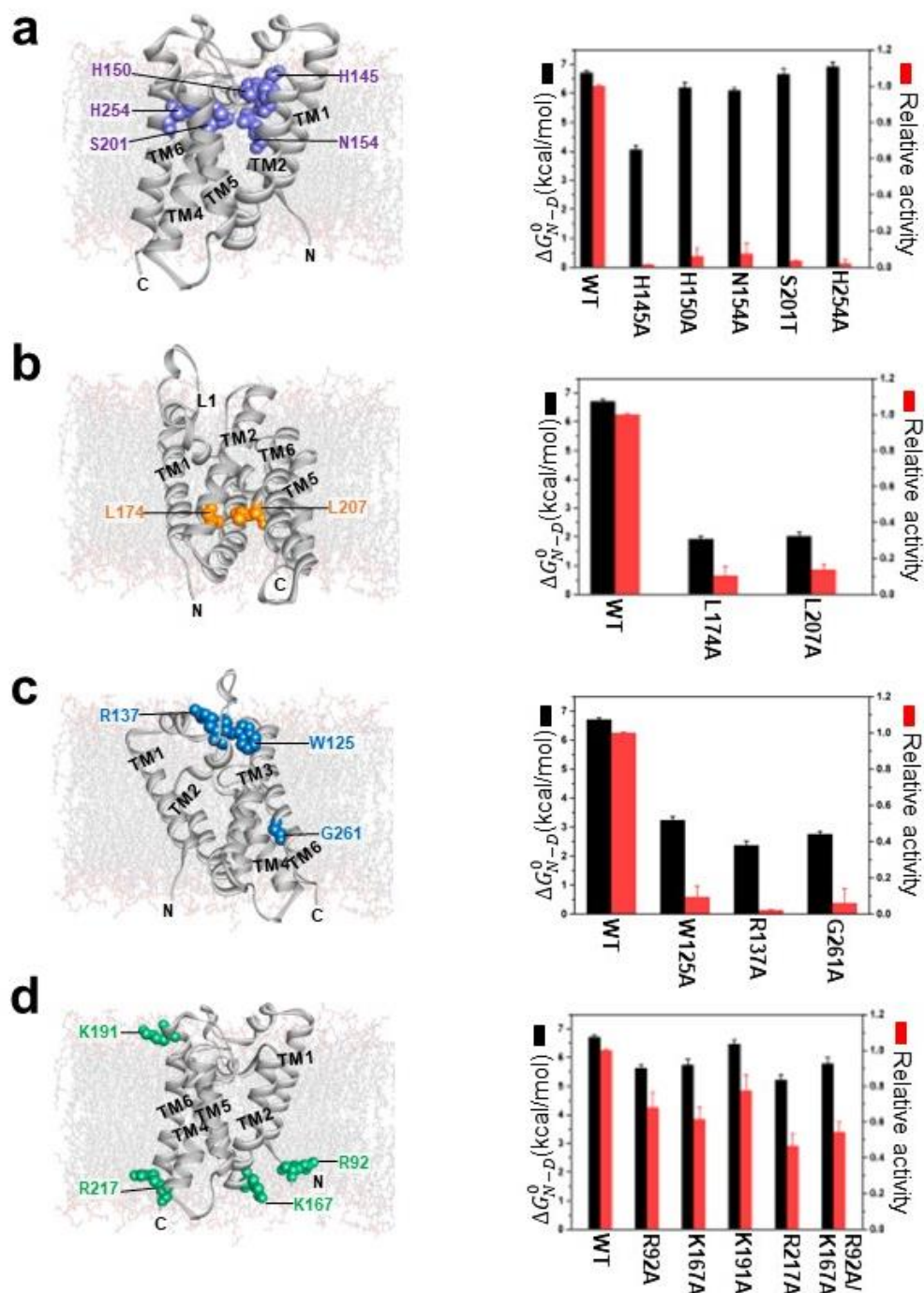


Figure 3.2. Mutational impact on the stability and activity of GlpG. (a-d) (Left) location of targeted mutational sites and (Right) the stability and relative activity of mutants, (a) catalytically important mutations, (b) destabilizing mutations, (c) destabilizing and functionally important mutations, (d) disrupting salt-bridge interactions, intra-protein or between lipid and protein.

Table 3.1. The change in thermodynamic stability ($\Delta\Delta G^{\circ}_{WT-Mut}$), activity (relative to wild type), and residue burial (f_{ASA} : fraction of buried side-chain area) of GlpG variants. Thermodynamic stabilities were measured in DMPC/CHAPS bicelles at 25 °C. Proteolytic activity was measured in DMPC/DMPG/CHAPS bicelles at 25 °C using SN-LacYTM2 as substrate.

Description of mutation	Mutation	f_{ASA}	C-subdomain (172/267 _c)	
			$\Delta\Delta G^{\circ}_{WT-mut}$	Rel.Activity
Destabilizing and functionally important	W125A	0.09	3.5 ± 0.1	0.09 ± 0.06
	R137A	0.04	4.3 ± 0.2	0.02 ± 0.00
	G261A	0	4.0 ± 0.1	0.06 ± 0.08
Destabilizing	L174A	0	4.8 ± 0.1	0.10 ± 0.05
	L207A	0	4.7 ± 0.2	0.14 ± 0.03
Catalytically important	H145A	0	2.6 ± 0.2	0.01 ± 0.00
	H150A	0.01	0.5 ± 0.2	0.06 ± 0.04
	N154A	0	0.6 ± 0.1	0.08 ± 0.06
	S201T	0	0.1 ± 0.2	0.03 ± 0.00
	H254A	0	-0.2 ± 0.2	0.02 ± 0.02
Disrupting salt-bridge interactions, intraprotein or between lipid and protein	R92A	0.98	1.1 ± 0.1	0.68 ± 0.08
	K167A	0.51	1.0 ± 0.2	0.61 ± 0.07
	K191A	0.59	0.2 ± 0.2	0.77 ± 0.09
	R217A	0.45	1.5 ± 0.2	0.47 ± 0.07
	R92A/K167A	N/A	0.9 ± 0.2	0.54 ± 0.06

To determine the kinetic stability of GlpG under native conditions, we tested limited proteolysis by proteinase K (ProK) under the same condition as the degradation assay (in DMPC/DMPG/CHAPS bicelles at 37 °C). ProK is a nonspecific endopeptidase known to proteolyze water-exposed unfolded regions in a protein, not cleaving the regions with secondary structures or buried in the membrane. Therefore, ProK selectively digest denatured proteins when the native and denatured states interconvert at equilibrium.¹ ProK is a highly reactive protease with $k_{cat} = \sim 10^3/\text{min}/\text{enzyme}$ (NEB, Molecular Biology Grade), which exceeds the rate scales of spontaneous refolding ($\sim 1/\text{min}$) and denaturation reaction ($\sim 10^{-3}/\text{min}$) of WT GlpG in bicelles.^{2,3}

Thus, when GlpG and ProK are mixed with an excess molar concentration of ProK (>10 relative to GlpG), the reaction will be further pushed towards the direction of proteolysis of the denatured state and the rate limiting step of GlpG digestion will be the spontaneous denaturation step of GlpG (that is, $k_{\text{proteol,ProK}} \sim k_{\text{denat}}$). Thus, by monitoring the digestion of GlpG by ProK, we can obtain the spontaneous denaturation rate of GlpG and mutants. Here, time-dependent proteolysis was monitored using SDS-PAGE and the spontaneous denaturation rate was determined by quantifying the decrease in gel intensity that corresponds to native GlpG. We determined the lifetime of the native state (*i.e.*, the inverse of the denaturation rate), $\tau_{\text{denat}} = 110 \pm 10$ h for WT (172_N267_M-BtnPyr₂) (**Figure 3.3**). Previously we determined the spontaneous denaturation of GlpG in the same buffer condition (in DMPC/DMPG/CHAPS at 37 °C) using steric trapping by monitoring GlpG activity as a denaturation readout.¹⁷ For C-terminal biotin pair $k_{\text{denat,app}} = 2.7 \pm 0.7 \times 10^{-4} \text{ min}^{-1}$ and for N-terminal biotin pair $k_{\text{denat,app}} = 1.8 \pm 0.5 \times 10^{-4} \text{ min}^{-1}$ was obtained. These denaturation rates are comparable to denaturation rates obtained by limited proteolysis by ProK. By the mutations, the kinetic stability of GlpG changed in a wide range of $\Delta\Delta G^{\ddagger}_{\text{N-D,WT-Mut}} = -0.5$ to 5.0 kcal/mol (**Figure 3.13**).

The plot of the mutation-induced changes in thermodynamic stability ($\Delta\Delta G^0_{\text{N-D,WT-Mut}}$) vs those changes in activation free energy ($\Delta\Delta G^{\ddagger}_{\text{N-D,WT-Mut}} = -RT\ln[k_{\text{denat,Mut}}/k_{\text{denat,WT}}]$) displayed strong correlation of $R^2 = 0.88$ and a slope close to 1 ($m = 1.00 \pm 0.12$) (**Figure 3.4**). That is, the decrease in GlpG stability by mutation is realized as the same decrease in energy barrier towards the denatured state. According to the transition state theory of protein folding using ϕ -value analysis, the slope of 1 in a $\Delta\Delta G^0_{\text{N-D,WT-Mut}}$ vs $\Delta\Delta G^{\ddagger}_{\text{N-D,WT-Mut}}$ plot is an indication that the conformation of the folding transition state resembles that of the denatured state.³²

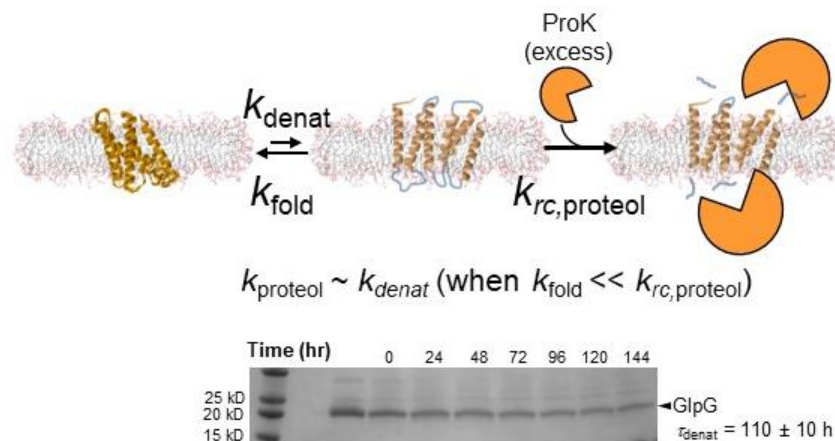


Figure 3.3. Spontaneous denaturation of GlpG monitored by ProK digestion. GlpG reacts with ProK for various incubation times in 2% DMPC/DMPG/CHAPS bicelles at 37 °C. The denaturation rates (or the lifetime of the denatured state, τ_{denat}) are obtained by analyzing the decrease of the band intensities that correspond to GlpG SDS-PAGE. $k_{\text{rc,proteol}}$: the rate constant for proteolysis of random coil; k_{proteol} : the rate constant for apparent proteolysis as detected by SDS-PAGE.

3.3.2. Degradation rates of the membrane protein GlpG are independent of the conformational stability.

The GlpG variants with differing thermodynamic stabilities and intrinsic denaturation rates allowed us to test how protein's conformational stability affects the rate of degradation by FtsH. To answer this question, degradation of GlpG was induced by fusing the protein to an FtsH-specific C-terminal degradation marker, SLLWS, known as 108 tag (GlpG-108) (**Figure 3.5.a**).^{17,33} Degradation rates were measured *in vitro* in DMPC/DMPG/CHAPS bicelles at 37 °C using the fluorescence-based assay (**Figure 3.5.b right**). GlpG-108 was conjugated to NBD-fluorophore at G172C in the middle helix TM3. As the NBD label is released from the bilayer phase to the aqueous phase upon degradation, the fluorescence intensity decreases. Degradation assays (Methods) were carried out at a fixed FtsH concentration varying the concentrations of NBD-

labelled GlpG WT or mutants in the mol fraction unit of GlpG, X_{GlpG} . Michaelis-Menten kinetic analyses yielded $K_{\text{m,deg}} = (1.7 \pm 0.2) \times 10^{-5} X_{\text{GlpG}}$ and $k_{\text{cat,deg}} = (2.4 \pm 0.0) \times 10^{-1} \text{ FtsH}_6^{-1}$ (**Figure 3.5.b right**).

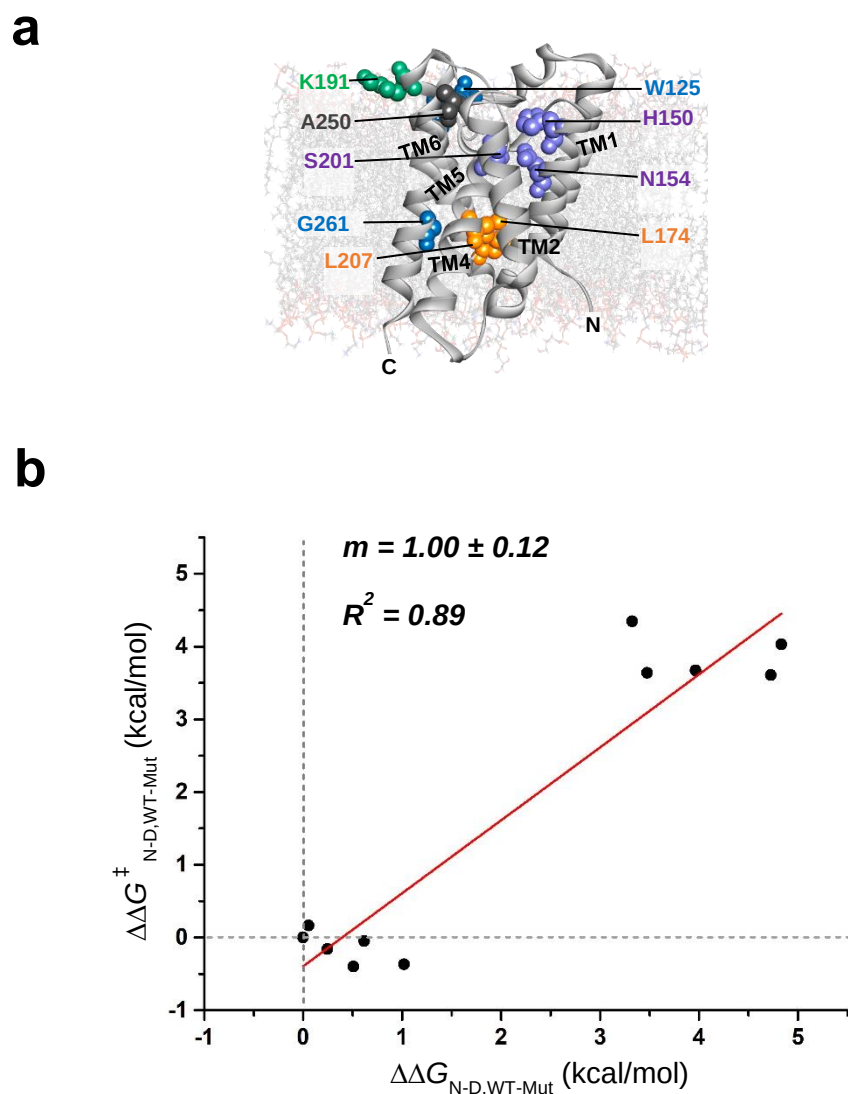
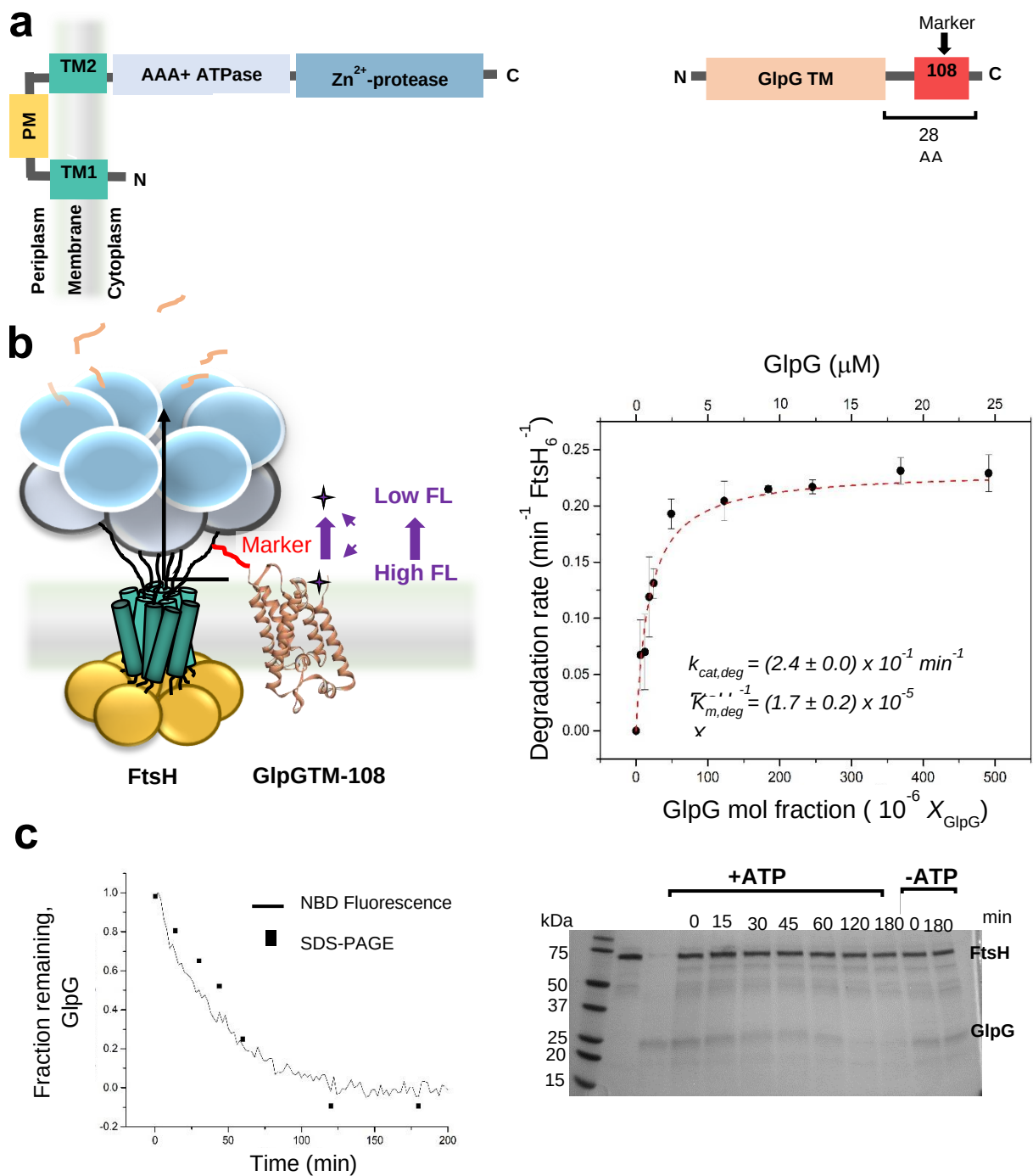


Figure 3.4. (a) Targeted mutations to study thermodynamic stability and spontaneous denaturation rates in bicelles. Mutations are color-coded as described in Fig 3.1. A250 in loop 5 at the membrane-water interface was mutated to lysine or arginine. **(b) The correlation plot of the change in thermodynamic stability ($\Delta\Delta G_{\text{N-D,WT-Mut}}$) vs the change in activation free energy of spontaneous denaturation ($\Delta\Delta G^{\ddagger}_{\text{N-D,WT-Mut}} = -RT \ln[k_{\text{denat, Mut}}/k_{\text{denat, WT}}]$)**



data of GlpG-108 (10 mM) by FtsH (2mM) in 3% DMPC/DMPG/CHAPS bicelles at 37 °C. (left) Comparison of the fluorescence and SDS-PAGE assays from right. (right) Degradation of GlpG monitored by SDS-PAGE and Coomassie blue staining.

The plot of the mutation-induced changes in thermodynamic stability *vs* those changes in activation free energy of degradation ($\Delta\Delta G^0_{N-D,WT-Mut}$ *vs* $\Delta\Delta G^{0\dagger}_{deg,WT-Mut} = -RT\ln[k_{cat,deg,WT}/k_{cat,deg,Mut}]$) yielded a slope of 0.028 ± 0.026 with poor correlation of $R^2 = 0.18$ (**Figure 3.6**). Also, the plot of the mutation-induced changes in activation free energy of spontaneous denaturation *vs* those changes in activation free energy of degradation ($\Delta\Delta G^{0\dagger}_{N-D,WT-Mut}$ *vs* $\Delta\Delta G^{0\dagger}_{deg,WT-Mut}$) yielded a slope of 0.034 ± 0.018 again with poor correlation, $R^2 = 0.27$ (**Figure 3.6**). Surprisingly, these results indicate that the conformational stability (both thermodynamic and kinetic) of a membrane substrate does not affect the rate of degradation mediated by the membrane-integrated ATP-driven protease FtsH, implying that the substrate denaturation is not the rate-limiting step in the degradation reaction. In contrast, for degradation of a stable water-soluble protein titin I-27 by ClpXP, the slopes were larger, and the correlation was stronger ($m = 0.451 \pm 0.133$ with $R^2 = 0.68$ in the $\Delta\Delta G^0_{N-D,WT-Mut}$ *vs* $\Delta\Delta G^{0\dagger}_{deg,WT-Mut}$ plot and $m = 0.309 \pm 0.097$ with $R^2 = 0.65$ in the $\Delta\Delta G^{0\dagger}_{N-D,WT-Mut}$ *vs* $\Delta\Delta G^{0\dagger}_{deg,WT-Mut}$ (**Figure 3.14**). Therefore, the dependence of degradation rate on substrate conformational stability fundamentally differs between membrane and water-soluble substrates.

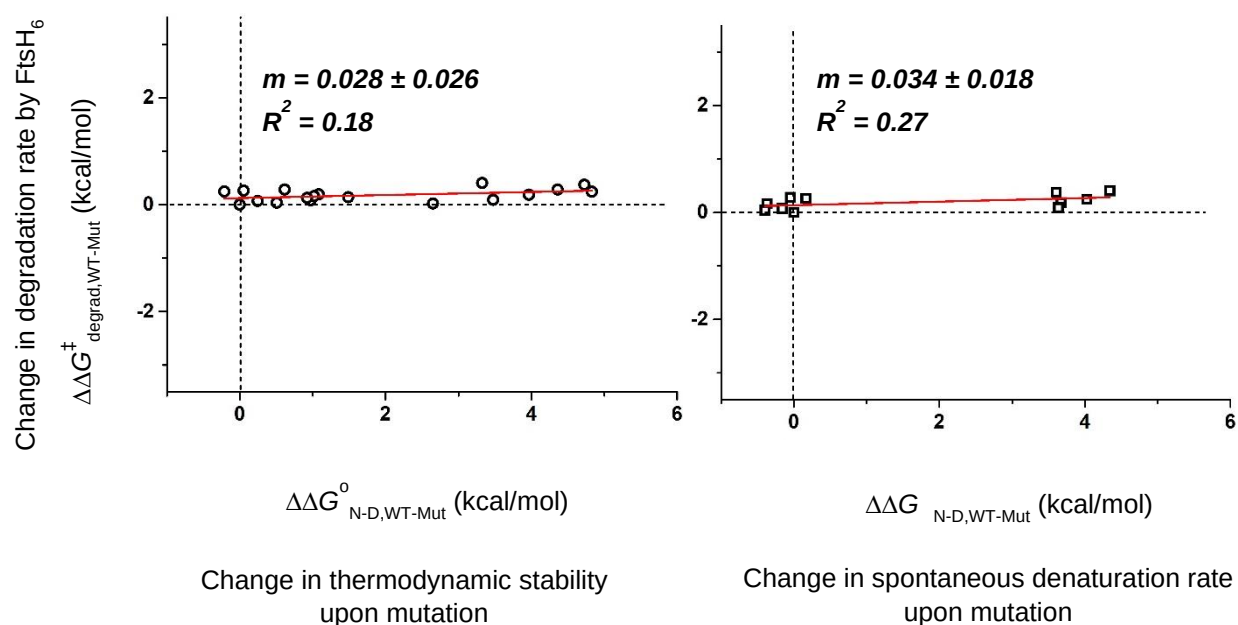


Figure 3.6. The correlation plots of degradation rate by FtsH₆ vs conformational stability of GlpG.

3.3.3. The hydrophobicity of TM segments and interfacial loop residues contributed to the degradation rate.

Next, we tested the effect of substrate hydrophobicity on the degradation rate, which would resist membrane dislocation during degradation. We designed mutations on two regions (**Figure 3.7**). First, the lipid-contacting residues in TM6 (G252L/A253V/A256L/G257L/A265V/S269L: “LVLLVL” and G252L/A253V/A265V/S269L: “LVVL”) not disrupting the tertiary packing interactions near the degradation marker.¹⁷ These sets of mutation increased the hydrophobicity of TM6 by 9.9 and 5.7 kcal/mol, respectively, which were estimated based on the Wimley-White water-octanol scale. The degradation was decelerated by 30 to 40% (*i.e.*, the decrease in $k_{\text{cat,deg,LVLLVL}}$). Second, we tested the impact of increasing the hydrophilicity of the periplasmic loop (L5) that precedes the C-terminal TM helix TM6. Either positively charged lysine or arginine

were introduced on the solvent-exposed neutral residues (F242, G246, and A250, **Figure 3.8**). The hydrophilicity increased for the resulting Arg triple mutants (F242R, G246R, and A250R) by 3.4 kcal/mol (based on Wimley-White interfacial scale) and for the Lys triple mutants (F242K, G246K, and A250K) by 3.9 kcal/mol. We expected that the introduction of the multiple positively charged residues in the periplasmic loop closest to the degradation marker would decrease the degradation rate because it will increase the energy barrier of membrane dislocation across the membrane toward the cytosolic protease domain of FtsH. Indeed, in both cases, $k_{\text{cat,deg}}$ decreased by 50%. In summary, increasing the hydrophobicity or increasing the hydrophilicity in the *cis* side of the protease domain of FtsH near the degradation marker showed resistance to degradation probably by decelerating the dislocation rate from the hydrophobic membrane. However, we note that the free energy changes by modifying the hydrophobicity or hydrophilicity were not fully realized in the increase in the activation free energy barrier of degradation.

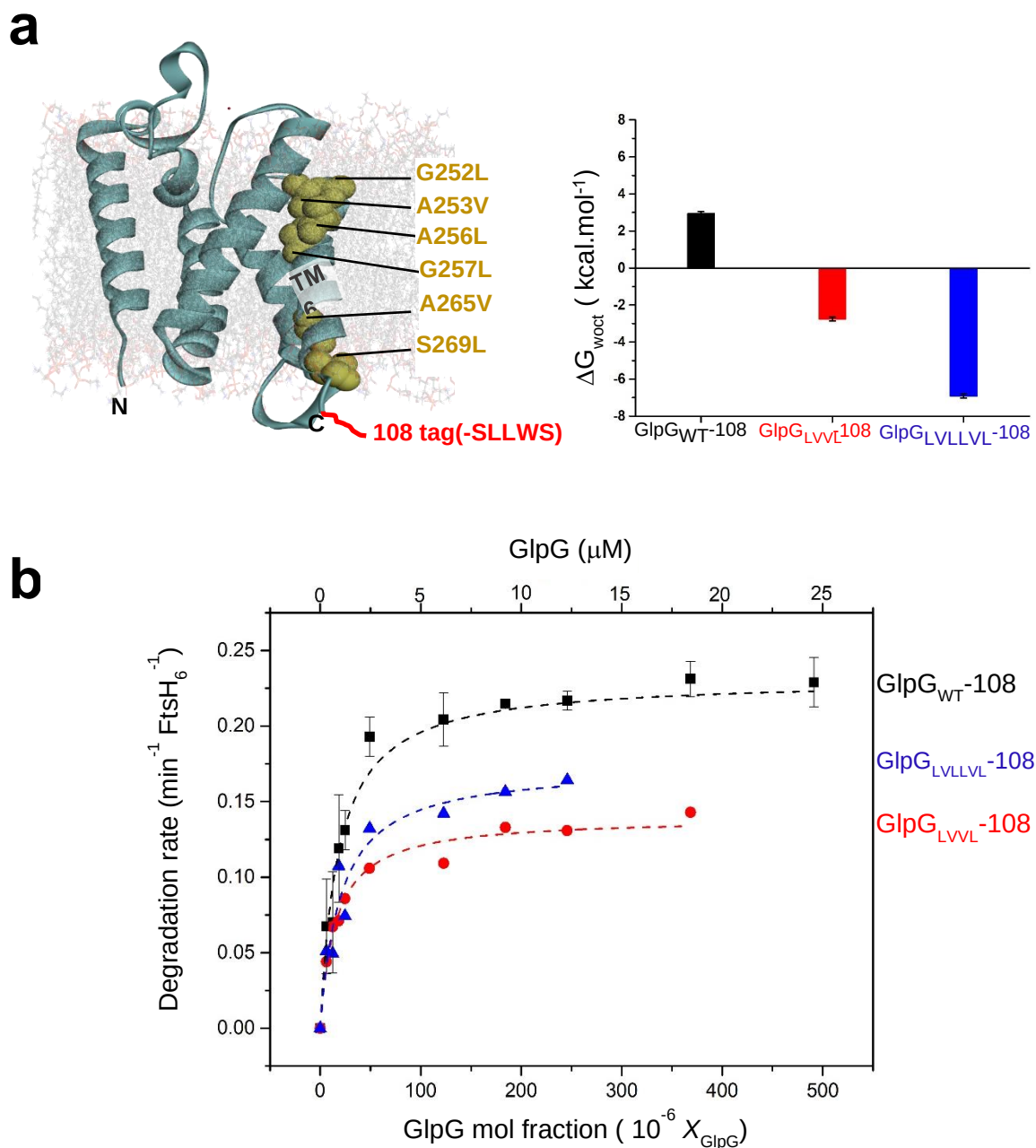


Figure 3.7. Effect of increased hydrophobicity on degradation: (a) (left) The residues for amino acid substitutions that increase the hydrophobicity of TM6 near the degradation marker. (right) The changes in hydrophobicity of TM6 as determined by the Wimley-White water-octanol scale. (b) The effect of the increased hydrophobicity on the degradation rate.

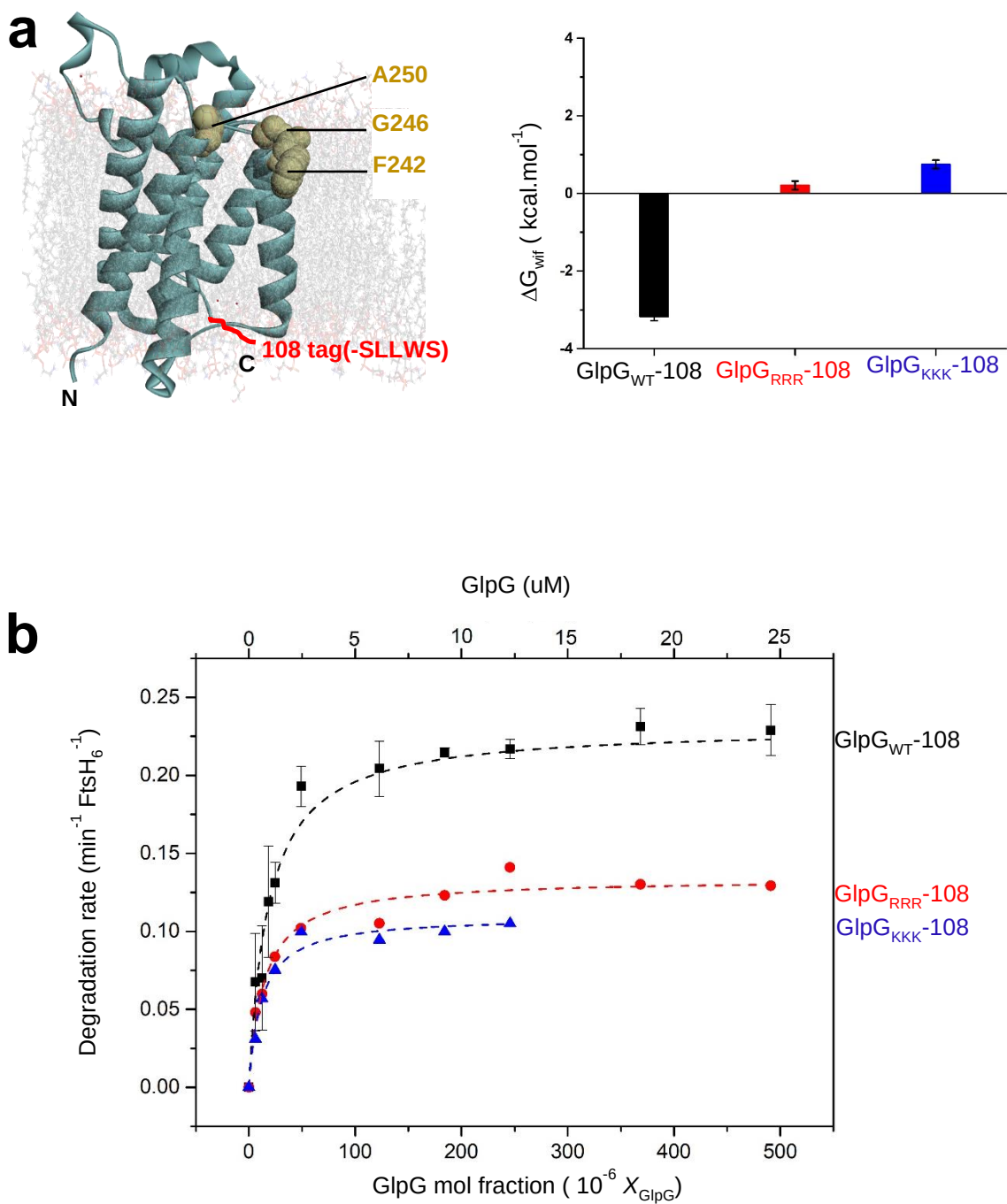


Figure 3.8. Effect of increased hydrophilicity on degradation: (a) (left) The residues for amino acid substitutions that increase the hydrophilicity of the periplasmic loop (L5, the *cis* position to the cytosol) preceding the C-terminal TM helix 6. (right) The change in hydrophobicity of loop 5 as determined by the Wimley-White interfacial scale. (b) The effect of the decreased hydrophobicity on the degradation rate.

3.3.4. Conserved prolines in iRhoms are highly destabilizing but its effect is compensated by a combination of inactivating mutations and evolved to be degraded similarly to active rhomboids.

Finally, we further investigated the mutational effects of the functionally important residues around the active site of GlpG on their degradation rate. We benchmarked the evolutionary traces of a subset of rhomboid homologs classified as “iRhoms”, which are proteolytically inactive rhomboids present in metazoans.³⁴⁻³⁶ Although their specific biological roles and mechanisms remain unclear, they are involved in multiple quality control or regulatory pathways, assisting degradation of signaling proteins,⁴⁷ chaperoning membrane-bound proteases (ADAM17/TACE in inflammatory signaling pathway),⁴⁸ or participating in the endoplasmic reticulum (ER) retro translocation complex for endoplasmic reticulum-assisted degradation (ERAD).³⁷

From multi-sequence alignment of nine metazoan iRhoms and *E.coli* GlpG (**Figure 3.9.**),³⁶ we identified several key sequence modifications on iRhoms relative to the sequence of *E. coli* GlpG. The conserved modifications were concentrated on the three catalytically critical residues, S201 and H254 (hereafter, the residue numbers designate those of *E. coli* GlpG) which form the catalytic dyad, and N154 which serves as an oxyanion hole during proteolysis. Interestingly, all iRhoms possess a proline residue at the position of L200 within one of the rhomboid consensus motifs (G₁₉₉X₂₀₀S₂₀₁G₂₀₂ in the periplasmic L3 loop, where X denote various residues and S201 is the catalytic serine). The sets of modified sequences included N154S/L200P/S201A (*Homo sapiens* iRhom’s 1 and 2, *Mus musculus* iRhom’s 1 and 2, and *Danio rerio* iRhom), N154S/L200P (*Caenorhabditis elegans* iRhom 1), N154T/L200P (*Caenorhabditis elegans* iRhom 2), N154T/L200P/H254L

(*Drosophila melanogaster* iRhom), and N154S/L200P/S201L/H254L (*Anopheles gambiae* iRhom). The phylogenetic analysis suggests that the proline residue has been acquired before the loss of the catalytic dyad residues.³⁰

We generated a series of single to quadruple mutants on *E.coli* GlpG mimicking the sequence modifications on iRhoms and measured their thermodynamic stabilities in DMPC/CHAPS bicelles ($q = 1.5$) using steric trapping. $\Delta\Delta G^{\circ}_{N-D,WT-Mut}$'s of the single mutants, N154S, N154T, S201L, and S201A, were 1.0 ± 0.2 , 1.7 ± 0.2 , 1.5 ± 0.1 , and 0.7 ± 0.1 kcal/mol, respectively, displaying moderate destabilization (**Figure 3.10**). L200P, the common modification on iRhoms induced the largest destabilization with $\Delta\Delta G^{\circ}_{N-D,WT-Mut} = 4.8 \pm 0.1$ kcal/mol. Interestingly, the substantial stability loss by L200P was compensated by the additional mutations on the catalytically important residues, N154S(or T), S201A(or L) or H254L. These mutations induced the stability gain of $\Delta\Delta G^{\circ}_{N-D,L200P-Mut} = -0.5$ to -1.1 kcal/mol relative to L200P except for the quadruple mutation N154T/L200P/S201L/H154L, which further destabilized the protein by $\Delta\Delta G^{\circ}_{N-D,L200P-Mut} = \sim 1.0$ kcal/mol. Probably, introducing bulky Leu residues on the structurally proximal S201 and H254L induced sterically unfavorable positioning of the two residues, which cannot compensate the stability loss by L200P.

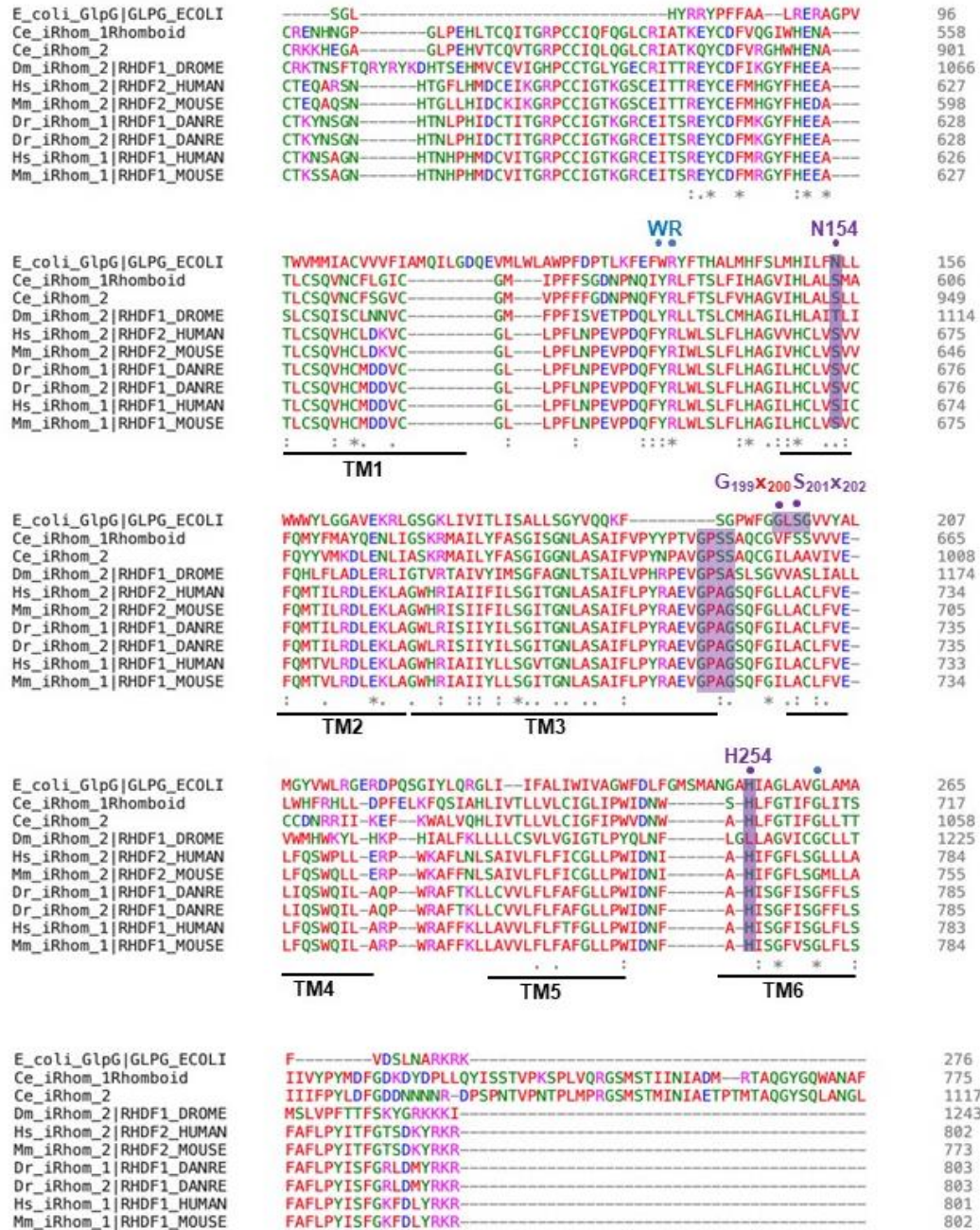


Figure 3.9. Multiple-sequence alignment of metazoan iRhoms to *E. coli* GlpG. Additional N and C-terminal domains in iRhoms are excluded from the alignment. The predicted transmembrane helices are based on the crystal structure of *E. coli* GlpG and underlined in black. The residues involved in the catalysis are labeled in purple and represented by purple dots. The highly conserved and functionally important residues of *E. coli* GlpG are represented by blue dots. Amino acid color

code reflects the physiochemical properties of side chains: Red; small + hydrophobic, Blue; acidic, Magenta; basic, Green; hydroxyl/sulphydryl/amine/glycyl. In iRhoms, the residue position 200 (from *E. coli* GlpG) is absolutely conserved as Pro. The database accession numbers are: *E. coli* GlpG, Swiss-Prot P909392; *Homo sapiens* (Hs) iRhom 1: Swiss-Prot Q96CC6; iRhom 2: Swiss-Prot Q6PJF5; *Mus musculus* (Mm) iRhom 1: Swiss-Prot Q6PIX5; iRhom 2: Swiss-Prot Q80WQ6, *Danio rerio* (Dr) iRhom 1: Swiss-Prot Q6GMF2; *Caenorhabditis elegans* (Ce) iRhom 1: GenBank NP_503013, iRhom 2: GenBank NP_001041013.

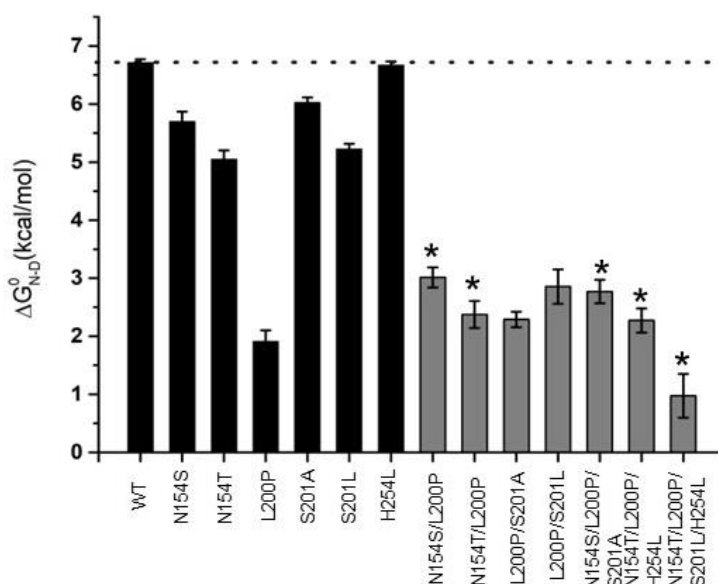
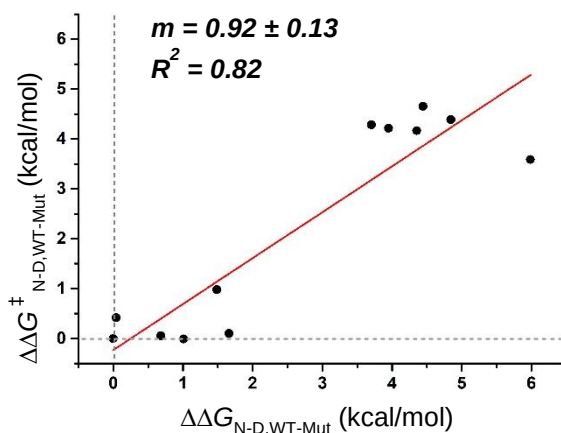


Figure 3.10. Thermodynamic stabilities of GlpG mutants measured by steric trapping. Black bars represent the point mutations while dark grey bars represent the double, triple, and quadruple mutations on the WT template (WT: 172_N267_M-BtnPyr₂). Asterisk marks (*) represent the naturally existing sequence modification of the active site residues in iRhoms.

Next, we challenged GlpG harboring the sequence modifications that are conserved in iRhoms to the FtsH degradation machinery. We again assessed the correlation between the thermodynamic and kinetic stability of the mutants and their degradation rate, respectively (**Figure 3.11**). The $\Delta\Delta G^{\circ}_{\text{N-D,WT-Mut}}$ vs $\Delta\Delta G^{\circ\dagger}_{\text{deg,WT-Mut}}$ plot yielded $m = 0.006 \pm 0.037$ with $R^2 = 0.08$, and the $\Delta\Delta G^{\circ\dagger}_{\text{N-D,WT-Mut}}$ vs $\Delta\Delta G^{\circ\dagger}_{\text{deg,WT-Mut}}$ plot yielded $m = 0.042 \pm 0.038$ with $R^2 = 0.37$. Therefore, despite the drastic disruption of conformational stability and activity by the mutations, the degradation rate remained constant.

a



b

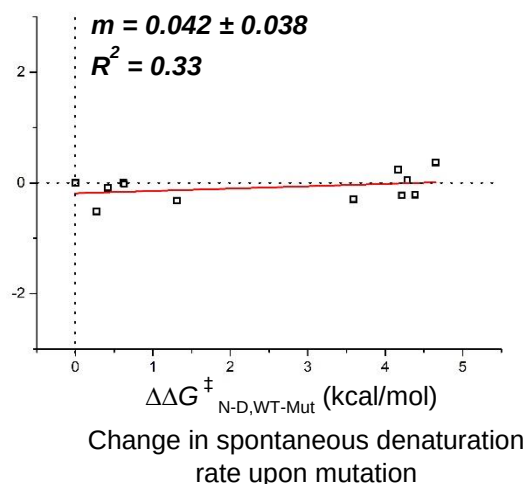
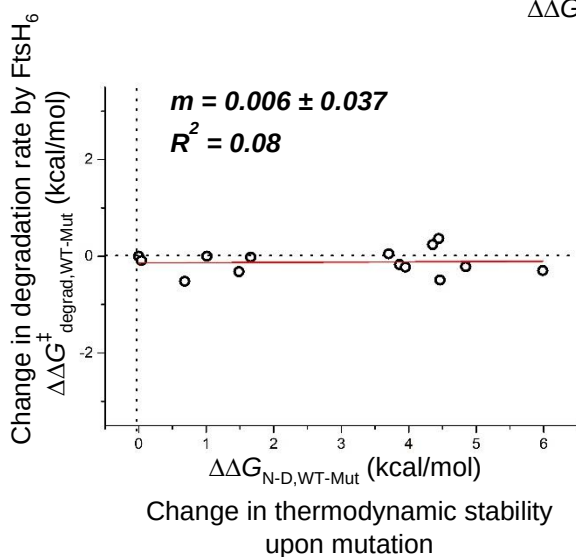


Figure 3.11. Correlation plots for iRhom sequence modifications: (a) The plot of spontaneous denaturation rate vs thermodynamic stability for iRhom sequence modifications. (b) The correlation plots of the degradation rate by FtsH vs conformational stability of GlpG.

3.4. Discussion

Folding status of proteins has been thought to be one of the major surveillance criteria of cellular degradation machinery. While this paradigm remains valid, the quantitative folding-degradation relationship has not been rigorously defined for a specific degradation system which preferentially degrades membrane proteins. Strikingly, we find that either the thermodynamic or kinetic conformational stability does not affect the degradation rate of the model helical-bundle membrane protein GlpG.

Previously, we have proposed a three-step model for membrane protein degradation:¹⁷ 1) engagement of the degradation marker to the pore loop of the AAA+ domain, 2) active denaturation of the membrane-bound substrate within the membrane, and 3) active membrane dislocation of the denatured substrate followed by proteolysis. It has not been clear which ATP-driven step is rate-limiting, that is, substrate denaturation or membrane dislocation that poses major thermodynamic and kinetic uphill in the free energy landscape of membrane protein folding. Our result strongly support that the rate-limiting step of membrane protein degradation is not substrate denaturation. Rather, the increase in hydrophobicity of the TM segments or the incorporation of positively charged residues into the periplasmic loop (the *cis* position to the proteolytic active site in the cytosol) significantly decelerates degradation. Therefore, it is likely that the rate limiting step is the membrane dislocation. The dependence of the hydrophobicity of a TM helix on the membrane-dislocation rate has observed for the ERAD pathway mediated by the AAA+ enzyme p97.³⁸

The unfolding energy landscape of membrane proteins is distinct from globular water-soluble proteins. While the native-to-denatured state transition occurs in the same aqueous phase for water-soluble proteins, that for membrane proteins involves the highly unfavorable transfer of

hydrophobic TM segments from the membrane to water in addition to the denaturation in the membrane.¹⁷ Single-molecule force spectroscopy, computational simulation, and steric trapping studies of GlpG folding show that the thermodynamic stability ($\Delta G^{\circ}_{\text{N-D}}$) is -4 to -9 kcal/mol^{22,31,39-42} with the energy barrier, $\Delta G^{\ddagger}_{\text{N-D}} = 12$ to 15 kcal/mol^{31,39} or the Arrhenius activation energy of denaturation, $E_{\text{a,N-D}} = 30$ to 40 kcal/mol,¹⁷ and the free energy change upon membrane dislocation, $\Delta G^{\circ}_{\text{dislocation}} = \sim 360$ kcal/mol with the energy barrier unknown.¹⁷ Other studies indicate that $\Delta G^{\circ}_{\text{N-D}}$'s of various helical membrane proteins fall in the range of -4 to -10 kcal/mol,⁴³ similar to that of globular proteins, and $\Delta G^{\circ}_{\text{dislocation}}$ amounts to 30 to 50 kcal/mol/TM helix.⁴⁴ Therefore, it may not be a surprise that membrane dislocation is the rate-limiting step since it requires ~ 50 times more work (ΔG°) than denaturation in the membrane, surpassing both $\Delta G^{\circ}_{\text{N-D}}$ and $\Delta G^{\ddagger}_{\text{N-D}}$. Our previous study quantifies that FtsH hydrolyzes 380 to 550 molecules of ATP (0.5 to 1.0 ATP hydrolysis/residue) to degrade a single copy of GlpG, which is comparable to the ATP cost for ClpXP or ClpAP in degrading water-soluble proteins (0.5 to 6.0 ATP hydrolysis/residue).^{17,18}

Taken together, these results indicate that FtsH is a molecular machine that degrades a given membrane protein at a constant rate regardless of their conformational stability and impressively, efficiently utilizes the free energies generated from ATP hydrolysis to overcome the dual energetic challenge, substrate denaturation and membrane dislocation. The result further supports our previous suggestion that FtsH has strong unfoldase activity such that a variation in conformational stability is not noticeably reflected in the apparent degradation rate.

How is the degradation activity of FtsH compared to that of water-soluble AAA+ proteases? Degradation rates of the water-soluble protein titin-I27 by ClpXP, which is known as a strong ATPase/unfoldase, moderately depend on the stability (**Figure 3.14.**)² In contrast, ClpXP degrades Arc variants with various thermodynamic stabilities ($\Delta G^{\circ}_{\text{N-D}} = -0.4$ to -14.6

kcal/mol) at a constant rate.³ Furthermore, ClpXP and ClpAP (both strong ATPase/unfoldase) degrade a circular-permuted variant (CP) of *E. coli* DHFR, CP-Pro25 (the new N and C-termini are created at Pro25), faster than CP-Lys38 although CP-Pro25 is more stable than CP-Lys38.²⁶ In this study, the two variants have an N-terminal degradation marker and the local structure to which the marker is fused have different structures due to the circular permutation. Also, ClpAP degrades inhibitor-bound barnase faster than inhibitor-bound mouse DHFR although the former is more stable than the latter.²⁶ Interpretation of these various degradation patterns displayed by the same enzyme have led to a set of rules for predicting the susceptibility of a given substrate to degradation:^{19,26,45} 1) The local stability of the region to which a degradation marker is fused is more important than the global stability; 2) When the secondary structure near the degradation marker is α -helical, which is stabilized by local hydrogen-bonding interactions, the degradation rates is facilitated regardless of the global stability; 3) When the secondary structures near the marker are similar, the protein that unfolds faster is degraded faster.

Interestingly, we have observed that FtsH degrades GlpG with the C-terminal degradation marker ~2 times faster than the same GlpG with the N-terminal marker.¹⁷ At the time of publication, we attributed this discrepancy to the difference in local stability (*i.e.*, C-subdomain is less stable than N-subdomain).^{17,22} This conclusion seems contradictory to our current result that the degradation rate of GlpG (with the C-terminal marker) is independent of conformational stability. Recently, we have shown that, when GlpG is denatured in the membrane, the less hydrophobic TM helices (TM3, TM4 and TM6) in C-subdomain transiently unfold and become water-exposed at the membrane surface while the more hydrophobic TM1 and TM2 in N-subdomain is largely integrated in the membrane.⁴⁶ When GlpG engaged with FtsH is denatured by ATP-hydrolysis *and* the degradation marker is fused to the C-terminus, C-subdomain will be

exposed to water and thus, FtsH does not have to pay much $\Delta G^{\circ}_{\text{dislocation}}$ for dislocating C-subdomain. In contrast, when the degradation marker is fused to N-subdomain, FtsH should pay more $\Delta G^{\circ}_{\text{dislocation}}$ to dislocate N-subdomain. Therefore, when the degradation is initiated, the degradation rate of GlpG with the C-terminal marker will be faster due to the smaller $\Delta G^{\circ}_{\text{dislocation}}$. By involving the conformational distribution of the denatured state ensemble, the apparent discrepancy regarding the stability dependence of degradation rates can be resolved.

Finally, we designed the mutations on the active site residues of *E. coli* GlpG following the evolutionary traces of iRhom's (proteolytically-inactive pseudo-rhomboids) and challenged the variants to the FtsH degradation machinery. We unexpectedly find that the mutation of Leu200 to Pro (in the rhomboid consensus sequence, G₁₉₉X₂₀₀S₂₀₁: the residue numbers are from *E. coli* GlpG), which is absolutely conserved in iRhom's, dramatically destabilizes and inactivates GlpG, but the stability loss is rescued by any mutations on the active site residues, Asn154, Ser201 and His254. This result implies that the activity loss is sufficient by substituting Pro in the consensus motif and the substitutions on N154, S201 and H254 are beneficial to stability for function. We cannot exclude the possibility that this stability-changing pattern may stem from the unique structural context of *E. coli* GlpG, which serves as a template for the mutations. However, since various combinations of residue substitution on the active site residues (N154T/S, S201T/L, or H254L) commonly lead to a similar degree of stabilization, it is likely that the same effect would occur in iRhom's. We initially expected that the stability-changing pattern by iRhom mutations would be correlated with their degradation rate by FtsH, mimicking the natural selection process of iRhom's during evolution. However, all mutants are degraded at the same rate. This result may imply that if energy-dependent degradation machinery participates in the evolutionary process of

membrane proteins, it would impose the same pressure toward degradation regardless of the impacts of amino acid substitutions on stability or function.

APPENDICES

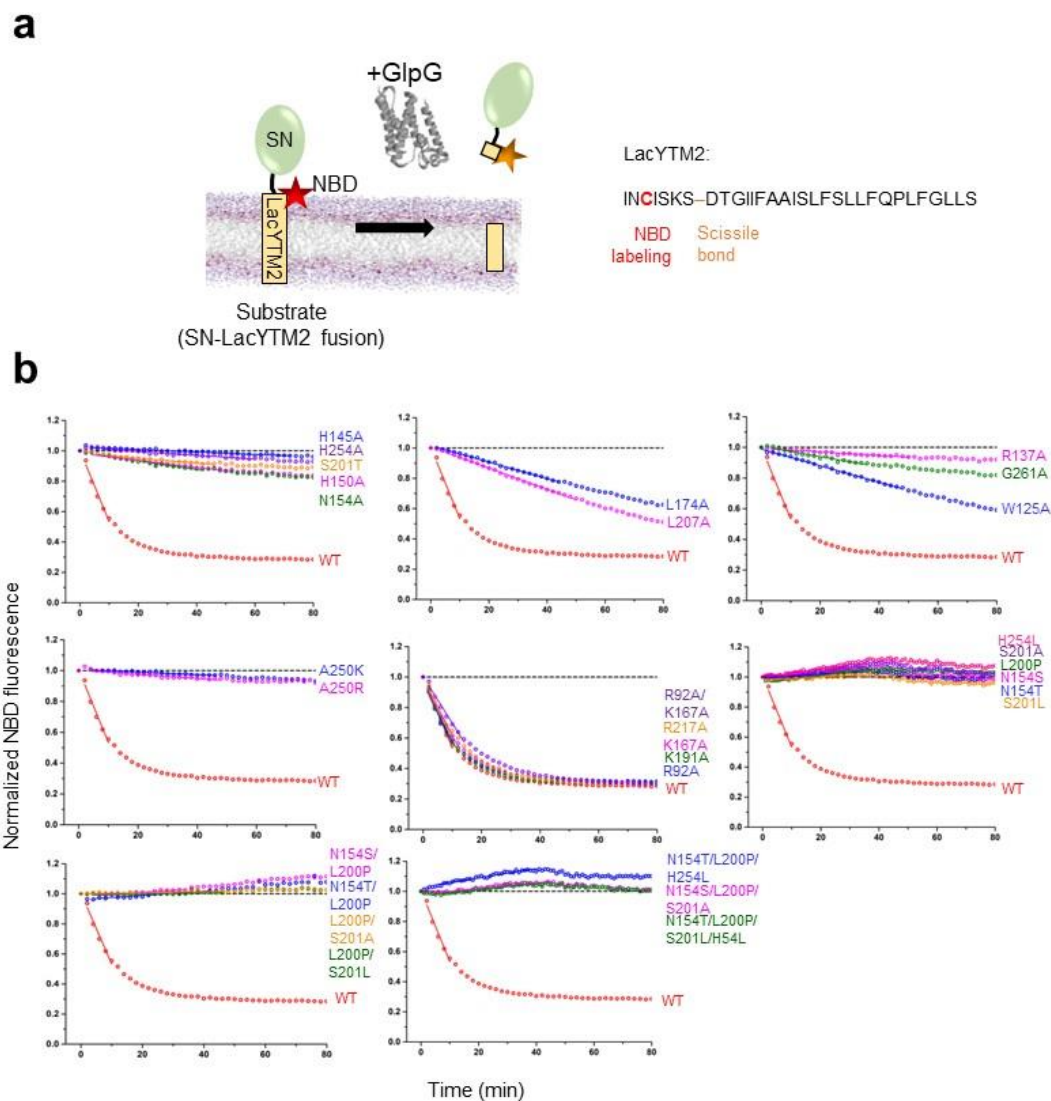


Figure 3.12. Proteolytic activity of GlpG variants: (a) NBD-fluorescence assay to measure the proteolytic activity of GlpG. SN-LacYTM2 labeled with the environment-sensitive NBD fluorophore in the P5 position upstream of the scissile bond. After cleaved by GlpG, NBD will be released into the aqueous phase, leading to the quenching of NBD fluorescence.

(b) Activity assay of GlpG variants. The activity was measured in 2% DMPC/DMPG/CHAPS bicelles at room temperature. The initial slope of the change in NBD fluorescence vs time indicates the proteolytic activity of GlpG.

Table 3.2. Kinetic parameters for GlpG degradation by FtsH. All measurements were performed with 2 mM FtsH in 3% DMPC/DMPG/CHAPS bicelles at 37 °C. $k_{\text{cat, deg}}$, and $K_{\text{m, deg}}$ values were obtained by fitting the data to the Michaelis-Menten equation.

Description of mutation	Mutation	$k_{\text{cat, deg}}$ ($\text{min}^{-1} \text{ FtsH}_6^{-1}$)	$K_{\text{m, deg}}$ (X_{GlpG})
	WT	$2.4 \pm 0.0 \times 10^{-1}$	$1.7 \pm 0.2 \times 10^{-5}$
Destabilizing and functionally important	W125A	$2.7 \pm 0.1 \times 10^{-1}$	$2.1 \pm 0.3 \times 10^{-5}$
	R137A	$3.8 \pm 0.3 \times 10^{-1}$	$7.3 \pm 1.8 \times 10^{-5}$
	G261A	$3.2 \pm 0.2 \times 10^{-1}$	$2.0 \pm 0.6 \times 10^{-5}$
Destabilizing	L174A	$3.6 \pm 0.4 \times 10^{-1}$	$11.2 \pm 3.5 \times 10^{-5}$
	L207A	$4.4 \pm 0.1 \times 10^{-1}$	$5.4 \pm 0.6 \times 10^{-5}$
Catalytically important	H145A	$2.4 \pm 0.1 \times 10^{-1}$	$1.3 \pm 0.4 \times 10^{-5}$
	H150A	$2.5 \pm 0.1 \times 10^{-1}$	$1.4 \pm 0.2 \times 10^{-5}$
	N154A	$3.8 \pm 0.5 \times 10^{-1}$	$9.6 \pm 3.5 \times 10^{-5}$
	S201T	$3.7 \pm 0.1 \times 10^{-1}$	$5.7 \pm 0.8 \times 10^{-5}$
	H254A	$3.6 \pm 0.2 \times 10^{-1}$	$3.4 \pm 0.8 \times 10^{-5}$
Disrupting the salt-bridge interactions, intraprotein or between protein and lipid	R92A	$3.3 \pm 0.1 \times 10^{-1}$	$2.7 \pm 0.5 \times 10^{-5}$
	K167A	$2.7 \pm 0.1 \times 10^{-1}$	$1.4 \pm 0.4 \times 10^{-5}$
	K191A	$2.7 \pm 0.1 \times 10^{-1}$	$1.8 \pm 0.4 \times 10^{-5}$
	R217A	$3.0 \pm 0.2 \times 10^{-1}$	$2.2 \pm 0.7 \times 10^{-5}$
	R92A/K167A	$2.9 \pm 0.4 \times 10^{-1}$	$1.8 \pm 1.1 \times 10^{-5}$
Increasing hydrophilicity of the loop 5	A250K	$3.1 \pm 0.3 \times 10^{-1}$	$3.3 \pm 1.1 \times 10^{-5}$
	A250R	$4.7 \pm 0.3 \times 10^{-1}$	$3.7 \pm 1.0 \times 10^{-5}$

Table 3.3. Kinetic parameters of GlpG degradation by FtsH with iRhom mutations. All measurements were performed with 2mM FtsH in 3% DMPC/DMPG/CHAPS bicelles at 37 °C. $k_{\text{cat, deg}}$ and $K_{\text{m, deg}}$ values were obtained by fitting the data to the Michaelis-Menten equation.

Description of mutation	Mutation	$k_{\text{cat,deg}}$ ($\text{min}^{-1} \text{ FtsH}_6^{-1}$)	$K_{\text{m,deg}}$ (X_{GlpG})
	WT	$2.4 \pm 0.0 \times 10^{-1}$	$1.7 \pm 0.2 \times 10^{-5}$
Single	S201A	$1.0 \pm 0.1 \times 10^{-1}$	$1.3 \pm 0.9 \times 10^{-5}$
	S201L	$1.4 \pm 0.3 \times 10^{-1}$	$1.5 \pm 1.3 \times 10^{-5}$
	L200P	$1.6 \pm 0.1 \times 10^{-1}$	$2.3 \pm 0.9 \times 10^{-5}$
	N154S	$2.4 \pm 0.4 \times 10^{-1}$	$5.4 \pm 3.4 \times 10^{-5}$
	N154T	$2.3 \pm 0.3 \times 10^{-1}$	$6.9 \pm 2.8 \times 10^{-5}$
	H254L	$2.0 \pm 0.3 \times 10^{-1}$	$4.0 \pm 1.9 \times 10^{-5}$
Double	L200P/S201A	$4.4 \pm 1.1 \times 10^{-1}$	$3.5 \pm 3.3 \times 10^{-5}$
	L200P/S201L	$1.8 \pm 0.2 \times 10^{-1}$	$1.8 \pm 0.8 \times 10^{-5}$
	N154S/L200P	$2.6 \pm 0.2 \times 10^{-1}$	$2.6 \pm 0.8 \times 10^{-5}$
	N154T/L200P	$3.5 \pm 0.4 \times 10^{-1}$	$4.6 \pm 2.1 \times 10^{-5}$
Triple	N154S/L200P/S201A	$1.6 \pm 0.2 \times 10^{-1}$	$2.4 \pm 1.4 \times 10^{-5}$
	N154T/L200P/H254L	$3.6 \pm 0.4 \times 10^{-1}$	$7.0 \pm 2.5 \times 10^{-5}$
Quadrupole	N154T/L200P/S201L/H254L	$1.4 \pm 0.1 \times 10^{-1}$	$1.8 \pm 0.6 \times 10^{-5}$

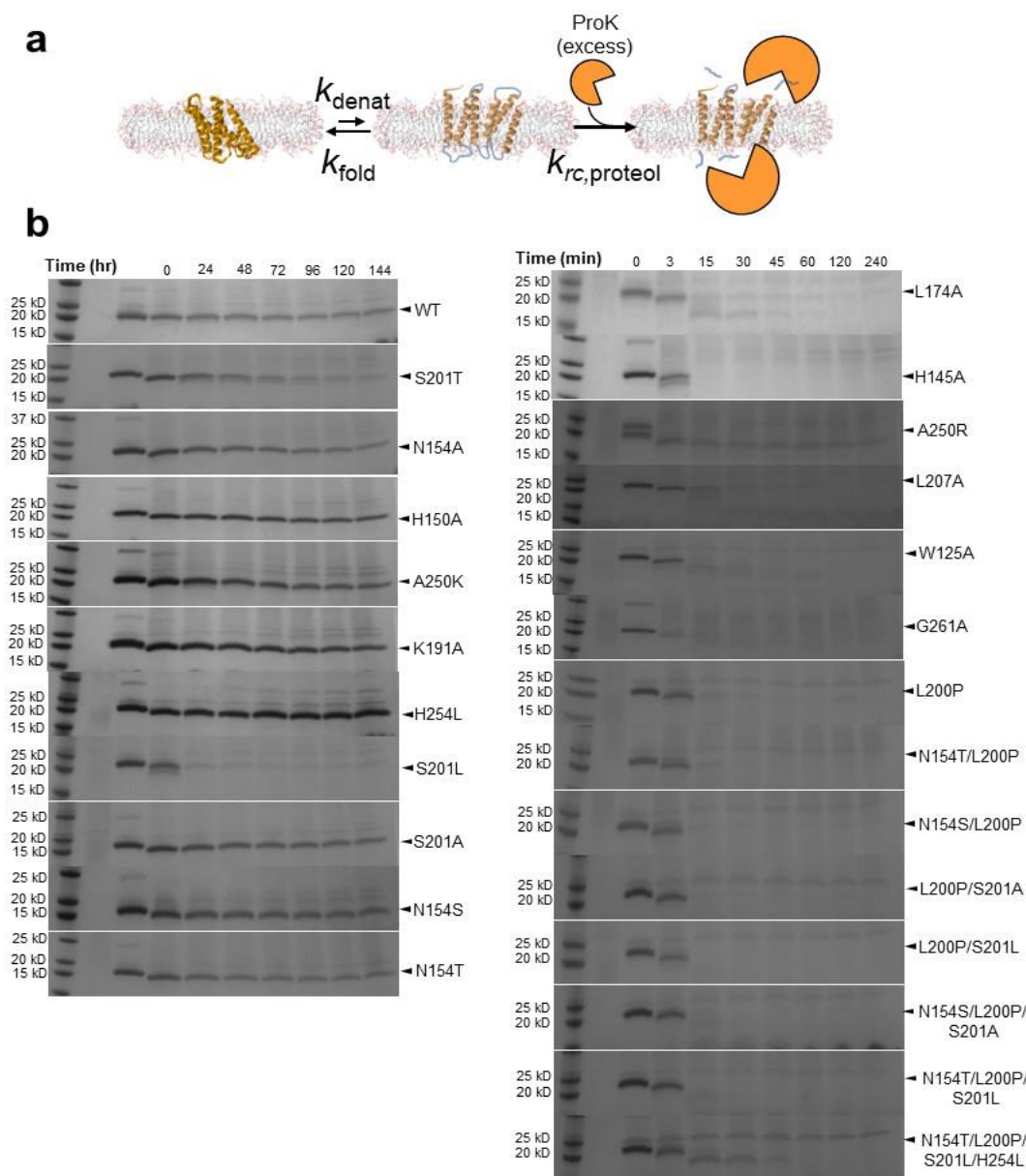


Figure 3.13. Spontaneous denaturation of GlpG monitored by Prok digestion. GlpG reacted with ProK for various incubation times in 2% DMPC/DMPG/CHAPS bicelles at 37 °C. The band intensities on SDS-PAGE that correspond to GlpG was analyzed by the ImageJ program.

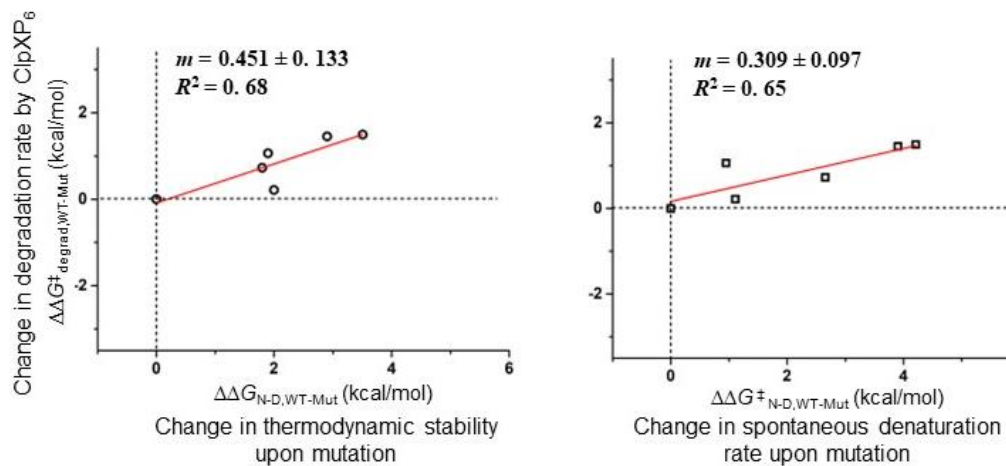


Figure 3.14. The correlation between the conformational stability of soluble substrate Titin-I27 and the degradation rate by the water-soluble AAA + protease, ClpXP₆ (Kenniston, Baker, Fernandez and Sauer, *Cell* 2003 114, 511-520).

REFERENCES

REFERENCES

1. Gottesman, S. Proteases and their targets in *Escherichia coli*. *Annu Rev Genet* **30**, 465-506, doi:DOI 10.1146/annurev.genet.30.1.465 (1996).
2. Sauer, R. T. & Baker, T. A. AAA+ proteases: ATP-fueled machines of protein destruction. *Annu Rev Biochem* **80**, 587-612, doi:10.1146/annurev-biochem-060408-172623 (2011).
3. Olivares, A. O., Baker, T. A. & Sauer, R. T. Mechanistic insights into bacterial AAA+ proteases and protein-remodelling machines. *Nat Rev Microbiol* **14**, 33-44, doi:10.1038/nrmicro.2015.4 (2016).
4. Maillard, R. A. *et al.* ClpX(P) generates mechanical force to unfold and translocate its protein substrates. *Cell* **145**, 459-469, doi:10.1016/j.cell.2011.04.010 (2011).
5. Cordova, J. C. *et al.* Stochastic but highly coordinated protein unfolding and translocation by the ClpXP proteolytic machine. *Cell* **158**, 647-658, doi:10.1016/j.cell.2014.05.043 (2014).
6. Bard, J. A. M., Bashore, C., Dong, K. C. & Martin, A. The 26S Proteasome Utilizes a Kinetic Gateway to Prioritize Substrate Degradation. *Cell* **177**, 286-298 e215, doi:10.1016/j.cell.2019.02.031 (2019).
7. Langer, T. AAA proteases: cellular machines for degrading membrane proteins. *Trends Biochem Sci* **25**, 247-251 (2000).
8. Kondadi, A. K. *et al.* Loss of the m-AAA protease subunit AFG₃L₂ causes mitochondrial transport defects and tau hyperphosphorylation. *Embo J* **33**, 1011-1026 (2014).
9. Malnoe, A., Wang, F., Girard-Bascou, J., Wollman, F. A. & de Vitry, C. Thylakoid FtsH protease contributes to photosystem II and cytochrome b6f remodeling in *Chlamydomonas reinhardtii* under stress conditions. *The Plant cell* **26**, 373-390, doi:10.1105/tpc.113.120113 (2014).
10. Akiyama, Y., Ogura, T. & Ito, K. Involvement of FtsH in protein assembly into and through the membrane. I. Mutations that reduce retention efficiency of a cytoplasmic reporter. *The Journal of biological chemistry* **269**, 5218-5224 (1994).
11. Di Bella, D. *et al.* Mutations in the mitochondrial protease gene AFG3L2 cause dominant hereditary ataxia SCA28. *Nat Genet* **42**, 313-321 (2010).
12. Nolden, M. *et al.* The m-AAA protease defective in hereditary spastic paraplegia controls ribosome assembly in mitochondria. *Cell* **123**, 277-289 (2005).

13. Koodathingal, P. *et al.* ATP-dependent proteases differ substantially in their ability to unfold globular proteins. *J Biol Chem* **284**, 18674-18684, doi:10.1074/jbc.M900783200 (2009).
14. Herman, C., Prakash, S., Lu, C. Z., Matouschek, A. & Gross, C. A. Lack of a robust unfoldase activity confers a unique level of substrate specificity to the universal AAA protease FtsH. *Mol Cell* **11**, 659-669 (2003).
15. Kim, Y. I., Burton, R. E., Burton, B. M., Sauer, R. T. & Baker, T. A. Dynamics of substrate denaturation and translocation by the ClpXP degradation machine. *Mol Cell* **5**, 639-648 (2000).
16. Kihara, A., Akiyama, Y. & Ito, K. Dislocation of membrane proteins in FtsH-mediated proteolysis. *EMBO J* **18**, 2970-2981, doi:10.1093/emboj/18.11.2970 (1999).
17. Yang, Y. *et al.* Folding-Degradation Relationship of a Membrane Protein Mediated by the Universally Conserved ATP-Dependent Protease FtsH. *J Am Chem Soc* **140**, 4656-4665, doi:10.1021/jacs.8b00832 (2018).
18. Yang, Y., Gunasekara, M., Muhammednazaar, S., Li, Z. & Hong, H. Proteolysis mediated by the membrane-integrated ATP-dependent protease FtsH has a unique nonlinear dependence on ATP hydrolysis rates. *Protein Sci* **28**, 1262-1275, doi:10.1002/pro.3629 (2019).
19. Kenniston, J. A., Baker, T. A., Fernandez, J. M. & Sauer, R. T. Linkage between ATP consumption and mechanical unfolding during the protein processing reactions of an AAA(+) degradation machine. *Cell* **114**, 511-520, doi:10.1016/S0092-8674(03)00612-3 (2003).
20. Olivares, A. O., Baker, T. A. & Sauer, R. T. Mechanical Protein Unfolding and Degradation. *Annu Rev Physiol* **80**, 413-429, doi:10.1146/annurev-physiol-021317-121303 (2018).
21. Aubin-Tam, M. E., Olivares, A. O., Sauer, R. T., Baker, T. A. & Lang, M. J. Single-molecule protein unfolding and translocation by an ATP-fueled proteolytic machine. *Cell* **145**, 257-267, doi:10.1016/j.cell.2011.03.036 (2011).
22. Guo, R. *et al.* Steric trapping reveals a cooperativity network in the intramembrane protease GlpG. *Nat Chem Biol* **12**, 353-360, doi:10.1038/nchembio.2048 (2016).
23. Hong, H. & Bowie, J. U. Dramatic destabilization of transmembrane helix interactions by features of natural membrane environments. *J Am Chem Soc* **133**, 11389-11398, doi:10.1021/ja204524c (2011).
24. 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* **107**, 19802-19807, doi:10.1073/pnas.1010348107 (2010).
25. Blois, T. M., Hong, H., Kim, T. H. & Bowie, J. U. Protein unfolding with a steric trap. *J Am Chem Soc* **131**, 13914-13915, doi:10.1021/ja905725n (2009).

26. Lee, C., Schwartz, M. P., Prakash, S., Iwakura, M. & Matouschek, A. ATP-dependent proteases degrade their substrates by processively unraveling them from the degradation signal. *Mol Cell* **7**, 627-637 (2001).
27. Gaffney, K. A. & Hong, H. The rhomboid protease GlpG has weak interaction energies in its active site hydrogen bond network. *J Gen Physiol* **151**, 282-291, doi:10.1085/jgp.201812047 (2019).
28. Baker, R. P. & Urban, S. Architectural and thermodynamic principles underlying intramembrane protease function. *Nat Chem Biol* **8**, 759-768, doi:10.1038/nchembio.1021 (2012).
29. Wang, Y., Maegawa, S., Akiyama, Y. & Ha, Y. The role of L1 loop in the mechanism of rhomboid intramembrane protease GlpG. *J Mol Biol* **374**, 1104-1113, doi:10.1016/j.jmb.2007.10.014 (2007).
30. Junker, M. *et al.* Pertactin beta-helix folding mechanism suggests common themes for the secretion and folding of autotransporter proteins. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 4918-4923, doi:10.1073/pnas.0507923103 (2006).
31. 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* **11**, 981-987, doi:10.1038/nchembio.1939 (2015).
32. Matouschek, A., Kellis, J. T., Jr., Serrano, L. & Fersht, A. R. Mapping the transition state and pathway of protein folding by protein engineering. *Nature* **340**, 122-126, doi:10.1038/340122a0 (1989).
33. Herman, C., Thevenet, D., Boulloc, P., Walker, G. C. & D'Ari, R. Degradation of carboxy-terminal-tagged cytoplasmic proteins by the Escherichia coli protease HflB (FtsH). *Genes Dev* **12**, 1348-1355 (1998).
34. Dulloo, I., Muliyl, S. & Freeman, M. The molecular, cellular and pathophysiological roles of iRhom pseudoproteases. *Open Biol* **9**, 190003, doi:10.1098/rsob.190003 (2019).
35. Freeman, M. The rhomboid-like superfamily: molecular mechanisms and biological roles. *Annu Rev Cell Dev Biol* **30**, 235-254, doi:10.1146/annurev-cellbio-100913-012944 (2014).
36. Lemberg, M. K. & Freeman, M. Functional and evolutionary implications of enhanced genomic analysis of rhomboid intramembrane proteases. *Genome Res* **17**, 1634-1646, doi:10.1101/gr.6425307 (2007).
37. Neal, S. *et al.* The Dfm1 Derlin Is Required for ERAD Retrotranslocation of Integral Membrane Proteins. *Mol Cell* **69**, 306-320 e304, doi:10.1016/j.molcel.2017.12.012 (2018).

38. Guerriero, C. J. *et al.* Transmembrane helix hydrophobicity is an energetic barrier during the retrotranslocation of integral membrane ERAD substrates. *Mol Biol Cell* **28**, 2076-2090, doi:10.1091/mbc.E17-03-0184 (2017).
39. Choi, H. K. *et al.* Watching helical membrane proteins fold reveals a common N-to-C-terminal folding pathway. *Science* **366**, 1150-1156, doi:10.1126/science.aaw8208 (2019).
40. 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* **112**, 7978-7983, doi:10.1073/pnas.1424751112 (2015).
41. Schafer, N. P., Truong, H. H., Otzen, D. E., Lindorff-Larsen, K. & Wolynes, P. G. Topological constraints and modular structure in the folding and functional motions of GlpG, an intramembrane protease. *Proc Natl Acad Sci U S A* **113**, 2098-2103, doi:10.1073/pnas.1524027113 (2016).
42. Wang, Z., Jumper, J. M., Freed, K. F. & Sosnick, T. R. On the Interpretation of Force-Induced Unfolding Studies of Membrane Proteins Using Fast Simulations. *Biophys J* **117**, 1429-1441, doi:10.1016/j.bpj.2019.09.011 (2019).
43. Sanders, M. R., Findlay, H. E. & Booth, P. J. Lipid bilayer composition modulates the unfolding free energy of a knotted alpha-helical membrane protein. *Proc Natl Acad Sci U S A* **115**, E1799-E1808, doi:10.1073/pnas.1714668115 (2018).
44. Preiner, J. *et al.* Free energy of membrane protein unfolding derived from single-molecule force measurements. *Biophys J* **93**, 930-937, doi:10.1529/biophysj.106.096982 (2007).
45. Matouschek, A. Protein unfolding--an important process in vivo? *Curr Opin Struct Biol* **13**, 98-109, doi:10.1016/s0959-440x(03)00010-1 (2003).
46. Gaffney, K. A. *et al.* Lipid bilayer induces contraction of the denatured state ensemble of a helical-bundle membrane protein. *Proc Natl Acad Sci U S A* **119**, doi:10.1073/pnas.2109169119 (2022).

CHAPTER 4

Concluding remarks and outlook

My dissertation research aims to elucidate the folding-degradation relationship of membrane proteins to understand cellular protein quality control mechanisms. It has been known that the cells selectively degrade misfolded and intrinsically unstable proteins to maintain functional protein concentration.¹ However, the chemical and physical principles of the selective degradation are not well understood.

In chapter 2, I established the steric trapping strategy in a bilayer environment provided by bicelles to study the folding of membrane proteins. The thermodynamic stability and cooperativity of intramembrane protease GlpG from *E.coli* has been studied by comparing these folding properties in neutral DDM micelles and DMPC/CHAPS bicelles. Although widely used for providing a hydrophobic environment for membrane protein research, detergents and lipids are fundamentally different with regards to the chemical structure and the mesoscopic morphology and physical property of their self-assembly:² Detergents (typically with one hydrocarbon chain- C_6 to C_{12}) self-aggregate to form micelles and are in a dynamic equilibrium between monomeric and micellar forms in aqueous solution. They bind to hydrophobic regions of membrane proteins, providing a permissive environment for membrane protein structure and function. On the other hand, lipids (typically with two long hydrocarbon chains- C_{12} to C_{20}) are assembled into the quasi-two-dimensional molecular layer (called the lipid bilayer) with the hydrocarbon tails of lipid molecules are facing each other to form a hydrophobic core. The lipid bilayer structure is maintained by repulsive (in the polar headgroup and hydrocarbon tail regions) and adhesive (at the water-hydrocarbon interface) lateral interactions. The balance between the repulsive and adhesive interactions changes depending on the structure of lipid molecules, generating a curvature stress in the bilayer.

I found that DMPC/CHAPS bicelles moderately stabilize GlpG compared to DDM micelles. In DDM micelles, ΔG_{N-D}^0 was -5.8 and -4.7 kcal/mol for N- and C-subdomain, respectively.³ However, the stability increased to -7.0 and -6.7 kcal/mol in DMPC/CHAPS bicelles. Interestingly, the free energy difference between the two subdomains became negligible in bicelles. I further elaborate on the role of the hydrophobic solvent environment in stabilizing GlpG by measuring the contribution of individual residue interactions to the stability. Strikingly, compared to micelles, the completely buried residues in the protein interior that do not contact the solvating amphiphiles contributed more to the stability in bicelles whereas the partially buried and exposed residues that contact the amphiphiles showed smaller or no destabilization in bicelles. These observations led to the conclusion that solvating lipids facilitate the folding into the native structure. The hydrophobic solvent environment also affect the residue interaction network of GlpG. The cooperativity profiling method that the Hong lab developed³ shows that cooperative interactions are found as multiple small clusters in micelles while they span the whole packed regions in the protein in bicelles. Through molecular dynamics simulations, we showed that the stability and cooperativity enhancement by lipids stems from strong lipid-lipid interactions that make the lipid molecules exchange fast and bind weak on the protein surface compared to the detergent–detergent or detergent–protein interactions in micelles. Thus, the weak lipid solvation on the protein allows the intraprotein interaction to outcompete the lipid-protein interaction, resulting in the additional stabilization of the protein and the strengthening of the residue interaction network. This result is in line with our recent study that the lipid bilayer induces contraction of the denatured state ensemble of GlpG, implying that lipids may not be a good solvent for membrane, that is, lipid-protein interactions are comparable or weaker than intraprotein interactions.⁴ The enhancement of intraprotein interactions by the weak lipid solvation in the

bilayer environment may support the claim that the amino acid sequence of membrane proteins is a major determinant of their conformation and provide a physical basis of the current sequence-based structural prediction efforts for membrane proteins.^{5,6}

In Chapter 3, I studied the quantitative relationship between the folding and degradation of membrane proteins mediated by the membrane-integrated AAA+ protease FtsH. We established an array of mutants with a wide range of stability and spontaneous denaturation rate values by mutations for the model membrane substrate GlpG and subjected it to degradation. Unexpectedly, I found that the conformational stability of membrane substrates did not contribute to degradation rates. Hence, the rate-limiting step is not substrate denaturation in the membrane, but the hydrophobicity of TM segments and the positively charged residues in the interfacial loop residues, both of which resist membrane dislocation, noticeably changed the degradation rate. These findings provide insights into the membrane protein degradation mediated by ATP-dependent proteolysis. This knowledge could be further expanded to the understanding of ATP-dependent degradation in ERAD, one of the key degradation pathways for membrane proteins in eukaryotic cells. In ERAD, misfolded membrane proteins are recognized by an ubiquitin ligase in the Hrd1p complex in the ER membrane, dislocated from the membrane by AAA+ p97, and finally targeted to the proteasome in the cytosol for degradation.⁷ This process appears to employ a similar logic (*i.e.*, the membrane dislocation by an AAA+ enzyme plays a key role), but is still poorly understood. Therefore, our results will provide useful knowledges for generating new hypotheses for quality control and disease mechanisms involving membrane proteins in eukaryotic cells.

Still, many questions are waiting to be answered regarding membrane protein folding and degradation. Although my study focuses on the role of lipid–protein interactions in folding, the role of water molecules in driving the second stage of membrane protein folding still remains

elusive. In the denatured state of membrane proteins, water-molecules may penetrate deeper into the hydrophobic core when TM segments contain polar residues, or the hydrophobic mismatch between individual TM segments and the bilayer induces lipid deformation creating bilayer defects around the protein that provide room for water molecules within the membrane.^{8,9} Those water molecules may be unusually ordered within the lipid bilayer, providing a substantial driving force for folding. Therefore, clarifying the role of water will be an important contribution to the advancement of the field.

My studies reveal that the rate-determining step is not substrate denaturation in the membrane. But, to understand the detailed molecular mechanism, it would be essential to employ a “divide and conquer” approach, that is, dissecting the degradation reaction into individual steps such as engagement, denaturation, membrane dislocation, and proteolysis, and obtain detailed kinetic and energetic information in each stage. Also, there are no high-resolution structures for full-length FtsH bound with substrates. The structural information will greatly help our understanding of how various AAA+ proteases are similar and differ in their degradation mechanisms and what aspects are unique about FtsH. For example, FtsH is the only membrane-anchored AAA+ protease among the five proteases in *E. coli*. This raises the question of whether the membrane integration of FtsH has any advantages over other AAA+ proteases in selectively targeting membrane substrates. Furthermore, how do different types of AAA+ proteases compete to degrade membrane substrates? Finally, most studies of FtsH focus on model substrates. However, extending this work to the natural substrates would reveal how the degradation of specific cellular proteins tunes to substrates’ physical properties and chemical modification such as oxidation or photodamage. These are challenging questions so far, but the tools and strategies developed in this study will provide a useful platform to address these questions in the future.

REFERENCES

REFERENCES

1. Balch, W. E., Morimoto, R. I., Dillin, A. & Kelly, J. W. Adapting proteostasis for disease intervention. *Science* **319**, 916-919, doi:10.1126/science.1141448 (2008).
2. Garavito, R. M. & Ferguson-Miller, S. Detergents as tools in membrane biochemistry. *J Biol Chem* **276**, 32403-32406, doi:10.1074/jbc.R100031200 (2001).
3. Guo, R. *et al.* Steric trapping reveals a cooperativity network in the intramembrane protease GlpG. *Nat Chem Biol* **12**, 353-360, doi:10.1038/nchembio.2048 (2016).
4. Gaffney, K. A. *et al.* Lipid bilayer induces contraction of the denatured state ensemble of a helical-bundle membrane protein. *Proc Natl Acad Sci USA* **119**, doi:10.1073/pnas.2109169119 (2022).
5. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589, doi:10.1038/s41586-021-03819-2 (2021).
6. Tunyasuvunakool, K. *et al.* Highly accurate protein structure prediction for the human proteome. *Nature* **596**, 590-596, doi:10.1038/s41586-021-03828-1 (2021).
7. Carvalho, P., Goder, V. & Rapoport, T. A. Distinct ubiquitin-ligase complexes define convergent pathways for the degradation of ER proteins. *Cell* **126**, 361-373, doi:10.1016/j.cell.2006.05.043 (2006).
8. Min, D. *et al.* Unfolding of a ClC chloride transporter retains memory of its evolutionary history. *Nat Chem Biol* **14**, 489-496, doi:10.1038/s41589-018-0025-4 (2018).
9. Chadda, R. *et al.* Membrane transporter dimerization driven by differential lipid solvation energetics of dissociated and associated states. *Elife* **10**, doi:10.7554/eLife.63288 (2021).