

QUANTITATIVE DISSECTION OF MOLECULAR DRIVING FORCES IN MEMBRANE
PROTEIN FOLDING

By

Jiaqi Yao

A DISSERTATION

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

Chemistry – Doctor of Philosophy

2025

ABSTRACT

Burials of ionizable residues inside hydrophobic core are unfavorable in general due to the high desolvation cost of transferring them from water to lipid bilayer. Nevertheless, these residues can form ion-pairs and play vital roles in cellular functions including proton/electron transfer, catalysis, and receptor activation. Proteins in thermophilic microbes maintain their fold and activity in extreme environments like volcanoes, ocean ridges and hot springs with high temperatures (80–110 °C). Previous studies suggested that thermophilic proteins achieve thermostability via an increased number of salt-bridges between ionizable residues compared to their mesophilic homologs. Thus, studies regarding the ionizable residues and thermostability of thermophilic proteins are essential for the fundamental understanding of protein stability, function, and engineering. While related studies have mostly concerned water-soluble proteins, these topics received less attention in membrane proteins. In my dissertation, I aim to bridge the knowledge gaps using the universally conserved rhomboid protease family as a model. I focused on addressing the following two questions: 1) What is the energetic consequence of burying ionizable residue pairs in the core of membrane protein? 2) What is the origin of the thermostability of membrane proteins in thermophilic organisms? In the first question, I found that bearing ionizable residues inside membrane protein induces destabilization of membrane protein kinetically and thermodynamically. Double mutant cycle analysis suggests paired internal ionizable residues form favorable interaction in micelle and bicelle environments. In the second question, my results demonstrate that the delipidated thermophilic rhomboids lose their stability and are fully inactivated at temperatures below their optimal growth conditions compared to mesophilic rhomboid. These results suggest lipids play critical roles in buried ionizable pairs and thermostability in membrane protein folding.

Dedicated to my dearest parents Yaping and Xu: I love you

ACKNOWLEDGEMENTS

My journey of exploring the world of scientific research and pursuing higher education at Michigan State University is an extraordinary adventure, made possible by the unwavering support, guidance, and kindness of countless individuals. To all of them, I extend my deepest gratitude.

First of all, I want to express my sincerest thankfulness to my advisor, Dr. Heedeok Hong, for offering me the invaluable opportunity to join his research group in the study of membrane protein folding. Beyond being a brilliant and experienced scientist, you have been a patient and supportive mentor who shaped not only my skills as a researcher but also my resilience and curiosity. Your ability to foster an inclusive, creative, and collaborative environment allowed me to thrive even in moments of challenge.

I thank my committee members: David Weliky, Dr. Jian Hu, and Dr. Xiangshu Jin for your insightful critiques, thought-provoking questions, and steadfast support throughout my candidacy exams, committee meetings, and research seminars. Your expertise across diverse fields broadened the scope of my work and inspired me to approach problems with fresh perspectives. I also extend my gratitude to Dr. Qiang Cui of Boston University for his pivotal collaboration and for sparking the ideas that led to the work in Chapter 2 of this dissertation.

To my lab mates in the Hong group: thank you for making the lab a place of camaraderie, learning, and shared purpose. A special acknowledgment goes to Dr. Miyeon Kim, who is more than a supportive coworker but also a patient listener and a heart-warming friend. My sincere gratitude to Dr. Ruiqiong Guo, Dr. Kristen Gaffney, Dr. Yiqing Yang and Dr. Shaima Nazaar for training me during my early stage in lab. Big thanks to Dr. Mihiravi Gunasekara, Manoj Rana, Saba Kanwal and Maya Khorrami for their positivity, encouragement, and shared laughter. I would like to thank Kai Lange, the undergraduate student who worked with me and contributed to Chapter 3.

To my dear friends I met at Michigan State University: your friendship has been a beacon of warmth through Michigan's icy winters. Thank you to Dr. Daoyang Chen, Dr. Zhilin Hou, Dr. Qianjie Wang, Dr. Fei Fang, Rosemary Augustine, Dr. Yijin Zhang, Dr. Mengxia Sun, Dr. Fangchun Liang, Dr. Zhichang Yang, Dr. Xiaoge Wang, Shuxin Li and

Kunwei Yang for the road trips, potluck parties, snack breaks, movie nights, and countless moments of solidarity.

Last but not least, I want to express my deepest gratitude to my family, especially my loving mom Yaping Wang and loving dad Xu Yao. Thank you for your unconditional love, constant support, and unshakable belief in me. If I were a vessel sailing afar, home with you would always be my tenderest star— not just a harbor, but a guiding light, anchoring my soul through the loneliest nights.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS	vii
CHAPTER 1: Introduction to membrane protein folding and stability	1
CHAPTER 2: Role of buried ionizable residues in the stability of one α -helical intramembrane protease	48
CHAPTER 3: Investigate the origin of thermostability and activity of thermophilic rhomboid proteases in comparison to their mesophilic homologs	95
CHAPTER 4: Concluding remark and outlook.....	131
REFERENCES.....	137

LIST OF ABBREVIATIONS

CHAPS	3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate
DMPC	Dipalmitoylphosphatidylcholine
DDM	N-dodecyl- β -D-maltopyranoside
<i>E. coli</i>	<i>Escherichia coli</i>
<i>PfuRh</i>	<i>Pyrococcus furiosus</i> Rhomboid
<i>TmRh</i>	<i>Thermotoga maritima</i> Rhomboid
<i>TpRh</i>	<i>Thermococcus profundus</i> Rhomboid
IPTG	Isopropyl β -D-1-thiogalactopyranoside
PMSF	PhenylMethylSulfonyl Fluoride
SDS-PAGE	Sodium Dodecyl Sulfate -Polyacrylamide Gel Electrophoresis
TCEP	Tris (2-carboxyethyl) phosphine
Tris-HCl	Tris(hydroxymethyl)aminomethane
T_m	Transition temperature
Ni-NTA	Nickel-charged affinity resin (nitrilotriacetic acid)
WT	Wild Type
TM	Transmembrane

CHAPTER 1: Introduction to membrane protein folding and stability

1.1. Overview of protein folding

Proteins play indispensable roles in cell life such as catalyzing chemical reactions, transducing/receiving signals, transporting materials and even providing structural support for cells. Function of proteins depends on their three-dimensional structure. Protein folding is the process by which a linear polypeptide chain is transformed into a unique three-dimensional (3D) structure. Misfolded proteins, unless degraded by proteases or refolded by molecular chaperones, tend to form aggregations or amyloid fibrils inside and outside cells, which can interfere with normal cellular function. Multiple lethal neurodegenerative diseases such as Creutzfeldt-Jakob disease (misfolding of prion protein)¹, Parkinson disease (misfolded tau and possibly α -synuclein²), Huntington diseases (misfolding induced aggregation of HTT^{ex1})³, and amyotrophic lateral sclerosis (misfolded transcriptional repressor TDP-43)⁴ are examples of protein misfolding diseases⁵. Thus, it is crucial to acquire detailed understanding in protein folding to find cures for life-threatening diseases.

The first known protein structure was water-soluble protein myoglobin (**Figure 1.1.**), which was solved by Perutz, Kendrew, and coworkers using X-ray crystallography⁶. At the resolution of 6 Å, this revolutionary result revealed complicated, asymmetric features of protein molecules and immediately posed the protein-folding problem: what are the fundamental physical and chemical principles behind the formation of this intricate protein structure? Over the past decades, this problem evolved into three major questions⁷: 1) What are the physicochemical properties of the one dimensional (1D) amino acid sequence that drive them to fold into 3D native structure? 2) What are the folding pathways that proteins follow to be folded at a remarkable speed (i.e., in a millisecond to second scale) despite the huge number of conformational combinations? and 3) Can computer-based algorithms assist accurate predictions of 3D structures from their 1D amino acid sequences?

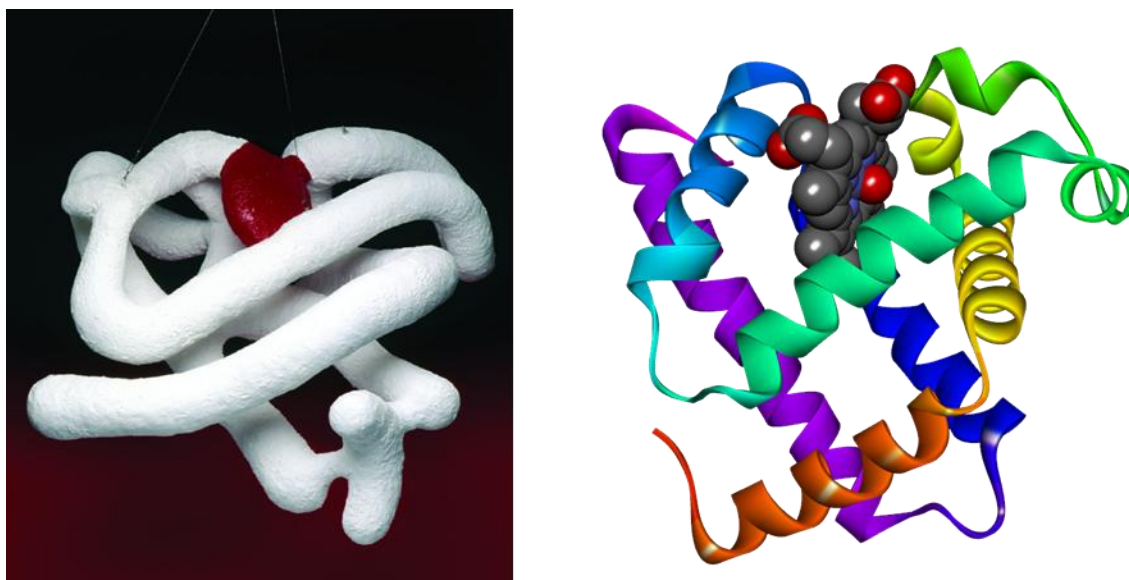


Figure 1.1. First protein structure Myoglobin⁷ with resolution of 6 Å and the myoglobin structure with resolution of 2 Å⁸. In a 6 Å resolution, shape and fold of the domain can be recognized while the secondary structure is still not clear. The typical chemical bond length is around 1-2 Å. To precisely assign the identity of sidechains, the atoms in the structure need to be visible, which requires resolution of 1.7-2.8 Å. In 1973 Kendrew and Waston solved myoglobin structure with a higher resolution of 2 Å (PDB 1MBN), the secondary structure and the iron containing porphyrin IX ligand can be recognized. Reprint permission from Dill *et al.*⁷ (license number: 1616289-1).

In 1961, Anfinsen's experiments showed that the denatured polypeptide chains of ribonuclease A whose cysteine residues are fully reduced can reversibly refold into the functional native state of the protein⁹. This result demonstrates that the native conformation of proteins is determined solely by the amino acid sequence and is optimized to a global free energy minimum¹⁰.

In 1968, Cyrus Levinthal developed a thought experiment to explain how a given polypeptide chain can fold at such a high speed despite the astronomic numbers of possible conformational arrangements⁹. For example, a small protein ubiquitin is composed of 76 amino acids which are connected by 75 peptide bonds with 150 ψ and ϕ angles in total. Assuming each angle has three stable rotamers, the conformational space of the polypeptide chain is composed of 3^{150} alternative configurations from which there is one properly folded combination. If ubiquitin can sample the conformational

space as effective as 10^{10} arrangements per second, it will take approximately 10^{54} years, which is about 10^{45} times of the age of the earth (i.e., 10^9 - 10^{10} years). This seemingly self-contradictory description is known as Levinthal's paradox and leads to the argument that proteins fold guided by certain pathways with limited native confirmations.

Anfinsen and Levinthal's experiments together led to the development of the "folding energy landscape" theory of protein folding¹¹. It is now widely recognized that the folding energy landscape of a single-domain protein is funnel-shaped. That is, many high energy, unstructured polypeptide chains fold into a global energy minimum representing the unique native structure with shallow valleys representing local energy minima and bumps representing kinetic barriers during folding. That is, starting from the unfolded chains with highly heterogeneous conformations, the progress of folding largely occurs in the ways to decrease the free energy of the chains (i.e., by forming "native contacts") and to reduce the degree of freedom of the chain configurations until they reach the native conformation at the free energy minimum.

Intrinsically Disordered Proteins (IDPs) are a unique class of proteins that lack a fixed three-dimensional structure under physiological conditions, existing instead as dynamic ensembles of conformations. Unlike compact, folded proteins, IDPs are characterized by high flexibility and lack of stable secondary structures, and they are usually fully expanded in water. Sequences of IDPs were previously suggested to be richer in polar and charged amino acids, such as glutamine, serine, and arginine, while being low in hydrophobic residues compared with ordered proteins¹². However, a more recent small angle x-ray scattering measurement on solvent quality for IDPs suggested that even IDP with a high hydrophobicity can be expanded in water¹³. Moreover, the abundance of IDPs is surprisingly high, bioinformatics studies predicted that around 30% of eukaryotic proteins are mostly disordered^{14,15}. The dynamic conformation enables IDPs to interact with multiple binding partners, making them crucial for cellular signaling, regulation, and molecular recognition. For example, p53, one of the most extensively studied IDPs, is a transcription factor showing multiple conformations when bind with different receptor proteins¹⁶. The prevalence and functional importances of IDPs challenges the generality of sequence-structure-function paradigm.

1.2. Overview of the cell membranes

Biological membranes are essential components of cells. By definition, membranes serve as a permeable barrier while allowing essential nutrients in and wastes out. In prokaryotic cells, gram-negative bacteria have two layers of membranes: cytoplasmic membrane and outer membrane. Gram-positive bacteria only have one layer of plasma membrane. For eukaryotic cells, membranes not just define the boundary of the whole cell but also encapsulate subcellular organelles such as nucleus, mitochondria, golgi apparatus, lysosomes and endoplasmic reticulum to separate them from cytoplasm.

As the fundamental scaffolding of cell membranes, the lipid bilayer presents inherent challenges for structural elucidation. A key dilemma lies in reconciling its non-crystalline state with the hydrated, thermal disordered, biologically relevant, fluid phase ($L\alpha$). To overcome the challenge, researchers have employed multiple methods, including X-ray and neutron diffraction¹⁷, molecular dynamic (MD) simulation¹⁸ to approximate the 3D structure of lipid bilayer in the $L\alpha$ phase. Collectively, these pioneering studies have characterized a consistent lipid bilayer structure featuring a hydrocarbon core flanked by two aqueous interfaces. In the distribution profile of quasi-molecular structural group of lipids, nonpolar tails exhibit a normal distribution centered within the hydrocarbon core, with a peak width of $\sim 10 \text{ \AA}$ ¹⁷. The carbon-carbon double bonds and methylene groups are distributed throughout the core, showing reduced probability density both at the center of the core, and near the interfacial regions¹⁷. Meanwhile, the polar head groups are predominantly localized outside of the hydrocarbon core, concentrated at the interfacial boundaries. The hydrocarbon core maintains consistently low polarity throughout its volume, while exhibiting a distance-dependent polarity gradient across the interfacial regions¹⁹. The structure suggested three key features of lipid bilayers¹⁹: 1) The broad distribution of the structural groups' probability reflects significant thermal fluctuations; 2) the thickness of $\sim 30 \text{ \AA}$ allows accommodation for most membrane proteins; and 3) the heterogeneous polar groups at the interfaces facilitate polypeptide partitioning by forming non-covalent interactions.

The formation of lipid bilayers is governed by well-established thermodynamic principles. When dispersed in an aqueous phase, lipid molecules aggregate to minimize the exposure of their hydrophobic to water, thereby satisfying the thermodynamic

preference. This process is primarily driven by the hydrophobic effect, contributed by the aliphatic lipid chains²⁰. The Gibbs free energy of bilayer formation, given by: $\Delta G = \Delta H - T\Delta S$ has two major contributions: enthalpic and entropic. Experimental data have suggested that the ΔG is dominated by a large entropic contribution (ΔS)²⁰. It has been believed that the aggregation of hydrophobic moieties disrupts the ordered, cage-like water shells around lipid tails, increasing the entropy of the system²¹. Consequently, ΔG becomes more negative ($\Delta G < 0$), rendering lipid bilayer formation a spontaneous process under physiological conditions.

Phospholipids are a major component of membrane lipids. The basic chemical structure of phospholipid includes (i) glycerol backbone, (ii) two acyl chains that are esterified to the first and second carbons of glycerol and (iii) polar headgroup link to phosphate group which is attached to the third carbon of glycerol. The polar headgroup determines the type and charge state of phospholipids. The net charge of headgroup can be negative: Phosphatidylserine (PS), Phosphatidylglycerol (PG), Phosphatidylinositol (PI) and neutral: Phosphatidylethanolamine (PE) and Phosphatidylcholine (PC). Acyl chains lengths vary in range from 12 to 24 carbons, and the chains can be saturated or unsaturated with one or few double bonds in cis- or trans- conformation. Cardiolipin (CL) is an inverted conical-shaped diphosphatidylglycerol lipid with two phosphatidic acids linked by a glycerol backbone in the center thus has four acyl chains and is highly anionic due to two phosphate groups. CL predominantly exists in the inner mitochondria membrane and causes the negative curvature of membrane²². Besides phospholipids, cell membranes also contain glycolipids, sphingolipids, and cholesterol, which modulate the membrane's fluidity, signal transduction, and cellular recognition. Cholesterol has unique structure with rigid ring headgroup and flexible hydrocarbon tail, which allow it exhibit dual role in decreasing and increasing the membrane fluidity. In areas where the phospholipids are loosely packed or at high temperatures, cholesterol reduces movement of hydrocarbon tails and thus reduces the membrane fluidity. On the other hand, in regions where phospholipids are tightly packed or at low temperatures, the presence of cholesterol disrupts the orderly arrangement. This disruption prevents the fatty acid chains of phospholipids from packing too tight, thereby increasing fluidity.

1.3. Overview of membrane protein structures and folding

Together with the lipid bilayer, membrane proteins are essential components of cell membranes. They take up 25-30% of the total open reading frames in various genomes²³ and occupy 30% of the total surface area and 50% of the total mass of the cell membranes²⁴. Regarding the secondary structures in the membrane-embedded parts, membrane proteins can be classified into two types: α -helical and β -barrel. α -helical membrane proteins are in the cytoplasmic membranes of prokaryotic cells or in the cytoplasmic and all organelle membranes of eukaryotic cells. β -barrels are mostly found in the outer membranes of gram-negative bacteria, the cell walls of gram-positive bacteria, and the outer membranes of mitochondria and chloroplasts. Beyond their distinct subcellular locations within biological membranes, α -helical and β -barrel proteins exhibit several well-defined structural characteristics. α -helices are right-handed, coiled, rod-like membrane structures, and their cores are tightly packed, minimizing the sizes of cavities or pores. Therefore, the overall polarity of α -helical membrane proteins is generally low, dominantly composed of non-polar residues. In contrast, β -barrel membrane proteins have dominantly nonpolar residues at the outer surface contacting the lipid bilayer and often contain a hydrophilic cavity lining polar or charged residues at the inner surface of the barrel. In both α -helical and β -barrel membrane proteins, the nonpolar residues traversing the hydrophobic core of the membrane are called “hydrophobic stretches”. The alpha helical hydrophobic stretches typically are composed of approximately 20 amino acid residues²⁵, whereas the beta barrel hydrophobic stretches have around 10 amino acid residues, followed by a more polar region within the hydrophilic pore or cavity²⁶.

Driven by the hydrophobic force, the folding of membrane protein is initiated by the energetic favorable insertion of nascent polypeptide, especially the transmembrane (TM) domain, from N-terminus to C-terminus, a process known as sequential mode insertion. To adopt proper secondary and tertiary structures within the lipid bilayer, the transmembrane domain (TMD) undergoes topological changes governed by several key rules and determinants²⁷: 1). membrane protein topology is strongly influenced by the enrichment of positively charged residues in the cytoplasmic extramembrane domain. Arg and Lys are four times more abundant on the cytoplasmic side than the extracellular

side; this phenomenon is often described as the “positive-inside rule”²⁸. 2). Studies showed that changing the charge balance of a single-pass protein can affect TMD orientation^{29,30}. This result suggests that the distribution of charged residues near the membrane spanning regions is crucial in protein orientation, as known as the “charge difference rule”. 3). Both sequential (N to C terminus) and non-sequential TMD insertion yield similar final topologies, suggesting the possibility of reorientation mechanisms to refine protein structure²⁷. In addition to these rules, the translocon machinery and membrane potential serve as key determinants of topological transformation²⁷. While these mechanisms remain elusive, specific events, such as post-translational modification-driven structural rearrangements and salt bridges that stabilize membrane protein conformation, provide insight into unique underlying principles governing membrane protein topology.

Membrane proteins play pivotal and versatile roles in cellular life including signal transduction, exchange of metabolites, catalysis and cell adhesion³¹. To carry out their functions, membrane proteins must be properly folded within the constraint of lipid bilayer. Membrane protein conformations are highly sensitive to changes in physical properties, influenced by solvation-related interactions with lipids, cofactors, and other molecules. Genetic mutations destabilize native fold through disrupting/eliminating the critical intra-or-inter protein interactions, leading to the protein misfolding. For example, nuclear magnetic resonance (NMR) study of the L16P mutant locate in the TM helix peripheral myelin protein 22 (PMP22) revealed its increased tendency to adopt partially unfolded conformation compared to the wild-type (WT) protein³². Disturbing backbone H-bonding for helix formation, proline mutation breaks TM helix, forming flexible hinge leading to the equilibrium shift from native to molten globule state, underlies Trembler-J syndrome, a form of Charcot-Marie-Tooth disease (CMTD)³². A 47-site mutagenesis study of the KCNQ1 channel's voltage sensor domain recognized certain mutations promote misfolding, leading to cellular mistrafficking and long QT syndrome, a cardiac disorder characterized by delayed heart repolarization³³. While mutation is a major cause of membrane protein misfolding, other factors—such as post-translational modifications³⁴ and oxidative stress³⁵ can also destabilize proteins. Although the pathological pathways of misfolded membrane proteins are complex and beyond the

scope of this discussion, their misfolding consistently relates to thermodynamic and kinetic energy barriers between folded and unfolded states. Thus, it is practically important to understand membrane protein folding to understand the mechanisms of these misfolding diseases and to discover the ways of their cures.

Despite the importance in organismal health, studying the folding of membrane proteins is a difficult task. Membrane proteins fold in a chemically heterogeneous lipid bilayer environment. The fully hydrated lipid bilayer has total thickness of $\sim 55\text{--}60\text{ \AA}$ ³⁶, which consists of chemically and physically distinct regions: the hydrophobic core composed of lipid hydrocarbon tails of $\sim 30\text{ \AA}$ ³⁶, and the two interfacial regions composed of lipid polar head groups contacting bulk water. Lipid bilayers impose vital environmental constraints that influence the shape of membrane proteins. The complex chemical and physical properties of the lipid bilayer make it challenging to develop the protocol to effectively control the folding and unfolding processes for folding studies.

A foremost task towards answering these questions is to elucidate the intrinsic folding energy landscape of membrane proteins. The elements of the folding energy landscape include the free energy levels of the states which are populated during folding, the energy barriers between the states, and the conformation of the states and transition states³⁷. In practice, these elements are obtained from equilibrium and kinetic experiments by shifting the population distribution between the states using various perturbants (e.g., chemical denaturant, pH, heat, and mechanical force) and by detecting the states that appear during the folding³⁸. The conformations of the states and transition states are determined by mapping the regions in which the native contacts are retained or lost, or by measuring the degree of compactness of each state^{13,39–41}.

For water-soluble proteins, a variety of methods have been developed to study the folding of water-soluble proteins in test tubes and cells, at the global and single-residue resolution, and in the ensemble and single-molecule scale^{42–44}. However, those methods have had limited success when applied to membrane proteins. For example, polar chaotropic agents such as urea and guanidine-hydrochloride (GdnHCl) can induce reversible unfolding and membrane extraction of β -barrel membrane proteins^{45–47}. But for helical membrane proteins, those methods are often ineffective in inducing the

reversible disruption of secondary or tertiary interactions except for a few types of the proteins^{48–50}. Heating the membrane proteins solubilized in amphiphilic assemblies such as detergent micelles or lipid bilayers typically leads to irreversible protein aggregation^{51,52}. The large size of lipid vesicles and limited accessibility of water to the hydrocarbon core of a bilayer render it difficult applying solution-NMR, hydrogen-deuterium exchange (HDX), and small-angle X-ray scattering. Notably, solid-state NMR and HDX combined with mass spectrometry have recently been successful for several membrane proteins^{53–55}.

Studies about structure and folding energetics of membrane proteins still lag those of water-soluble proteins (**Figure 1.2**). It was 1985, almost 30 years after first reported water-soluble protein myoglobin, when the first structure of membrane protein photosynthetic reaction center was solved experimentally⁵⁶. The first quantitative studies of membrane protein folding energetics^{57,58} were reported in early to mid-1990s, which is also 30 years after Anfinsen's experiments.

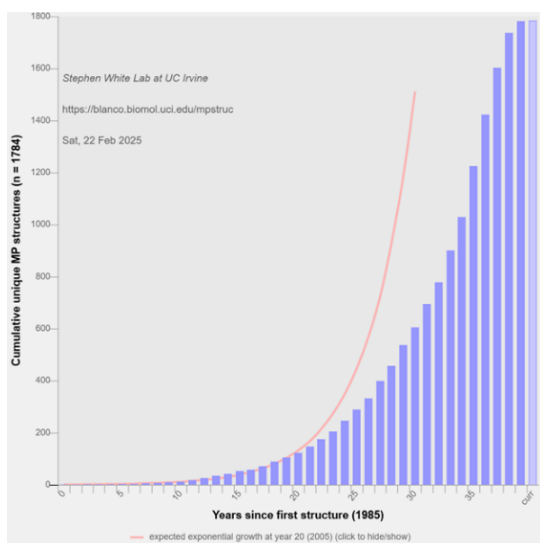


Figure 1.2. Cumulative unique membrane protein structures. Red line represents the expected exponential growth curve according to the first 20 years of unique membrane protein structure accumulation trend. The actual accumulation growth after 2005 lagged far behind the expected trend. Unique proteins in the database are 1784; Coordinate files in the database are 8343; Published reports of membrane protein structures in database are 3591. Source: <http://blanco.biomol.uci.edu/mpstruc/>

1.3.1. Two stage model of helical membrane protein folding

In 1990, Popot and Engelman proposed the “two-stage model” to explain the folding of α -helical membrane proteins (**Figure 1.3 a**).⁵⁹ In **Stage I**, nonpolar segments in a polypeptide chain are inserted into the membrane to form TM helices. In **Stage II**, the inserted TM helices associate to form a compact native state.

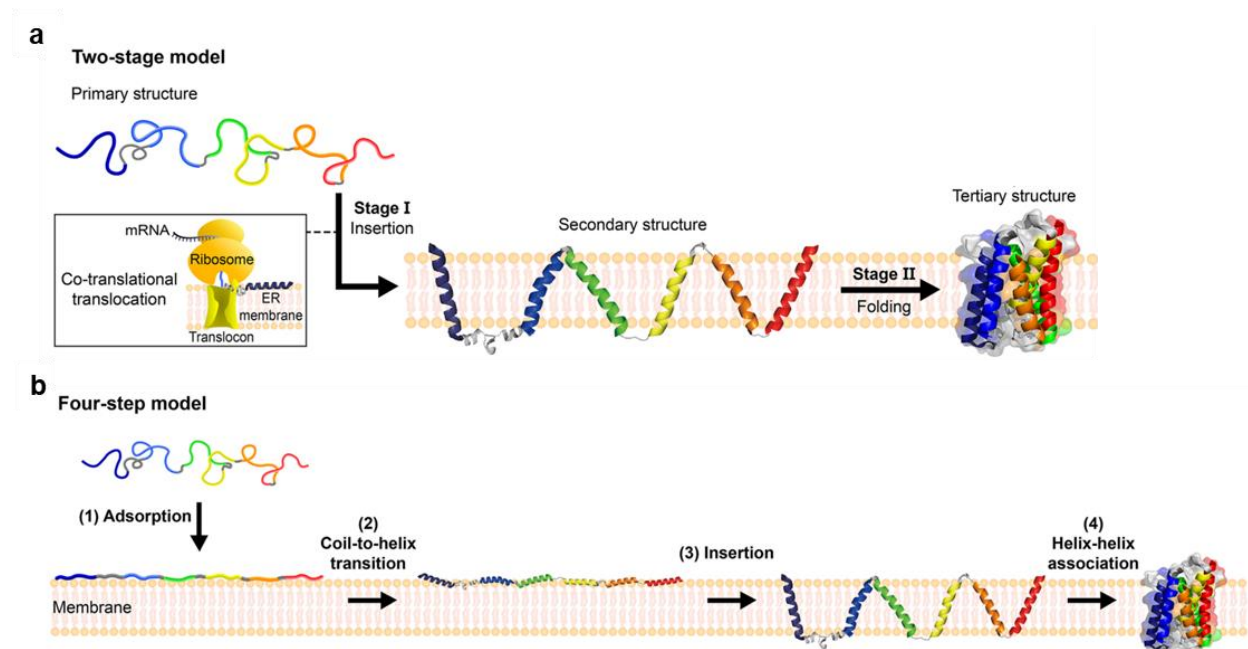


Figure 1.3. Two models of helical-bundle membrane proteins folding⁶⁰. **(a)** Two-stage model. In Stage I, nascent polypeptide chain is inserted into a lipid bilayer from aqueous environment and forms separated TM helices as individual secondary structures. The inset describes the *in vivo* situation. The mRNA is translated by ribosome and insertion of polypeptide chain mediated by translocon. In Stage II, inserted helices are associated to form tertiary structure **(b)** Four-step model. The entire folding is further dissected into (1) adsorption of polypeptide chain onto the membrane surface, (2) coil-to-helix transition on the membrane surface, (3) insertion of helices into membrane, and (4) helix–helix association to the folded state. The tertiary structure model protein here is from *E.coli* rhomboid protease GlpG (PDB: 2XOV)⁶¹. Reprint permission from Elsevier (license number: 6040820811553).

The experimental observations that led to the development of the two-stage model are based on the structural and folding studies of bacteriorhodopsin (bR), which

is a proton pump activated by the light driven *cis-trans* isomerization of covalent bound retinal: 1) bR denatured by the depletion of retinal using hydroxylamine or the addition of the chemical denaturant SDS was refolded into mild detergents or liposomes. The proton-pumping activity and structure of refolded bR were identical to those of the native proteins⁶². Thus, the folding of bR is reversible and its native conformation is a thermodynamically stable state; 2) The treatment of bR with chymotrypsin yields two fragments, the N-terminal segment with two TM helices and the C-terminal segment with five TM helices. When reconstituted in liposomes, the mixture of the two fragments associated to form the native structure with the native level activity⁶². This result indicates that the native tertiary structure of bR is formed by the lateral association of the separated TM helices within the membrane; 3) Individual TM segments of bR were largely hydrophobic and formed TM helical conformation except for TM4 which inserted into the membrane in a pH-dependent manner and for TM7 which aggregated in the membrane. The two-stage model is a “thermodynamic” model rather than a description of the folding mechanism such that this model hypothesizes that each stage is driven by distinct set of molecular forces.

In **Stage I**, hydrophobic segments in a polypeptide chain are inserted into the membrane as stable and separated helices. In this stage, the hydrophobic effect is the primary driving forces. Additionally, since the unpaired backbone hydrogen (H)-bonds in the nonpolar bilayer core is highly unfavorable, the backbone H-bonds are stabilized in the bilayer, providing another driving force for the formation of TM helices ($\Delta G^{\circ}_{\text{bilayer}} = -3$ to -5 kcal/mol per H-bond ⁶³).

In the cells, the targeting and insertion of nascent polypeptide chains to the membranes are mediated by an elaborate array of the conserved factors ^{64,65}. In contrast to the two-stage model postulating that TM helices are established upon membrane insertion; the formation of a hydrophobic helix is known to occur within the ribosome exit tunnel during translation⁶⁶. The N-terminal hydrophobic helix emerging from the exit tunnel is captured by a signal recognition particle (SRP), which arrests the elongation of the NPC and brings the NPC-ribosome complex to the SRP receptor (SR) on the cytoplasmic (FtsY in *Escherichia coli*) or endoplasmic reticulum (ER) membrane (SR α/β in eukaryotes) ⁶⁷. The receptor docks the NPC-ribosome complex onto a

translocon, followed by the release of the SRP upon GTP hydrolysis to resume translation⁶⁸. The translocon is a heterotrimeric membrane protein (SecYEG in *E. coli* or Sec61 $\alpha\beta\gamma$ in mammals). The core protein in the translocon (SecY or Sec61 α) forms a protein conduction channel, allowing the passage of an elongating NPC in two directions: A hydrophobic helical segment partitions into the membrane through a lateral gate while hydrophilic extracellular loops and domains translocate across the membrane through a vertical gate^{69,70}.

In *E. coli*, the translocon is further assembled to a larger protein complex, SecYEG/SecDF-YajC/YidC in *E. coli*⁶⁴. Those additional components are known to prevent the backsliding of a substrate during translocation (SecDF), to facilitate the membrane insertion of marginally hydrophobic TM segments (YidC), and to act as a chaperone with an unknown mechanism (YidC)⁶⁴.

Recent studies on the mammalian translocon complex have revealed an intriguing biogenesis mechanism of multispanning membrane proteins^{71,72}. While the TM segment of a single-spanning membrane protein inserts to the membrane near the lateral gate of Sec61 α , the N-terminal TM segment in a multi-spanning membrane protein, which is followed by a short interhelical loop, inserts to the membrane on the backside of Sec61 α (i.e., the opposite side of the lateral gate).⁷³ The first TM segment induces the recruitment of a series of the membrane proteins, BOS (Back Of Sec61: THEM147/Nicalin/NOMO), GEL (GET-and EMC-Like: TMCO1/C20Orf24), and PAT (Protein Associated with Translocon: CCDC47/Asterix)^{73,74}.

Based on cryo-EM, structural modeling, and biochemical data, the resulting “multipass translocon” has been modeled to form a lipid-filled cavity whose size can accommodate 6 to 7 TM helices (**Figure 1.4**)⁷⁵. This model suggests that GEL and PAT present a polar surface facing the lipid cavity such that the newly inserted TM helices possessing polar residues can be stabilized in the lipid environment via hydrogen bonding⁷⁵. The lipid cavity surrounded by the multipass translocon excludes the subsequently inserted helices from the outside milieu in the membrane, thereby allowing the TM helices to fold in the isolated space⁷². The model proposes that the multipass translocon serves as both a translocase and chaperone, protecting the NPCs

from premature intra- or intermolecular collapse during their membrane insertion and folding⁷².

These studies suggest an intricate biogenesis mechanism of membrane proteins orchestrated by multiple cellular factors. Then, how does these cellular factors modify the intrinsic folding energy landscape of membrane proteins? Inversely, how does the features of the intrinsic folding energy landscape necessitate the involvement of a specific cellular factor in the folding? Answering these questions may provide a comprehensive understanding of membrane protein folding. Notably, the partition free energies of amino acid residues measured between the translocon and ER-derived membranes are highly correlated with those between water and octanol as well as between water and the bilayer ^{45,76}, suggesting that the thermodynamic principle of membrane insertion (i.e., Stage I of membrane protein folding) is still valid in the cellular context.

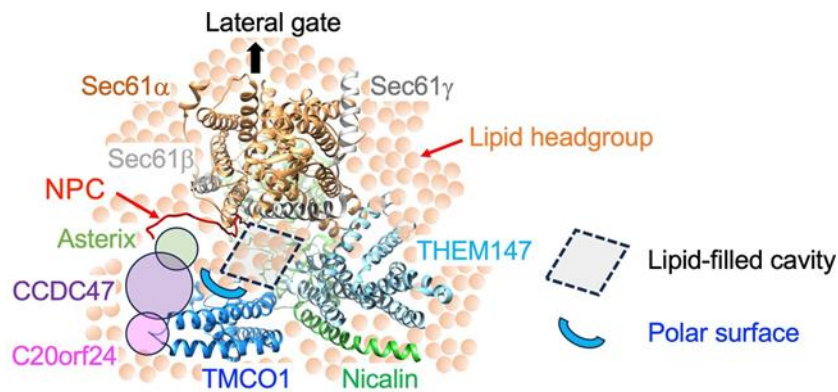


Figure 1.4. The scheme and structure of human multipass translocon⁷⁷.

Translocon mediates the membrane insertion and, potentially, folding of multispinning membrane protein (top view from the cytosolic side; PDB code: 6W6L)⁷⁸. The original structure contains a ribosomal complex (deleted here) with a bound nascent polypeptide chain (NPC). The NPC is expected to insert into the membrane at the opposite side of the lateral gate of Sec61a. The lipid headgroups in the membrane are only hypothetical and schematic. This figure was adapted from Borner and Weissman 2018 publication ⁶⁵ after modification.

In **Stage II**, individually inserted helices associate side-by-side and pack into a compact native structure. This stage is the net outcome between two competing molecular forces: **1)** the folding overcomes the conformational entropic costs (i.e., of the backbone and side chains) involved in the protein compaction; **2)** Due to the scarcity of water molecules in the hydrophobic bilayer core, the hydrophobic effect cannot drive MP folding at this stage. Therefore, to drive folding, van der Waals packing interaction, H-bonding, salt bridges, or other weak polar interactions such as π - π , π -cation and π -anion interactions that are regarded less important in the folding of water-soluble proteins should emerge as important contributors in stage **II**. Later in 2003, based on advancing knowledge of membrane protein folding, Poppt and Engleman added **Stage III**, which describes potentially oligomerizing, binding of prosthetic groups or partitioning of coil regions⁷⁹.

Although the two-stage model is still considered as valuable framework in studying the folding of α -helical membrane proteins, more evidence have shown that this model oversimplified the conformational states involved in folding, especially, the conformations (i.e., the denatured state ensembles)^{80,81} before the compaction in Stage II. For a more realistic thermodynamic description, Wimley and White suggested a four-step model (**Figure 1.3 b**) dissecting the whole folding process into four sequential steps, each of which is experimentally accessible and can be described using free energy terms¹⁹: 1) the adsorption of an unstructured coil from water to the membrane interface; 2) the transition from the random coil to structured helices on the interface; 3) the insertion of the helices from the interface to the core of the membrane; and 4) the association of the separated helices to a native helical bundle. However, in the insertion stage of membrane protein folding, the free energy derived by hydrophobic effect is already largely expended, leaving the question of what the driving forces for Stage II are open.

1.4. Molecular driving forces in membrane protein folding

1.4.1. Hydrophobic effect

Hydrophobic effects are widely accepted as a major driving force in water-soluble protein folding, as evidenced by the enrichment of nonpolar residues in the core of the folded proteins and the large increase in heat capacity upon unfolding⁸². However, in the

insertion stage of membrane protein folding, the free energy derived by hydrophobic effect is already largely expended, leaving the driving force for the inserted transmembrane helices folding unclear.

Studies have focused on analyzing the amino acid residue composition and hydrophobicity of the TM segments aiming for understanding the role of the hydrophobic effect in membrane folding. Stevens and Arkin compared the hydrophobicity of the interior and exterior of 61 individual transmembrane helices from seven structures and found them to be similar⁸³. In another study, Adamian and coworkers performed an analysis of amino acid residue composition of 29 membrane protein structures and revealed that the transmembrane protein interior exhibits lower hydrophobicity than the exterior contacting the surrounding lipids⁸⁴. However, the relative prevalence of polar residues in the interior does not imply that the hydrophobic effect is the primary driving force. A proposed hypothesis for this discrepancy in hydrophobicity is that evolution favors optimal functionality and stability, rather than solely maximizing stability. This hypothesis is supported by the presence of hydrophilic cores or pores in ion channels or transporters, which are essential for the transportation of ionic or polar solutes.

Researchers also developed the experimentally-determined hydrophobicity scales of individual amino acid residues. Wimley and White determined the partition free energies ($\Delta G^{\circ}_{\text{transfer}}$) of 20 amino acid residues using a short pentapeptide host-guest system (Ac-WL-x-LL, x can be any 20 amino acid residues of interest as a guest) from water to the hydrophobic environment, octanol (“the whole residue hydrophobicity scale” including the side chain and the peptide bond)⁸⁵. With the same pentapeptide system, they determined the $\Delta G^{\circ}_{\text{transfer}}$ of amino residues transfer from water to the water-membrane interfaces of the zwitterionic 1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine bilayer⁸⁶. Hessa-von Heijne hydrophobicity scale is developed with the Sec61 translocon-mediated insertion of *E.coli* leader peptidase (Lep)⁷⁶. The H segment contains amino acids of interest and is installed between two glycosylation sites on the P2 domain of Lep, which locate in the ER luminal side. Once H segment is integrated into the ER membrane, only the site remaining in luminal side will be glycosylated. The apparent membrane insertion equilibrium constant from translocon to can bilayer is described as comparison between the singly (inserted, f_{1g}) and doubly (translocated, f_{2g})

glycosylated Lep fraction (i.e., $K_{app} = f_{1g}/f_{2g}$), which is quantified with SDS-PAGE gel⁷⁶. Free energy of H segment membrane insertion ΔG_{app} can be calculated with the K_{app} to directly compare the membrane integration tendency of residues (i.e., the more favorable ΔG_{app} , the higher hydrophobicity). Measurements of ΔG_{app} with varying the position of amino acid and total length of H segment reveals the position and length dependence⁸⁷. Moon and Flemming reported the first experimentally determined side-chain hydrophobicity scale measured with beta-barrel membrane protein outer membrane phospholipase (OmpLA) in large unilamellar vesicles composed of 1,2-dilauroyl-sn-glycero-3-phosphocholine (DLPC)⁴⁵. In their study, Ala210 of OmpLA, which is lipid-exposed and located in the center of the hydrophobic core of the bilayer was chosen as the site of interest for substituting 19 other residues. Upon denaturation by GdnHCl, OmpLA and its variants in the membrane partition to the aqueous phase so that the difference in the free energy change of unfolding induced by each substitution relative to WT can be measured and interpreted as the difference in the water to lipid bilayer partition energy between the substituting residue and alanine⁴⁵.

In summary, Wimley-White scale estimates water-bilayer partition free energy of whole-residue within protein mimicking small peptide. Hessa-von Heijne scale is obtained from a more biologically relevant context (i.e., translocon to bilayer) revealing translocon recognition of amino acids in TM segment. While the former two scales employ peptide and transmembrane helix as study model, Moon-Fleming scale measured contribution of sidechain in native membrane protein context. Nonetheless, the partition energies of acidic residues (i.e., Asp and Glu) are underestimated in Moon-Fleming scale since the bilayer to water transfer happen at low pH, which is close to the sidechain pKa of Asp and Glu, making their desolvation costs less significant compared with physiological pH.

1.4.2. Van der Waals interactions

Van der Waals (VdW) interactions arise from temporal fluctuation in electron distribution, which create instantaneous electric fields⁸⁸. This instantaneous electric fields are felt by other nearby atoms and molecules, which in turn adjust the spatial distribution of their own electrons⁸⁸. Thus, VdW packing interaction occurs between nonpolar residues such as Gly, Ala, Val, Leu, Ile, and Met. Inside the lipid environment, the hydrophobic effect

cannot be the major driven force because of the scarcity of water. Instead, studies have shown that VdW packing interaction between separated helices becomes a more dominant force for helical association to overcome the entropically favorable helical separation and to compete with VdW between helices and nonpolar lipid tails^{89–91}. The energetic contribution of VdW interaction to membrane protein stability was quantitatively studied for the dimerization of the single membrane-spanning glycoporphin A transmembrane domain (GpATM)⁹². Subsequent mutagenesis and structural studies discovered a GxxxG glycine zipper motif in GpATM that allows the helices to closely packed, facilitating the formation of a helix-helix interface and promoting dimerization^{93–95}. Another example of VdW packing is the Lxxlxxx motif found in the homopentameric membrane protein phospholamban (PLN). A ~100 ns time-scale molecular dynamics (MD) simulation of the PLN pentamer showed that nonpolar Lxxlxxx motif in the C-terminal TM region is more rigid (backbone RMSF = 0.53Å) than the polar N-terminal (backbone RMSF = 1.53Å)⁹⁶. Based on the thermostability and MD simulation studies on the modified Lxxlxxx motifs demonstrated that the VdW packing between nonpolar residues alone is adequate for stabilizing the designed membrane protein without polar interactions⁸⁹.

1.4.3. Hydrogen bonding interactions

Using N-methylacetamide (NMA) as a model molecule, Ben-tal and coworkers computed the free energy required to form hydrogen bonds in vacuum, water and liquid alkane mimicking aqueous and hydrophobic environments⁹⁷. They discovered that transferring a pair of the unbonded polar groups in water (representing the unfolded proteins in an aqueous phase) to the hydrogen bonded pair in the non-polar phase (representing folded proteins in the membrane) is energetically unfavorable (+2.5 kcal/mol). This is primarily due to the significant cost of water- desolvation of the unbonded polar groups upon their transfer from water to the nonpolar core of the membrane. Nonetheless, the dissociation of a H-bonded pair within the membrane is highly unfavorable because the loss of favorable H-bonds cannot be compensated for the separated polar groups (~-5 kcal/mol).

While hydrogen bonding was expected equally or even more crucial for the stability of membrane protein folding, studies from the Engelman⁹⁸ and DeGrado⁹⁹

groups showed the importance of polar residues such as Asn, Asp, Gln, Glu and His mediated hydrogen bonding and oligomerization of single membrane-spanning helices. However, they also noted that the energetic contribution from H-bonds and vdW packing are comparable. The Bowie group further demonstrated that substituting Ala for either polar or nonpolar residues had similar effects on the stability of bacteriorhodopsin¹⁰⁰. Through double mutant cycle analysis, they found that the contribution of side-chain H-bonds between TM helices is only about ~ -0.6 kcal/mol on average, suggesting a modest influence of H-bonds to membrane protein stability¹⁰⁰. Bowie suggested that the discrepancy between experiment and theory arise from the lack of consideration of the membrane protein's structural context (i.e., the polarity of peptide backbones)¹⁰¹. The abundance of competing hydrogen bonds from the backbone, whose hydrogen bonding potential is not fully saturated, can compensate the need for hydrogen bonds in the denatured state where the TM helices are separated. Additionally, the dielectric constant within the lipid bilayer may be higher than in pure alkane liquid due to the penetration of water molecules into the bilayer forming competitive hydrogen bonds.

1.4.4. Salt Bridge interaction

Two oppositely charged amino acid side chains within 4 Å are regarded being engaged with a salt bridge interaction¹⁰². Based on the analysis of 222 unique salt bridges from 36 different high-resolution structures of globular proteins, Nussinov and coworker suggested that the strength of salt bridge highly depends on the geometry of the charged side chains¹⁰³. Interestingly, from their database, the most stabilizing and destabilizing salt bridge interactions were found in the same protein with a similar degree of burial while the significant difference was the angle between the lines connecting the C α atoms and the centroids of the two charged groups¹⁰³. A “good” salt bridge geometry allows more H-bonding interaction form between the side chains. Analysis of the database also demonstrated that most of the salt bridges are formed between residues that are relatively near each other in the sequence¹⁰³.

Several studies report that salt bridges participate in modulating both protein stability and functionality. In the water-soluble protein, Arc repressor, the pairwise interactions within the buried salt-bridge triad Arg31-Glu36-Arg40 stabilize the protein by -1.7 to -4.7 kcal/mol¹⁰⁴. Interestingly, substituting the charged residues in the triad with

larger nonpolar residues substantially enhanced stability by -2.1 to -3.8 kcal/mol without disrupting structure and function¹⁰⁴. This result indicates that the nonpolar residues is more beneficial to stability than the charged residue pair in the hydrophobic core of a globular protein.

In a study of the β -barrel channel protein Outer Membrane Protein A (OmpA), a series of double mutant cycle analyses were carried out to investigate the impact of the “charge tetrad” (Glu52-Arg138-Glu128-Lys82) occluding the central lumen of the small β -barrel¹⁰⁵. Interestingly, the tetrad is “insulated” by three aromatic residues that may further stabilize the tetrad by π -cation or π -anion interactions. The charge tetrad revealed favorable salt-bridge interactions with a board range from -0.6 to -5.6 kcal/mol¹⁰⁵. Notably, disrupting the central “trans” salt bridge Glu52-Arg138 significantly increased the probability of ion conduction through OmpA and disrupting the “cis” salt bridge reduced the probability. This suggests a dynamic switching of salt bridging partners—from the primary Glu52-Arg138 to the alternative pair, Glu52-Lys82 and Arg138-Glu128, mediating the gate-opening mechanism in OmpA. These findings highlight the influence of salt bridges on membrane proteins’ stability and function. Unfortunately, for α -helical membrane proteins, the low abundance of charged residues in the membrane-embedded region leads to a low occurrence of salt bridge interactions¹⁰⁶. Furthermore, it is not a trivial task to analyze the protonated states of charged residues of even in the structures of high resolution¹⁰⁷. Thus, the effect of salt-bridge interactions on stability and function has not been comprehensively studied for helical membrane proteins.

1.4.5. Entropic effect

During folding, the side chain and backbone of a polypeptide chain form intramolecular interactions and become less flexible. This process is entropically unfavorable since conformational entropy is directly related to how free bond can rotate, while interaction/bond formation will constrain it. For example, MD simulations of the native and denatured states of a small globular protein ubiquitin showed that the total change in entropy is $T\Delta S_{\text{Total}} = 1.4 \text{ kcal}\cdot\text{mol}^{-1}$ per residue at 300 K with only 20% from the loss of side-chain entropy¹⁰⁸.

Similarly, for membrane proteins, entropy can be considered as a negative contributor to their thermodynamic stability. Any strategies that mitigate significant entropy loss can therefore facilitate membrane protein folding.

One strategy is to increase the membrane protein motions in the native state and therefore reduce an entropic cost during the folding. In the comparative studies of β -barrel membrane protein OmpX in detergent, lipid bicelles and nanodiscs, NMR relaxation rates showed that transmembrane β -strands of OmpX experienced more conformational exchange in a lipid environment than in detergents¹⁰⁹. For the light-activated receptor rhodopsin, the inter-spin distance distributions measured by double electron-electron resonance (DEER) spectroscopy in detergent and nanodisc were significantly different¹¹⁰. In DDM micelles, rhodopsin exhibited one conformation, while in lipids, a more heterogeneous conformational ensemble is observed. O'Brien and coworker found that based on the distribution of NMR-derived order parameters, both helical and β -barrel membrane proteins have more dynamic side chain motions in ns-timescales compared to water soluble proteins¹¹¹. Order parameter describes the motional amplitudes of methyl-bearing sidechains in the scale of 0 to 1. When the order parameter is 0, side chains are completely disordered while the order parameter of 1 meaning sidechains are completely aligned. In this study, for both membrane proteins, the distribution of the order parameters showed a new class of higher amplitude motions (i.e., 0-0.2, which are usually rare or absent in water soluble proteins), which were not observed for water-soluble proteins.

Another strategy is to reduce the entropic cost by having residues with smaller side chains, which have fewer possible rotamers than residues with larger side chains. MacKenzie and Engelman's structure-based prediction of GpATM dimerization suggested that restrictions on the sidechain rotameric freedom are critical for GpA helix stabilization⁹⁴. Their prediction on GpATM dimer stability found that point mutation L75A is stabilizing only because WT Leu lost the larger rotamer freedom during dimerization while substituting Ala side chain did not.

1.5. Methods to study membrane protein folding

In the following sections I will describe several methods to study membrane protein folding *in vivo* and *in vitro*.

1.5.1. SDS denaturation

Sodium Dodecyl Sulfate (SDS), a strong anionic detergent, induces the disruption of non-covalent interactions in proteins, leading to their unfolding and formation of the SDS-polypeptide chain complex. This linearized denatured state solvated by SDS molecules allows proteins to be separated based on their molecular weight by electrophoresis.

When solubilized in mild detergents, the native conformation can be retained. By increasing the content of SDS in the micellar system (i.e., the mol fraction of SDS, $X_{\text{SDS}} = [\text{SDS}]/[\text{total detergents}]$), the protein can be denatured reversibly. Using an indicator reporting the folding status of the protein, an equilibrium denaturation curve can be constructed by plotting the fraction unfolded vs. SDS mol fraction. Such indicator could be intrinsic tryptophan fluorescence, circular dichroism, or other properties that can sensitively report the conformational properties of the protein. In some cases, the denaturation curves can be well represented by the two-state model involving only the folded and denatured states. The free energy of denaturation ($\Delta G^{\circ}_{\text{N-D}}$) can be obtained by fitting the denaturation curve to a model function. One of the important assumptions in the fitting model is that $\Delta G^{\circ}_{\text{N-D}}$ is linearly dependent on the SDS mol fraction and the $\Delta G^{\circ}_{\text{N-D}}$ under the native condition (i.e., without SDS) by linear extrapolation of $\Delta G^{\circ}_{\text{N-D}}$ to $X_{\text{SDS}} = 0$.

SDS-denaturation method has been employed to study thermodynamic stability $\Delta G^{\circ}_{\text{N-D}}$ of several helical membrane proteins such as bR (20 kcal/mol)¹¹², various rhomboid proteases (2.1-4.5 kcal/mol depends on their origin)⁵¹, DsbB (4.4 kcal/mol)¹¹³, and diacylglycerol kinase (16 kcal/mol)¹¹⁴. Although important insights in driving forces for membrane protein folding were obtained by this method, limitations exist as well. Detailed mechanisms of how SDS denature proteins are not clearly understood and the validation of linear extrapolation especially at lower SDS mole fraction ($X_{\text{SDS}} = 0-0.4$) has been questionable^{115,116}. For example, in the study of rhodopsin denaturation with SDS, the conformational change of rhodopsin has distinctive stages depending on SDS concentration¹¹⁷. At lower SDS concentrations (i.e., w/v 0.05-0.3%, $X_{\text{SDS}} = 0.5 - 0.85$), the increase in tryptophan fluorescence and buried cysteine reactivity suggests an opening of rhodopsin conformation¹¹⁷. At elevated SDS concentration (i.e., w/v 0.3-3%), the lowered tryptophan fluorescence and cysteine reactivity suggest a more compact

conformation during denaturation probably due to the compaction of the denatured state by the decrease in the micelle volume enriched by SDS¹¹⁷. Additionally, strong anionic SDS modifies the structure of the native lipid bilayer causing a pore formation¹¹⁸. Thus, this method is not optimal to be applied to the native lipid bilayer system.

1.5.2. Urea and GdnHCl denaturation

Urea and GdnHCl are polar chaotropic agents inducing protein denaturation by disrupting the non-covalent interactions, increasing the solubility of denatured polypeptide chains in solutions. For water soluble protein folding, both denaturants are extensively used to reversibly control denaturation. Although commonly used, the molecular level understanding in GdnHCl mediated denaturation mechanism is not fully understood. For urea-mediated denaturation, experimental and MD simulation studies suggest two mechanisms: 1) a “direct mechanism”: through the electrostatic interactions with the backbone and polar side chains of protein^{119–121}; and 2) “indirect mechanism”: via perturbing the structure of the solvent water therefore weakening the hydrophobic effect to facilitate the exposure of buried nonpolar residues^{122,123}.

Urea and GdnHCl can induce reversible unfolding and membrane extraction of several β -barrel membrane proteins (e.g., OmpA, PagP, OmpLA, and OmpW) enabling the thermodynamic analysis of their folding^{45–47}. But for helical membrane proteins, those methods are often ineffective inducing the reversible disruption of secondary or tertiary interactions. For example, bR cannot be efficiently denatured by urea and GdnHCl. The absorbance spectrum of covalently bounded retinal on bR indicates that the tertiary structure is maintained at high denaturant concentration (i.e., [Urea] = 8M, [GdnHCl] = 6M)¹²⁴. Furthermore, renaturation of bR denatured by SDS into CHAPS/DMPC mixed micelle is not significantly inhibited with the presence of 7M urea¹²⁴. Nonetheless, the folding reversibility has been achieved for a few helical membrane proteins such as GlpG¹²⁵ in micelles, GalP⁵⁰ in micelles, and LeuT in both micelles and liposomes. Unlike bR (ΔG°_{N-D} = 10 to 12 kcal/mol), those proteins have moderate stability (ΔG°_{N-D} for GlpG = 4 to 5 kcal/mol; ΔG°_{N-D} for GalP = 2.5 kcal/mol; ΔG°_{N-D} for LeuT = 2.5 to 3.8 kcal/mol), which likely allow the effective denaturation by the polar chaotropic reagents.

1.5.3. Force profile analysis

Force profile analysis (FPA) focuses on co-translational folding *in vivo* utilizing the property of a translational arrest peptide (AP) which stalls or decelerates the synthesis of a nascent peptide chain on a ribosome. The SecM-arrest peptide in gram-negative bacteria is one of the well-characterized examples¹²⁶. AP binds on the exit tunnel of the ribosome and halts the translation when its last codon in the sequence reaches the acceptor (A) site on a ribosome¹²⁷. Resolution of the arrested state by AP depends on the pulling force, which is generated by the folding of the preceding N-terminal region of the nascent polypeptide chain during translation (i.e., co-translational folding). Thus, an AP serves as a folding sensor whose ability to arrest translation can be engineered to achieve an optimal stalling/sensing efficiency¹²⁸.

The FPA has been applied to measure the folding of TM helices of helical membrane proteins^{128,129}. If the hydrophobic interaction between a nascent polypeptide segment and the membrane is strong, the membrane insertion and helix formation generate a strong force. Resultantly, the strong force facilitates the resolution of the translational arrest by AP, allowing the complete translation of the full-length nascent polypeptide chain. In contrast, a less hydrophobic TM helix generates a weak pulling force and thus, the translation of a full-length nascent chain is suppressed. The population ratio of the fully translated protein to the protein fragment can be detected by SDS-PAGE. With the FPA method, residue-by-residue analyses on three polytopic α -helical membrane proteins with different levels of topological complexity (i.e. EmrE, GlpG and BtuC) were reported¹²⁹. Combining the mutagenesis of force generating TM helices and coarse-grained MD simulation, the impacts of charged residues, hydrophobicity of periplasmic surface helix and re-entrant helix on membrane insertion have been studied¹²⁹. Thus, FPA is considered as a powerful tool for studying the propensity of a given TM segment to insert into the membrane (i.e., Stage I of membrane protein folding) *in vivo*. However, this method does not seem sensitive to the possible force generated by the lateral association between the TM helices (Stage II).

1.5.4. Single molecule force spectroscopy

Single-molecule force spectroscopy disrupts the native structure of membrane proteins by applying mechanical force on one or both termini of the proteins. In general, this

method allows folding and stability studies of infinitely diluted membrane protein while maintaining the membrane environment. There are two major types of method: atomic force microscopy (AFM) and magnetic tweezer.

In the setup of AFM, membrane protein in lipid is deposited to mica substrate on a piezo stage while a cantilever adsorbs to one end of protein and retracts away from the surface vertically. As the pull force increases, the secondary structure of protein is gradually disrupted in a stepwise manner. The generated distance versus pulling force showed a “saw tooth” pattern representing the unfolding event of the secondary structures embedded in the membrane. Yu and coworkers utilized AFM with ultrashort cantilevers providing an enhanced time resolution and force precision to study unfolding of bR¹³⁰. They observed the previously undetected intermediates separated by only two to three amino acids. However, in AFM the direction of pulling is not relevant to biological unfolding process.

In the experimental set up of magnetic tweezers, DNA handles bind protein termini between the magnetic bead and surface so that the sample membrane protein and lipids surface is not directly contact with surface. Compared with AFM, magnetic tweezers unfold membrane protein in the lateral direction so that once the pulling force is well controlled, only the tertiary interactions are disrupted while the secondary structure remain intact. This property makes magnetic tweezer a suitable tool to study the second stage of helical membrane protein folding. Decreasing the force will refold the protein. Magnetic tweezers have been employed to investigate force-induced unfolding and refolding events of helical-bundle membrane proteins at either constant or varying force^{131–135}. Compared with AFM, magnetic tweezers unfold membrane protein in the lateral direction so that once the pulling force is well controlled, only the tertiary interactions are disrupted while the secondary structure remain intact. This property makes magnetic tweezer a suitable tool to study Stage II of helical membrane protein folding.

When magnetic tweezers are applied to a membrane protein whose N- and C-termini are exposed in the same side of the membrane, the pulling force exerted in the direction parallel to the membrane allows the control of unfolding and refolding within the membrane, as measured by the one-dimensional extension of the protein⁶⁰.

Remarkably, at a series of discrete force levels (<5 pN), transitions are observed between discrete states of the protein, from which the transition rates and probability of each state can be obtained. The transition rate constants measured as a function of force are fitted to the Bell-Zhukrov model (the force and Δx^\ddagger , the distance difference between each transition state and the kinetically connected state, are independent with one another) or the Dudko-Hummer-Szabo model (Δx^\ddagger depends on force) to yield Δx^\ddagger and the free energy barrier (ΔG^\ddagger) at zero force, which are then used to reconstruct the free energy landscape of folding^{131,133,135}.

Min and coworkers studied unfolding of monomeric CIC chloride transporter from *E.coli* in DMPC/CHAPS bicelle with magnetic tweezers¹³². The CIC transporter monomer is known for having inverted symmetrical N-C subdomain topology. By gradually increasing and decreasing the force, they found there is an intermediate state between two identical unfolding steps, which suggested N and C half of CIC transporter unfold separately. This result supports the previous hypothesis that homodimeric membrane transporters with the same or opposite topologies likely evolved from ancient gene duplication¹³⁶.

1.5.5. Native mass spectrometry

Native nano-electrospray ionization ion-mobility/mass spectrometry (IM-nESI-MS) is a powerful tool (**Figure 1.5.**) to study specific protein-lipid and protein-protein interactions^{137–140}. In this state-of-the art technic, protein-lipid complexes protected by detergent droplets are gently charged and transferred into gas phase ions by a nESI capillary. Ion beam then passes through quadrupole mass filter to rapidly separate weakly bound bulk lipid and detergent molecules from the protein in a ms time scale. When the ionized proteins with selectively bound lipids are guided into collision cells, they are accelerated by the electric field applied on the cell and collide with inert gas molecules. By controlling the collision voltage, the protein-lipid complexes can be preserved or dissociated, and the native protein conformation can be maintained or expanded. The ionized protein-lipid complexes are then further separated depending on their mass, charge, and shape in an ion-mobility cell. Finally, the protein-lipid complexes are transferred into the time-of-flight mass analyzer generating mass-per-charge m/z spectra. At a moderate voltage which preserves the native protein-lipid complexes, the

signal intensity and distribution in the native MS spectra can be used to identify selectively bound lipid species and their number bound to the protein. The spectra can further be fitted into ligand binding model to obtain the dissociation constants of the lipids from the protein. By a gradual increase of voltage, the protein-lipid complexes can become subject to unfolding, from which the unfolding free energy can be obtained.

This method addresses protein-lipid interactions in membrane protein folding by identifying various phospholipids that tightly bind and their effects on the stability of helical membrane proteins such as MscL (a mechanosensitive channel of large conductance), AqpZ (aquaporin Z) and AmtB (the ammonia channel)¹³⁸. Native MS can be also applied to investigate the thermodynamics of lipid-membrane protein interactions by temperature-dependent nESI-IM-MS, leading to the evaluation of enthalpy and entropy contributions to the binding of individual lipids to AmtB¹³⁹. This method has been validated with multiple water-soluble proteins yielding similar results to conventional methods such as isothermal titration calorimetry and surface plasmon resonance¹³⁹. nESI-IM-MS has also been used to determine the dissociation constants of the AmtB-GlnK complex and further to reveal the allosteric modulation of the AmtB-GlnK interaction by lipid binding¹⁴⁰. A major advantage of this technique is that native MS requires less protein samples (0.1–1 μ M) compared with other methods to quantitatively measure lipid binding. The limitation of this method is that lipid binding thermodynamics is measured in gaseous phase instead of a real membrane environment. This method measures protein-lipid interactions in the background of detergent micelles, that is, lipids compete with detergents for binding to the protein rather than with lipids as in the cell membranes.

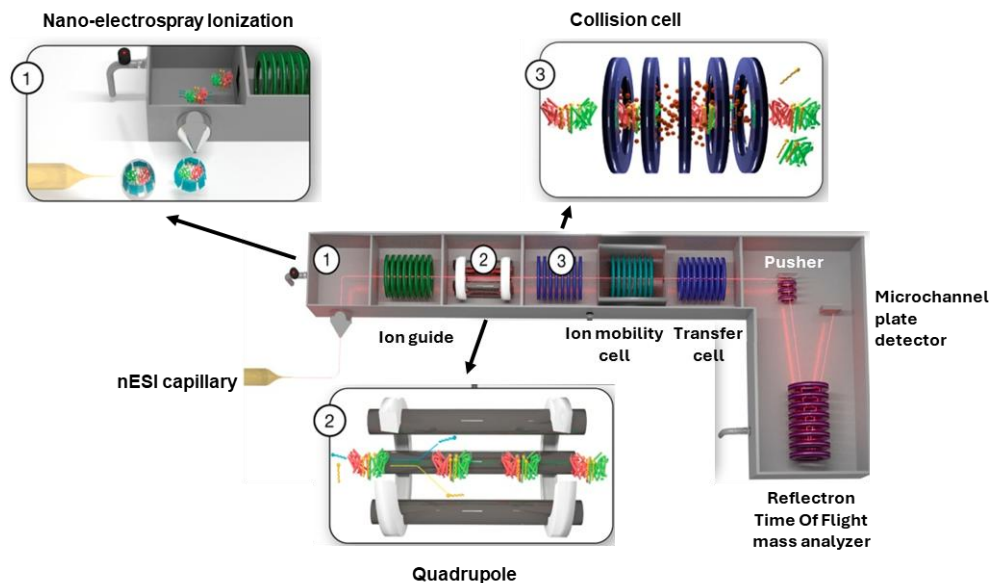


Figure 1.5. Schematic set up of protein-lipid complexes in native mass

spectrometry¹³⁷. Protective detergent micelles (cyan shells) protein–lipid complexes (protein, red and green rods; lipid, yellow sticks) droplets are charged and evaporated into gas phase by a high potential difference applied to the cone (inset 1). Quadrupole isolates the protein–lipid complex from detergent and bulky, loosely bound lipids (inset 2). Collision with inert gas in the collision cell dissociates the protein–lipid complex and further depends on the collision voltage (inset 3). Reprint permission from Springer Nature (license number: 6040830678314).

1.5.6. Steric trapping

Steric trapping enables the study of various aspects of membrane protein folding such as thermodynamic stability, spontaneous denaturation rate, conformational features of the denatured states, and cooperativity with minimal perturbation of native protein–lipid and protein–water interactions. The underlying principle of steric trapping is to couple the spontaneous denaturation of a doubly biotinylated membrane protein to the simultaneous binding of two bulky monovalent streptavidin (**mSA**) molecules (**Figure 1.5**)^{80,115,141–144}.

As an engineered version of streptavidin, mSA maintains the tetrameric structure of streptavidin while possessing only one effective biotin binding subunit (**Figure 1.6a**)¹⁴⁵. The biotin binding subunit (“alive”, the wild-type streptavidin polypeptide chain with

His₆ tag) retains the high biotin affinity of wild-type streptavidin ($K_{d,\text{biotin}} = \sim 10^{-14}$ M) and the three inactive subunits (“dead”, a triple mutant N23A/S27D/S45A without His₆ tag) has a dramatically reduced biotin affinity ($K_{d,\text{biotin}} = \sim 10^{-3}$ M)¹⁴⁵. Thus, at working concentrations of mSA (up to $\sim 10^{-4}$ M), binding of biotin to the dead subunits is negligible such that one mSA molecule can bind to one biotin molecule at the alive subunit.

To apply the steric trapping strategy, biotin tags are conjugated to a target protein at two specific residues that are spatially close in the folded state and distant in the amino acid sequence (**Figure 1.6b**)^{115,116}: A first mSA binds to either biotin label with an intrinsic binding affinity ($\Delta G^{\circ}_{\text{Bind}}$). Binding of a second mSA is inhibited due to the steric hindrance with the first bound mSA but allowed when the native tertiary contacts are transiently unraveled. Thus, the binding of the second mSA is attenuated depending on the thermodynamic stability of the protein ($\Delta G^{\circ}_{\text{Bind}} + \Delta G^{\circ}_{\text{N-D}}$).

We have engineered mSA to control the affinity to biotin and the off-rate of bound biotin in a broad range ($K_{d,\text{biotin}} = 5 \times 10^{-11}$ M to 8×10^{-6} M; $t_{\text{off},\text{biotin}} = \text{seconds to hours}$) by amino acid substitution on the biotin binding pocket in the alive subunit^{115,146}. Thus, the mSA binding and denaturation reactions can be reversibly controlled by using a mSA variant with a reduced biotin affinity and by changing its concentration. A mSA variant with a higher biotin affinity is adequate when the stability of the target protein is high. A mSA variant with a lower biotin affinity is used when the stability is low. In this way, the second mSA binding effectively competes with the refolding of the target protein, yielding an optimal attenuation of the second binding (in the range up to 40-80 mM [mSA]). The $\Delta G^{\circ}_{\text{N-D}}$ can be obtained by fitting the attenuated second binding phase or the degree of protein denaturation monitored as a function of mSA concentration to the model function^{115,116,142}. The fitting model has been derived from the coupled equilibria between the spontaneous denaturation of the target protein with one bound mSA and the second mSA binding^{142,144}.

The application of the model is validated by demonstrating that the first binding of mSA does not affect the conformation and stability of the target protein and that the second mSA binding and protein denaturation are coupled^{116,144}. Steric trapping has a broad dynamic range of measurable stability ($\Delta G^{\circ}_{\text{N-D}} = -2$ kcal/mol to -8 kcal/mol)¹¹⁵.

This method has enabled the stability measurements of strong TM helix–helix interactions in lipid bilayers^{141,143}, bR in bicelles, and GlpG in micelles and bicelles^{115,141,147–149}.

Steric trapping can also be used as a tool to measure the spontaneous denaturation rate of a target protein^{144,150}. Binding of biotin to wild-type (WT) streptavidin is known to have a high on-rate ($k_{\text{on,biotin}} = \sim 10^8 \text{ M}^{-1} \text{ s}^{-1}$) and low off-rate ($k_{\text{off,biotin}} = \sim 10^{-5} \text{ s}^{-1}$), which results in an enormously high affinity ($K_{\text{d,biotin}} = \sim 10^{-14} \text{ M}$)¹⁵¹. By increasing the concentration of mSA WT to the level at which the timescale of the second mSA binding is shorter than the lifetime of the denatured state ($\tau_{\text{on,biotin}} \ll \tau_{\text{D}}$, that is, $k_{\text{on,biotin}} [\text{mSA}] \gg k_{\text{F}}$), the denatured state can be captured as soon as it is formed. Because the maximal speed limit in the formation of a helical hairpin is estimated to be $\sim 10^{-2} \text{ s}$ in the bilayer and the binding of biotin to mSA rapidly occurs in $\sim 10^{-4} \text{ s}$ (at $[\text{mSA}] = 10 \text{ }\mu\text{M}$)^{151,152}, it is expected that the transiently denatured protein is effectively captured by the second binding of mSA. Consequently, the overall reaction rate is determined by the spontaneous denaturation rate of the target protein ($\sim \text{min}$ to $\sim \text{days}$)^{144,150}. By design, steric trapping detects the opening of the region encompassing two specific biotinylated sites in a target protein. Therefore, the local thermodynamic and kinetic stability of a protein can be measured by moving the position of a biotin pair within the protein¹¹⁵.

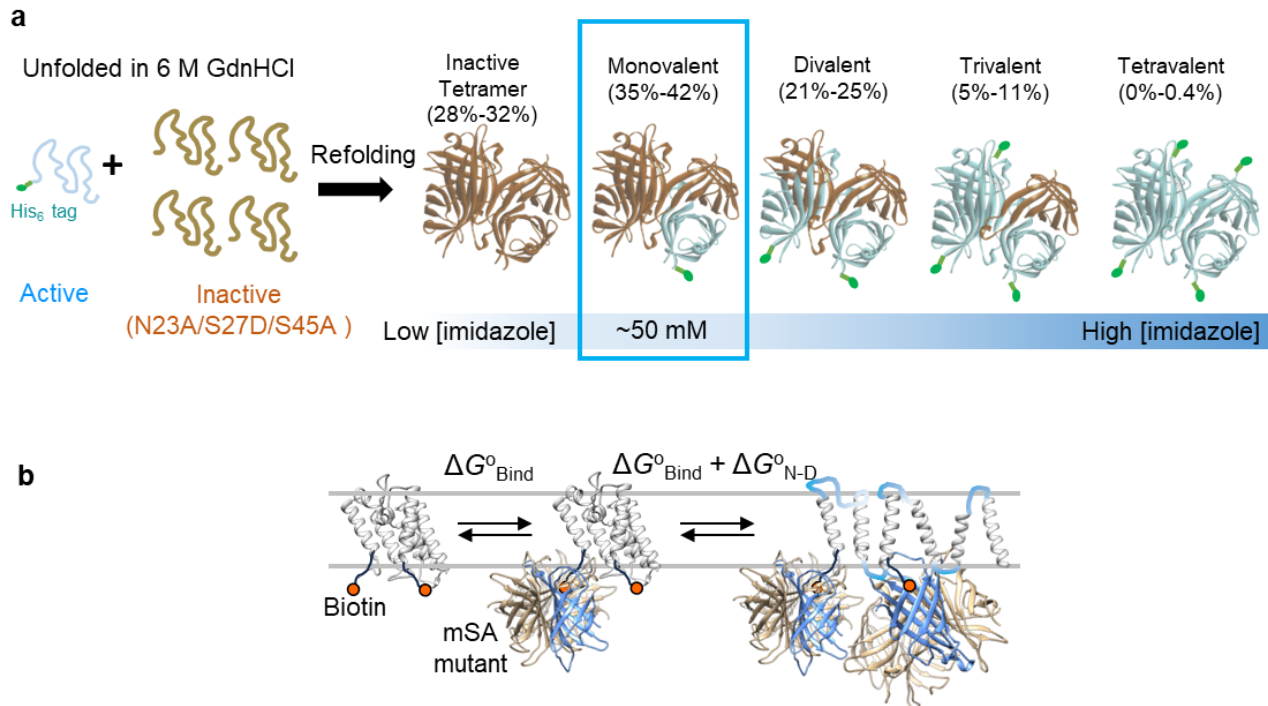


Figure 1.6. Schematic set up of monovalent streptavidin preparation and steric trapping strategy⁷⁷. (a) A mixture of denatured active (with His₆-tag) and inactive (the N23A/S27D/S45A mutant without His₆-tag) streptavidin subunits (molar ratio = 1:4) in 6 M GdnHCl is refolded in pH 7.5 sodium phosphate buffer solution. The resulting refolding products contain inactive, monovalent, divalent, trivalent, and tetravalent tetramers, which have differential affinities to a Ni-NTA affinity resin. Monovalent streptavidin is selectively eluted at ~50 mM of imidazole concentration. Distribution of each tetramer species is from binomial prediction and densitometry data (PDB ID: 1STP). (b) The reaction scheme for measuring thermodynamic stability ($\Delta G^{\circ}_{\text{N-D}}$) of α -helical membrane protein GlpG (PDB ID: 3B45).

1.6. Intramembrane rhomboid protease as study model

Proteases, also known as peptidase or proteinase, are enzymes that catalyze the hydrolysis of peptide bond present in protein substrates. Intramembrane proteases are embedded in cell membranes and involved in protein quality control, protein processing, and activation of signaling pathways through regulated proteolysis^{153,154}. So far, there are four major families of intramembrane proteases: 1) site-2 metalloproteases (S2P)

with two His residues coordinated to zinc to catalyze proteolysis¹⁵⁵; 2) γ -secretases which are aspartyl proteases with two conserved TM Asp residues in the active site¹⁵⁶; 3) rhomboid serine proteases with a unique Ser-His catalytic dyad rather than the more common Ser-His-Asp catalytic triad in water-soluble serine proteases^{157,158}; 4) the most recently described glutamyl proteases with the conserved Glu catalytic residue¹⁵⁹.

In this dissertation, the rhomboid proteases are the main model membrane proteins for studying helical membrane protein folding. The name, 'rhomboid', originates from *Drosophila* genetics screening. Genes are typically named after the altered phenotype during embryo development disturbed by mutation. Thus, rhomboid earned the name from the mis-shaped, rhombus-like head skeleton of *Drosophila* larva¹⁶⁰. The rhomboid family proteins are found in all branches of life¹⁶¹. There are four major types of function that are carried out by rhomboid proteases (**Figure 1.7.**): 1) Activating signaling pathways¹⁶²: Spitz, a membrane-anchored growth factor in endoplasmic reticulum, is transported to Golgi Apparatus, where it is recognized and cleaved by Rhomboid-1. The cleaved Spitz is then released from the mother cell and activates the epidermal growth factor (EGF) pathway of the neighboring cells, facilitating *Drosophila* development; 2) Mediating selective gene expression in bacteria¹⁶³: In *Providencia stuartii*, rhomboid proteases AarA target TatA, a protein that is activated upon the removal of a small amino terminal extension. Activated TatA engages in the assembly of twin-arginine translocation machinery, promoting protein export. The process enables communication between cells and leading to selective gene expression, a phenomenon known as quorum sensing; 3) Regulation of mitochondrial homeostasis¹⁶¹: In human mitochondria, rhomboid PARL cleaves mitochondrial kinase PINK1 thus prevents it from recruiting Parkin ubiquitin ligase. Cleavage of PINK1 leads to disruption of PINK1/Parkin pathway and consequentially reduces the rate of selective degradation of damaged mitochondria; 4) Parasite invasion: During the invasion of host blood cells by the Malaria parasite *Plasmodium*, adhesion occurs between the parasite and the host cell's membrane protein, adhesin¹⁶⁴. A rhomboid protease cleaves adhesin, enabling the parasite to detach to ensure its survival. It is expected that there are still other rhomboids with unknown functions.

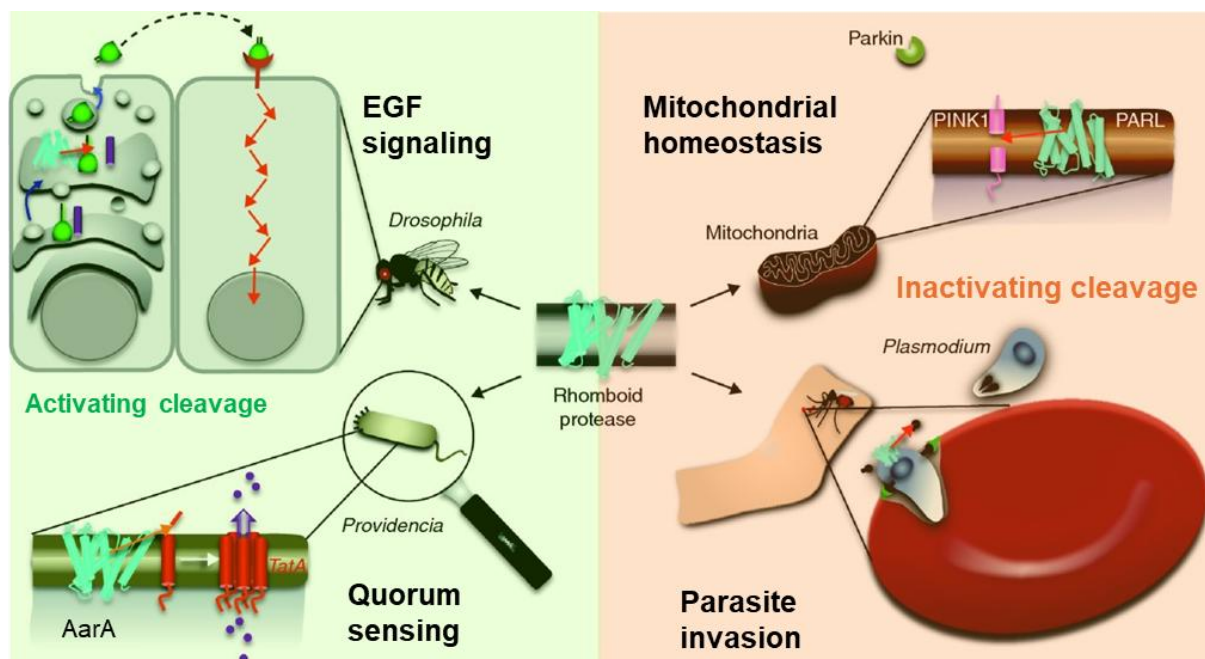


Figure 1.7. Four general categories of rhomboid protease cellular roles¹⁶¹. In the left green shadowed roles, cleavage of the substrate activates certain pathways. In the right orange shadowed roles, rhomboid cleavage inactivates the target protein or terminate process. Upper left: Initiating EGF signaling during *Drosophila* development by Rhomboid-1, which is localized in the Golgi apparatus. Rhomboid-1 cleaves Spitz (green) after it is transported from the endoplasmic reticulum by Star (purple). Bottom left: In *Providencia stuartii*, rhomboid protease AarA activates TatA (red) by cleaving its short terminal extension. Cleaved TatA assemble into the twin-arginine translocation machinery contributing to quorum-sensing. Top right: Rhomboid PARL in mitochondria cleaves PINK1 to suppress Parkin recruiting to mitochondria and slow down mitophagy. Bottom right: Rhomboid protease from Malaria parasite *Plasmodium* cleaves adhesins (black) to disassemble the connection between parasite and host cell (red) at the end stage of parasite invasion. Reprint permission from Springer Nature (license number: 6040830979368).

The rhomboid protease *E. coli* GlpG, is the first intramembrane protease whose structure has been solved¹⁶⁵, and its structure and the catalytic mechanism are relatively well characterized. Later, the crystal structure of *Haemophilus influenzae* GlpG has been solved.¹⁶⁶ So far, *E. coli* and *Haemophilus influenzae* GlpG are still the only

two rhomboid proteases with known structures. Both structures show that the tightly packed six TM helices form a catalytic core. Crystal structures revealed that the catalytic dyad, Ser201 from the GxSG motif in TM4 and His254 from the (A/G)H motif in TM6, is buried ~10 Å below the membrane surface¹⁶⁵. Subsequent structural and mutational studies have identified several highly conserved motifs as essential components for catalysis other than the catalytic dyad¹⁶⁷. His150 and Asn154 from the HxxN motif stabilizing the Ser201-formed oxyanion hole during catalysis. Structurally, the G(A)xxG(A) motifs in both TM4 and TM6 are highly conserved, enabling the tight packing of the two helices harboring the catalytic dyad. These findings have provided critical insights into the structural basis of intramembrane proteolysis and laid the groundwork for further exploration of the still incompletely understood catalytic mechanisms of the rhomboid protease family.

1.6.1. Rhomboid protease substrate sequence specificity

Although the details of rhomboids' proteolysis mechanism are not fully understood, the structural features shared by its preferred substrates provide useful information. The substrates are typically biotopic membrane proteins, which typically form TM helices in a hydrophobic membrane environment. However, such rigid secondary structure is typically not subject to proteolysis, and local unfolding is required for proteolysis to occur¹⁶⁸. To be successfully recognized and proteolyzed, the substrate should be partially unstructured near the scissile bond¹⁶⁹. As expected, helix-destabilizing motifs such as the Gly-Ala motif in Spitz from *Drosophila* Rhomboid-1, are observed in multiple other rhomboid substrates¹⁶⁹. Notably, Spitz can be cleaved by multiple rhomboid protease orthologs, suggesting the substrates share conserved sequence/structural characteristics¹⁷⁰.

Strisovsky and coworkers have found that three bacterial rhomboid proteases with different predicted membrane topologies (AarA from *P. stuartii*, GlpG from *E. coli* and YqgP from *B. subtilis*) cleave four different substrates (TatA, Gurken, Spitz and LacYTM2) at the same exact peptide bond in the TM domains of the substrates¹⁷¹. A further positional mutagenesis screening and the crystal structures of GlpG bound with substrate-derived peptidyl inhibitors show that small amino acids without branched

sidechain (e.g., Ala and Cys) are preferred on the P1 position of the substrate while the negatively charged residue (e.g., Asp) is not preferred anywhere from P1 to P4¹⁷².

1.6.2. Proteolytic mechanism of rhomboid protease GlpG

Studies agree that the binding of the TM substrate involves the flexible TM5 helix.

Previously, it has not been clearly understood how the scissile bond of the TM substrates contacts the active site due to the lack of GlpG structures bound with a substrate. Two hypotheses have been suggested: The first is that TM5 with ~8 residue-flanking loops undergoes lateral motions within the membrane acting as a substrate gate. The lateral gate opens to allow the substrate to deeply enter and contact the active site. This hypothesis is supported by experimental evidence that replacing three aromatic and bulky nonpolar residues at the TM2-TM5 interhelical site (i.e., Trp236, Phe232, and Leu229) with valine dramatically increases proteolytic activity of GlpG compared to WT¹⁷³. However, when three pairs of TM2-TM5 interfacial residues (Phe153/Trp236, Trp157/Phe232 and Tyr160/Leu229) are engineered into Cys residues and tethered by chemical cross-linkers, GlpG proteolytic activity does not change¹⁷⁴.

This result indicates that the substrate binding might not require large movement of TM5. The second is the loop 5 (L5) cap hypothesis with the TM2-TM5 interface still serving as a potential docking site. This hypothesis suggests that rather than TM5 helix, L5 opens up and the region containing the scissile bond in the substrate bends towards the active site upon docking. This hypothesis is supported by the conformational plasticity of L5 in the crystal structures¹⁷⁵. Urban and coworkers have solved ten time-resolved structures of *E. coli* GlpG spanning the near-entire steps of the proteolytic reaction in the presence of an aldehyde inhibitor. These structures suggest that the movements of both TM5 and L5 play important roles in the substrate binding and proteolysis¹⁷⁶.

1.6.3. Folding studies of rhomboid protease GlpG

Besides the catalytic mechanism, the folding and stability of GlpG have been carried out in various hydrophobic environments by several groups using different methods. Urban and Baker performed a comprehensive thermal denaturation and activity study with 151 mutants in DDM micelles⁵¹. The changes in melting temperature and proteolytic activity by mutations were used to identify critical structural elements for stability and function.

The resultant structural mapping led to the discovery of four ‘key stone’ regions (**Figure 1.8.a**): 1) The H-bonding network that stabilizes the hairpin conformation of the L1 loop on the extracellular side; 2) The H-bonding network along TM2, L2 and TM3 on the cytoplasmic side; 3) the tight helix-helix packing between the asymmetric “glycine zipper,” GxxxGxxxA on TM6 and GxxxAxxG on TM4 helix; 4) another packing core mainly mediated by large nonpolar residues (i.e., Phe197, Leu200, Val204 and Leu 207) and small residues (i.e., Gly194 and Gly199) in the N-terminal TM helices (TM1 to TM4) and L3. Interestingly, despite the suggested importance of the H-bond networks in the first two “key stones”, the double mutant analysis showed that the individual H-bonds are weak (-0.58 ± 0.2 kcal/mol for Glu166-Thr97 and -0.58 ± 0.2 kcal/mol for Asp268-Lys173).

Otzen and coworkers performed Φ -value analysis on 69 residue sites in the TM domain of GlpG¹⁷⁷ in DDM/SDS mixed micelles (**Figure 1.8b**). Here, increasing SDS fraction denatures the protein whereas increasing DDM renatures protein. Φ -value represents the ratio of the change in activation free energy of folding to the change in free energy of folding between WT and a mutant. $\Phi = 0$ indicates that the mutant site is unfolded in the transition state; $\Phi = 1$ indicates that the mutant site is folded structure in the transition state. The Chevron plot (i.e., the unfolding and refolding rates as a function of denaturant concentration) displays a V-shape indicating that only two states (i.e., folded and unfolded) exist during folding. The Φ -values obtained were categorized into three classes: 1) Mutated residues with large positive Φ values mostly populated at the cytoplasmic side of TM1 and TM2, which were identified as the folding nucleus; 2) Residues on TM3-TM6 yield near-zero Φ values, indicating that this part of the protein is unfolded in the transition state. This observation is reasonable since the catalytic dyad Ser201-His254 is in TM4 and TM6, and TM5 is thought to be flexible to allow the substrates to approach the catalytic dyad. Residues with uncommon negative Φ values are distributed along the loops 1-3 and this region is proposed to experience conformational rearrangements (“backtracking”) to correctly position the folding nucleus (i.e., cytoplasmic side of TM1 and TM2) while assisting the folding of the rest of the domain.

Yoon and Bowie group performed the single-molecule studies with magnetic tweezers for GlpG in DMPC/CHAPSO bicelles¹³⁵. This study focused on Stage II of GlpG folding since the magnetic tweezers can reversibly control the unfolding and refolding in the membrane. As the pulling force increases, GlpG remains compact until the force reaches 25 pN and then unfold with the length extension by 40 nm. This result suggests unfolding of GlpG is highly cooperative without major intermediate. In the 'force-jump' experiment, the pulling force is increased rapidly and maintained at the constant force, 21 pN. While a major fraction of unfolding events exhibited a single step (i.e., 56% from the native to unfolded state), unfolding via intermediates (I_1 , I_2 , or both) were also observed. The dwell times of the two intermediate states were much shorter compared to the dwell time of the unfolded state (τ_1 and $\tau_2 < 2\%$ of τ_u), implying there is only one major energy barrier during unfolding, thus, further providing evidence for largely cooperative unfolding of GlpG. Direction of the mechanical unfolding is tested by introducing mutations that can locally destabilize GlpG (i.e., L155A for N-terminal region and A206G for C-terminal region). The changes in probability of I_1 and I_2 caused by the mutations suggest that the unfolding starts from the C-terminal region and then propagates to the N-terminal region as suggested by the previous Φ -value analysis¹⁷⁷. To construct the folding energy landscape, GlpG was unfolded and refolded over the different ranges of applied forces (13–33 pN to unfold, 1–7 pN to refold). The kinetic barrier for unfolding from the folded state was high indicating that the folded state has a long lifetime ($t_{1/2} \sim 3.5$ h). Putting together, these observations suggested that the high folding cooperativity and high unfolding kinetic barrier of GlpG lowers the probability of partially unfolded states during folding.

The Hong group studied thermodynamic stability, cooperativity, packing interactions, and the compactness of the denatured state of GlpG in micelles, bicelle, or lipid bilayers using steric trapping^{115,147,148,178}. They developed novel biotin tags possessing the thiol reactive and spectroscopic (fluorophore or spin-label) reporter groups¹¹⁵. By changing the position of the biotin pairs and measuring the thermodynamic stability, they demonstrated distinct folding properties of the two subdomains (i.e., the rigid N-subdomain and the flexible C-subdomain). They further developed the "cooperativity profiling" method to quantify the degree of propagation of

mutation-induced structural perturbation. Through the mapping of the cooperativity profiles on the structure (**Figure 1.8.c**), they discovered the networked nature of the residue interactions in GlpG and that the stability of GlpG is maintained by the cooperative, localized, and overpropagated residue interactions in micelle and bicelle environments^{115,147}. The cooperativity map revealed that cooperative and localized interactions are clustered in three distinct regions in micelles: the buried region in N-subdomain near the center of the bilayer and the structured L1 loop, the residues near the water-retention site close to the catalytic dyad, and the TM4/TM6 interface harboring the catalytic dyad. Interestingly, the cooperativity map in DMPC/CHAPS lipid bicelle shows that the residues in the nearly whole packed regions are engaged with cooperative interaction, indicating that the lipid environment in bicelles can facilitate the propagation of local structural perturbations throughout protein¹⁴⁷.

Steric trapping was applied to study the impact of the native structural cavities on the stability and activity of GlpG¹⁴⁸. Using cavity-filling mutations, they found that improving packing can stabilize the protein at the expense of compromising activity. This finding suggests that the cavities have evolved to balance stability and activity of membrane proteins.

An advantage of steric trapping is that the denatured state ensemble (DSE) of GlpG can be captured enabling the detailed characterization of the DSE using various biophysical and biochemical techniques such as double electron-electron resonance (DEER) spectroscopy, limited proteolysis, and mass spectrometry in micelles, bicelles, and liposomes¹⁷⁸. By combining the experimental results with MD simulations for generating the reference DSE's, they found that the DSE of GlpG in the lipid bilayer is partially expanded but not collapsed. This suggests that the lipid bilayer is not either a good or bad solvent for the DSE of membrane proteins, allowing some degree of association of the TM helices.

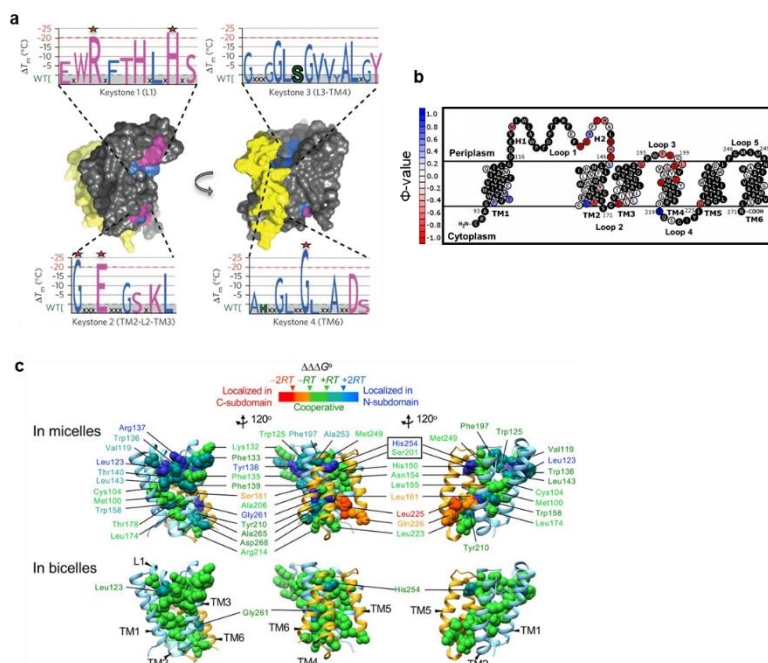


Figure 1.8. Representative folding and stability features of GlpG^{51,115,147,177}. (a) Four “key stone” regions⁵¹ depicted with space-filling model highlighting architectural properties underlying GlpG function. Gray represents areas of packing interactions which contribute to the low thermodynamic stability and high environmental responsiveness of rhomboid proteases. Regions with predominant hydrogen-bonding residue interactions are in pink. Strong residue-packing interactions are illustrated in blue. Yellow highlights the critical regions for dynamic functions during proteolysis. The destabilizations of each mutation in each region are denoted quantitatively by the size of each letter (catalytic serine and histidine are in green, and the five most important stabilizing residues are highlighted with stars). Reprint permission from Springer Nature (license number: 6041170879165). (b) 2-D topological diagram of GlpG¹⁷⁷. Residues in are colored according to their ϕ -values as indicated by the color bar on the left (c) Cooperativity profiles mapped within detergent and lipid on GlpG structure^{115,147}. The color code of each residue’s cooperativity profile: “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, $-2RT > \Delta\Delta\Delta G$).

1.7. Thermophilic proteins

1.7.1. Characteristics of thermophiles

The word “thermophile” comes from two Greek words, “thermotita”(means: heat) and “philia” (means: love). Thermophiles refer to microbes that endure and thrive at relatively high temperature (up to 122 °C) and high pressure (200–500 atm). The anaerobic, chemoautotrophic thermophiles are thought to be the first microbe to thrive on earth about 4 billion years ago¹⁷⁹. Studies of thermophiles were mainly initiated by Thomas Brock in the 1960s, after the ground-breaking discovery of microbes from the hot springs of Yellowstone National Park¹⁸⁰. A phylogenetic tree (**Figure 1.9.**) has been built based on the sequences of small-subunit ribosomal RNAs. Starting from the universal common ancestor occupying the root, the tree has tripartite divisions into bacterial, archaeal and eukaryotic domains¹⁸¹. The bold branches represent thermophiles. Thermophilic organisms widely exist in all three domains of life. Thermophiles occupy mainly in the deeper and shorter branches suggesting that they emerged in early times and evolved at a slow rate.

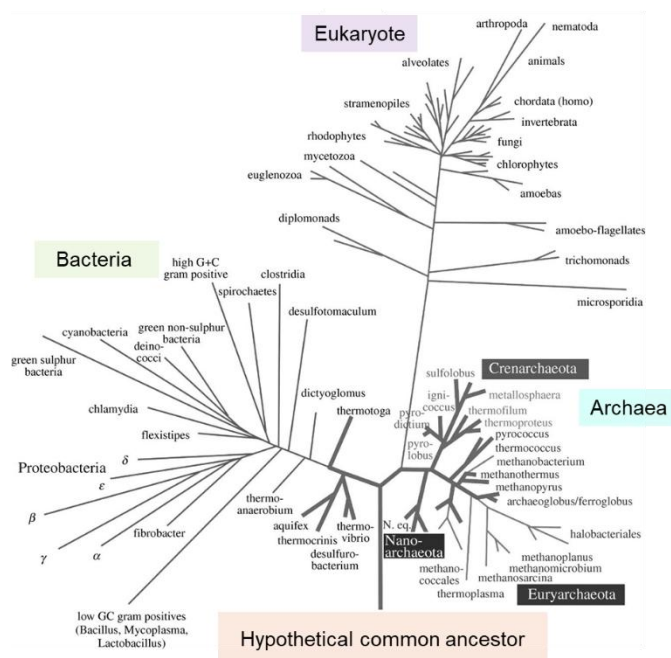


Figure 1.9. Small subunit ribosomal RNA-based phylogenetic tree¹⁸¹. The root of the tree represents the common ancestor of all species. Length of branches correlate to rate of evolution: the longer the branch, the faster the evolution rate. The bold branches in this tree represent thermophiles.

1.7.2. Thermophiles and their biotechnological applications

Due to the tolerance to high temperatures, proteins from thermophiles have garnered many interests for their industrial applications. One of such examples is the widely used high-fidelity DNA polymerase (i.e., the error rate of 1 out of 1.3 million base pairs) discovered from thermophilic archaea *Pyrococcus furiosus*¹⁸². Enzymes isolated from anaerobic thermophiles (i.e., *Clostridium thermocellum* and *Caldicellulosiruptor saccharolyticus*) have a capability to convert lignocellulosic biomass into hydrogen at elevated temperatures^{183,184}. Some thermophiles can also degrade petroleum hydrocarbons¹⁸⁵. Two strains of thermophilic bacteria from the *Bacillus* family can concentrate the radioactive strontium on both sides of a bioreactor into one side, a process that can be used to remove the heavy metal contaminations¹⁸⁶. Natural compounds isolated from thermophiles have been found to be promising drug candidates. Two natural compounds isolated from *Aspergillus terreus* significantly repress the growth of ABCG2-expressing breast cancer cells¹⁸⁷ and inhibit the inflammation caused by acute kidney injury¹⁸⁸. Dimeric ferritin from *Thermotoga maritima* can be engineered to form diverse nanocontainers of drug molecules for the delivery and targeting to cancer cells^{189,190}.

1.7.3. Features of membrane lipids from thermophiles

Temperature is one of the most important environmental factors that impact life by changing the behaviors and stability of biomolecules. High temperatures close 100 °C can denature nucleic acids and proteins and increase membrane fluidity to a lethal level¹⁹¹.

To stay functional as a permeability barrier and mediator of membrane proteins' function, the cell membranes in thermophiles should be maintained to the liquid crystalline phase at high temperatures. In the liquid crystalline phase of the membranes, the lipid molecules are highly mobile like liquid but oriented in a range of direction like solid. To maintain these properties, thermophiles have membrane lipid compositions highly distinguishable from mesophiles, reflecting their adaptation strategies to high temperatures. As the temperature increases, the membranes shift from the gel phase to fluid and eventually to a nonlamellar form with a negative monolayer curvature. Longer and saturated (**Figure 1.10.a**) fatty acid chains packed better than shorter and

unsaturated ones thus will be more rigid at higher temperatures. The cyclopentane rings in fatty acid chain (**Figure 1.10.d**) restrict the chain motions (analog to cholesterol molecule) therefore stabilize membrane at higher temperatures.

Other than fluidity adjustment, the stability of individual lipid molecules is also critical to maintaining the functional membranes at high temperatures. The lipids with ether backbones (**Figure 1.10.b**) found in thermophilic archaea (e.g., *Pyrococcus furiosus*) and bacteria (e.g., *Aquifex pyrophilus*) are known to be more chemically stable than the lipids with ester backbones, which are prevalent in mesophiles^{192,193}. The presence of bipolar tethered lipids (**Figure 1.10.c,d**) is considered as another outstanding feature of membrane lipids in thermophilic archaea. Phospholipid molecules with one polar, two fatty acyl chain pack in an end-to-end manner to form a bilayer. In contrast, the bipolar lipids contain two polar head groups that are tethered by two long tails to form a single layer with a similar thickness to a bilayer in mesophiles. Previous studies with thermophilic archaea have shown a strong correlation between the living temperature and the fraction of the tethered lipids^{194–196}. Due to the presence of condensed tail, the motions of hydrocarbon chains are extensively restricted. Thus, the leakage caused by high temperatures can be effectively reduced as shown by a carboxylfluorescein permeability assay¹⁹⁷. MD simulations revealed that tethered lipids in the monolayer possess a less torsional entropy compared to the phospholipids at elevated temperatures¹⁹⁷.

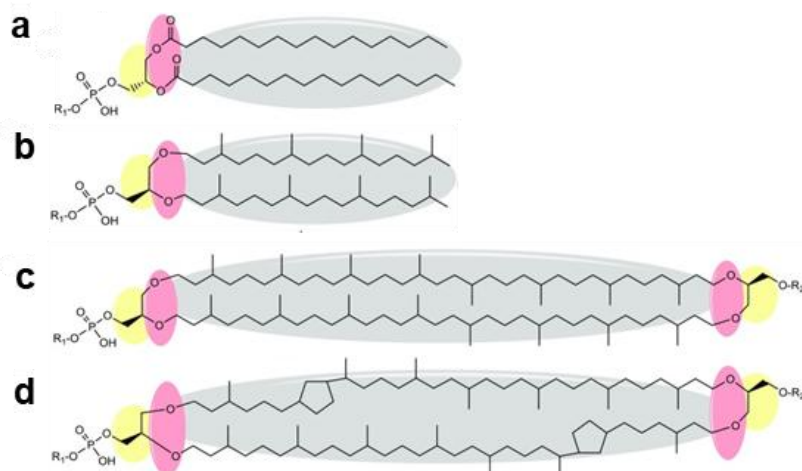


Figure 1.10. General examples of lipids from thermophiles compare with mesophiles¹⁹². Structural elements of lipid molecules are highlighted by three colors. Hydrocarbon chains in grey are represented by fatty acids in bacteria and isoprenoid chains in archaea. Ester or ether bond link the hydrocarbon chains to the glycerol backbone are in red. In archaea, hydrocarbon chains are attached to the glycerol backbone by ether-bonds. The backbone moiety highlighted by yellow represents glycerol-3-phosphate in bacterial lipids and glycerol-1-phosphate in archaeal lipids. Head groups *R1* represent phosphate polar heads and *R2* represents single or multiple hexoses. **(a)** representative phosphate lipid in mesophilic bacteria **(b)** lipid in thermophile with phytanyl chains **(c)** archaeal bipolar glycerol dialkyl glycerol tetraether with phytanyl chains spans the membrane to form lipid monolayers **(d)** archaeal bipolar glycerol dialkyl glycerol tetraether contains 2 cyclopentane rings in the phytanyl chains spans the membrane to form lipid monolayers.

1.7.4. Features of thermophilic proteins

Thermophilic and mesophilic proteins have distinguishable amino acid compositions. In general, on the protein surface, there are more positively and negatively charged residues in thermophilic proteins than their mesophilic homologues¹⁹⁸. Resultantly, the formation of salt bridges at the protein surface could be a stabilization strategy for thermophilic protein. Moreover, aromatic residues such as Phe, Tyr, and Trp are significantly more abundant in thermophilic proteins than in mesophilic proteins, playing a crucial role in the enhanced thermal stability¹⁹⁹. These aromatic residues are typically

clustered in the protein surface and contribute to thermostability by forming favorable aromatic-aromatic interactions²⁰⁰.

The Bowie group compared the structural properties (e.g., side-chain burial, packing, hydrogen bonding, transmembrane kinks, loop lengths and hydrophobicity) of 25 nonhomologous helical membrane proteins from thermophiles and 101 helical membrane proteins from mesophiles²⁰¹. Interestingly, while most of these properties are similar, two properties are significantly different: 1) A slightly smaller number of interhelical H-bonds is found in thermophiles compared to that of mesophiles; 2) Thermophilic membrane proteins are overall more hydrophobic in TM helices than thermophilic membrane proteins. These results agree with the sequence analysis of a larger membrane protein database showing the suppression of polar (i.e., Asn, Gln, Tyr) and ionizable (i.e., Asp, Glu, Arg) residues and the increase of small and nonpolar residues (i.e., Ala, Gly, Phe, Leu) in thermophilic membrane proteins. This study suggests that ensuring the membrane embedment of TM helices is more important than forming interhelical H-bonds to increase thermostability of membrane proteins.

1.8. Project Description

In this dissertation, I tackled two unresolved problems regarding the stability of helical membrane proteins using rhomboid proteases as a study model.

In chapter 2, I investigated how buried ionizable residues impact the thermodynamic and kinetic stability of GlpG of *E. coli* in two different hydrophobic environments, detergent micelles and lipid bicelles. While TM segments of helical membrane proteins are mainly composed of aliphatic or aromatic residues, polar and ionizable (Asp, Glu, and Lys) residues are found with relatively low frequencies²⁰². Despite the low abundance, those types of residues are known to be important for membrane protein function such as formation of proton channel²⁰³, activation of receptor²⁰⁴ and mediating transporter activity²⁰⁵. Here, we aim to answer three specific questions: 1) How do ionizable residues buried in the interior of a membrane protein impact stability compared to nonpolar residues with similar sizes? 2) Is the paired ionizable residues (i.e., Glu-Lys or Asp-Lys) buried in the protein interior stabilizing the protein? 3) Does the paired ionizable residues form an ion-pair (i.e., a salt bridge) or a charge-neutral hydrogen bond? To answer these questions, I first identified cavities in

the interior of GlpG and engineered two spatially proximal nonpolar residues contacting the cavity into ionizable residues individually and simultaneously. To create potential ion-pair interactions, the two nonpolar residues were replaced by acidic (Glu or Asp) and basic (Lys) residues. I first measured the kinetic stability (i.e., spontaneous denaturation rates) of WT and the variants using a proteolysis assay combined with the SDS-PAGE. I found that in both micelle and bicelle, the introduction of single or double ionizable residues in the protein interior dramatically decreased the kinetic stability of GlpG.

Next, I employed the steric trapping strategy to measure the thermodynamic stability of WT and variants in micelles and bicelles. I found that ionizable residues can be tolerated in the protein interior with a moderate destabilization by $\Delta\Delta G^{\circ}_{WT-Mut} = 1.4$ to 3.4 kcal/mol in micelles and 1.2 to 2.7 kcal/mol in bicelles despite their dramatically negative impacts on the kinetic stability. Double mutant cycle analysis indicates that the pair of acidic and basic residues are overall stabilizing with the interaction free energies ($\Delta\Delta G^{\circ}_{int}$) ranging from -3.6 ± 0.3 to -1.2 ± 0.2 kcal/mol in micelles and from -1.5 ± 0.3 to -0.5 ± 0.2 kcal/mol in bicelles. Thus, the ionizable residue pairs were less favorably engaged with each other in lipid environments. Substituting the acidic residues (Glu) with a polar residue (Gln) had a similar impact on stability, indicating that the ionizable residues exist as a neutral form and are engaged with each other by H-bonds rather than a salt-bridge.

In Chapter 3, we asked what is the physical origin of the unusual heat-resistance of stability and function of membrane proteins in thermophilic organisms. We addressed this question by comparing the thermostability and activity of thermophilic rhomboid proteases to those of mesophilic *E. coli* GlpG. In the phylogenetic tree of the rhomboid protease family (**Figure 1.11.**), rhomboids from some thermophilic bacteria, archaea, and eukarya exist together in the same branches, suggesting a horizontal gene transfer during the evolution of rhomboids. Many thermophilic rhomboids occupy shorter branches close to the root of the phylogenetic tree, implying that they may represent prototype rhomboids. Although previous studies indicate some differences in hydrophobicity and the frequency of interhelical H-bonds between mesophilic and thermophilic membrane proteins, we still need experimental data that can support this suggestion or provide a new physical principle of thermostabilization. To achieve the

goal, I employed three rhomboids from thermophilic archaea (*Thermococcus profundus* and *Pyrococcus furiosus*) and bacteria (*Thermotoga maritima*) as study models in addition to mesophilic *E.coli* GlpG. We successfully cloned the genes and expressed and purified them for further characterizations.

Using a thermal denaturation assay, I found that delipidated thermophilic rhomboids are inactivated at temperatures lower than the optimal living temperatures of their host organisms. Interestingly, those thermophilic rhomboids have lower thermostability than mesophilic *E. coli* GlpG. This result led us to the hypothesis that the thermophilic rhomboids are not stabilized by their intraprotein, residue-specific interactions but stabilized by their lipid environments. Using the proteolytic activity assay for TM and water-soluble substrates, I found that thermophilic rhomboids display various activities, which are higher or lower than mesophilic *E.coli* GlpG. Interestingly, a multiple sequence alignment indicates that the activity variation is highly correlated with the lengths of the flanking loops of the gating helix TM5. This result leads to the hypothesis that the activity level of rhomboids is determined by the lengths of those flanking loops, which affect the dynamics of the gating helix. Using *Ecoli* GlpG as a starting template, we tested this hypothesis by increasing and decreasing the lengths of the flanking loops. I provided some preliminary results that partially agree with our hypothesis.

CHAPTER 2: Role of buried ionizable residues in the stability of one α -helical intramembrane protease

2.1. Summary

Ionizable residues buried in proteins play vital roles in cellular functions including electron transfer, catalysis, and receptor activation. Therefore, studying the energetics of side-chain interactions of the ionizable residues in the protein interior would deepen our fundamental understanding of protein stability, provide insights into protein functions, and potentially guide protein engineering. While related studies have dominantly focused on water-soluble proteins, relevant analyses in membrane proteins are scarce even though membrane proteins constitute one-third of total proteins. The interior of membrane proteins is known to have a similar hydrophobicity to the hydrophobic core of water-soluble proteins. Unlike water-soluble proteins, the solvent-accessible cavities and surface of the TM region of a membrane protein contact nonpolar lipid tails instead of water. Thus, if ionizable residues are buried within the protein, their environment in the native and denatured states is expected to be highly hydrophobic. The limited number of studies on buried ionizable residues in membrane proteins has hindered the discovery of new biological mechanisms, leaving significant gaps in our understanding of their conformational and functional roles.

Here, I aim to bridge the gap with GlpG, a member of the universally conserved rhomboid protease family as a model. I address three major questions: 1) will buried ionizable residues be tolerated in the core of the membrane protein? 2) will the lipid environment favor or disfavor the burial of ionizable residues relative to detergent micelles? 3) Will ionizable residue pairs form charged interaction or just charge-neutral polar interaction in a membrane protein? To answer these questions, two pairs of acidic/basic residues (Glu/Lys) are installed into the two nonpolar cavities of GlpG, respectively. Thermodynamic stability of GlpG and the engineered variants are quantified using steric trapping, and spontaneous denaturation rates were measured using proteolysis. I find that the burial of ionizable residues induces destabilization of the protein kinetically and thermodynamically. However, double-mutant cycle analysis indicates favorable interactions between the Glu and Lys residues. Notably, the strength of the favorable interaction is much smaller than the expected strength of the salt-bridge interaction formed between the buried acidic and basic residues in water-soluble proteins. Further analysis suggests that the Lys/Glu residue pairs in GlpG are likely to

form hydrogen bonds in neutral forms instead of charged pairs. Our results provide fundamental understanding on energetic consequences of buried ionizable residue membrane protein.

2.2. Introduction

The internal ionizable residues play a variety of critical roles in protein functions including regulation of ATP hydrolysis²⁰⁷, proton transfer²⁰³, membrane transport²⁰⁵ and so on. However, harboring ionizable residues inside the hydrophobic core of protein is energetically unfavorable due to the desolvation cost, i.e., the free energy increase during the transfer of polar or charged residues from the aqueous phase to the nonpolar environment. In Wimley and White's experimentally determined hydrophobicity scale²⁰⁸, transferring acidic and basic residue in their charged forms from water to octanol or from water to water-membrane interfacial regions are highly unfavorable (i.e., 1.0-2.8 kcal/mol for positively charged lysine and 2.0-3.6 kcal/mol for negatively charged glutamate)²⁰⁸. When ionizable residues are in neutral forms, the partitioning becomes moderately to marginally unfavorable (i.e., 0-0.1 kcal/mol for neutral glutamate)²⁰⁸. When oppositely charged residues are simultaneously transferred from water to the hydrophobic medium, two energetic contributions compete with each other. One is the high desolvation cost for burying individual charged residues, which is highly unfavorable. The other is the strong coulombic attraction between two oppositely charged residues in the nonpolar medium with a low dielectric constant, which is highly favorable. The attractive electrostatic interaction is also called a salt bridge interaction, which is formed when the charge centers of two oppositely charged residue groups are within 4 Å¹⁰². The net transfer free energy of forming a charged pair in a nonpolar medium depends on the balance between these two opposing energetic contributions. With a model peptide having both acidic and basic residues in the same molecule, the free energy change from the unpaired charged Lys and Arg residues in water to the formation of a Lys-Arg salt bridge in octanol were experimentally determined to be -4 kcal/mol²⁰⁹. In the context of protein structures, the balance could be more complicated because the two other factors should be taken into account: the unfavorable side chains entropy loss due to the formation of a well-defined salt-bridge as well as interactions between the ion-pair with the rest of the protein.

Interestingly, the distributions of ionizable residues in membrane protein structures closely follow the trends of thermodynamic partition free energies between water and a hydrophobic medium. The Ulmschneider group plotted the distribution of each amino acid residue in the available structures of helical membrane proteins along the membrane depth²¹⁰. The plot shows that in the hydrocarbon core of the membrane, nonpolar residues are enriched and the acid/basic residues including Arg, Lys, Asp, Glu, and His are rare. The probabilities of the negatively charged residues, like Asp and Glu, are similar between the cytoplasmic and periplasmic membrane-water interfaces, while those of the positively charged residues, including Arg and Lys, are higher on the cytoplasmic side. The biased distribution of Arg and Lys residues is called the “positive-inside rule”^{211,212}, which is a critical determinant for the membrane topology of a transmembrane helices.

For water soluble proteins, the energetic contributions of the pair of buried acidic/basic residues have been evaluated with the model protein Arc repressor¹⁰⁴. The three ionizable residues (i.e., Arg31, Glu36, and Arg40) form two pairs of stabilizing salt bridge interactions (-1.7kcal/mol for Arg31-Glu36 and -4.7kcal/mol for Glu36-Arg40)¹⁰⁴. However, mutating them into large, hydrophobic residues such as Met, Tyr, Leu or Val further stabilizes the protein by 2.1 to 3.9 kcal/mol without disrupting the activity¹⁰⁴. These observations suggested that although buried ion-pairs can be favorable, they would not bring more stabilization than nonpolar residues with comparable sizes. Over the decades, the Garcia-Moreno group has extensively studied the impacts of internal ionizable residues on protein stability focusing on the model water-soluble protein, staphylococcal nuclease^{213–217}. Around 25 internal hydrophobic residues of staphylococcal nuclease have been engineered into residue with ionizable side chains: Glu, Lys, and Arg one at a time^{213,214,216}. The side chain pK_a 's of those residues measured by changes in unfolding free energy as a function of pH display a substantial shift from their intrinsic pK_a 's in water. The pK_a of Glu is 4.5 in water while increasing up to 9 in the protein interior. The pK_a of Lys is 10.5 in water but decreases down to 5 in the protein's hydrophobic core. The direction of the pK_a shift suggests that when a Glu or Lys residue is buried, a neutral state is preferred over the ionizable state. In contrast, the insensitivity of buried Arg over pH titration demonstrates that the guanidinium side

chain with pK_a 12 in water experience no shift, implying that a buried Arg is always at its charged state²¹⁴. Structural studies of the variants indicate that an ionizable group can be tolerated without causing significant structural consequences disrupting the native state^{213,214,216}. No loss in enzymatic activities of the most ionizable variants further supports that replacing internal hydrophobic residues with ionizable residues has no destructive impacts on the structure and function of SNase. The effects of replacing internal Leu and Val of Staphylococcal nuclease with Lys and Glu, respectively, have been extensively studied²¹⁵. The crystal structures of WT, single mutants, and double mutants demonstrate that the substituting ionizable residues can be tolerated without significantly disrupting the native structure²¹⁵. The polarity of the microenvironment around the ionizable residues increases as two water molecules penetrate and form H-bonding network with the ionizable residues. The stability of single and double ionizable variants indicate that harboring ionizable residues inside the protein destabilizes the protein in general. The measurements of side-chain pK_a and stability as a function of pH demonstrate that when buried together, Lys and Glu are more likely to exist as charged forms. Further double mutant cycle analysis indicates that although bearing the ionizable residues are destabilizing, the acidic/basic residue pair is engaged with a highly favorable interaction of -4.9 kcal/mol²¹⁷.

Studies show that the interaction between ionizable residues are important to function and stability of membrane proteins^{204,218}. For example, in the interior of OmpA, a small β -barrel membrane protein ion channel, four ionizable residues are clustered in the center of the narrow lumen and form strong to marginally (-5.6 to -0.6 kcal/mol) favorable salt bridge interactions. The switchable salt-bridges within the charge tetrad are known to control the open and close of the channel gate²¹⁸. For OmpA, three aromatic residues are surrounding the tetrad, further stabilizing it through π -cation or π -anion interactions. Moreover, the structure showed presence of water molecules at two vestibules of the ion channel partially contacting the tetrad²¹⁸. The water molecules increase local dielectric constant as well as further making the burial of the charges favorable.

The energetic contribution of ionizable residues in helical membrane protein remains unclear because relevant energetic analysis is scarce due to the low their

occurrence in helical membrane proteins^{210,219}. Buried ionizable residues in the helical membrane proteins are expected to be highly insulated from water not only by the hydrophobic residues in the protein interior, but also by the nonpolar tails of membrane lipids. Thus, it is less likely for water molecules to stabilize ionizable residues inside the proteins than for those in water-soluble proteins. Here I aim to bridge the knowledge gap by answering three key questions: 1) will buried ionizable residues be tolerated in the core of the membrane protein? 2) will the lipid environment favor or disfavor the burial of ionizable residues relative to detergent micelles? 3) Will ionizable residue pairs form charged interaction or just charge-neutral polar interaction in a membrane protein?

I chose *E.coli* GlpG, a member from the universal conserved rhomboid protease family as a study model. In the rhomboid protease family, *E.coli* GlpG possesses 6 TM helices and is well characterized with more than twenty solved high-resolution structures. I targeted and engineered four buried nonpolar residues into ionizable residues Lys and Glu to create two potential charged pair interactions located in two different regions within the protein. Employing the novel steric trapping strategy¹¹, we measured the thermodynamic stability in two hydrophobic environments that are widely used for membrane protein research: micelles, the spherical aggregates of detergent molecules; and bicelle, the disc-shaped lipid bilayer with edge stabilized by detergents. I find that these buried ionizable mutants largely destabilize GlpG yet a double-mutant cycle analysis suggests that marginally to moderately favorable interactions are formed between Lys and Glu. Overall, the lipid environment enhances the accommodation of ionizable residue interactions compared to detergent micelle. The similar energetic effects of ionizable Glu and its neutral proxy Gln suggest that when buried, ionizable residues prefer a neutral form engaged in polar interactions.

2.3. Materials and Methods

2.3.1. Mutagenesis, expression and purification of GlpG

Site-directed mutagenesis was performed on the gene encoding the TM domain (87 - 276) of GlpG with a N-terminal His₆ affinity tag in pET15b vector. The main mutations were: 1) introducing double-cysteine residues, P95C/G172C and G172C/V267C, respectively, for conjugating the thiol-reactive biotin derivatives; 2) replacing single or double buried nonpolar residues (Met100/Thr178 or Leu207/Val165) by ionizable or

polar residues Lys/Glu or Lys/Gln, respectively. The proteins were expressed in *E. coli* BL21(DE3) RP competent cells (Agilent Technologies) and was induced with 0.5 mM IPTG (Isopropyl β -D-1-thiogalactopyranoside) for 16-18 h at 15°C. Cells were harvested and resuspended in 30 mL/L-culture of 50 mM Tris-HCl (tris (hydroxymethyl)aminomethane Hydrochloride) buffer (pH 8.0) with 5 mM EDTA (Ethylenediaminetetraacetic acid), 1 mM DTT (Dithiothreitol) and 1 mM PMSF (phenylmethylsulfonyl fluoride). The resuspended cells were lysed with a pressure homogenizer (Avestin) for 4 times. Cell lysates were centrifuged for 20 min at 6,000 rpm, 4°C in a FS-34 rotor using a Sorvall RC6+ centrifuge (Thermo Scientific). Supernatant was collected and centrifuged to obtain the total membrane fraction for 2 h at 24,000 rpm, 4°C in the 45Ti rotor using an ultracentrifuge (Beckman-Coulter). Membrane pellets were resuspended in 20 mL/L-culture of 50 mM Tris-HCl buffer (pH 8.0) with 200 mM NaCl, 0.5 mM TCEP (Tris(2-carboxyethyl)phosphine) using a tissue homogenizer (Fisher Scientific). The membrane resuspension was solubilized by adding 0.8% (w/v) DDM (n-Dodecyl β -D-maltoside) and aggregates were removed by ultracentrifugation at 12,000 rpm for 20 min. The supernatant containing detergent-solubilized GlpG was incubated with 2 mL of Ni-NTA (Nickel Nitriloacetic acid) resin (Qiagen, 50% w/v) by rotating at 4°C for 1 h. GlpG was eluted with 50 mM Tris-HCl buffer (pH 8.0) and 200 mM NaCl in 0.1% DDM containing 300mM imidazole. The eluted fraction was concentrated and desalted with Amicon centrifugal filter unit (Millipore Sigma, 30 kDa MWCO) and desalting column (Bio-Rad Econo Pac 10DG desalting column) respectively. Concentration of GlpG was measured by UV absorbance at 280nm (the extinction coefficient, ϵ =69,440 M⁻¹cm⁻¹) with a Nanodrop (Thermo Scientific).

2.3.2. Labeling of GlpG and determination of labeling efficiency

The concentrations of the purified double cysteine variants of GlpG were adjusted to 25-50 μ M in 0.1% DDM, 50 mM Tris-HCl and 200 mM NaCl (pH 8.0). The protein solution was incubated with 2 mM TCEP for 1 h at room temperature, followed by the addition of 40-molar excesses of the thiol-reactive biotin derivative with fluorescent pyrene (BtpPyr-IA) dissolved in DMSO (~20% w/v) while vortexing. The labeling reaction was incubated by gently rotating in dark at room temperature for 16-18 h. Aggregation formed during

the overnight reaction was then removed by centrifugation at the 3,500 rpm for 10 min. Excessive free labels in the supernatant were removed by washing GlpG bound to Ni-NTA resin with 0.05% DDM, 50 mM Tris-HCl and 200 mM NaCl (pH 8.0) by ~50 residue volume and then GlpG was eluted with 50 mM Tris-HCl and 200 mM NaCl (pH 8.0), 0.1% DDM containing 300 mM imidazole. Residual free labels and imidazole were removed by desalting (BIO-RAD Econo Pac 10DG column).

Labeling efficiency of GlpG variants was determined by measuring the pyrene absorption at 346nm ($\epsilon = 42,000 \text{ M}^{-1}\text{cm}^{-1}$) and the protein concentration with a 660 nm colorimetric assay (Thermo-scientific). The pyrene-to-protein molar ratio ranged from 1.2 to 1.8 (ideally 2.0). To further confirm the labeling efficiency and the removal of free biotin labels, SDS-PAGE was set up as follows: 10 μL of 5 μM GlpG was incubated with the SDS sample buffer for 30 min to denature GlpG. 10 μL of 25 μM WT-mSA was added and incubated for another 30 min to bind the biotin-labels on GlpG. The SDS-PAGE was run at 100 V for 90 min on ice to prevent mSA tetramer from heat-induced dissociation. Labeling efficiency was calculated by quantifying the band intensities of single-mSA bound and double-mSA bound GlpG, which agree with the results obtained from UV absorption and 660 nm assays.

GlpG labeled with BtnPyr was injected into gel filtration (GE superose, 10/300 GL)-Fast Protein Liquid Chromatography (BioRad, Biologic Duo flow) for further purification and the checking of the oligomeric state. The column was equilibrated with 2 column volumes (60mL) of 50 mM Tris-HCl buffer (pH 8.0) and 200 mM NaCl containing 0.08 % DDM and 1 mM TCEP at the flow rate of 0.5 mL/min. The UV absorbance of elution peaks are detected at 280 nm. As molecular weight references, the FPLC was performed for 125 μL of 36 mg/mL gel filtration standard proteins on the same column.

2.3.3. Expression, purification and labeling of TM substrate SN-LacYTM2

As an indicator of folding, the activity of GlpG is probed with the site-specific cleavage of SN-LacYTM2, a membrane-bound substrate derived from the second TM domain of the lactose permease (LacYTM2, sequence: INCISKSDTGIIFFAISLFSLLFQPLFGLLS with the scissile bond denoted as “-” between S and D) of *E. coli* fused to staphylococcal nuclease (SN). The gene of the fusion protein has a C-terminal His₆ tag and is encoded in the pET30a vector with the SN domain in the N-terminal region and LacYTM2 in the

C-terminal region. A TEV cleavage site is engineered in the linker between SN and LacYTM2 (SN-TEVcut-LacYTM2-His6). In LacYTM2, the residue five residues upstream of the scissile bond (P5 position) was mutated to cysteine to conjugate with the thiol-reactive, environmentally sensitive fluorophore, iodoacetyl-7-nitrobenz-2-oxa-1,3-diazol (IA-NBD amide, Setareh Biotech). The fusion protein was expressed in the *E. coli* BL21(DE3) RP strain. Detailed procedures for expression, purification, and fluorescent labeling of SN-LacYTM2 are described in the previous literature¹¹⁵.

2.3.4. Preparation of bicelles

A stock of 25% (w/v) DMPC (1,2-dimyristoyl-*sn*-glycero-3-phosphocholine)/CHAPS (3-[(3-Cholamidopropyl) dimethylammonio]-1-propanesulfonate) (the lipid-to-detergent molar ratio, $q = 1.5$) bicelles were prepared by first hydrating DMPC lipids with water and then adding 25% (w/v) CHAPS was to the desired q value. Bicelle samples were homogenized through three cycles of freeze-thaw using liquid nitrogen and a 42 °C water bath. Bicelle stock solutions with 0.05% NaN₃ were kept at -20 °C for long term use.

2.3.5. Determination of GlpG activity

GlpG activity was probed with NBD-labeled SN-LacYTM2 as a substrate¹¹⁵ at the GlpG-to-substrate molar ratio of 1:10. Upon cleavage, the environmentally sensitive fluorophore NBD is transferred from the hydrophobic to the aqueous environment leading to a decrease of fluorescence. Time-dependent changes of NBD fluorescence were monitored at 37 °C in 96-well plate using a SpectraMax M5e plate reader with the excitation and emission wavelengths of 485 nm and 535 nm, respectively. The activity was represented by the initial slope of fluorescence change over time.

Another way to measure GlpG activity was with the peptide substrate²²⁰ (MMPS-024, CPC scientific) as a model substrate in molar ratio of 1:20 (GlpG: substrate). This 9-residue peptide (mca-RPKPYAv/WM-K(dnp), where lowercase v stands for norvaline, a non-natural amino acid; “/” represents the position of scissile peptide bond) has been proven to be proteolyzed by GlpG.²²¹ Here, the fluorophore 7-methoxycoumarin (mca) is on the amino terminus and is internally quenched by dinitrophenol (dnp) conjugated to the Lys at the carboxyl-terminus. Upon cleavage of the peptide, mca fluorophore is dequenched leading to an increase of fluorescence. Time-dependent changes of mca

fluorescence were monitored at 37 °C on a SpectraMax M5e plate reader with the excitation and emission wavelengths of 320 nm and 430 nm, respectively. GlpG activity was represented by the initial slope of fluorescence change over time.

2.3.6. Determination of GlpG intrinsic denaturation rate with Proteinase K (ProK) digestion assay in micelle and bicelle

The 5 μ M GlpG (172_M267_C-BtnPyr) in 5 mM DDM micelle or incorporated in 2% (w/v) DMPC/CHAPS bicelles in 20 mM HEPES (pH 7.5) and 200 mM NaCl was incubated on ice for 15 min. 3 mM CaCl₂ is added as a stabilizer of for Proteinase K (ProK, PCR recombinant grade, Roche). The proteolysis reaction was initiated by adding 200 μ g/mL of ProK (the final concentration) and incubated for different time lengths. 15 μ L aliquot of each sample was taken at each time point and thoroughly mixed with 5 mM PMSF followed by further incubation for 10 mins. Proteolysis was visualized by SDS-PAGE (4 to 20% gradient gels, Bio-Rad). The remaining fraction of GlpG after ProK digestion was quantified by measuring the band intensities of GlpG using ImageJ²²². The band intensity at each time point was normalized to that of the control (GlpG without ProK) fitted to the first-order exponential decay function to obtain the intrinsic denaturation rate.

$$y = y_0 + Ae^{-k_{\text{denat}}t} \text{ (Eq1)}$$

$$\tau_D = \frac{1}{k_{\text{denat}}} \text{ (Eq2)}$$

$$\tau_{1/2} = \frac{\ln 2}{k_{\text{denat}}} \text{ (Eq3)}$$

y : the remaining GlpG fraction after ProK digestion; y_0 : the final GlpG fraction; A : the amplitude of the total intensity change; k_{denat} : the intrinsic denaturation rate; τ_D : the time constant of intrinsic denaturation; $\tau_{1/2}$: the half-life of GlpG; t : the time.

2.3.7. Preparation and labeling of monovalent streptavidin (mSA)

Active and inactive forms of streptavidin encoded in pET15b were transformed into *E. coli* BL21(DE3) RP strain (Agilent) and overexpressed in terrific broth (TB) media with 0.5 mM IPTG for 16-18 h at 37°C. The “active” subunits include wild-type streptavidin and other lower biotin binding affinity variants (S27R, N23A/S45A, W79M, S45A, S27A, and E51S) with a C-terminal His₆ tag. The “inactive” subunit refers to the triple mutant, N23A/S27D/S45A (low biotin affinity, $K_{d,\text{biotin}} = 1.2 \times 10^{-3}$ M) streptavidin without C-

terminal His₆ tag²²³. Harvested cells were resuspended in 30 mL/liter-culture of 50 mM Tris-HCl (pH 8.0), 0.75 M sucrose buffer with 1 mg/mL hen egg lysozyme. The resuspended cells were lysed 5 to 7 times using a pressure homogenizer (Avestin). Inclusion bodies were collected by centrifuging the cell lysates at 12,000 rpm for 15 min at 4 °C. The pellets were washed by 40 mL of 50 mM Tris-HCl (pH 8.0), 1.5 M NaCl, 0.5% Triton X-100 (Sigma) buffer with a tissue homogenizer, and centrifuged at 12,000 rpm for 15 min at 4 °C. The washing of pellets with Triton X-100 was repeated ~3 times. After the detergent washing, the pellets were washed with 35 mL of 50 mM Tris-HCl (pH 8.0), 1.5 M NaCl for one time. The pellets were 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. The optical density of the supernatant was measured with UV absorbance at 280nm by a Nanodrop (Thermo Scientific).

To refold streptavidin, the active and inactive subunits of streptavidin in GdnHCl were mixed at a molar ratio of 1:4 and added dropwise to the final GdnHCl concentration less than 0.3 M in 20 mM sodium phosphate buffer (pH 7.5), 200 mM NaCl, and 0.5 mM TCEP containing 15% glycerol while vortexing vigorously on ice. Aggregates were removed by centrifuging at 6,000 rpm for 30 min at 4 °C. The clear supernatant was incubated with Ni-NTA resin (~1 ml resin per 100 ml buffer, Qiagen) 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). The eluted fractions were dialyzed (10 kD MWCO, Thermo Scientific) against 0.5 mM TCEP, 20 mM sodium phosphate, 200 mM NaCl (pH 7.5) buffer to remove imidazole for next purification. A second affinity chromatography in Ni-NTA resin was performed with the same washing and eluting steps. To remove imidazole, the eluted mSA fractions were desalted and concentrated with the desalting column (BIO-RAD Econo Pac 10DG) and the Amicon centrifugal filtration units (Millipore Sigma, 30 kDa MWCO) respectively. The concentration of mSA was measured by UV absorbance at 280nm ($\epsilon = 167,760 \text{ M}^{-1}\text{cm}^{-1}$) with a Nanodrop (Thermo Scientific).

To label mSA with a thiol-reactive quencher, Tyr83 near the biotin-binding pocket of the active subunit was mutated to Cys (Y83C). 50 μM of mSA was incubated with 5

times molar excess of TCEP for 1 h at room temperature. A 20-time molar excess of dabcyI-maleimide (AnaSpec) stock in water (1% w/v) was added dropwise while vortexing and was incubated overnight in dark at 4°C. Excess free labels were removed three times on a desalting column (Bio-Rad Econo Pac 10DG) equilibrated with 20 mM Na₂HPO₄ (pH 7.5), 200 mM NaCl. The concentration of mSA_{DAB} was measured with UV absorbance at 484nm ($\epsilon = 3,200 \text{ M}^{-1}\text{cm}^{-1}$) by a Nanodrop (Thermo Scientific).

2.3.8. Determination of the biotin binding affinities of mSA variants in micelle and bicelle

The biotin binding affinities of mSA variants were determined by a FRET based assay with single-biotinylated GlpG (GlpG-BtnPyr), which contains one Cys residue (P95C)^{115,147}. Next, 200 nM of GlpG-BtnPyr solubilized in 5 mM DDM or 2% (w/v) DMPC/CHAPS bicelles ($q = 1.5$) was titrated by mSA variants with a lower biotin affinity: mSA_{DAB}-S27R, mSA_{DAB}-N23A/S45A. The titrated samples were incubated overnight and transferred to a fluorescence cuvette (Hellma, ultra-micro). Pyrene fluorescence spectra were acquired from 340nm to 500nm with the excitation wavelength at 345 nm on a SpectroMax M5e spectrometer (Molecular Devices). For each sample, pyrene fluorescence between 375nm and 405nm is averaged. Excessive biotin was added to the final concentration of 2 mM and incubated for 4h to dissociate bound mSA from biotinylated GlpG and pyrene fluorescence were acquired. This data with biotin was used as a baseline to be subtracted from the data without biotin. The baseline-subtracted data were fitted to the following equation to obtain $K_{d, \text{biotin}}$ of mSA_{DAB}-S27R and mSA_{DAB}-N23A/S45A in micelle and bicelles¹¹⁵.

$$F = A1 \times \frac{(P_T + [mSA] + K_{d, \text{biotin}}) - \sqrt{(P_T + [mSA] + K_{d, \text{biotin}})^2 - 4P_T[mSA]}}{2P_T} + A2 \quad (\text{Eq4})$$

F : the measured fluorescence intensity; P_T : the total GlpG concentration; $[mSA]$: the total mSA concentration; $K_{d, \text{biotin}}$: the dissociation constant for the biotin-mSA_{DAB} complex; $A1$: the net change in fluorescence; $A2$: the fluorescence level without mSA_{DAB}. Fitted values included $A1$, $A2$, and $K_{d, \text{biotin}}$ while the other values were fixed.

2.3.9. Measuring GlpG thermodynamic stability with binding isotherm in micelle and bicelle

The 1 μM of GlpG doubly labeled with BtnPyr was titrated by thiol-reactive dabcyI (AnaSpec) conjugated mSA (mSA_{DAB}) at various concentrations in 20 mM sodium phosphate (pH 7.5), 200 mM NaCl, 0.25 mM TCEP buffer containing 5 mM DDM. For measurements in bicelles, 1 μM of the same doubly biotinylated GlpG was reconstituted in 3% DMPC/CHAPS ($q = 1.5$) in 20 mM HEPES (pH 7.5), 200 mM NaCl, 1 mM DTT buffer by direct injection. The mSA variants with a range of biotin affinity (mSA_{DAB}-S27R, N23A/S45A, W79M, S45A, S27A, and E51S) were tested until a second mSA binding phase optimally attenuated in a mSA concentration window of 0 to 60 μM was found. The reaction mixtures were transferred to a 96-well microplate (Greiner), sealed with polyolefin film (VWR), and incubated at room temperature for 24-48h for equilibration.

As the indicator of binding, quenching of pyrene fluorescence was measured at 390 nm with the excitation wavelength of 345 nm on a SpectroMax M5e plate reader (Molecular Devices). Data were averaged from three readings. To obtain the thermodynamic stability, the attenuated second binding phase was fitted with a 2-state scheme (i.e., the equilibrium between the native and denatured states) to the equation¹¹⁵:

$$K_{d, \text{biotin}} = \frac{[\text{U} \cdot \text{mSA}][\text{mSA}]}{[\text{U} \cdot 2\text{mSA}]}$$

$$K_u = \frac{[\text{U} \cdot \text{mSA}]}{[\text{F} \cdot \text{mSA}]}$$

$$F = \frac{1}{\left[1 + \left(K_{d, \text{biotin}} + \frac{K_{d, \text{biotin}}}{K_u} \right) \frac{1}{[\text{mSA}]} \right]} (F_{\infty} - F_0) + F_0 \quad (\text{Eq5})$$

F : fluorescence intensity; F_0 : the fluorescence intensities of pyrene from the BtnPyr labels on GlpG with no mSA binding; F_{∞} : the fluorescence intensities at the saturated bound level; $[\text{mSA}]$: the total mSA concentration; $K_{d, \text{biotin}}$: the intrinsic biotin affinity of mSA; K_u : the equilibrium constant for protein denaturation. After obtaining the fitted K_u , the thermodynamic stability was calculated using the equation, $\Delta G^{\circ}_{\text{N-D}} = -RT \ln K_u$.

2.3.10. Double mutant cycle analysis

To measure the pairwise interaction energies of the engineered residues, double-mutant cycle analysis was employed²²⁴. For this analysis, the wild-type protein (WT), two single mutants, and the corresponding double mutants and the free energy changes upon mutation are used:

$$\begin{aligned}\Delta\Delta G^{\circ}_{\text{int}} &= \Delta\Delta G^{\circ}_{XY'-X'Y'} - \Delta\Delta G^{\circ}_{XY-X'Y} = \Delta\Delta G^{\circ}_{X'Y-X'Y'} - \Delta\Delta G^{\circ}_{XY-XY'} \\ &= \Delta G^{\circ}_{X'Y} - \Delta G^{\circ}_{XY} - \Delta G^{\circ}_{X'Y'} + \Delta G^{\circ}_{XY'} = \Delta G^{\circ}_{XY'} - \Delta G^{\circ}_{XY} - \Delta G^{\circ}_{X'Y'} + \Delta G^{\circ}_{X'Y}\end{aligned}\quad (\text{Eq 6})$$

where X and Y are WT residues and X' and Y' are residues substituting X and Y, respectively. If the change in stability of the double mutation ($\Delta\Delta G^{\circ}_{XY-X'Y'}$) is different from the sum of the changes brought by the single mutations ($\Delta\Delta G^{\circ}_{XY-XY'} + \Delta\Delta G^{\circ}_{XY-X'Y}$), the two residues in WT are coupled and the magnitude of the difference represents the strength of interaction between them (the interaction free energy, $\Delta\Delta G^{\circ}_{\text{int}}$).

2.3.11. Cooperativity profiling

This method measures the degree of spatial propagation of the structural perturbation induced by a point mutation throughout the protein. This is quantified by measuring the mutation-induced stability changes at two different biotin pairs located in distinct regions of a protein. First, perturbation is induced by introducing a single point mutation in the background of the biotin pair variants 95_N172_M-BtnPyr₂ ($\Delta G^{\circ}_{\text{N-D,WT}^{\text{N}}}$) or 172_M267_C-BtnPyr₂ ($\Delta G^{\circ}_{\text{N-D,WT}^{\text{C}}}$), which are set as 'WT'. The superscripts "N" and "C" denote the location of the biotin pair, i.e., the N-terminal subdomain (N-subdomain) and C-terminal subdomain (C-subdomain). Then, the stability change induced by the same mutation is measured by steric trapping with two WT backgrounds ($\Delta\Delta G^{\circ}_{\text{N-D,WT-Mut}^{\text{N}}} = \Delta G^{\circ}_{\text{N-D,WT}^{\text{N}}} - \Delta G^{\circ}_{\text{N-D,Mut}^{\text{N}}}$ or $\Delta\Delta G^{\circ}_{\text{N-D,WT-Mut}^{\text{C}}} = \Delta G^{\circ}_{\text{N-D,WT}^{\text{C}}} - \Delta G^{\circ}_{\text{N-D,Mut}^{\text{C}}}$). Then, the differential effect of the mutation on the stability of the two subdomains is quantified as follows:

$$\begin{aligned}\Delta\Delta\Delta G^{\circ} &= \Delta\Delta G^{\circ}_{\text{N-D,WT-Mut}^{\text{N}}} - \Delta\Delta G^{\circ}_{\text{N-D,WT-Mut}^{\text{C}}} \\ &= [\Delta G^{\circ}_{\text{N-D,WT}^{\text{N}}} - \Delta G^{\circ}_{\text{N-D,Mut}^{\text{N}}}] - [\Delta G^{\circ}_{\text{N-D,WT}^{\text{C}}} - \Delta G^{\circ}_{\text{N-D,Mut}^{\text{C}}}] \quad (\text{Eq7})\end{aligned}$$

Four cut-off values, $\Delta\Delta\Delta G^{\circ} = -2RT$, $-RT$, RT and $2RT$ (R : gas constant and T : absolute temperature) were used to resolve the degree of cooperativity of each residue interaction. For a given $\Delta\Delta\Delta G^{\circ}$ value, a cooperativity profile is assigned for a given mutated site as following: $\Delta\Delta\Delta G^{\circ} > +2RT$: highly localized in N-subdomain; $+RT < \Delta\Delta\Delta G^{\circ}$

$\leq +2RT$: moderately localized in N-subdomain; $-RT \leq \Delta\Delta\Delta G^\circ \leq +RT$: cooperative; $-2RT \leq \Delta\Delta\Delta G^\circ \leq -RT$: moderately localized in C-subdomain; $\Delta\Delta\Delta G^\circ < -2RT$: highly localized in C-subdomain.

2.4. Results

2.4.1. Introducing ionizable residues in the hydrophobic interior of GlpG

I first identified proper nonpolar residue pairs that will be replaced by ionizable residues based on the two criteria: 1) the WT residues are buried within the protein interior with a low fraction of solvent accessible surface area (fASA); 2) the sizes of WT nonpolar residues that are to be replaced have similar side-chain volumes to the ionizable Lys-Glu pair; 3) the WT nonpolar residues are contacting internal cavities such that there is room for accommodating possible size mismatching between the WT and substituted residues. The fASA's of residues in the GlpG structure were evaluated on the GetArea server (<http://www.scsb.utmb.edu/getarea/>) using the crystal structure of GlpG (PDB id: 3B45) and the probe with a radius of 1.4 Å. Lys has an average side-chain volume of 171 Å³, similar to the nonpolar residues Met, Leu, and Ile (the side-chain volumes: 171 Å³, 168 Å³, 168 Å³, respectively)²²⁵. For the same practice, Val (142 Å³) has a similar volume to Glu (155 Å³)²²⁵.

Based on these criteria, we selected ten residues in GlpG as potential substitution sites: Val96, Met100, Leu161, Val165, Leu174, Ile175, Thr178, Leu200, Val204 and Leu207. Six of these sites (Leu161, Val165, Leu174, Leu200, Val204, and Leu207) were completely buried (i.e., fASA = 0) while the other four were partially exposed ($0 < \text{fASA} < 0.2$). To designate optimal nonpolar residue pairs, the WT GlpG structure was used as a template assuming that the amino acid substitutions maintain the overall fold of the protein. We identified five initial mutation candidates, V165E/L174K, L161K/V204E, V165E/L207K, V96E/I175K, and M100E/L200K. However, the screening for the expression and purification of these variants indicated that a reasonable yield was achieved only the variant L207K/V165E (0.1 mg/L-culture) for further characterization (**Figure 2.1.a**). Although Thr (122 Å³) to Glu mutation is not ideal under the criteria of the nonpolar-ionizable residue substitution and the minimal change in side-chain volume, the M100K/T178E substitutions satisfied the criterion of cavity contact. The yield was satisfactory for this variant (0.2 mg/L-culture) (**Figure**

2.1.b). Taken together, I chose the variants L207K/V165E (hereafter, denoted as L207K/V165E) and M100K/T178E (hereafter, denoted as M100K/T178E) for subsequent tests for kinetic and thermodynamic stability.

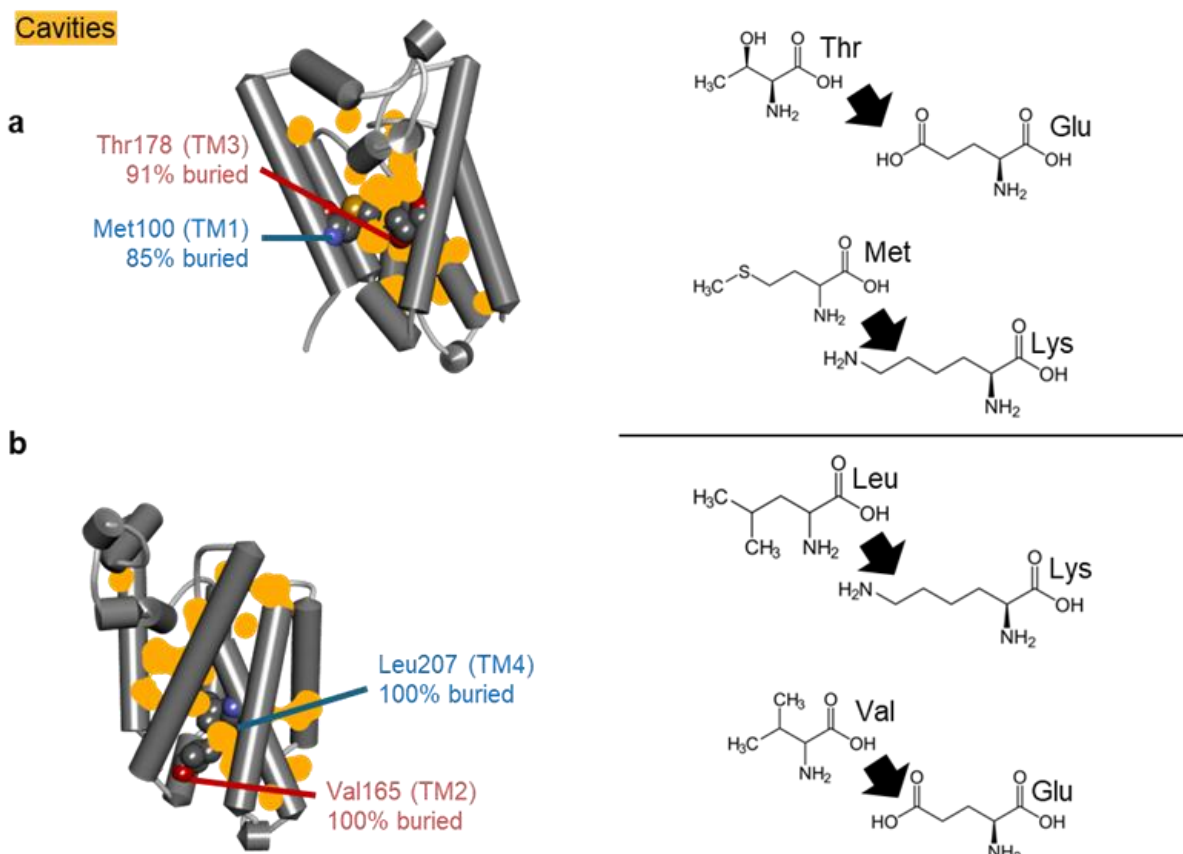


Figure 2.1. Structural representation of double mutants demonstrating solvent accessibility, cavities, and designed sidechain shape changes. (a) The double mutant M100K/T178E. **(b)** The double mutant L207K/V165E. The internal cavities were identified on the DEPTH server²²⁶. Yellow meshed spheres correspond to the probe water molecules (radius of 1.4 Å) filling the cavities.

2.4.2. Burial of the ionizable residues substantially decreases the kinetic stability of GlpG in DDM micelles

To investigate the impact of buried ionizable residues on unfolding kinetics of GlpG, we employed proteolysis, a technique that distinguishes folded and denature state of protein and quantify protein stability^{227,228} (**Figure 2.2 a**). The denaturation rate of RNaseH and maltose binding protein determined by proteolysis agree well with the kinetic parameters of denaturation determined by circular dichroism^{228,229}. In this study,

we utilized the nonspecific protease Proteinase K (ProK) with high digestion rate ($k_{\text{cat}} = \sim 10^4/\text{min}$) and low sequence specificity (typically nonpolar residues)²³⁰. If ProK exists at enough concentration, the proteolysis rate of GlpG is much higher than the refolding rate of GlpG ($k_{\text{refolding}} = 10/\text{min}$)²³¹. Under the assumption that GlpG is at dynamic equilibrium between the native and denatured states, ProK selectively proteolyzes water-exposed coil regions as soon as they appear due to denaturation. Because we added GlpG at an excess concentration of ProK (>10 relative to GlpG) in this assay, the reaction irreversibly moved towards the direction of proteolysis of the denatured state (**Figure 2.2.a**). Analogous to the EX1 condition in hydrogen/deuterium (H/D) exchange where the apparent rate of H/D exchange is limited by the “open” rate (k_{open}) of the protein²³², the rate-limiting step of ProK digestion is the spontaneous denaturation of GlpG ($k_{\text{denat}} \sim k_{\text{proteol}}$). I monitored the time-dependent proteolysis using SDS-PAGE to determine the k_{denat} . The amount of intact GlpG from ProK treatment (i.e., the native state) was determined by measuring the corresponding band intensities on the SDS-PAGE gel.

As shown in **Figure 2.2.b**, WT GlpG remained intact in the presence of ProK for 48-72 hours (the half-life, $t_{1/2} = \sim 90$ hr) in detergent micelles, implying a high kinetic stability of the native state. In stark contrast, the single mutants (M100K, T178E, L207K, and V165E) and double mutants (M100K/T178E and L207K/V165E) exhibited a substantially high susceptibility to ProK digestion being digested within 2 min. Thus, the half-lives of GlpG with the ionizable residue pairs decreased by 2,700-to-18,000-fold compared to WT. Among all mutants, the most kinetically unstable variant was the single mutant, L207K (**Figure 2.2.f**). Interestingly, the single mutant M100K had the highest kinetic stability among all variants ($t_{1/2} = \sim 2$ hr) (**Figure 2.2.c** and **Table 2.1**). These results indicate that introducing an ionizable residue in the protein interior dramatically induces a dramatic reduction of kinetic stability of GlpG compared to WT.

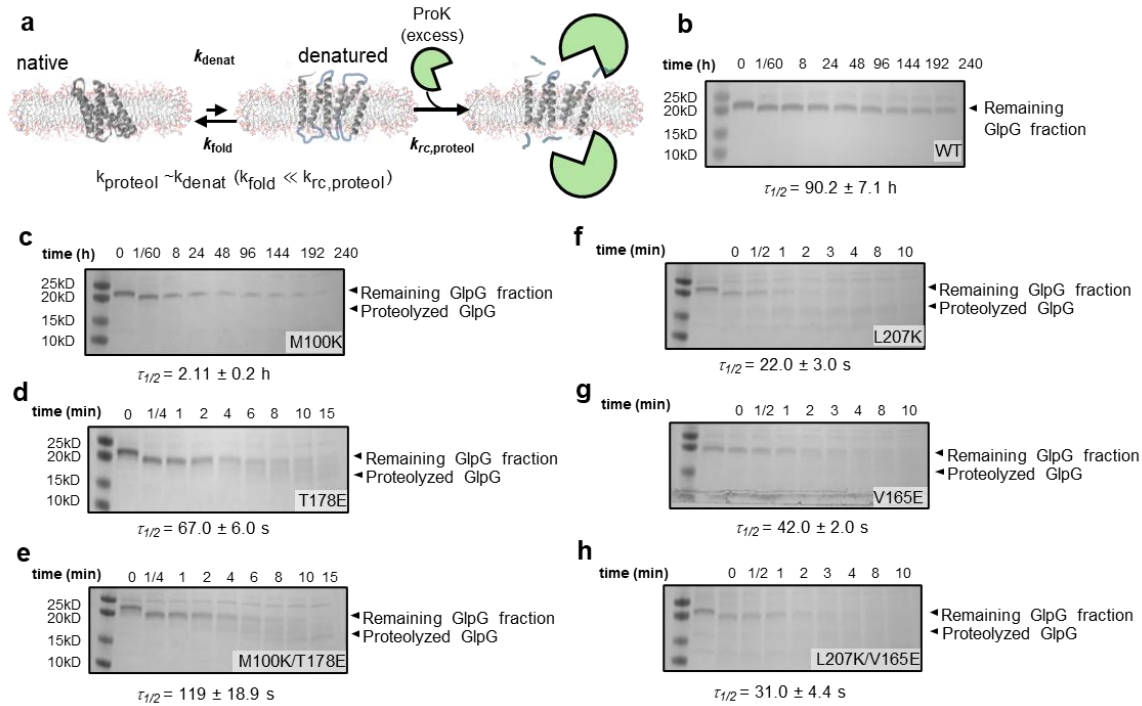


Figure 2.2. Determination of the spontaneous rate k_{denat} of GlpG with a ProK

digestion assay. (a) The reaction scheme of the ProK digestion assay. GlpG was treated with an excess concentration of ProK. k_{fold} : spontaneous folding rate; $k_{\text{rc,proteol}}$: the rate constant of proteolysis for random coil peptides; k_{proteol} : the apparent rate constant of proteolysis measured by SDS-PAGE. When the random coil is proteolyzed much faster than folding, the denaturation rate is approximated to the apparent proteolysis rate. (b) SDS-PAGE result of the ProK digestion assay for WT in DDM. The spontaneous denaturation rate (k_{denat}) was obtained by fitting the time-dependent changes of the remaining GlpG fractions with the model of first order reaction $y = y_0 + Ae^{-k_{\text{denat}} \cdot t}$. (c)-(h) SDS-PAGE gels of the ProK digestion assays in DDM for M100K, T178E, M100K/T178E, L207K, V165E, and L207K/V165E respectively.

Table 2.1. The spontaneous denaturation rates, half-lives, denaturation time constant of WT and the ionizable single, double mutants in DDM micelles.

	k_{denat} in micelle (sec^{-1})	$\tau_{1/2}$ in micelle (sec)	τ_D in micelle (sec)
WT	$2.1 \pm 0.2 \times 10^{-6}$	$3.3 \pm 0.3 \times 10^5$	$3.7 \pm 0.4 \times 10^5$
M100K	$9.1 \pm 0.9 \times 10^{-5}$	$7.6 \pm 0.8 \times 10^3$	$1.1 \pm 0.1 \times 10^4$
T178E	$1.0 \pm 0.1 \times 10^{-2}$	67.2 ± 5.7	97.0 ± 8.3
M100K/T178E	$5.8 \pm 0.9 \times 10^{-5}$	119 ± 19	172 ± 27
L207K	$3.2 \pm 0.5 \times 10^{-2}$	21.7 ± 3.3	31.3 ± 4.8
V165E	$1.7 \pm 0.1 \times 10^{-2}$	41.9 ± 2.0	60.4 ± 2.9
L207K/V165E	$2.2 \pm 0.3 \times 10^{-2}$	31.0 ± 4.4	44.7 ± 6.4

2.4.3. Oligomeric states of GlpG with ionizable residues in the protein interior

Kinetic stability results indicate that burying ionizable residue inside GlpG significantly compromises the half-lives of the folded state. Thus, it is possible that during denaturation ($t_{1/2} = \sim 2$ min), the buried ionizable residues are exposed and mediate oligomerization through intermolecular polar interactions. Also, it is possible that oligomerization in the denatured states contributes to GlpG stability.

To further test this stability, I employed size exclusion chromatography (SEC) to resolve the oligomeric states of GlpG variants. The underlying principle of SEC is that molecules with smaller size (i.e., a lower molecular weight, condensed structure) can dive into the pores on stationary phase which results in a longer path than the molecules with larger size (i.e. a higher molecular weight, expanded conformation). Therefore, larger molecules migrate early due to relatively shorter paths. Coupling with UV detectors and absorption spectrum at 280 nm, distribution of protein sample with different sizes can be differentiated.

In **Figure 2.3.a**, the superposition of the chromatogram of WT GlpG with that of the standard proteins indicates that WT GlpG at 20 μM exists as monomer (see the legends). For all the variants with buried ionizable residue mutants (**Figure 2.3.b-f**), the chromatograms indicate that at relatively higher injected protein concentrations (20-40 μM), the ionizable variants exist as a monomer-oligomer mixture while oligomeric forms are preferred. After the first separation, the monomer fractions were collected,

concentrated, incubated for 2-3 hours at room temperature, and injected for the second gel filtration. The superimposed chromatograms show that the isolated monomeric GlpG fractions remain as monomer and slowly exchange with oligomer. Interestingly, for the double mutant L207K/V165E, oligomer and monomer coexist even at a low concentration while still preferring the monomeric state. Overall, our gel filtration results demonstrate that oligomeric states of GlpG variants are concentration dependent. In other words, at lower concentrations the equilibrium of oligomerization shifts to a monomeric state. Therefore, under my steric trapping conditions using a low GlpG concentration (1 μ M GlpG), it is expected that GlpG largely exists as monomer.

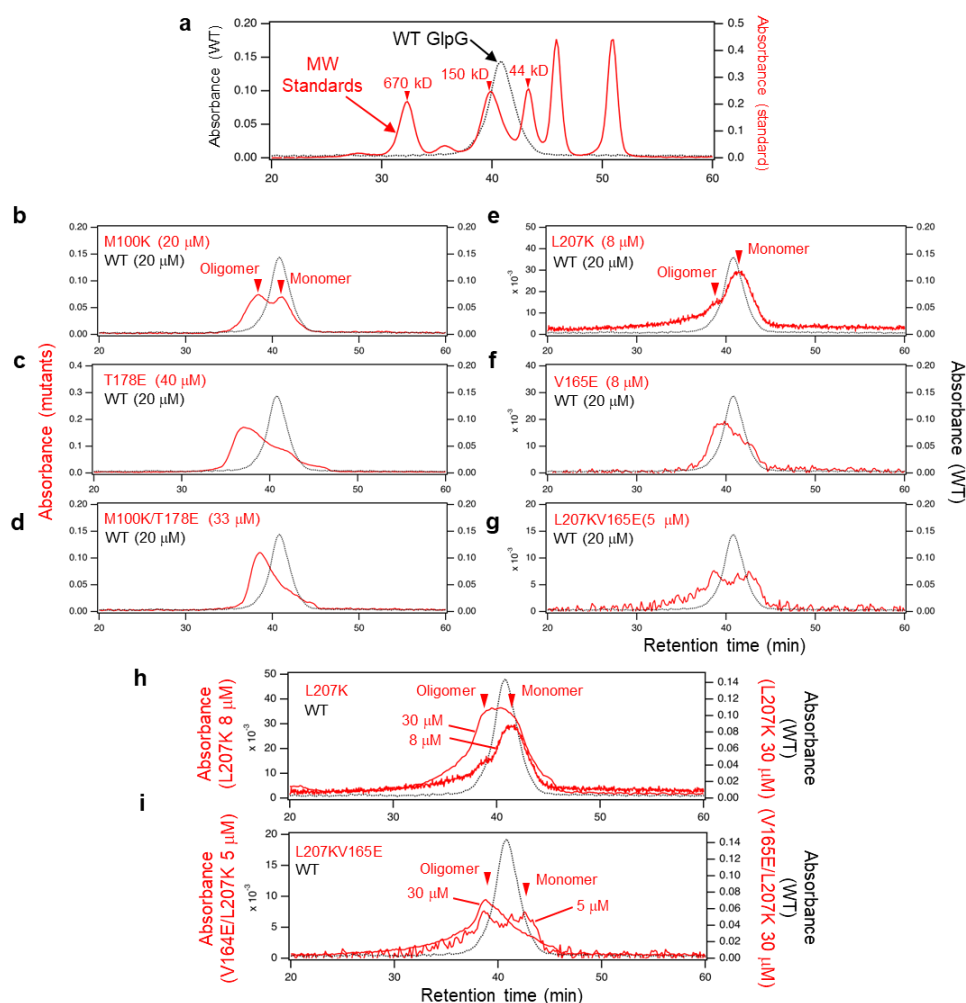


Figure 2.3. Determination of GlpG oligomeric state with size exclusion

chromatography. (a) Gel-filtration chromatograms of WT GlpG (black dotted line) superimposed with protein standards (red solid line). The molecular weights of standard peaks were labeled. The aggregation number per DDM micelle is ~150 and the molecular weight of a DDM is 0.5106 kD. Thus, the predicted molecular weight of a DDM micelle- GlpG monomer complex is 99.6 kD ($150 \times 0.5106 \text{ kD} + 23.4 \text{ kDa}$), which agrees with the observed peak. (b)-(g) Chromatograms of WT GlpG (black dotted line) superimposed with GlpG possessing single and double buried ionizable residues at various concentrations (red solid line) (h)-(i) Chromatograms of WT GlpG (black dotted line) superimposed with L207K and L207K/V165E (red solid lines), respectively. The chromatogram for the lower injected-sample concentration represents the monomer fractions isolated from the oligomer-monomer mixture obtained with the higher injected-sample concentration.

2.4.4. Impact of buried ionizable residues on the thermodynamic stability of GlpG in DDM micelles

Next, we measured the thermodynamic stabilities (ΔG°_{N-D}) of WT and variants using steric trapping to quantify the degree of destabilization induced by the burial of ionizable residues. Steric trapping is a method to capture the spontaneously denatured state of a doubly biotinylated protein through the simultaneous binding of two bulky monovalent streptavidin (mSA) molecules (**Figure 2.4.a**). Compared with conventional stability measurements using chemical denaturants (i.e., GdnHCl or urea), this method is advantageous because protein stability can be directly measured in the native solvent and lipid environment.

Previously, two pairs of optimal double-biotinylation sites on the TM domain of GlpG had been identified, which did not affect the activity and stability of and engineered into Cys (P95C/G172C and G172C/V267C) to probe GlpG stability at N-terminal (95_N172_M) and C-terminal (172_M267_C) subdomains (the subscripts N, M, and C denote the locations of biotinylated sites in the N-terminal helix TM1, Middle helix TM3, and the C-terminal helix TM6)¹¹⁵. The double cysteine variants, P95C/G172C and G172C/V267C, are labeled with the thiol-reactive biotin derivative carrying a pyrene fluorophore (BtnPyr). To measure ΔG°_{N-D} of GlpG, a binding isotherm is generated by titrating the doubly-biotinylated GlpG with an increasing concentration of mSA labeled with the dabcyI quencher (mSA_{DAB})¹¹⁵. That is, the quenching of pyrene fluorescence from the BtnPyr label on GlpG. In the binding isotherm, there are two clearly distinguishable binding phases (**Figure 2.4.b**). The first tight binding phase represents the intrinsic binding of the first mSA to either of the BtnPyr labels on GlpG. The second attenuated binding phase is coupled to GlpG denaturation. By fitting the second binding phase to Eq5, ΔG°_{N-D} of GlpG can be obtained. In DDM micelles (**Figure 2.4.b, Figure 2.5.**), the ΔG°_{N-D} of WT GlpG were 5.9 ± 0.1 kcal/mol with the 95_N172_M biotin at N-subdomain and 4.5 ± 0.1 kcal/mol with the 172_M267_C biotin pair at for C-subdomain. The values agreed with the previous determined results^{115,149}.

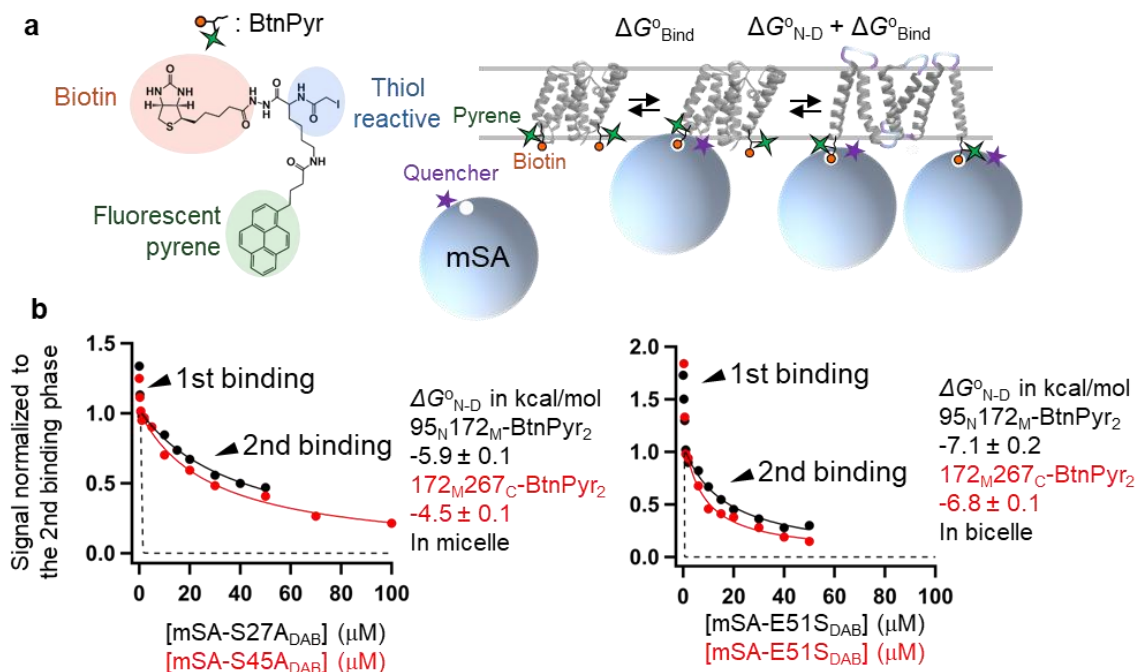


Figure 2.4. Scheme of steric trapping for measuring the thermodynamic stability of WT GlpG. (a) The principles of steric trapping. GlpG is doubly labeled with biotin tags at two specific residues which are spatially close in the native state but distant in the amino acid sequence. The biotin tag (BtPyr) is composed of three parts: a thiol-reactive group, biotin, and fluorescent pyrene. The first monovalent streptavidin (mSA) binds to either biotin label (ΔG°_{Bind}) without steric hindrance. The second mSA binds to the other biotin label only when the tertiary structure of GlpG is transiently unraveled (ΔG°_{N-D}) because of the steric crash between bulky mSA molecules. The coupling of mSA binding to denaturation attenuates the apparent binding affinity of the second mSA ($\Delta G^{\circ}_{Bind} + \Delta G^{\circ}_{N-D}$). (b) Binding isotherms between the double-biotin variants of GlpG and mSA in micelles (left) and bicelles (right). GlpG were labeled with BtPyr-IA at N-subdomain (95_N172_M-BtPyr₂) or C subdomain (172_M267_C). When a mSA variant with a weaker biotin affinity is used (here mSA_{DAB}-S45A in micelles and mSA_{DAB}-S51A in bicelles), the separation of two binding phases is observed. The more attenuated second binding indicates the higher stability (*i.e.*, larger ΔG°_{N-D}). In each plot, the fluorescence intensity was normalized to the intensity change of the second binding phase and ΔG°_{N-D} of WT GlpG was measured in micelle or bicelle by fitting the second binding phase with Eq.5.

The single mutants with the completely buried L207K and V165E were highly destabilized, as measured at both N- (**Figure 2.5.a**) and C-subdomains (**Figure 2.5.b**). However, the magnitudes of ΔG°_{N-D} of the single mutants with the completely buried ionizable residues were significantly higher than the magnitudes of ΔG°_{N-D} of the single mutants with partially buried ionizable residues M100K and T178E, implying that the ionizable residues buried in the interior induced the larger destabilization of the protein.

We also tested how the amphipathic residue Met100 and small polar residue Thr178 located in N-subdomain of the lipid-contacting region contribute to stability compared to the nonpolar residues with similar sizes, M100L and T178V. When stability was measured for N-subdomain with the biotin pair at 95_N172_M (**Figure 2.5.a**), the single mutation T178V destabilized GlpG by 1.2 ± 0.1 kcal/mol. The crystal structure of GlpG displays the side chain-backbone H-bond between Thr178 and Leu174. Thus, the T178V substitution may disrupt the side chain-backbone H-bond, leading to destabilization. On the other hand, the other single mutation M100L as well as the double mutation M100L/T178V only mildly destabilized GlpG by 0.6 ± 0.1 kcal/mol. Interestingly, when stability was measured for C-subdomain with the biotin pair at 172_M267_C, the ΔG°_{N-D} of M100L, T178V and M100L/T178V mutants were almost identical to WT. These results indicate that in the nonpolar environment contacting lipids, amphipathic Met has an energetic contribution to stability similar to Leu, and polar Thr can be stabilizing by forming a H-bond with the backbone. However, the latter energetic effect is highly local to the region (i.e., N-subdomain) where Thr is located as this stabilization does not bear any effect on the local stability of the other region (i.e., C-subdomain).

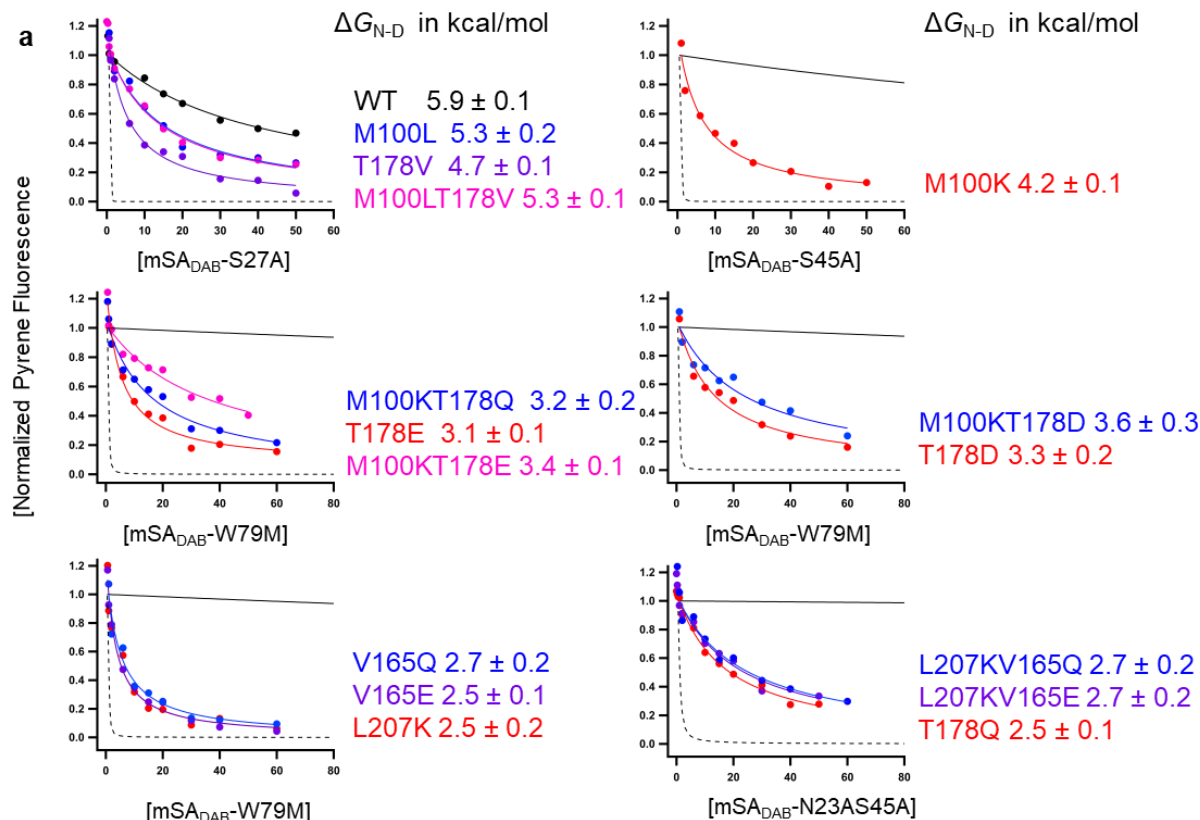
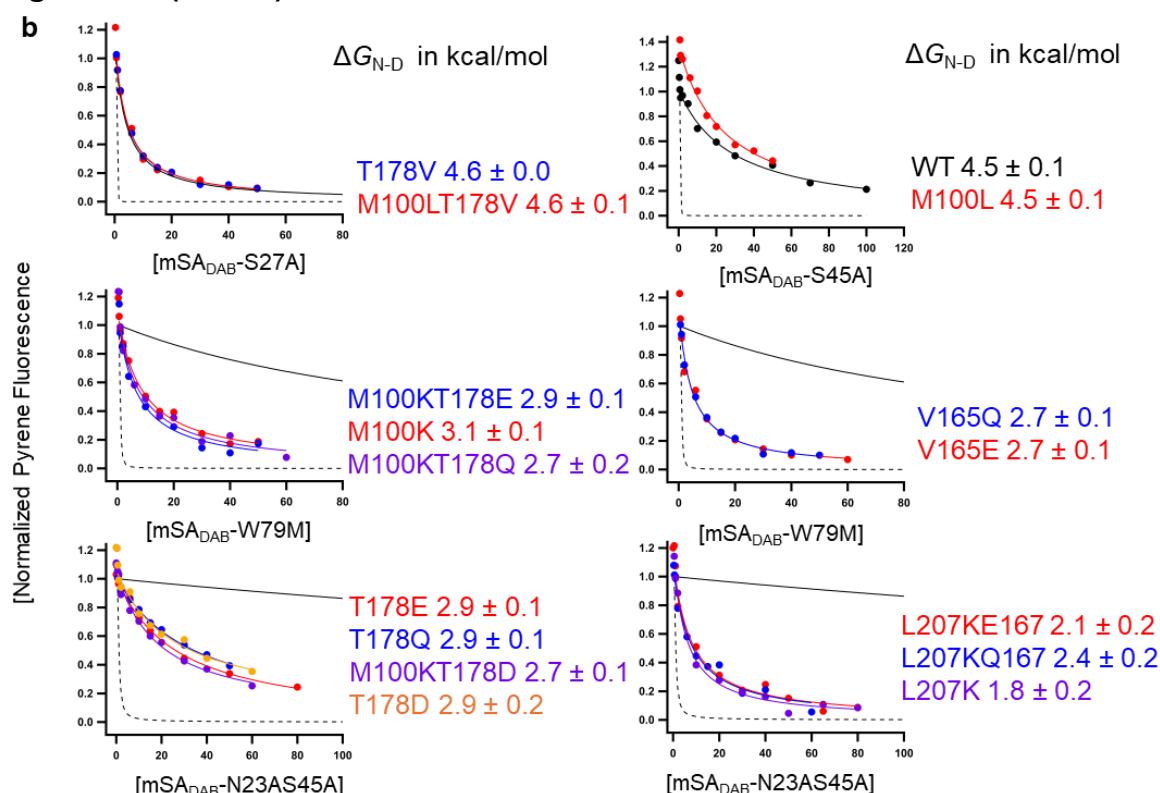


Figure 2.5. Binding isotherms between the double-biotin variants of GlpG and monovalent streptavidin (mSA) to determine the thermodynamic stability of WT and variants using steric trapping in DDM micelles. Binding was measured by quenching of pyrene fluorescence from the BtnPyr labels, which was induced by the dabcyI quencher conjugated to mSA (mSADAB). The double-cysteine variants of GlpG were labeled with BtnPyr at N-subdomain (95_N172_M) or C subdomain (172_M267_C). In each plot, black dashed lines indicate the intrinsic, unhindered binding of mSA variants without the steric trapping effect. ΔG°_{N-D} was obtained by fitting the attenuated second binding phase to Eq.5. In each plot, the fluorescence intensity was normalized to the intensity change of the second binding phase. The weaker second binding indicates the higher stability (*i.e.*, the larger ΔG°_{N-D}). For reference, the second binding phase to WT GlpG is shown when a given mSA variant was used (solid black line) in each plot. **(a)** Binding isotherms measured at N-subdomain (95_N172_M). **(b)** Binding isotherms measured at C-subdomain (172_M267_C)

Figure 2.5. (cont'd).



2.4.5. Buried ionizable residues form favorable interactions in DDM micelles

Double mutant cycle analysis was employed to study the interaction energy between the engineered basic and acidic residues that are close to each other in the core of GlpG. The two residues can be regarded as coupling or interacting with each other if the sum of degree of destabilization by individual single mutations is different from the degree of destabilization by their corresponding double mutation²²⁴. Thus, according to the definition of the interaction energy ($\Delta\Delta G^{\circ}_{\text{int}}$, **Fig. 2.6 a**), negative $\Delta\Delta G^{\circ}_{\text{int}}$ represents a favorable interaction, $\Delta\Delta G^{\circ}_{\text{int}}$ of zero means no interaction, and positive $\Delta\Delta G^{\circ}_{\text{int}}$ implies unfavorable interactions. In micelles, the partially exposed ionizable pair M100KT178E formed a favorable interaction with $\Delta\Delta G^{\circ}_{\text{int}} = -2.0 \pm 0.3$ kcal/mol when measured at N-subdomain and $\Delta\Delta G^{\circ}_{\text{int}} = -1.6 \pm 0.2$ kcal/mol at C-subdomain. The fully buried pair L207K/V165E also formed a favorable interaction with $\Delta\Delta G^{\circ}_{\text{int}} = -2.1 \pm 0.2$ kcal/mol at N-subdomain and $\Delta\Delta G^{\circ}_{\text{int}} = -3.6 \pm 0.3$ kcal/mol at C-subdomain. The difference between the interaction energies measured at N and C subdomains were overall larger for the completely buried L207K/V165E pair compared with the partially

exposed M100K/T178E pair. Overall, both the buried ionizable pairs form favorable interaction as part of compensation for the highly unfavorable energy cost of having the ionizable residues in the hydrophobic core of GlpG.

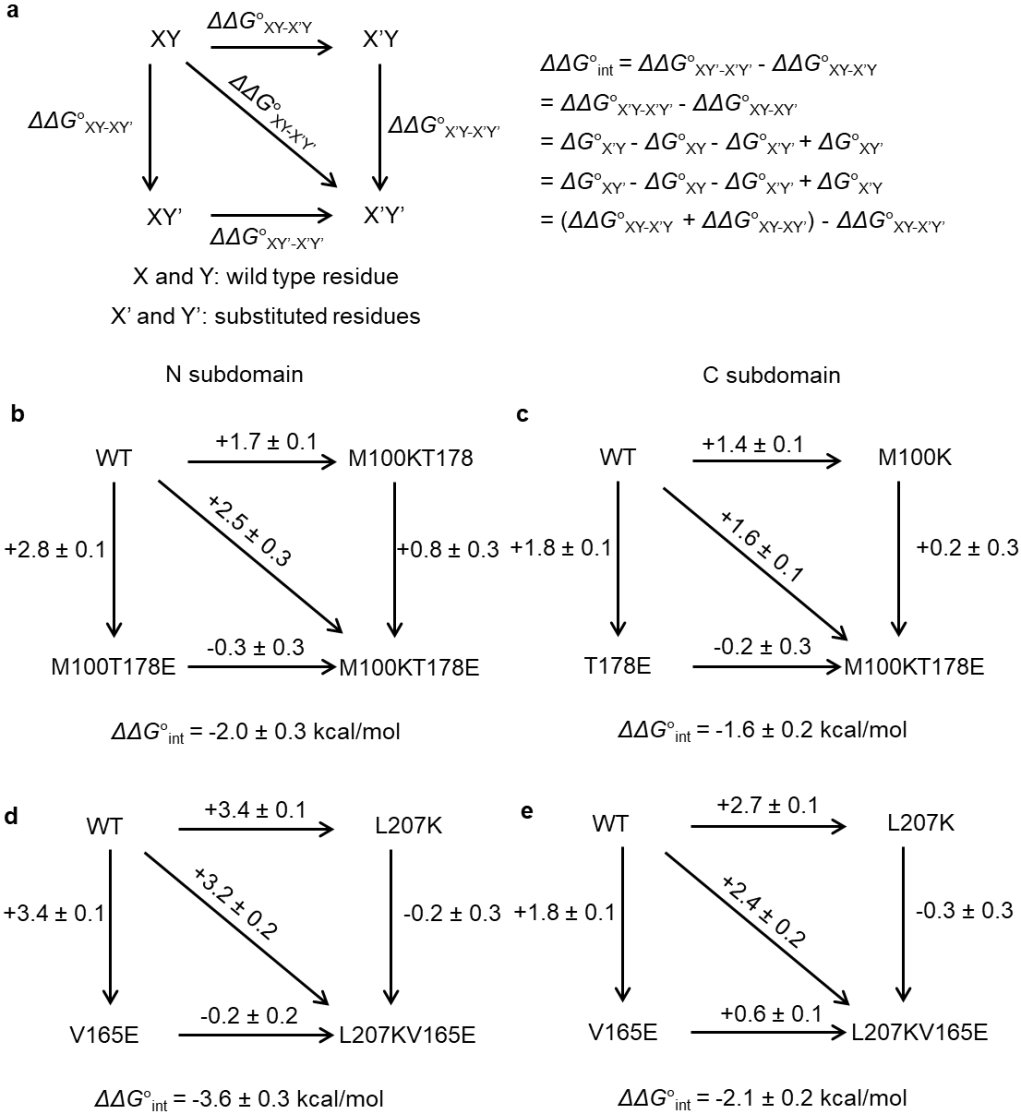


Figure 2.6. Double-mutant cycle for calculating the interaction energy ($\Delta\Delta G^{\circ}_{\text{int}}$) of the buried ionizable pairs in DDM micelles. (a) The scheme of double-mutant cycle and equations for calculating the interaction energy, $\Delta\Delta G^{\circ}_{\text{int}}$. (b) and (c) The double-mutant cycles for M100K/T178E contain the stability changes upon single and double mutations and the interaction energy measured at N-subdomain and C-subdomain. (d) and (e) The double-mutant cycles for L207K/V165E.

2.4.6. Impact of buried ionizable residues on the kinetic stability of GlpG in DMPC/CHAPS neutral bicelles

Recent studies have highlighted the critical roles of lipid-enriched hydrophobic environments in maintaining the conformational stability of membrane proteins¹⁴⁷. However, it is not well understood how different hydrophobic environments precisely impact the thermodynamic stability, kinetic stability, and functionality of membrane proteins bearing buried ionizable residues. Here we compare the impacts between a lipid-based system and a detergent-based system in their ability to accommodate ionizable residues. I first measured the kinetic stability of single and double mutants in DDM micelle and zwitterionic 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC) and 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS) neutral lipid bicelles. Cryo-electron microscopy structure revealed that at a molar ratio of DMPC to CHAPS of 1.5, uniform oblate spheroidal bicelle particles with an average diameter of 90 Å were formed, suggesting the formation of lipid-segregated bicelle with a planar bilayer phase in the center rather than mixed micelles¹⁴⁷. In contrast, detergent DDM formed near-spherical micelles with a smaller average diameter of 60 Å¹⁴⁷.

In parallel to the study in DDM, we evaluated the kinetic stability of GlpG in the lipid bilayer environment provided by bicelles by a proteolysis assay. The susceptibility of membrane protein to nonspecific proteolysis by ProK served as an indicator of differentiating the denatured state from the folded state. As shown in **Figure 2.7 a**, over 80% of WT GlpG remained intact after 12 days at room temperature. Although the current setup cannot determine the absolute half-life or the intrinsic denaturation rate of WT GlpG due to the loss of ProK activity during prolonged incubation, the fact that the half-life of WT GlpG in detergent was ~90 h allowed us to confidently conclude that the WT GlpG has higher kinetic stability in lipid bicelles than in detergent micelles. This conclusion is further supported by a previous study, which reported an intrinsic denaturation rate of approximately 110 hours for WT GlpG in DMPC/CHAPS bicelle at 37°C²³³. Additionally, the half-life of the single mutant M100K was approximately 25 hours in bicelles, which is 12-fold longer than its half-life in micelles. However, the bicelle environment did not provide further kinetic stabilization to the other mutants. In

fact, the intrinsic denaturation rate of the other mutants such as T178E, M100K/T178E, L207K, V165E and L207K/V165E were even faster in bicelles compared to micelles (**Figure 2.6.c-g** and **Table 2.2.**). These results indicate that, for most of the buried ionizable residue mutants, the bicelle environment does not provide additional enhancement of the kinetic stability compared to the micelle environment.

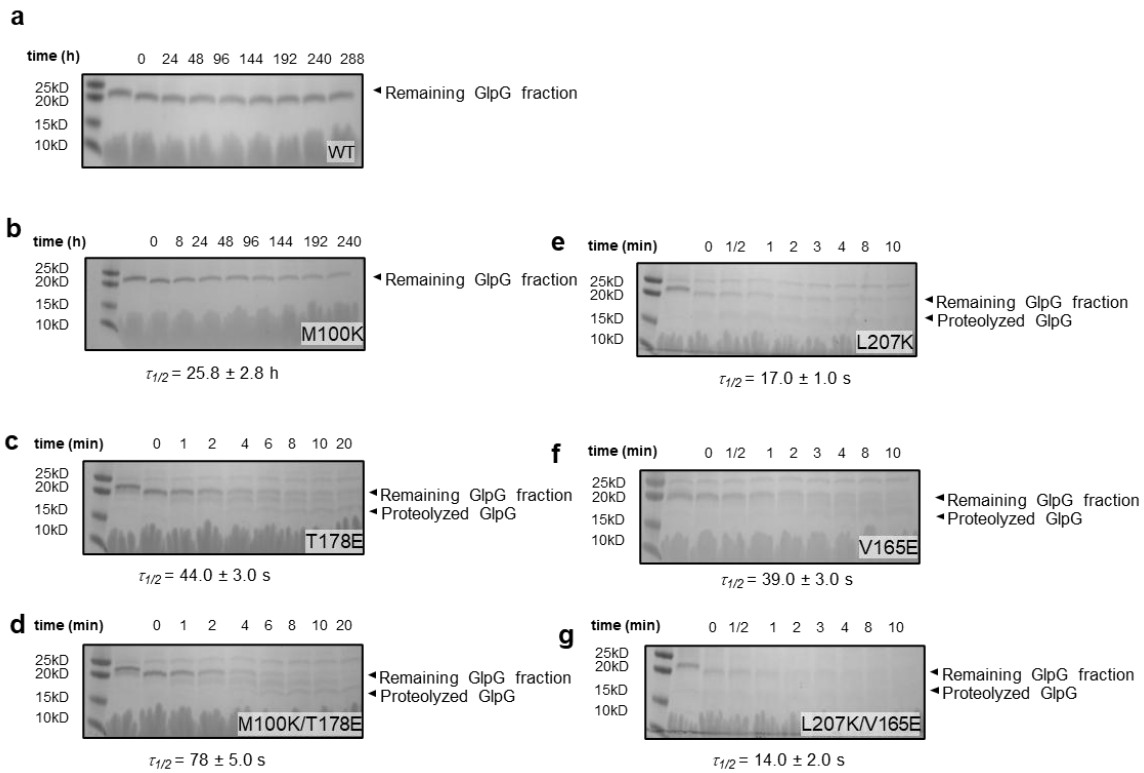


Figure 2.7. Determination of the spontaneous denaturation rate k_{denat} using ProK digestion assay. (a)-(g) SDS-PAGE results obtained by ProK digestion assays in 2% neutral DMPC/CHAPS bicelles (the molar ratio of DMPC to CHAPS = 1.5) for WT GlpG, M100K, T178E, M100K/T178E, L207K, V165E, and L207K/V165E variants respectively.

Table 2.2. Intrinsic denaturation rate, half-life, denaturation time constant of wild type and ionizable single, double mutants measured in DMPC/CHAPS bicelle.

	k_{denat} in bicelle (sec^{-1})	$t_{1/2}$ in bicelle (sec)	t_D in bicelle (sec)
WT	N/D	N/D	N/D
M100K	$7.5 \pm 1.0 \times 10^{-6}$	$9.3 \pm 1.0 \times 10^4$	$1.3 \pm 0.2 \times 10^5$
T178E	$1.6 \pm 0.1 \times 10^{-2}$	44.0 ± 3.0	64.0 ± 4.0
M100KT178E	$8.9 \pm 0.5 \times 10^{-3}$	78.0 ± 5.0	110 ± 7.0
L207K	$4.0 \pm 0.3 \times 10^{-2}$	17.0 ± 1.0	25.0 ± 2.0
V165E	$1.8 \pm 0.0 \times 10^{-2}$	39.0 ± 3.0	56.0 ± 4.0
L207KV165E	$5.1 \pm 0.1 \times 10^{-2}$	14.0 ± 2.0	20.0 ± 2.0

2.4.7. Impact of buried ionizable residues in thermodynamic stability of GlpG in the DMPC/CHAPS neutral bicelles

In neutral DMPC:CHAPS bicelles ($q = 1.5$), the measured $\Delta G^{\circ}_{\text{N-D}}$'s of WT GlpG were 7.1 ± 0.2 kcal/mol at N-subdomain and 6.8 ± 0.1 kcal/mol at C-subdomain. The values agree with the previous determined results¹⁴⁷ (**Figure 2.8.a**). Compared with the stabilities measured in DDM micelle, the mutants, M100L, T178V, and M100LT178V, are better tolerated in lipid bicelles, that is, the stability changes induced by these mutations ($\Delta\Delta G^{\circ}_{\text{N-D, WT-Mut}}$'s) in micelles were smaller than those in bicelles. At N-subdomain, $\Delta\Delta G^{\circ}_{\text{N-D, WT-Mut}}$ of the partially exposed single mutants M100K and T178E were $+1.3 \pm 0.3$ kcal/mol and $+1.9 \pm 0.3$ kcal/mol respectively, moderately destabilizing the protein (1.7 and 2.8 kcal/mol in micelles, respectively) (**Figure 2.8.a**). At C-subdomain, $\Delta\Delta G^{\circ}_{\text{N-D, WT-Mut}}$ of M100K and T178E were $+1.2 \pm 0.1$ kcal/mol and $+1.5 \pm 0.2$ kcal/mol respectively, destabilizing the protein (1.4 and 1.8 kcal/mol in micelles, respectively). $\Delta\Delta G^{\circ}_{\text{N-D, WT-Mut}}$ of M100K/T178E double mutant were $+1.9 \pm 0.2$ kcal/mol and $+1.6 \pm 0.1$ kcal/mol when measured in N and C subdomain respectively (**Figure 2.8.b**). The double mutation destabilized GlpG yet the destabilization by the two single mutations were not additive, implying the coupling of the two residues.

At N-subdomain, $\Delta\Delta G^{\circ}_{\text{N-D, WT-Mut}}$ of the completely buried single mutant L207K and V165E were both $+1.6 \pm 0.2$ kcal/mol, destabilizing the protein (3.4 kcal/mol in micelles). At C-subdomain, $\Delta\Delta G^{\circ}_{\text{N-D, WT-Mut}}$ of L207K and V165E were $+1.9 \pm 0.1$

kcal/mol and $+1.7 \pm 0.1$ kcal/mol, respectively (2.7 and 1.8 kcal/mol in micelles, respectively), similar to the destabilization observed at N-subdomain. The $\Delta\Delta G^{\circ}_{N-D, WT-Mut}$ of the double mutant L207K/V165E were $+2.7 \pm 0.2$ kcal/mol and $+2.1 \pm 0.2$ kcal/mol when measured at N and C subdomain, respectively. Overall, the degree of destabilization was smaller in bicelles than in micelles, indicating that the buried ionizable residues are better tolerated in bicelles than in DDM micelles.

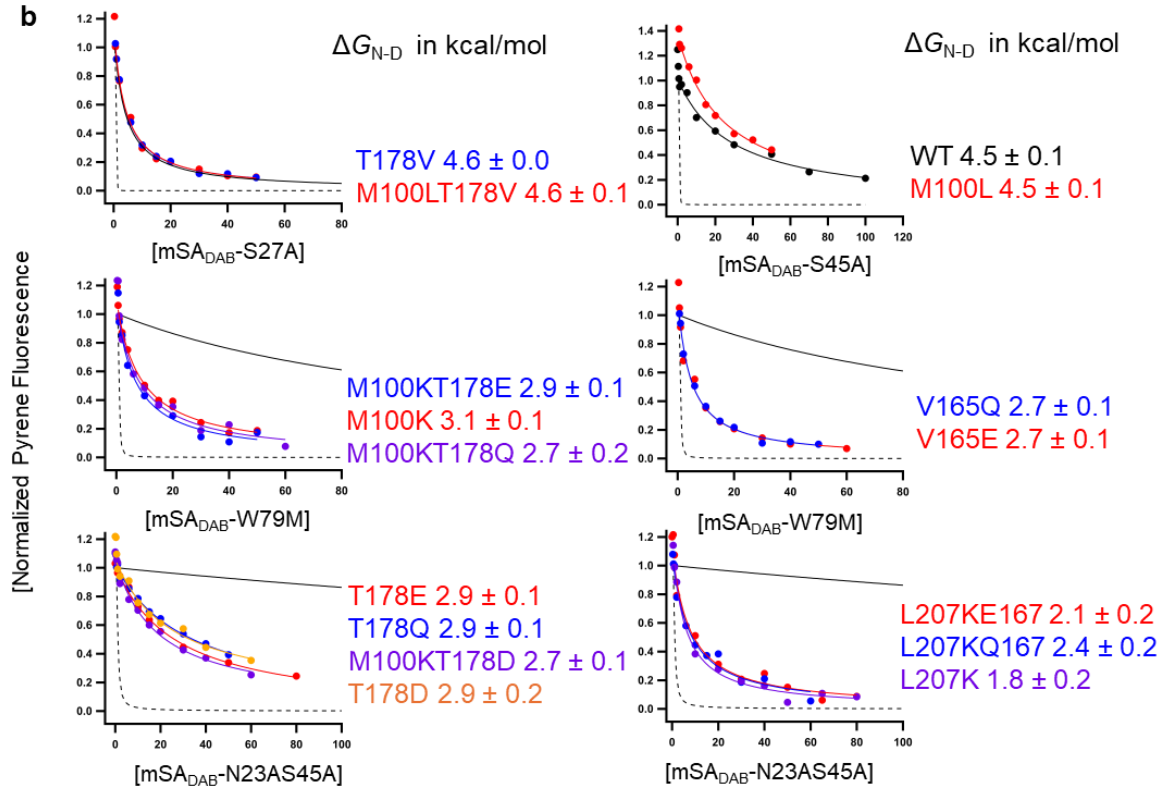
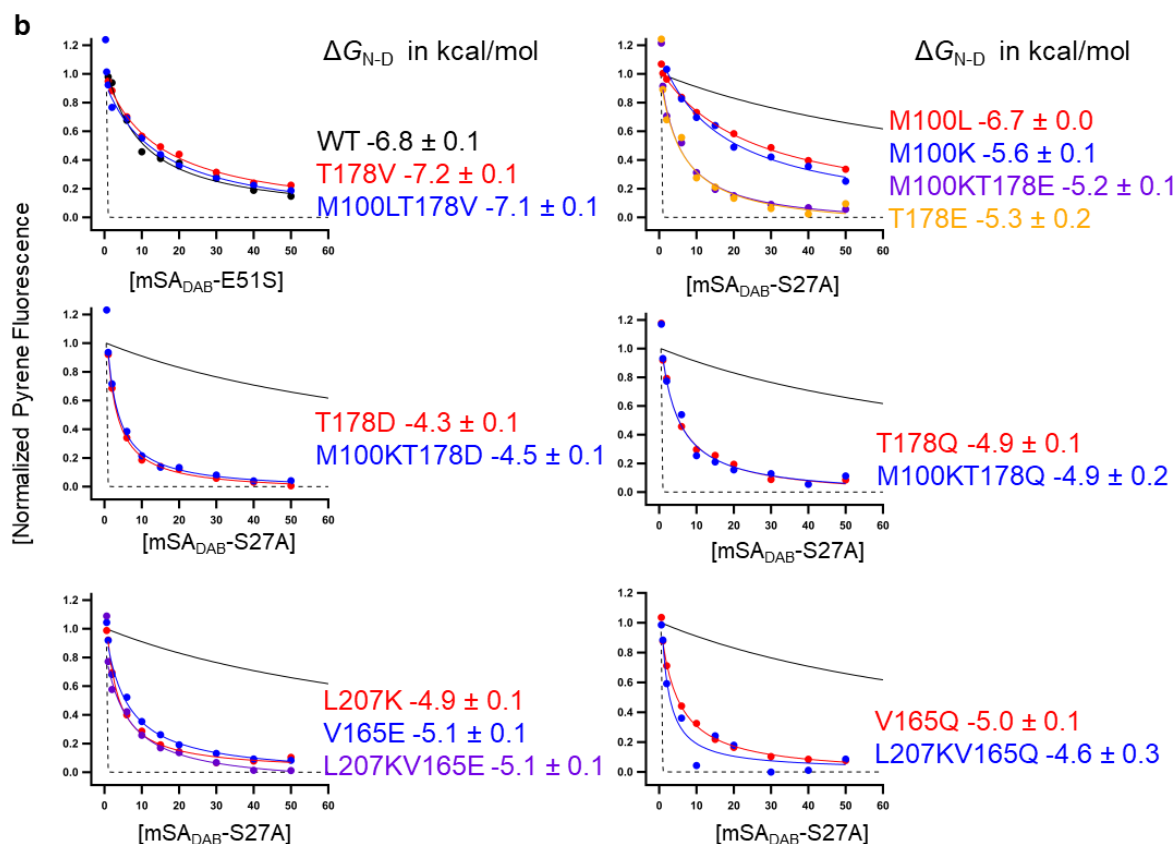


Figure 2.8. Binding isotherms between doubly-biotinylated GlpG variants and monovalent streptavidin (mSA) variants to determine the thermodynamic stability of GlpG using steric trapping in DMPC/CHAPS bicelles. In each plot, the fluorescence intensity was normalized to the intensity change of the second binding phase. The fitted stabilities of WT and mutants are shown. For reference, the simulated second binding phase for WT GlpG (solid black line) and the intrinsic unhindered binding phase of mSA (dotted black line) are shown in each plot. **(a)** Binding isotherms measured at N-subdomain (95_N172_M). **(b)** Binding isotherms measured at C-subdomain (172_M267_C-BtnPyr₂)

Figure 2.8.(cont'd)



2.4.8. Buried ionizable residues form favorable interactions in neutral bicelles

Next, we quantified the interaction strengths ($\Delta\Delta G^{\circ}_{int}$) between the buried ionizable residues in lipid bicelles using double mutant analysis (**Figure 2.9.**). Overall, the Lys100/Glu178 and the Lys207/Glu165 pairs form favorable interactions, with $\Delta\Delta G^{\circ}_{int}$ values ranging from -0.5 ± 0.3 to -1.5 ± 0.2 kcal/mol. Interestingly, however, when comparing these interaction energies between bicelles and micelles, the interactions between the buried ionizable residues in bicelles were weaker than in micelles. An outstanding example is the completely buried L207K/V165E pair in N-subdomain in lipid bicelle with the $\Delta\Delta G^{\circ}_{int}$ of -0.5 ± 0.3 kcal/mol weaker than $\Delta\Delta G^{\circ}_{int}$ of -3.6 ± 0.3 kcal/mol in micelles.

Across all the interaction energy of double mutants obtained from the different subdomains and hydrophobic environments, $\Delta\Delta G^{\circ}_{int}$ values spanned a range of -3.6 to -0.5 kcal/mol (**Figure 2.10.a**). We further categorized these $\Delta\Delta G^{\circ}_{int}$ values based on the

hydrophobic environment (micelle vs. bicelle) and subdomains. The student *t*-test revealed a significant difference ($p < 0.05$) in $\Delta\Delta G^{\circ}_{\text{int}}$ obtained between N and C subdomains in micelles, but no difference ($p > 0.05$) in $\Delta\Delta G^{\circ}_{\text{int}}$ obtained between N and C subdomains in bicelles (**Figure 2.10.b**). Thus, the interaction energies were more uniform between different subdomains in the lipid environment provided by bicelles than in micelles. Notably, the pooled $\Delta\Delta G^{\circ}_{\text{int}}$ values from micelles and bicelles were significantly different ($p < 0.05$) (**Figure 2.10.b**), indicating that the aforementioned smaller interaction energies between the ionizable residue pairs in bicelles than those in micelles had a statistical significance. This discrepancy may arise from the relatively enhanced conformational stability of C-subdomain in bicelles (i.e., the uniform interaction energies across the subdomains) and a larger unfavorable contribution from bicelles to the interaction energies between the ionizable groups (i.e., the smaller interaction energies in bicelles than in micelles). This unfavorable contribution is likely stemmed from the poorer partition of the groups into the more dehydrated hydrophobic core of the bicelles compared to that of the micelles.

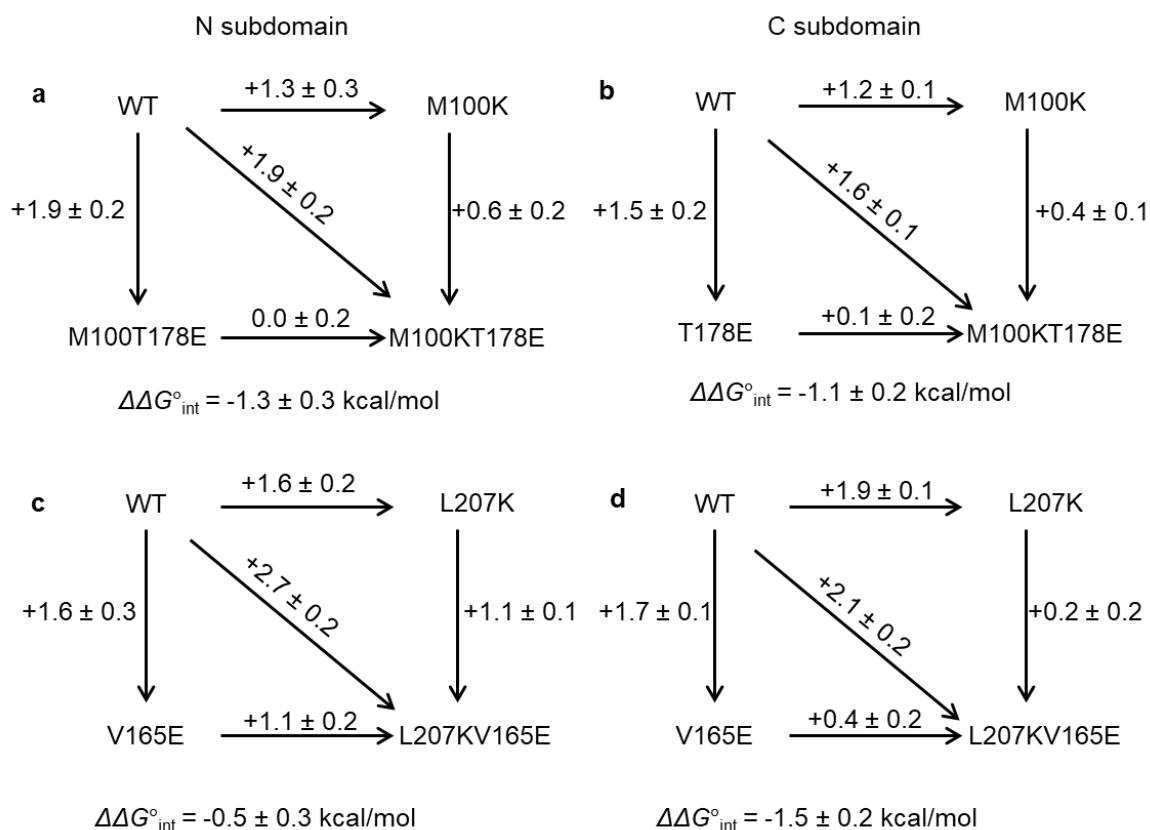


Figure 2.9. Double-mutant cycle analysis for calculating the interaction energies of buried ionizable pairs in DMPC/CHAPS bicelles. (a) and (b) Double-mutant cycles for the M100K and T178E substitutions and the associated their stability changes and interaction energies at N-subdomain (a) and C-subdomain (b); (c) and (d) Double-mutant cycles for L207K and V165E substitutions and the associated stability changes and interaction energies at N-subdomain (c) and C-subdomain (d).

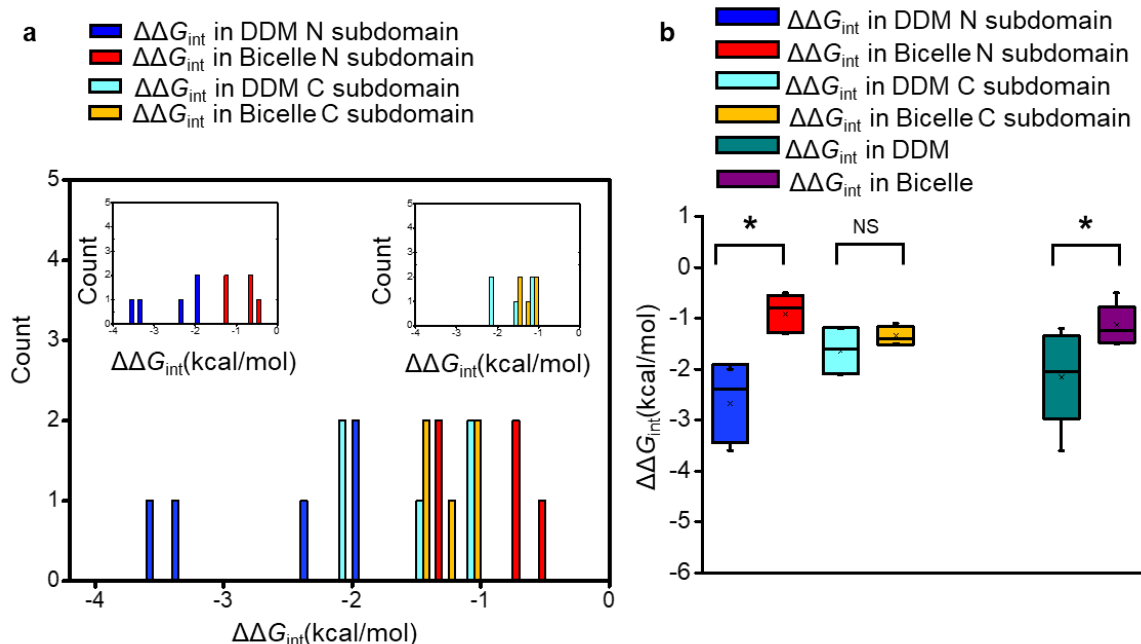


Figure 2.10. Statistical analysis of the interaction energies measured in different environments and at different subdomains. (a) Histogram of all interaction energies ($\Delta\Delta G^{\circ}_{\text{int}}$) in N-subdomain and C-domain in micelles and bicelles, respectively. **(b)** Box plots of $\Delta\Delta G^{\circ}_{\text{int}}$'s and the *t*-test results between two relevant data sets. Asterisks indicate the statistically different data pair with $p < 0.05$.

2.4.9. Ionizable residues are more likely to exist as neutral form in GlpG

In principle, to unequivocally assign the charged state of buried ionizable residues, pK_a of buried Glu and Lys should be determined by measuring the stabilities of WT and mutants at different pHs and by creating a titration curve. However, this strategy is not applicable with our current steric trapping method due to the difficulties in determining the intrinsic biotin affinities of mSA variants at different pH's and the poor behavior of mutant proteins (i.e., aggregation at lower pH's). As an alternative strategy, I replaced ionizable residues (e.g., Glu) with the neutral polar residue having a similar sidechain and volume (e.g., Gln) as a nonionizable proxy representing the charge-neutral state. Accordingly, I measured the stability of T178Q, V165Q, M100K/T178Q, and L207K/V165Q variants in micelles and bicelles (**Figure 2.5.** and **Figure 2.8.**). I expected when buried individually, the charged Glu residue would be much less tolerated than with its charge-neutral proxy, Gln. Overall, when measured under the same conditions

(i.e., the same position, the same environment, and the same subdomain), the single Glu-to-Gln mutation yielded a similar stability. I also expected when buried simultaneously with charged Lys, the charged Glu form would interact more strongly with Lys than the Lys-Gln pair. However, the double mutant cycles yielded a similarly favorable interaction energy between Lys-Glu and Lys-Gln pairs (**Figure 2.11.**, **Figure 2.6.**, **Figure 2.9.**, and **Figure 2.12.**).

Together, these results suggest that the buried Glu residues are more likely to exist as a charge-neutral form and form a polar-polar instead of a charge-charge interaction with Lys. Previous study indicates the strength of salt bridge interaction depends on the geometry of the two oppositely charged residues¹⁰³. It is agreed that according to the Coulomb's law, the attraction between opposite charges is inversely proportional to distance between the two charges. I expected that the increase in distance between the charged centroids of oppositely charged side chains would cause a decrease in the strength of salt bridge interaction if there are no extra favorable H-bonding interaction formed. Compared to Glu, Asp has a shorter side-chain length that will increase the distance from Lys. The double mutant cycle analysis (**Figure 2.5.**, **Figure 2.8.** and **Figure 2.12.**) indicates that M100K/T178D pair and M100K/T178E pair yielded a similar favorable interaction under the same condition. Therefore, it is hard to differentiate contributions of the side chain distance by comparing the stability changes and interaction energies.

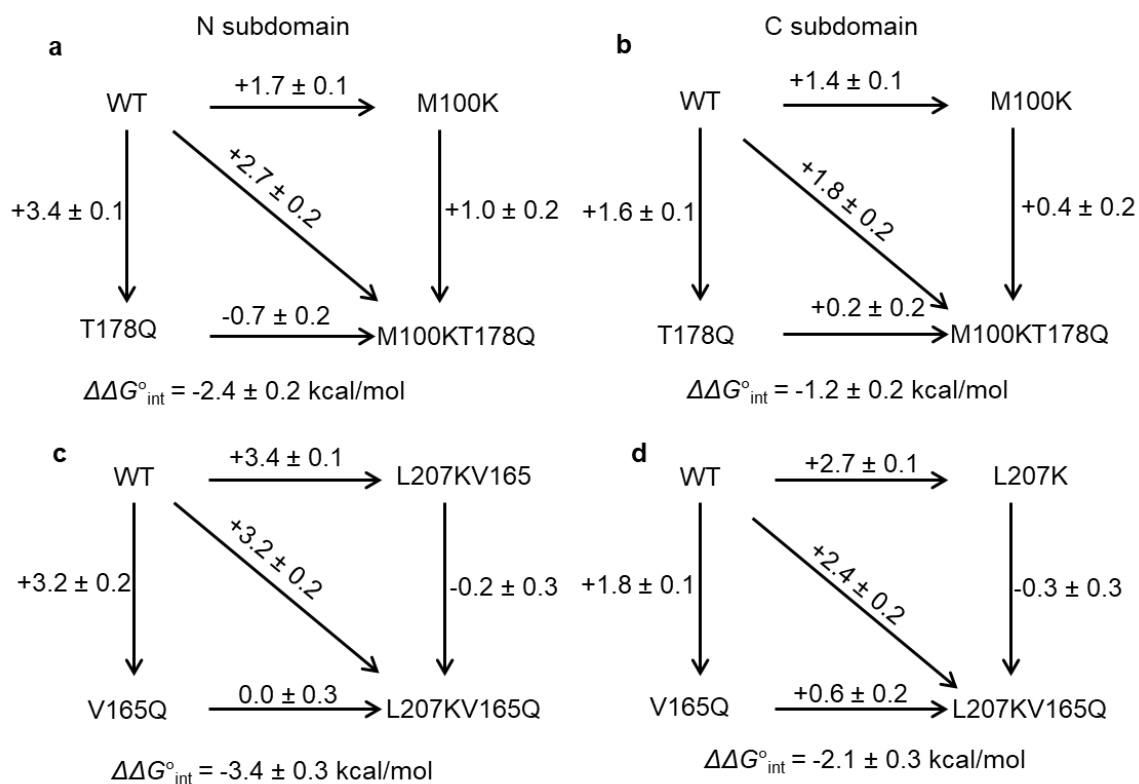


Figure 2.11. Double-mutant cycles for calculating the interaction energy of buried ionizable/polar residue pairs in DDM micelles. (a) and (b) The double-mutant cycles for the M100K/T178Q substitutions at N- and C-subdomains. (c) and (d) The double-mutant cycles for the L207K/V165Q substitutions at N- and C-subdomains.

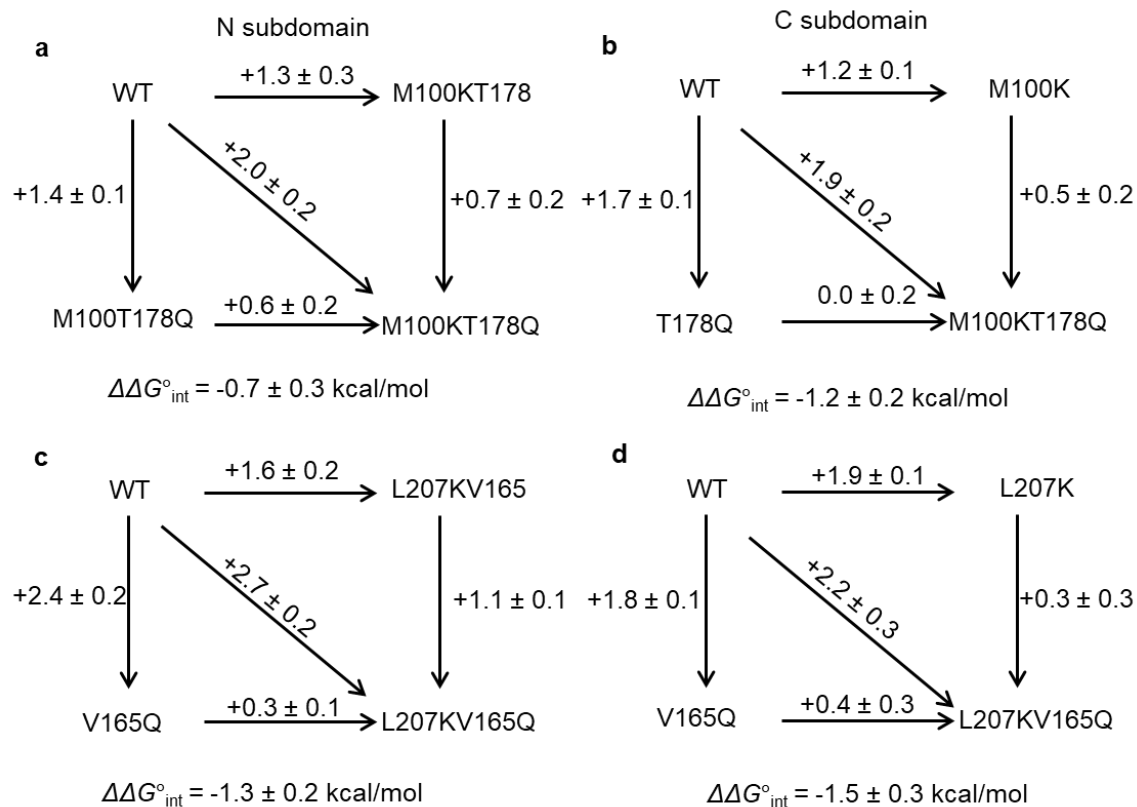


Figure 2.12. Double-mutant cycles for calculating the interaction energy of buried ionizable/polar residue pairs in bicelles. (a) and (b) The double-mutant cycles for the M100K/T178Q substitutions at N- and C-subdomains. (c) and (d) The double-mutant cycles for the L207K/V165Q substitutions at N- and C-subdomains.

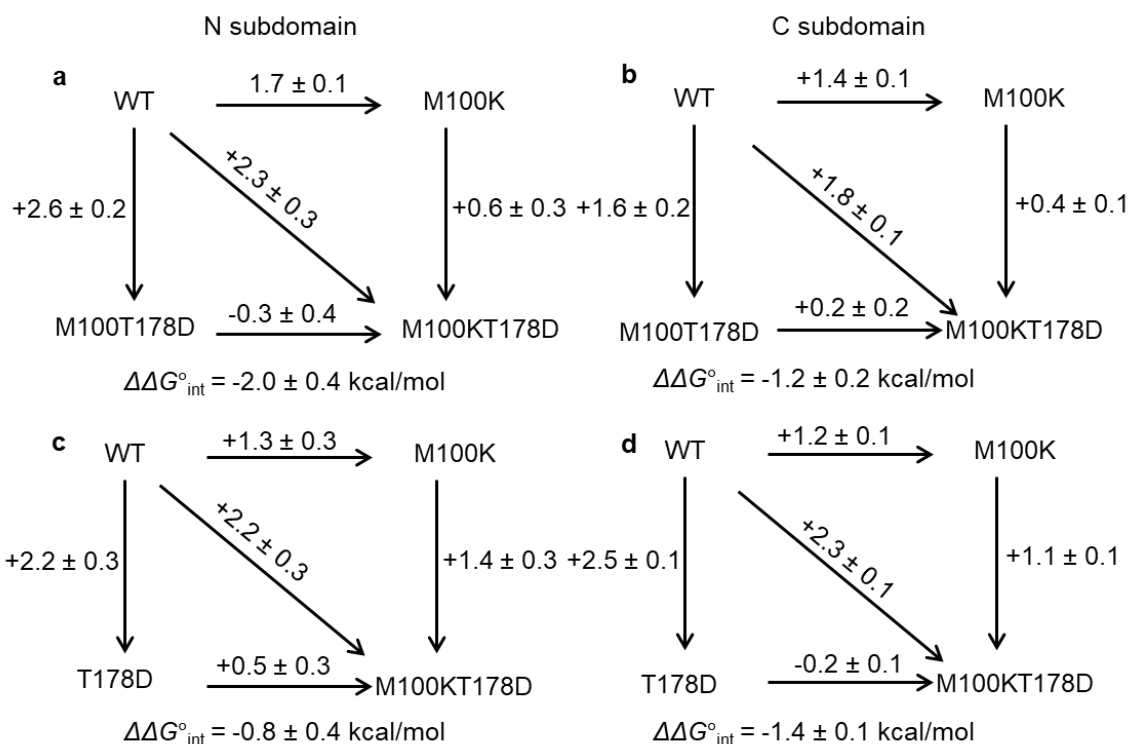


Figure 2.13. Double-mutant cycles for calculating the interaction energy of buried ionizable/polar residue pairs in different environments. (a) and (b) The double-mutant cycles for the M100K/T178D substitutions at N- and C-subdomains in micelles. (c) and (d) The double-mutant cycles for the M100K/T178D substitutions at N- and C-subdomains in bicelles.

2.4.10. Cooperativity of buried ionizable residues in micelle and bicelle

Next, we analyzed how the destabilization (i.e., the perturbation of side-chain interactions) caused by the buried ionizable/polar residues propagates within GlpG using the cooperativity profiling based on the steric trapping strategy¹¹⁵. Steric trapping captures the unraveling of the tertiary contacts around the region containing a specific biotin pair, allowing the measurement of local stability of a protein depending on the position of a biotin pair. By measuring the stability changes induced by a point mutation with the biotin pairs located in different regions of GlpG, the degree of propagation of the perturbation caused by the mutation can be quantitatively evaluated.

Here, the effect of a single mutation on stability ($\Delta\Delta G^{\circ}_{\text{N-D, WT-Mut}}$) is measured with two biotin pairs located in N and C subdomains (i.e., $\Delta\Delta G^{\circ}_{\text{N-D, WT-Mut}}^{\text{N}}$ and $\Delta\Delta G^{\circ}_{\text{N-D, WT-Mut}}^{\text{C}}$,

respectively). If the difference in the measured stability changes ($\Delta\Delta\Delta G^\circ = \Delta\Delta G^\circ_{u,WT-Mut^N} - \Delta\Delta G^\circ_{u,WT-Mut^C}$) is smaller than the thermal fluctuation energy ($|\Delta\Delta\Delta G^\circ| \leq RT = 0.6$ kcal/mol), it means the impact of the point mutation is uniformly propagated throughout GlpG. Thus, the WT residue is involved in “cooperative” interactions with the surrounding. When the mutation preferentially destabilizes the subdomain containing the point mutation ($|\Delta\Delta\Delta G^\circ| > RT$), the WT residue is engaged in “localized” interactions,. Finally, when the more destabilized subdomain does not contain the point mutation, ($|\Delta\Delta\Delta G^\circ| > RT$), the perturbed WT residue is engaged in an “over-propagated” interaction.

In micelle, four out of nine single mutations corresponded to “localized” interactions in N-subdomain ($RT < |\Delta\Delta\Delta G^\circ| < 2RT$) (**Table 2.3.**). For the M100L and M100K point mutations (WT residue was Met), the perturbation of the side-chain interaction was evenly propagated (i.e., “cooperatively” engaged). The mutation on L207K caused over-propagated interaction. Thus, in micelles, the residues are engaged with the surrounding with various cooperative profiles.

On the other hand, most of the mutated residues are engaged in cooperative interactions in bicelles (**Table 2.4.**). These results are consistent with our previous study that most of the residue interactions in bicelles are cooperatively engaged¹⁴⁷.

Table 2.3. Summary of the thermodynamic stability, change in thermodynamic stability upon mutation relative to WT, and cooperativity measured in micelles.

	N-subdomain (95 _N 172 _M)		C-subdomain (172 _M 267 _C)			
Mutants	ΔG°_{N-D} (kcal/mol)	$\Delta\Delta G^{\circ}_{N-D,WT-Mut}$ (kcal/mol)	ΔG°_{N-D} (kcal/mol)	$\Delta\Delta G^{\circ}_{N-D,WT-Mut}$ (kcal/mol)	$\Delta\Delta\Delta G^{\circ}$ (kcal/mol)	Cooperativity
WT	5.9 ± 0.1		4.5 ± 0.1			
M100L	5.3 ± 0.2	+0.6 ± 0.2	4.4 ± 0.1	+0.1 ± 0.1	+0.5 ± 0.1	Cooperative
T178V	4.7 ± 0.1	+1.2 ± 0.1	4.6 ± 0.0	-0.1 ± 0.1	+1.3 ± 0.1	Localized N-subdomain
M100K	4.2 ± 0.1	+1.7 ± 0.1	3.1 ± 0.1	+1.4 ± 0.1	+0.3 ± 0.1	Cooperative
T178E	3.1 ± 0.1	+2.8 ± 0.2	2.7 ± 0.1	+1.8 ± 0.2	+1.0 ± 0.3	Localized N-subdomain
T178D	3.3 ± 0.2	+2.6 ± 0.2	2.9 ± 0.2	+1.6 ± 0.2	+1.0 ± 0.3	Localized N-subdomain
T178Q	2.5 ± 0.1	+2.3 ± 0.1	2.9 ± 0.1	+1.6 ± 0.1	+0.7 ± 0.3	Localized N-subdomain
L207K	2.5 ± 0.2	+3.4 ± 0.2	1.8 ± 0.2	+2.7 ± 0.2	+0.7 ± 0.3	Overpropagated
V165E	2.5 ± 0.1	+3.4 ± 0.1	2.7 ± 0.1	+1.8 ± 0.1	+1.6 ± 0.1	Localized N-subdomain
V165Q	2.7 ± 0.2	+3.2 ± 0.2	2.7 ± 0.1	+1.8 ± 0.1	+1.4 ± 0.2	Localized N-subdomain

Table 2.4. Summary of the thermodynamic stability, change in thermodynamic stability upon mutation relative to WT, and cooperativity measured in bicelles.

	N-subdomain (95 _N 172 _M)		C-subdomain (172 _M 267 _C)			
Mutants	ΔG°_{N-D} (kcal/mol)	$\Delta\Delta G^{\circ}_{N-D,WT-Mut}$ (kcal/mol)	ΔG°_{N-D} (kcal/mol)	$\Delta\Delta G^{\circ}_{N-D,WT-Mut}$ (kcal/mol)	$\Delta\Delta\Delta G^{\circ}$ (kcal/mol)	Cooperativity
WT	7.1 ± 0.2		6.8 ± 0.1			
M100L	6.5 ± 0.1	+0.6 ± 0.2	6.7 ± 0.0	-0.1 ± 0.1	+0.5 ± 0.2	Cooperative
T178V	7.2 ± 0.2	-0.1 ± 0.3	7.2 ± 0.1	-0.4 ± 0.1	+0.3 ± 0.3	Cooperative
M100K	5.8 ± 0.2	+1.3 ± 0.3	5.6 ± 0.1	+1.2 ± 0.1	+0.1 ± 0.3	Cooperative
T178E	5.2 ± 0.2	+1.9 ± 0.3	5.3 ± 0.2	+1.5 ± 0.2	+0.4 ± 0.4	Cooperative
T178D	4.9 ± 0.2	+2.2 ± 0.3	4.3 ± 0.1	+2.5 ± 0.1	-0.2 ± 0.3	Cooperative
T178Q	5.7 ± 0.1	+1.4 ± 0.2	4.9 ± 0.1	+1.9 ± 0.1	-0.5 ± 0.2	Cooperative
L207K	5.5 ± 0.1	+1.6 ± 0.2	4.9 ± 0.1	+1.9 ± 0.1	-0.3 ± 0.2	Cooperative
V165E	5.5 ± 0.2	+1.6 ± 0.2	5.1 ± 0.1	+1.7 ± 0.1	+0.1 ± 0.2	Cooperative
V165Q	4.7 ± 0.1	+2.4 ± 0.2	5.0 ± 0.1	+1.8 ± 0.1	+0.6 ± 0.2	Localized N-subdomain

2.4.11. Effect of buried ionizable residues on enzymatic activity of GlpG

Finally, we investigated the effect of burying ionizable residues on the proteolytic activity of GlpG. In the case of staphylococcal nuclease with buried ionizable residues, the enzymatic activity is not disrupted for most of the sites unless the engineered sites are contacting the active site²¹⁶. As an indicator of folding and conformational integrity of GlpG¹⁴⁸, the enzymatic activities of the variants with buried ionizable residues were measured in micelles and bicelles using the model TM substrate LacYTM2^{234–236} and the model water-soluble substrate MMPS-042. The two substrates probe different aspects of the conformational integrity of GlpG. The TM substrate LacYTM2 interacts with GlpG through the binding to the “TM2-TM5 substrate docking site” (i.e., “the lateral

gate”) within the membrane^{167,176}. The water-soluble substrate MMPS-042, which is a 10-mer peptide²²¹, is more likely to access the active site directly from the aqueous phase to the opening of “L5 cap” on the catalytic dyad.

We observed the improvement in enzymatic activity for all mutants toward the TM substrate LacYTM2 in bicelles over micelles (**Figure 2.14.b**). The improvements were 2- to 3-folds for the mutations to the nonpolar residues including M100L, T178V and M100LT178V, and were near ten-fold for the mutations to the ionizable residues. In contrast, the improvement of bicelle over micelle toward the water-soluble substrate MMPS-024 was much lower (**Figure 2.14. b,d**). Most of the buried ionizable single or double mutations significantly inactivated GlpG, suggesting that either the folding or the conformation of the active site was perturbed. The mutation M100K was an exception maintaining $70 \pm 7\%$ and $94 \pm 7\%$ of the proteolytic activity toward the TM substrate in micelles and bicelles respectively, and $47 \pm 4\%$ and $61 \pm 9\%$ toward the water-soluble substrate MMPS-024 in micelles and bicelles respectively. Although the possibility that the ionizable residues in the protein interior perturbed the active site, it is also likely that the dramatic inactivation stemmed from the short lifetime of the folded state (~ 2 min) caused by the buried ionizable residues. The timescale of the substrate turnover by GlpG is 10^0 to 10^1 min, which overlaps with the lifetime of the folded state for the mutants with the buried ionizable mutations. That is, the mutants do not have enough time to carry out the catalysis reaction.

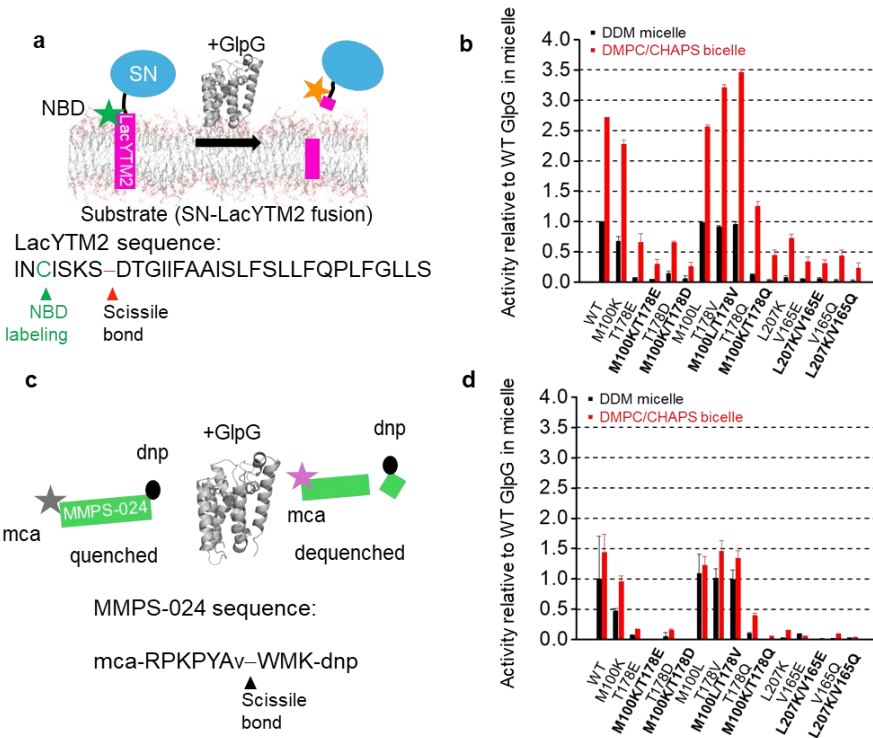


Figure 2.14. Activity assays to measure the impact of buried ionizable residues on GlpG activity. **(a)** The scheme for measuring the proteolytic activity of GlpG with the TM model substrate SN-LacYTM2 (SN: staphylococcal nuclease fusion; LacYTM2: the second TM segment of *E. coli* lactose permease). SN-LacYTM2 is labeled with the environment-sensitive fluorophore NBD on the five-residue upstream of the scissile bond. The cleavage of LacYTM2 induces the transfer of NBD from the hydrophobic bicelles to the aqueous phase. The transfer induces the decrease in NBD fluorescence. **(b)** Proteolytic activity (mean \pm s.d., $N = 3$) of GlpG variants toward LacYTM2. **(c)** The scheme for measuring the proteolytic activity of GlpG with the water-soluble model substrate MMPS-024. The peptide before proteolysis is an internally quenched state. The N-terminus of MMPS-024 is labeled with the fluorophore mca (7-methoxycoumarin). The quencher dnp (dinitrophenol) is conjugated on the C-terminus. The lower-case v stands for the nonnative amino acid norvaline. The cleavage of the scissile bond induces the increase in mca fluorescence. The size of the substrate relative to GlpG is not realistic, just for illustration. **(d)** Proteolytic activity (mean \pm s.d., $N = 3$) of GlpG variants toward MMPS-024.

2.5. Discussion

Although the ionizable residue pairs (i.e., the interaction between acidic and basic residues) inside membrane proteins has been proven to be important in membrane protein function and stability, their energetic impacts on stability and the influence of different solvation environments have been rarely studied. In this study, I tackled these problems through systematic site-specific mutagenesis, steric trapping, proteolysis, and double mutant cycle analysis using a TM helical protein, *E. coli* GlpG. My findings demonstrate that the incorporation of ionizable residues in the core of GlpG destabilizes the protein, leading to a moderate-to-substantial reduction in both thermodynamic and kinetic stability. Despite the energetic disfavor of burying the ionizable residues inside the protein, we observed that the interactions between the acidic and basic residues are moderately favorable with the $\Delta\Delta G^{\circ}_{\text{Inter}}$ of (-0.5 kcal/mol to -3.6 kcal/mol). Further comparative stability studies in different hydrophobic environments (i.e., in micelles and bicelles) show that bicelles more effectively maintain the conformational stability and enzymatic activity of GlpG. Interestingly, the interaction between the ionizable acid and basic residues were weaker in bicelles than it is in micelles. Although our bicelle preparation (3% w/v-% DMPC:CHAPS, $q = 1.5$) contains a significant portion of detergent CHAPS, our previous physical characterizations of bicelles (cryo-EM, small-angle X-ray scattering, fluorescence anisotropy, and Laudan fluorescence) indicate that the hydrophobic thickness, the gel-fluid phase transition temperature, and the strength of amphiphile-amphiphile packing of our bicelles are very similar to those of a pure fluidic DMPC bilayer. Thus, the lipid segregation into the center of bicelles effectively occurs in our bicelles and thus, our bicelles provide a reasonable bilayer mimicking physical environment for membrane proteins.

Here, we carefully selected two inter-helical nonpolar residue pairs neighboring the internal cavities with the consideration of the residue distance (preferably less than 4 Å) to form two potential ion pairs. The increased tendency to oligomerization shown from SEC profiles and the significantly higher intrinsic denaturation rate observed for the variants bearing ionizable residues compared to those for WT suggest that when the internal residues are exposed upon denaturation, the ionizable residues have a significant tendency to induce inter-chain oligomerization. We observed that lipid

bicelles improve the kinetic stability of GlpG. However, for the single and double mutants bearing the ionizable residues, the intrinsic denaturation rates in bicelles are similar to those in detergent micelles. Interestingly, the decrease in activation free energy ($\Delta\Delta G^{\ddagger\circ}_{N-D, WT-Mut}$) upon mutation are overall larger than the decrease in stability ($\Delta\Delta G^{\circ}_{N-D, WT-Mut}$) in micelles (**Figure 2.15.**). That is, buried ionizable residues are highly detrimental to the maintenance of a kinetically stable native state. It is surprising that the M100K residue that is partially lipid-exposed is relatively better tolerated compared to L207K buried in the protein interior. It is possible that the conformational perturbation caused by the Lys residue at the position 100 is energetically less extensive thus can be better accommodated in the lipid environment. This accommodation of the partially lipid-exposed Lys could be due to the deprotonation of the amine group to the charge neutral state and to the formation of a side chain-backbone H-bond similar to the stabilization of the Thr178 side chain by the side chain-backbone H-bond.

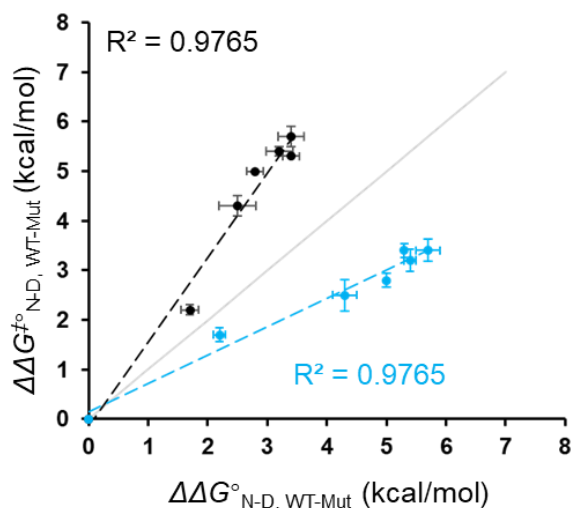


Figure 2.15. Comparison of GlpG kinetic stability and thermodynamic stability

changes caused by burial of ionizable residues in micelles. Comparison of the

mutational impacts on the kinetic stability (quantified as activation energy to transition state during denaturation $\Delta\Delta G^{\ddagger\circ}_{N-D, WT-Mut} = \Delta G^{\ddagger\circ}_{N-D, WT} - \Delta G^{\ddagger\circ}_{N-D, Mut}$,

$\Delta G^{\ddagger\circ}_{N-D} = -RT \ln k_{denat}$) and thermodynamic stability change ($\Delta\Delta G^{\circ}_{N-D, WT-Mut} = \Delta G^{\circ}_{N-D, WT} -$

$\Delta G^{\circ}_{N-D, Mut}$) in micelles. The solid lines with the slope (m) = 1 are shown as guide to

indicate the equal change in activation energy $\Delta\Delta G^{\ddagger\circ}_{N-D, WT-Mut}$ and thermodynamic stability change $\Delta\Delta G^{\circ}_{N-D, WT-Mut}$

Additionally, we observed that the introduction of ionizable residues into the core of GlpG induced a dramatic loss of activity. Based on the correlation between kinetic instability and the degree of inactivation, it is highly likely that the loss of activity stems from the kinetic instability (i.e., the short lifetime of the functional native state) rather than the perturbation of the native state by the buried ionizable residues (**Figure 2.16.a-b**).

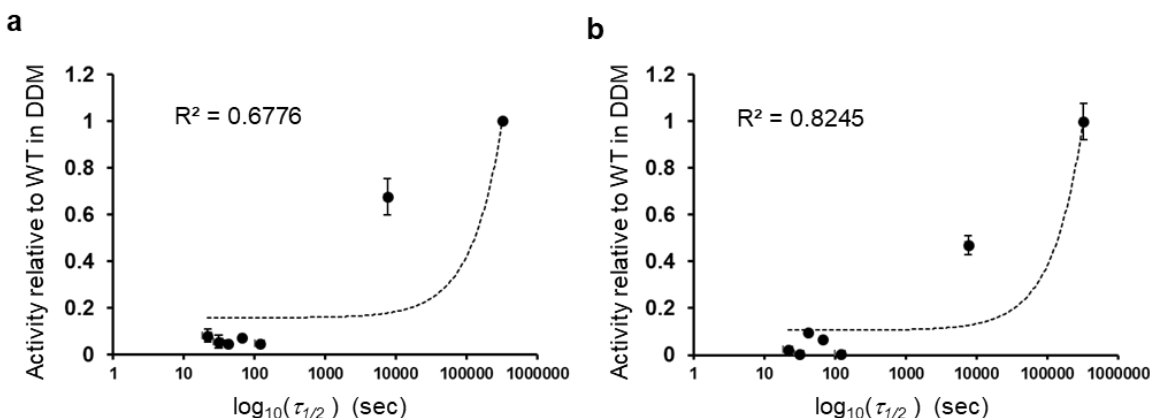


Figure 2.16. Correlation between the half-lives of the native states of GlpG WT and the variants vs their proteolytic activities in micelles. The linear trend lines are curved due to the scale of the native state half-life in the logarithmic format. All activities were normalized to the activity of WT GlpG. **(a)** Correlation with the proteolytic activity toward the TM substrate SN-LacY^{TM2}. **(b)** Correlation with the proteolytic activity toward the water-soluble model substrate MMPS-024.

It is interesting that the thermodynamic stability of the ionizable residue-bearing is compromised, but to a limited level. For those mutants, $\Delta\Delta G^{\circ}_{N-D,WT-Mut}$ ranges 1.7 to 3.4 kcal/mol for N-subdomain and 1.4 to 3.4 kcal/mol for C-subdomain in micelle; $\Delta\Delta G^{\circ}_{N-D,WT-Mut}$ ranges from 1.3 to 2.7 kcal/mol for N-subdomain, and 1.2 to 2.5 kcal/mol for C-subdomain in bicelle. For comparison, Leu207, one of the target residues for substitution to Lys, has been recognized as a critical residue since it is involved in the tight inter-helical side-chain packing interaction (i.e., with Leu161 on TM2)⁵¹. As a result, the destabilization caused by replacing Leu with Ala disrupting the TM2-TM4 interhelical packing (i.e., a “large deletion” mutation) was dramatic in both micelles and bicelles with ^{115,147} $\Delta\Delta G^{\circ}_{N-D,WT-Mut} = 3$ to 4 kcal/mol in micelle and $\Delta\Delta G^{\circ}_{N-D,WT-Mut} = \sim 5$ kcal/mol in

bicelle^{115,147}. Counterintuitively, the destabilization caused by L207A mutant is severer than the destabilization by L207K and L207Q. This result would imply the critical importance of van der Waals packing interaction in membrane protein stability over polar interactions. While the nonpolar-to-polar substitution destabilizes the protein by the increased desolvation cost of burying a polar residue in the nonpolar protein core, the Leu-to-Lys substitution reasonably maintains the packing constraints around the substitution site, not leading to severe destabilization.

These scenarios (i.e., the tolerance of lipid-contacting Lys and the less-than-expected degree of destabilization by the substituted ionizable residues) would be valid if the ionizable residues become charge-neutral in the protein interior. My result that the substitutions from Thr178 and Val165 to Glu or Gln have a negligible difference in GlpG stability partially support this premise. That is, Glu is highly likely to exist as the neutral state within the hydrophobic core.

CHAPTER 3: Investigate the origin of thermostability and activity of thermophilic rhomboid proteases in comparison to their mesophilic homologs

3.1. Summary

Proteins in thermophilic microbes maintain their fold and activity at high temperatures (80–110°C) and pressures (200–500 atm). In the phylogenetic tree, thermophilic bacteria and archaea dominate deep and short branches visually indicating their early emergence and slow rates of evolution. Thus, studies regarding the thermostability and activity of thermophilic proteins are generally beneficial to the fundamental understanding of protein stability and evolution. While such studies have largely focused on water-soluble proteins, the principles of thermostability and maintenance of activity under extreme conditions are not well understood for membrane proteins. Here, I tackle these problems using the universally conserved rhomboid protease family as a model. Three rhomboids from thermophilic bacteria (*TmRh* from *Thermotoga maritima*) and archaea (*PfuRh* from *Pyrococcus. furiosus* and *TpRh* from *Thermococcus profundus*) were cloned, expressed, and purified in *E. coli* cells.

Interestingly, the delipidated thermophilic rhomboids solubilized in micelles were fully inactivated below the optimal growth temperatures of their origins with their thermostability no greater than the mesophilic rhomboid GlpG from *E. coli* (*EcGlpG*). Additionally, the thermophilic rhomboids displayed varied activities towards transmembrane (TM) and water-soluble substrates. The activity variation is correlated with the lengths of the loops (L4 and L5) around flanking TMH (TM Helix) 5, which serves as a lateral gate for TM substrates. These results lead to two hypotheses: i) Thermostability of thermophilic membrane proteins are not acquired solely by their intraprotein interactions but the interactions with their native lipids are critical; ii) The lengths of the TMH4 and TMH5-flanking loops modulate the activity of rhomboids by controlling the mobility of their lateral gates.

3.2. Introduction

Life thrives in nearly every corner of earth, from the intense heat of deep-sea hydrothermal vents to the icy peaks of the Himalayas, and even in boiling hot springs or the frozen plains of Antarctica²³⁷. Organisms that adapt to extreme conditions are often grouped by the specific challenges they overcome—like temperature extremes (psychrophiles at low temperatures and thermophiles in intense heat), high-salt habitats (halophiles), acidic or alkaline conditions (acidophiles and alkaliphiles), and high-pressure zones (barophiles). Thermophiles are microbes that adapt to higher temperatures (80-110 °C). The phylogenetic tree based on the sequence of small subunit ribosomal RNA's has tripartite divisions that correspond to bacterial, archaeal and eukaryal domains harboring the root as a universal common ancestor²³⁸. Thermophilic organisms dominantly occupy deeper and shorter branches in the phylogenetic tree, suggesting their early emergence during evolution and slower evolution rates. Due to the heat resistance of their native conformation and activity, the proteins from thermophiles have garnered particular interests regarding protein engineering and industrial applications. For example, the widely used high-fidelity Pfu DNA polymerase, which maintains activity at 90 °C, was originally discovered and isolated from the thermophilic archaeon *Pyrococcus furiosus*²³⁹. The naturally dimeric ferritin found in thermophilic bacteria *Thermotoga maritima* forms protein nanocages and can be further engineered into various types of nanostructures for biomedical applications such as DNA protection against heat²⁴⁰. Thus, it is critical to understand fundamental principles behind how proteins maintain their folding and activity at high temperatures.

Heat increases molecular vibrations, consequently disrupting the non-covalent interactions such as H-bonding, VdW packing, and salt-bridge interactions which are essential for stabilizing the secondary and tertiary structures of proteins. To better understand the thermodynamic principles that govern protein stability under heat stress, Becketl and Schellman used the temperature dependent stability curve (**Figure 3.1.a**) that can be described by the modified Gibbs-Helmholtz equation (Eq1). This equation relates the heat capacity change upon unfolding (ΔC_p), melting temperature (T_m), and enthalpy change (ΔH_m) to the thermodynamic stability (ΔG_u) of a protein²⁴¹. Nojima and

coworkers proposed three strategies of how proteins modulate their stability (**Figure 3.1.b**) to adapt to higher temperatures (i.e., greater protein stability ΔG_U and melting temperature T_m)²⁴²: (I) an upshift of the stability curve to achieve a higher maximum stability ΔG_S , accompanying an increase of T_m (II) a broadening of the stability curve (i.e., a reduction of ΔC_p) to achieve a higher T_m (III) a shift of the stability curve towards the right direction for a greater T_m . The second derivative (Eq 2) of the Gibbs-Helmholtz equation (i.e., the curvature of the stability curve) is directly translated into ΔC_p . Thus, the ΔC_p should be smaller to achieve a higher stability,.

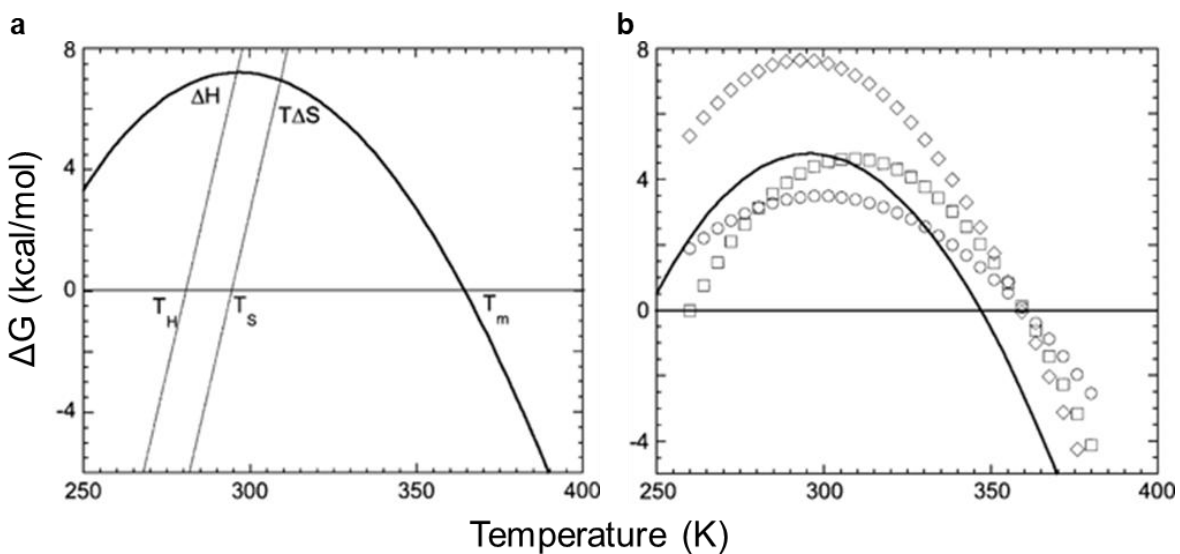


Figure 3.1. Stability curves for hypothetical proteins²³⁷. **(a)** Stability curve of a hypothetical protein plotted as a function of temperature. The curve can be described by the modified version of Gibbs-Helmholtz equation (Eq 1). T_S represents the temperature at maximal stability ΔG_S , T_m is the melting temperature when $\Delta G = 0$, T_H is the habitat temperature for the organism having the protein. **(b)** Stability curves show three strategies to achieve higher thermostability (i.e., the higher melting temperature T_m). The reference stability curve of a hypothetical mesophilic protein is in solid line. The T_m can increase by shifting the curve upward, which increases the overall stability, ΔG (strategy I depicted in diamonds); by broadening the curve (strategy II depicted in circles), or by shifting the curve to the right (strategy III depicted in squares). Reprint permission from John Wiley and Sons (license number: 6040840345436).

$$\Delta G = \left(1 - \frac{T}{T_m}\right) \Delta H_m - \left[(T_m - T) + T \ln\left(\frac{T}{T_m}\right)\right] \Delta C_p \# \text{ (Eq1)}$$

$$\frac{\partial^2 \Delta G(T_s)}{\partial^2 T} = -\frac{\Delta C_p}{T_s} \text{ (Eq2)}$$

The database that compares the experimentally determined thermodynamic parameters of homologous proteins from thermophiles and mesophiles only contains water-soluble proteins, many of which are DNA or RNA-interacting proteins²³⁷. Interestingly, the comparison between a histone protein *MfB* from thermophilic archaea and its mesophilic homologue *MfoB* (80% sequence identity) showed that all three strategies (i.e., the lower ΔC_p , higher ΔG_s , and higher T_s) are applied in *MfB* with the higher melting temperature T_m (113°C vs 74.8°C)²⁴³.

A global, non-redundant structural analysis of 64 mesophilic and 29 thermophilic proteins from 25 protein families show that thermophiles tend to have more ion pairs whose centroids of the charged groups are separated within 4 Å, suggesting the prevalence of salt bridge interactions in thermophilic proteins²⁴⁴. The comparative analysis of the amino acid distributions from eubacteria proteins suggests that there are more positively and negatively charged residues (i.e., Arg, His, Lys, Asp, and Glu) on the surface of thermophilic proteins compared to their mesophilic homologues²⁴⁵. Consistent with these analyses, Wong and coworkers also show that the water-soluble thermophilic protein, ribosomal L30e, from *Thermococcus celer* enhances its thermostability through the ion-pair interactions at the surface²⁴⁶. Through the Ala-scanning and double mutant cycle analysis, they determine that the two ion-pairs, Glu6/Arg92 and Glu62/Lys46, stabilize the protein by 2 kcal/mol and 5 kcal/mol in ΔG , respectively²⁴⁶. The Gibbs-Helmholtz analysis further shows that abrogating the two ion-pairs via Ala substitutions leads to an increase in ΔC_p (i.e., 5.3 kcal/mol/K for WT and 6.7 kcal/mol/K for Ala6/Ala92 double mutant), reducing the T_m . In the comprehensive survey of 29 thermophilic proteins and 64 of their mesophilic homologues from 25 protein families, the correlation between the number of ion pairs and the optimal growth temperature is observed²⁴⁷. Taken together, the surface ion-pair interaction is one of the key strategies for thermophilic water-soluble proteins to achieve their thermostability.

While the studies described above majorly focus on the comparisons of thermodynamic parameters, structure, and amino acid composition of water-soluble proteins in thermophile and mesophiles, studies on thermophilic membrane proteins have gained less attention partially due to the scarce information of their structure and stability. A comparative sequence analysis on the predicted TM helices of 8 thermophilic and 12 mesophilic membrane proteins showed a preference of small residues such as Gly, Ser, and Ala suggesting tighter packing of thermophilic membrane proteins²⁴⁸. Later, the Bowie group built a database of 25 unique helical membrane proteins from thermophiles and their 101 homologues from mesophiles. A comparative analysis of the structural properties such as side-chain burial, packing, H-bonding, TM helical kinks, loop lengths, and hydrophobicity²⁰¹ did not find a noticeable difference for most properties except for 1) a slightly decrease in the number of interhelical H-bonds in thermophiles; 2) a slightly larger hydrophobicity of the TM helices in thermophilic membrane proteins. These two properties agree with the broader sequence analysis showing the depletion of polar (i.e., Asn, Gln, Tyr) and ionizable (i.e., Asp, Glu, Arg) residues and the increase of small and nonpolar residues (i.e., Ala, Gly, Phe, Leu) in thermophilic membrane proteins²⁰¹. Notably, the increase in the number of small nonpolar residues would contribute to stabilization by reducing the entropic cost during helix-helix packing since the smaller residues have fewer possible rotamers, and by allowing tighter packing between helices²⁰¹.

Besides the structural features and amino acid composition of proteins, the lipid compositions of thermophilic organisms are also substantially different from mesophiles. It is expected that the lipids in thermophiles contribute to thermal adaptation by effectively responding their membrane properties to high temperatures impacting the function and behaviors of membrane proteins. Cells maintain their membrane integrity and homeostasis through a mechanism called 'homeoviscous adaptation'¹⁹². The electron paramagnetic resonance study first demonstrated that the *E. coli* membrane lipids extracted at various incubation temperatures (i.e., 15-43°C) exhibit a similar viscosity when assembled into a bilayer while the fractions of long and saturated fatty acyl chains in phospholipid increase as the growth temperature increases²⁴⁹. To maintain a proper level of fluidity and permeability of the membranes, thermal

adaptation strategies involve an adjustment of the rigidity and ordering of the hydrocarbon chains. In thermophilic bacteria, the increase in the ratios of *i*) branched chain iso-fatty acids, *ii*) saturated fatty acids, *iii*) long-chain fatty acids, and *iv*) polar carotenoid content is commonly observed as the optimal growth temperature increases^{250–253}. The lipids in archaea can be differentiated from those of bacteria by the presence of archaeol, a diether composed of two phytanyl chains²⁵⁴. In thermophilic archaea, macrocyclic archaeols, which have isoprenoid chains that cross-link the two tail-ends improve the membrane integrity against water by restricting lipid motions²⁵⁵. Membrane-spanning tetraether lipids, which form rigid monolayers, are frequently observed with a high abundance in archaea^{194–196}. Experimental and simulation studies demonstrate that such structure can efficiently reduce the heat-induced membrane leakage¹⁹⁷. Incorporation of pentacyclic rings into saturated acyl chains can further increase the gel-fluid phase transition temperature of the monolayers²⁵⁶. Interestingly, when the optimal living temperatures are beyond 90°C, the tetraether membrane spanning lipids are detected in both bacteria and archaea¹⁹².

Rhomboid proteases are a family of integral membrane, which activate membrane-anchored effector proteins via the cleavage of a peptide bond near the membranes¹⁶¹. They regulate various biological processes, including cell signaling, quorum sensing, mitochondria remodeling, and protein quality control¹⁶¹. Their universal occurrence in all life domains suggest that the rhomboid family may have evolved from the last universal common ancestor²⁵⁷. However, phylogenetic analysis suggests that rhomboids are more likely to have bacterial origin and then spread to archaea and eukaryotes through horizontal gene transfers during the early stage of tripartition²⁵⁸. Considering the early emergence of thermophiles in the phylogenetic tree, it would be reasonable to regard thermophilic rhomboids as the prototypes of the family. Thus, studying the structure-stability-activity relationship will be a valuable effort for advancing our understanding of the thermal adaptation strategy and evolutionary pathway of this universal protein family. Here, we chose the rhomboids from three thermophilic organisms as our study models: *i*) the rhomboid from the archaeon, *Thermococcus profundus*, which was isolated from a deep-sea hydrothermal vent with the optimal growth temperature of 80°C²⁵⁹; *ii*) the rhomboid from *Pyrococcus furiosus*,

which was originally found in a geothermal sediment with the optimal growth temperature of 100°C²⁶⁰; *iii*) the rhomboid from the thermophilic bacteria, *Thermotoga martima*, which was discovered in the sediment of marine geothermal area (i.e., hot spring) with an optimal growth temperature of 80°C (livable in a broad temperature range from 55°C to 90°C²⁶¹). The mesophilic homologue, GlpG from *E. coli* has been extensively studied regarding the structure, dynamics, and proteolytic mechanism as a control for this comparative study.

After successful cloning, expression, and purification, these thermophilic rhomboids were characterized using proteolytic activity and thermal denaturation assays. Interestingly, the thermal denaturation assay showed that the delipidated thermophilic rhomboids had lower thermostability than *E. coli* GlpG. This result leads to the hypothesis that the thermophilic rhomboids are not only stabilized by their intraprotein interactions but also their native lipids. In the proteolytic activity assay, the thermophilic rhomboids were tested towards various model substrates in micelles and bicelles. The rhomboids cleaved a given substrate with different activities while generally displaying a lower activity in neutral phospholipid bicelles. A multiple sequence alignment suggests that the activity variation of the different rhomboids stems from the different lengths of the flanking loops of the gating helix TM5. This result leads to the hypothesis that the activity level of rhomboids is controlled by the lengths of those flanking loops, which affect the mobility of the gating helix and the amplitude of the gating motions. Using *E. coli* GlpG as an engineering template, I tested this hypothesis by shortening or extending the lengths of the flanking loops.

3.3. Materials and Methods

3.3.1. The cloning of thermophilic rhomboids

The genes encoding the rhomboids from *T. maritima* (*TmRh*, UniProtKB: Q9X0H3), *T. profundus* (*TpRh*, UniProtKB: A0A2Z2ME71), and *P. furiosus* (*PfuRh*, UniProtKB: Q8U1H9) were amplified using a colony PCR (BIO-RAD) from their original genomes with the restriction sites of *NdeI* at the N-terminus and *XhoI* or *HindIII* at the C-terminus. The amplified genes and expression vectors (pET28a or pET30a) were digested with the restriction enzymes. The digested vectors were treated with calf-intestine phosphatase followed by ligation with T4-DNA ligase. XL1-blue competent cells were transformed with the ligation products, plated on LB agar plates containing kanamycin (pET30) and ampicillin (pET28), and incubated overnight at 37 °C. Isolated colonies were picked and further grown overnight. The final plasmids were extracted with the QIAspin miniprep kit (Qiagen).

3.3.2. Expression, purification, and identification of thermophilic rhomboids

The codon-optimized *TmRh* gene with a C-terminal His₆-tag encoded in pET30a was expressed in *E. coli* C43 (DE3) pLysS competent cells (Sigma Aldrich). 50 mL of LB cultures incubated overnight at 37°C were inoculated in 1 L TB media with 0.05 g/L kanamycin at 37°C until OD_{600nm} approached 0.8 to 1.0. The inoculated culture was chilled on ice before induction. IPTG was added to the final concentration of 0.5 mM for inducing protein expression and the culture was further incubated for 16-18h at 225 rpm, 15°C.

TpRh or *PfuRh* with a N-terminal His₆-tag encoded in pET 28a was expressed in *E. coli* C43 (DE3) pLysS competent cells (Sigma-Aldrich). 25 mL of LB cultures incubated overnight at 37°C were inoculated in 1 L LB media with 0.1 g/L ampicillin at 37°C until OD_{600nm} approached 1.0 to 1.2. The inoculated culture was chilled on ice before induction. IPTG was added to the final concentration of 0.5 mM for inducing protein expression and the culture was further incubated for 16 h to 18 h at 225 rpm, 15°C.

1 L of the LB culture was harvested and resuspended in 30 mL of 50 mM TrisHCl buffer (pH 8.0) with 5 mM EDTA (ethylenediaminetetraacetic acid), 1 mM DTT (dithiothreitol), and 1 mM PMSF. The cell resuspensions were lysed with a pressure

homogenizer (Avestin) 4 times. The cell lysate was centrifuged for 20 min at 4 °C, 6,000 rpm in a FS-34 rotor using a Sorvall RC6+ centrifuge (Thermo Scientific). Supernatant was collected and centrifuged to obtain the total membrane fraction at 24,000 rpm for 2 h with an ultracentrifuge (Beckman-Coulter). Membrane pellets were resuspended in 20 mL of 50 mM Tris-HCl buffer (pH 8.0) with 200 mM NaCl, 0.5 mM TCEP using a tissue homogenizer (Fisher Scientific). The membrane resuspension was solubilized by adding 1% (w/v) DDM followed by the removal of insoluble aggregates using ultracentrifugation at 18,000 rpm for 25 min. Supernatant was incubated with 2 mL of Ni-NTA resin (Qiagen, 50% w/v) by slowly rotating at 4°C for 1 h. The target protein was eluted at 300 mM imidazole in 50 mM Tris-HCl buffer (pH 8.0) and 1 M NaCl in 0.1% DDM. The eluted fraction was concentrated and desalted with the Amicon centrifugal filter units (Millipore Sigma, 30 kDa MWCO) and the desalting columns (BIO-RAD Econo Pac 10DG desalting column). The concentrations of thermophilic rhomboids were measured by UV absorbance at 280nm (*TmRh* ϵ =63,830 M⁻¹cm⁻¹; *TpRh* ϵ =45,505 M⁻¹cm⁻¹; *PfuRh* ϵ =31,400 M⁻¹cm⁻¹) with a nanodrop (Thermo Scientific).

After desalting, the rhomboids were injected into a gel filtration column (GE superose, 10/300 GL) for further purification and testing the oligomeric state with FPLC (BioRad, Biologic Duo flow). The gel filtration column was equilibrated with a 2-column volume (60mL) of 50 mM Tris-HCl buffer (pH 8.0) and 200 mM NaCl containing 0.08 % DDM and 1 mM TCEP. The flow rate was 0.5 mL/min. As a reference, 125 μ L of 36 mg/mL gel filtration standard proteins were injected. Before running, the column was equilibrated with the 2-column volume (60mL) of 50 mM sodium phosphate buffer (pH 7.2) and 150 mM NaCl. The elution was detected by absorbance at 280 nm.

For immunodetection, purified rhomboids were mixed with 20 μ L of the protein sample buffer with (w/v) 2% SDS, 1% (v/v) β -mercaptoethanol (β -ME) in 50 mM Tris-HCl buffer at pH 6.8. Samples were loaded on SDS-PAGE (4 to 20% gradient gel, Bio-Rad) at 180 V for 40 min. Western blotting was performed against the N-terminal or C-terminal His₆-tag epitope depends on the protein construct. The proteins separated on SDS-PAGE were transferred to a polyvinylidene difluoride (PVDF) membrane (Bio-Rad) at 100 V for 1 h. Rhomboids with the His tag were detected using rabbit monoclonal anti-His₅ 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).

3.3.3. Preparation of bicelles

A stock of 20% (w/v) DMPC (1,2-dimyristoyl-*sn*-glycero-3-phosphocholine) /CHAPS (3-[(3-Cholamidopropyl) dimethylammonio]-1-propanesulfonate) (lipid-to-detergent molar ratio, $q = 2.0$) bicelles were prepared by adding water into DMPC powder to hydrate lipids. 20% (w/v) CHAPS stock was added to obtain the desired q value. Bicelle samples were homogenized through three to five freeze-thaw cycles using liquid N₂ and a 37°C water bath. Bicelle stocks were stored at -20 °C.

3.3.4. Preparation of multiple rhomboid substrates

Plasmids of multiple substrates (Gurken, LacYTM2, Spitz and TatA) were gifts from Professor Kvido Strisvosky. The substrate constructs contain maltose-binding protein (MBP) in the N terminal region and thioredoxin (Trx) followed by a His₆ tag in the C-terminal region. The procedures for expression and purification were adopted from the previous literatures^{262–264}.

The TM substrates (Gurken, LacYTM2, and Spitz) were expressed in the *E. coli* KS55 (MC4100;glpGKO::Cat) strain in the presence of ampicillin and chloramphenicol. The TM substrate TatA was expressed in *E. coli* KS47(MC4100 WT +araD139) strain in the presence of ampicillin. 25mL of LB cultures grown overnight was inoculated in 1 L LB media at 37 °C until OD_{600nm} reached 0.8 and induced with 0.5 mM IPTG for 2–3 hr. The cell pellets were harvested using centrifugation and resuspended with 20 mL of 20mM HEPES buffer (pH 7.4) with 100 mM NaCl, 5 mM MgCl₂, 10% glycerol, and protease inhibitor (cOmplete Mini, Sigma-Aldrich). The cells were lysed with a pressure homogenizer (Avestin) for 3-4 times. The lysates were centrifuged by an ultracentrifuge at 10,000 rpm for 30 min (Beckman-Coulter). Pellets were isolated, resuspended with a tissue homogenizer (Fisher Scientific), and solubilized with 0.75% DDM. After ultracentrifugation of the solubilization products at 10,000 rpm for 30 min (Beckman-Coulter), the supernatant was incubated with 2 mL of Ni-NTA resin (Qiagen, 50% w/v) by slowly rotating at 4 °C for 1 h. Target proteins were eluted with 200 mM imidazole in 50mM HEPES buffer (pH 7.4) with 200 mM NaCl, 1 mM MgCl₂, 10% glycerol, and 0.1%

DDM. Imidazole was removed by a desalting column (Bio-Rad Econo Pac 10DG desalting column). The proteins were concentrated using an Amicon centrifugal filter unit (Millipore Sigma, 50 kDa MWCO) and finally quantified by the 660 nm detergent compatible quantitation kit (Bio-Rad).

3.3.5. Measuring the rhomboid activity with various transmembrane substrates

To test the proteolytic activities of the rhomboids (*EcGlpG*, *TmRh*, *TpRh*, and *PfuRh*) towards different substrates, 25 μ M of the rhomboids and a substrate were mixed in 5 mM DDM or 2% DMPC/CHAPS ($q = 2$) bicelles and incubated at room temperature for 6 h. Reactions were ceased by adding 15 μ L of the protein sample buffer with 2% (w/v) SDS, 1% (v/v) β -ME in 50 mM Tris-HCl at pH 6.8 into 15 μ L of the reaction mixture.

For immunodetection, the samples were loaded on SDS-PAGE (4 to 20% gradient gel, Bio-Rad) at 180 V for 40 min. Western blotting analysis was performed against the N-terminal MBP epitope. The proteins separated on SDS-PAGE were transferred to a nitrocellulose membrane (Bio-Rad) at 100 V for 1 h. Substrates with the MBP fusion were detected using mouse monoclonal anti-MBP 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).

3.3.6. Measuring the thermostability of rhomboids with thermal inactivation assay

The thermostability of delipidated rhomboids in DDM micelles were tested by heat induced aggregation and inactivation. 1 μ M rhomboid containing 5 mM DDM in 40 mM Tris-HCl (pH 8.5) with 200 mM NaCl and 0.25 mM TCEP was incubated at the temperatures from 25°C to 90°C with an interval of 5°C using the heating program in thermocycler (Bio-Rad). The samples were heated at a rate of 4°C/min to a designated temperature for 5 min and cooled back to 15°C for another 15 min. All samples were transferred to a 96-well plate. Light scattering is measured as optical density (OD) at 320nm on a SpectraMax M5e plate reader (Molecular Device) was measured at room temperature to detect the degree of aggregation. The transition temperatures of heat inactivation were obtained from fitting the normalized OD with a logistic function

$$y = \frac{1}{1 + e^{-k(x-x_0)}} \quad (\text{Eq 3})$$

Here x is the temperature, k is the steepness of the curve, x_0 represents the temperature at midpoint of the function (i.e., transition temperature) and y is the normalized optical density.

To measure the content of the native proteins after incubation at each temperature, the activity of rhomboids was monitored using the water-soluble model substrate, casein-BODIPY (EnzCheck™ protease assay kit, Invitrogen) at the molar ratio of 2:5 (rhomboid: substrate). Time-dependent dequenching of BODIPY fluorescence was monitored at 37°C using a SpectraMax M5e plate reader (Molecular Device) with the excitation and emission wavelengths of 480 nm and 530 nm, respectively. The initial slope of the time-dependent fluorescence change was interpreted as activity.

3.3.7. Engineering of the TM5 flanking loops in GlpG with substitution, deletion and insertion mutations

For substitution, site-directed mutagenesis on the L4 and L5 loops was performed using QuickChange Kit (Agilent). Insertion and deletion mutations modifying the lengths of L4 and L5 were performed using Q5 Site-directed mutagenesis Kit (NEB).

3.3.8. Determining GlpG activity with engineered L4 and L5

Detail procedures for the preparation of the substrate NBD-labeled SN-LacY™₂ are described in Chapter 2. GlpG activity was measured at the GlpG:substrate molar ratio of 1:10. Time-dependent changes of NBD fluorescence were monitored at 37°C using a SpectraMax M5e plate reader with the excitation and emission wavelengths of 485 nm and 535 nm, respectively. GlpG activity was also monitored with the water-soluble substrate casein-BODIPY (EnzCheck™ protease assay kit, Invitrogen) at the GlpG:substrate molar ratio of 1:5. Time-dependent changes of BODIPY fluorescence were monitored at 37°C using a SpectraMax M5e plate reader with the excitation and emission wavelengths of 480 nm and 530 nm, respectively. GlpG activity was finally probed with the peptide substrate MMPS-024 (CPC scientific) at the GlpG:substrate molar ratio of 1:20²²¹. Time-dependent changes of mca fluorescence were monitored at 37°C using a SpectraMax M5e plate reader with the excitation and emission

wavelengths of 320 nm and 430 nm, respectively. For all substrates, GlpG activity was represented by the initial slope of fluorescence change over time

3.4. Results

3.4.1. Purification of thermophilic rhomboids

With several rounds of optimization for the expression and purification (**Figure 3.2.a**), the final yields for thermophilic rhomboids were approximately 0.16 mg/L for *TmRh* (MW =28.1 kDa), 0.40 mg/L for *TpRh* (MW =25.2kDa), and 0.23 mg/L for *PfuRh* (MW =23.7kDa), still lower than the ~0.8mg/L of *EcGlpG*. The SDS-PAGE for purification products shows a limited purity for thermophilic rhomboids as evidenced by multiple bands on the gels (**Figure 3.1.b**). The identification of rhomboids was further confirmed through the immunodetection against His₆-tag. Western-blotting results (**Figure 3.1.c**) overall agreed with SDS-PAGE results but revealed the partial dimer formation of *EcGlpG* and oligomer formation of *TpRh*. Noticeably for *TmRh*, under the monomer band, the lower molecular weight bands were observed, suggesting the possibility of auto-proteolysis. The faint band of *PfuRh* on Western Blotting might suggest a tail truncation in the N-terminal tail region which contains His₆-tag as the corresponding band from Coomassie staining was denser than other thermophilic rhomboids.

For further purification and analysis, size exclusion chromatography (SEC) was performed (**Figure 3.2.d-g**). The broad, merged elution peaks of *EcGlpG* whose positions spanned those of the standard peaks in the range from 44kD to 670 kD, indicate the co-existence of monomers and oligomers at a high injection concentration ([*EcGlpG*] = 65 μ M). Purified thermophilic rhomboids were run on SEC at comparable concentrations (40 to 80 μ M). Elution peaks of *TmRh* also showed the existence of oligomers, while monomers are likely to be a dominant species in the merged peaks. In contrast, the single, broad elution peak of *TpRh* at ~150 kD (monomeric *EcGlpG* at 60 to 80 kD) indicated dominant formation of oligomers. It is possible that the loading concentration is so high that oligomer formation was favored. Thus, I collected the lower molecular weight fractions of *TpRh*, diluted them to a lower concentration, loaded them again on the SEC column. Those fractions are largely eluted as monomers, indicating that the oligomer formation depends on the protein concentration. The single, monodisperse elution peak of *PfuRh* indicates that the dominant species is monomer.

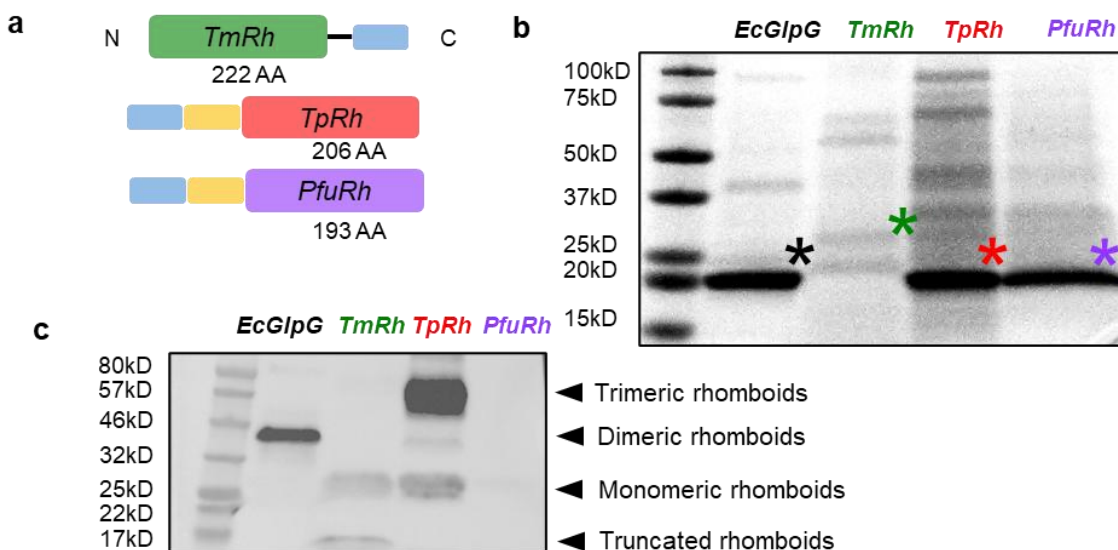
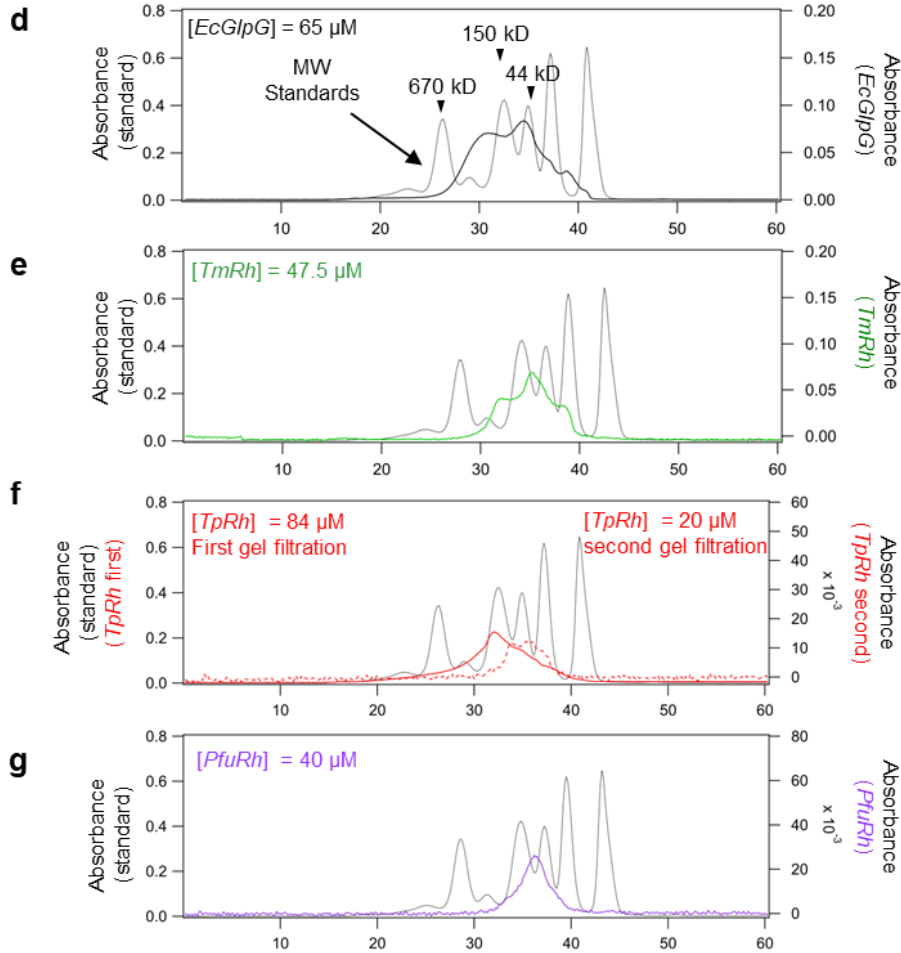


Figure 3.2. Expression construct and identification of rhomboid proteases.

(a) construct designs of thermophilic rhomboid proteases. Each blue block represents His₆-tag and each yellow block represents a thrombin cleavage site. (b) SDS-PAGE of purified rhomboids. The asterisks indicate monomeric rhomboids. The molecular weights for the rhomboid proteases (including the purification tag and cleavage site in the expression vector) are: *EcGlpG*: 23.4 kDa, *TmRh*: 28.1 kDa, *TpRh*: 25.2 kDa, and *PfuRh*: 23.7 kDa (c) Western blotting of the rhomboids using an anti-His₅-tag epitope primary antibody. (d-g) Gel-filtration chromatography of the rhomboids superimposed with the gel filtration protein standards (black dotted line). For the standard peaks, the molecular weights were annotated. The aggregation number per DDM micelle is ~150, the molecular weight of a DDM molecule is 0.5106 kDa. The predicted molecular weight of the DDM micelle-monomeric *EcGlpG* complex is $150 \times 0.5106 \text{ kD} + 23.4 \text{ kDa} = 99.6 \text{ kD}$. The molecular weight of the DDM micelle-monomeric *TmRh* complex is $150 \times 0.5106 \text{ kD} + 28.1 \text{ kDa} = 104.7 \text{ kD}$. The molecular weight of the DDM micelle-monomeric *TpRh* complex is $150 \times 0.5106 \text{ kD} + 25.2 \text{ kDa} = 101.8 \text{ kDa}$. The molecular weight of the DDM micelle-monomeric *PfuRh* complex is $150 \times 0.5106 \text{ kDa} + 23.7 \text{ kDa} = 100.3 \text{ kDa}$. There are two gel filtration runs for *TpRh*: the first run was for concentrated *TpRh* (solid red line), second run (dashed red line) was for the monomeric fractions from the first run.

Figure 3.2. (cont'd)



3.4.2. Thermophilic rhomboid showing various activities towards multiple given transmembrane substrates

We then tested proteolytic activities towards different transmembrane substrates including Gurken, LacYTM2, Spitz, and Tata (**Figure 3.3.a**). The chimeric substrate substrate was developed by the Strisovsky group¹⁷⁰ containing the N-terminal maltose-binding protein (MBP, 43 kDa) fused to the TM substrate (3-4 kDa) followed by thioredoxin (12 kDa) and His₆-tags. Among them, Gurken and Spitz are two natural substrates of *Drosophila* Rhomboid-1²⁶⁵. The naturally driven but not bona fide substrate LacYTM2 (LYTM2 hereafter) was derived from the second TM segment of the lactose permease from *E.coli* and proved to be efficiently proteolyzed by *EcGlpG*²³⁶. The natural substrate Tata is cleaved by the rhomboid protease AarA to mediate quorum-sensing in *Providencia stuartii*²⁶². The overall sequences of the four substrates are divergent but all have transmembrane helix destabilizing residues (i.e., Gly, Pro,

Gln, Ser). The previously recognized scissile bond of the four substrates is exposed to the extracellular side by several residues (**Figure 3.3.a**). Due to the molecular weight difference between the N and C terminal fusion tags flanking the TM substrate region, the full length and cleaved bands can be easily resolved on SDS-PAGE or the other separation tool such as size exclusion chromatography.

The proteolytic assays are performed in both detergent micelle (DDM) and neutral lipid bicelles (1 w/v-% DMPC:CHAPS, and $q = 2.0$) by incubating the substrate for 6 hours at room temperature. Overall, the SDS-PAGE and western blotting (**Figure 3.3.b-c**) demonstrated that all four substrates can be cleaved by the rhomboids in micelles with various efficiencies. For TatA and Gurken, the three rhomboids besides *TmRh* showed a comparable cleavage efficiency. The proteolytic efficiencies of *EcGlpG*, *TpRh* and *PfuRh* towards Spitz were relatively low compared to *TmRh*. *EcGlpG* and *TmRh* can proteolyze LYTM2 efficiently while the other two archaeal rhomboids showed a limited proteolytic efficiency. Spitz seems to be the poorest substrate for *EcGlpG*, *TpRh* and *PfuRh* showing the cleavage efficiencies of <10%.

Overall, *TpRh* and *PfuRh* have relatively low enzymatic activity in micelle, while *TmRh* showed the highest overall proteolytic efficiency. By the actions of *TmRh*, the SDS-PAGE and Western blotting results for the intact full-length substrates mostly disappeared. It is also interesting that for substrate TatA, multiple cleaved bands at lower molecular weight (i.e., between 15 and 20 kDa) other than the prevalent cleaved band (i.e., with the N-terminal MBP fusion tag around 50kDa) are observed for all four rhomboid proteases, suggesting that there could be more than one scissile bond in TatA.

In neutral bicelles (**Figure 3.4.**), the proteolytic efficiencies of all rhomboid proteases changed compared to those in micelles. *EcGlpG* was highly efficient in proteolyzing LYTM2. However, the proteolytic efficiencies of *TmRh* in bicelles dramatically decreased towards all substrates, especially towards Spitz. Similarly, the archaeal rhomboids showed reduced proteolytic activity towards Gurken and Spitz.

Building on the initial survey on the cleavage efficiency of the four rhomboids towards various TM substrates, we further compared the proteolytic activities of the thermophilic rhomboids relative to mesophilic *EcGlpG* in a more quantitative way.

Towards this goal, we employed a high-throughput, fluorescence-based assay, which was previously developed by our group. In this assay, staphylococcal nuclease (SN) is fused to the N terminus of the substrate LYTM2 and the environment-sensitive fluorophore NBD is labeled at the engineered cysteine residue, which is located at the fifth residue upstream from the scissile bond (**Figure 3.5.a**). Upon proteolysis, fluorophore is transferred from the hydrophobic environment to the aqueous phase exhibiting a decrease in fluorescence intensity. By monitoring the proteolysis-induced fluorescence changes over time, the activities of proteases can be quantified for direct comparison. In parallel, the activity towards the water-soluble substrate casein-BODIPY was tested. The water-soluble substrate is considered to approach the active site of the rhomboids from the aqueous phase (i.e., through the opening of the L5 cap). On the other hand, the TM substrates are expected to approach the active site instead diffusing from the hydrophobic core of the membrane (i.e., through the opening of the TM5 gate).

Towards the TM substrate LYTM2 (**Figure 3.5.c**), *TpRh* and *PfuRh* showed lower activity, ~20% of *EcGlpG*, while *TmRh* showed a higher activity. This result aligned well with the SDS-PAGE and Western blotting results (**Figure 3.3.b-c**). All three thermophilic rhomboids showed significantly higher proteolytic activity towards the water-soluble casein (**Figure 3.4.c**) compared to *EcGlpG*. *TpRh* and *PfuRh* activities are ~three-fold higher while *TmRh* showed almost ~20-fold relative to *EcGlpG*.

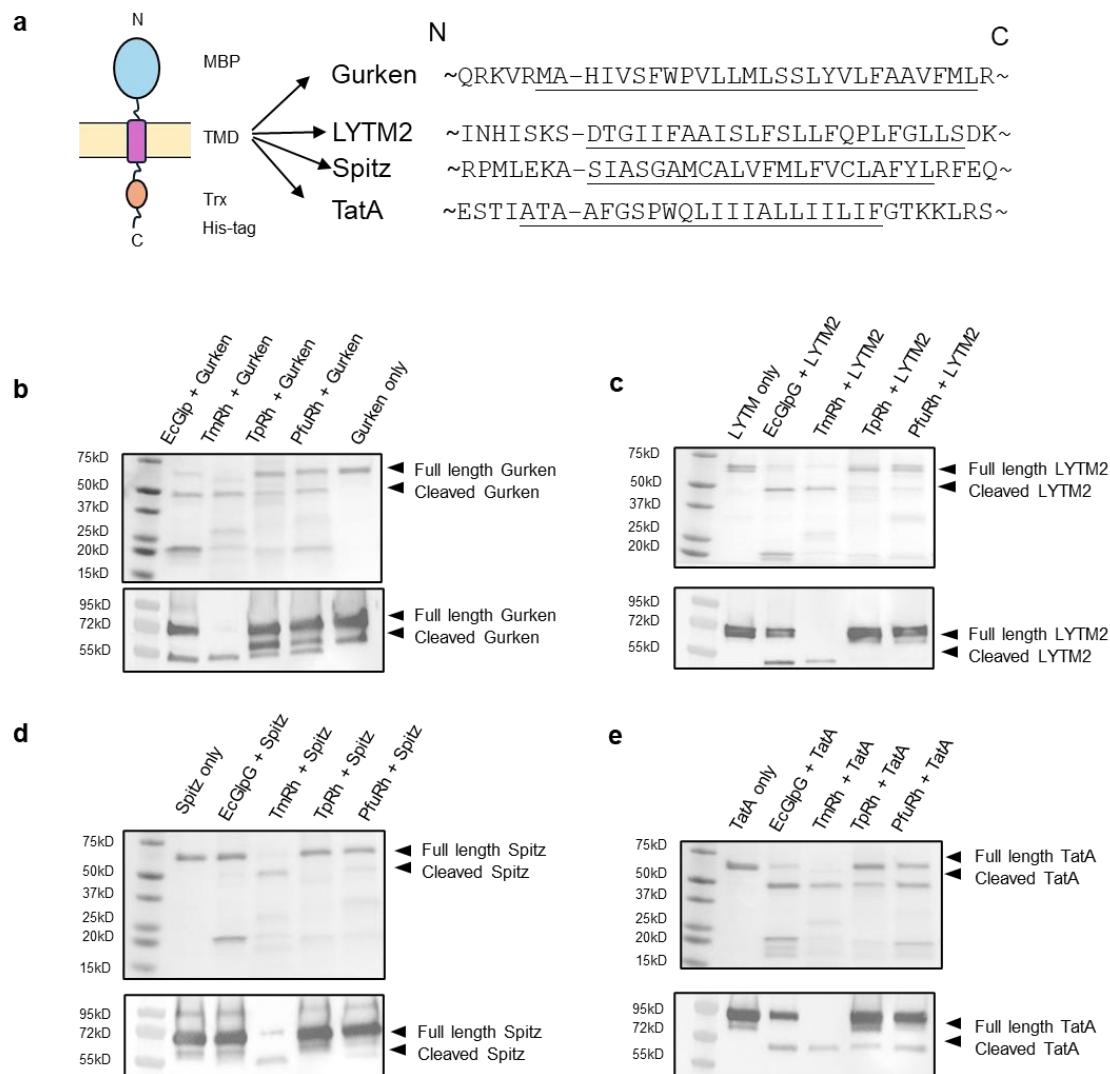


Figure 3.3. Proteolytic activity of the rhomboid proteases toward various transmembrane substrates in micelles. (a) The constructs of the TM substrates (magenta) with the N-terminal maltose binding protein (blue) and the C-terminal thioredoxin (orange). The sequence of each substrate with the scissile bond (dash) commonly cleaved by the bacterial rhomboids *EcGlpG*, AarA from *P. stuartii* and YgpG from *B. subtilis*. The predicted TM segments are underlined²⁶⁶. The molecular weights of the substrates (the full-length constructs): Gurken 69.8kD, LYTM2 67.7kD, Spitz 69.6kD, and TatA 69.2kD. SDS-PAGE (top) and western blotting (bottom) results of Gurken proteolysis **(b)**, LYTM2 proteolysis **(c)**, Spitz proteolysis **(d)**, and TatA proteolysis **(e)** by the rhomboids. For Western blotting, anti-MBP primary antibody was used.

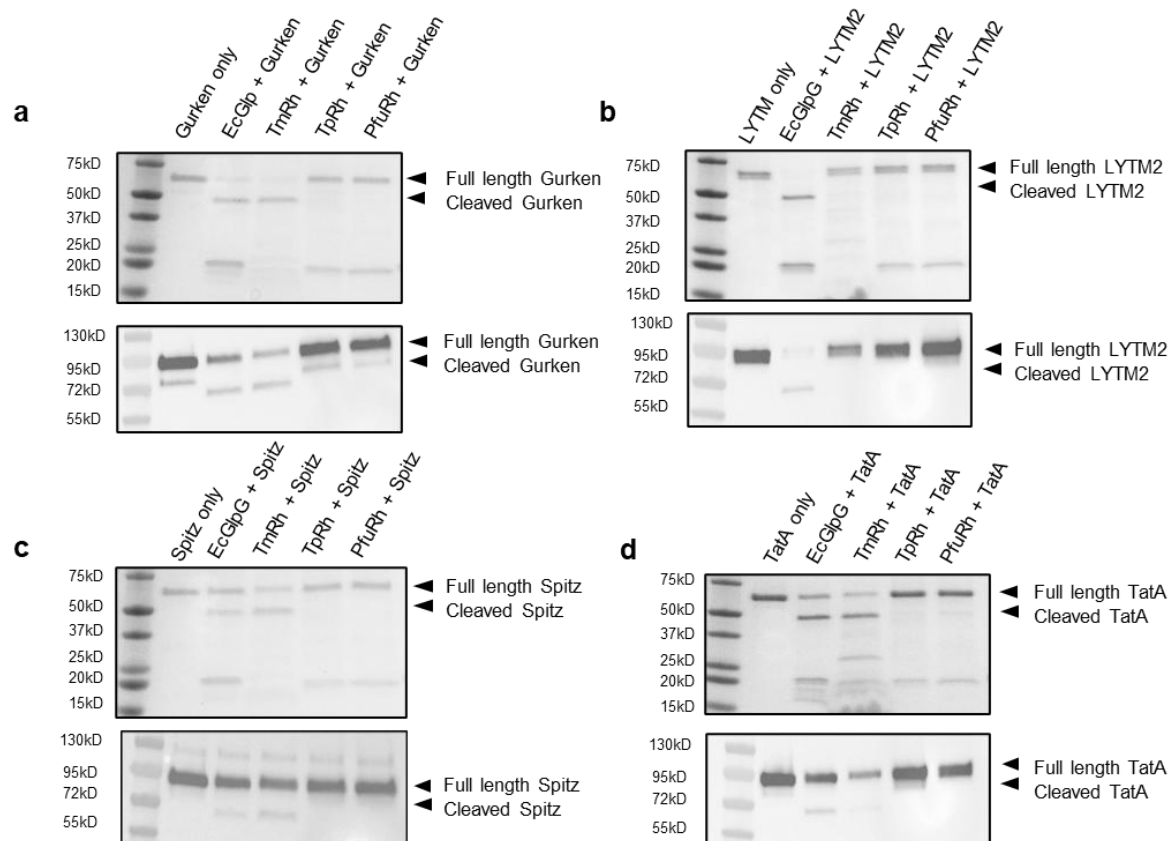


Figure 3.4. Proteolytic activity of the rhomboid proteases toward various transmembrane substrates in bicelles. SDS-PAGE (top) and Western blotting (bottom) results of Gurken proteolysis (**a**), LYTM2 proteolysis (**b**), Spitz proteolysis (**c**), and TatA proteolysis (**d**) by the rhomboids. Anti-MBP antibody was used as a primary antibody for Western blotting.

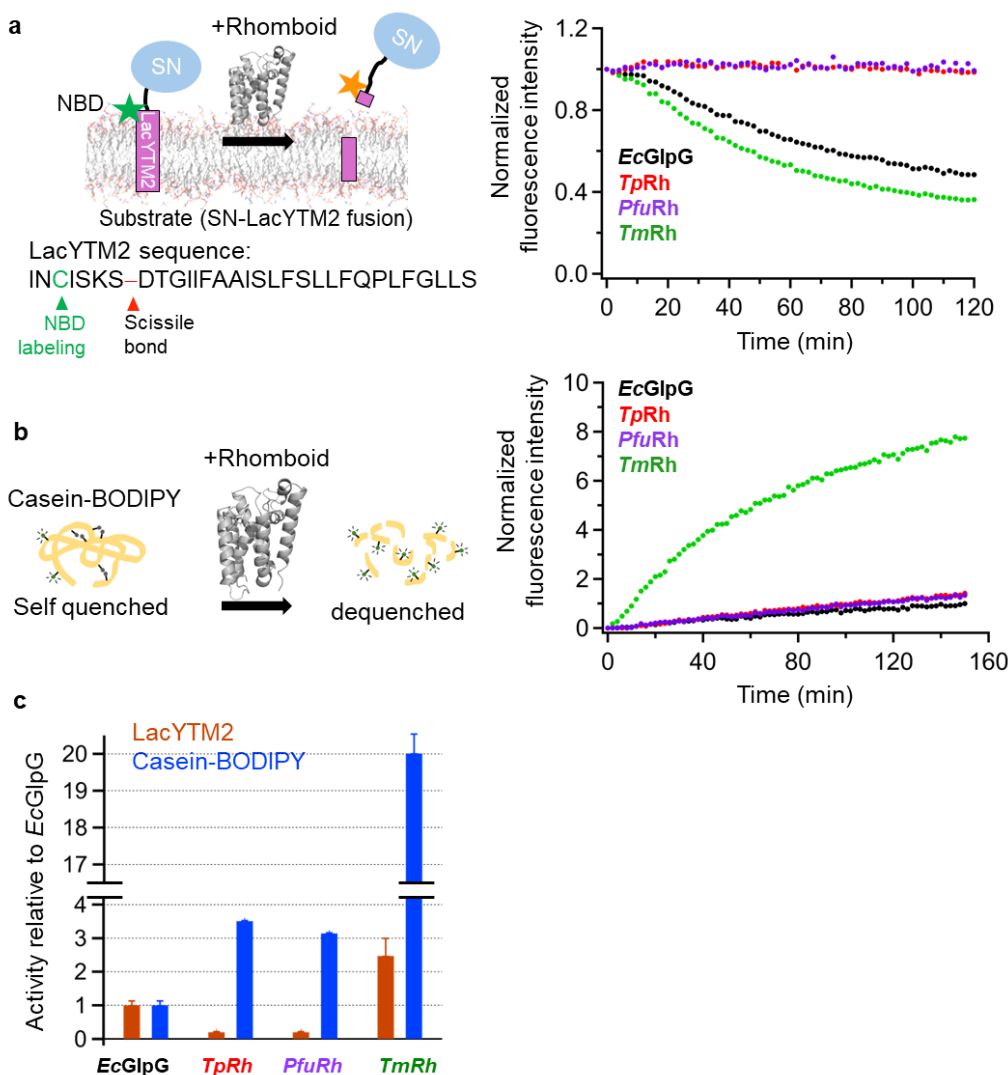


Figure. 3.5. Quantitative proteolysis assays for the mesophilic and thermophilic rhomboids toward the TM and water-soluble model substrates. (a) Scheme of the assay measuring the proteolytic activity of rhomboid proteases with a TM substrate such as SN-LYTM2 (SN: staphylococcal nuclease fusion tag; LacYTM2: the second TM segment of *E. coli* lactose permease). SN-LYTM2 is labeled with the environment-sensitive fluorophore NBD on the five-residue upstream of the scissile bond. **(b)** Scheme of the assay measuring the proteolytic activity of rhomboid proteases with a water-soluble model substrate such as casein-BODIPY. Casein is conjugated with the self-quenching fluorophore BODIPY conjugated to multiple Lys residues in casein. **(c)** Proteolytic activity of a given rhomboid relative to that of *EcGlpG* (mean \pm s.d., $N = 3$) towards LYTM2 and casein-BODIPY.

3.4.3. Thermostability of delipidated rhomboids in micelles

To test the functionality and stability of the thermophilic rhomboids in DDM micelles, I measured their thermostability by measuring the degree of heat-induced protein aggregation and inactivation (**Figure 3.6.a**). In general, heat induces denaturation and aggregation of membrane proteins. Aggregated proteins scatter more incident lights into random directions than unaggregated proteins. Therefore, the amount of light transmitted to the detector is reduced and the degree of reduction of transmitted light is proportional to the amount of aggregation. Therefore, we use optical density (OD), the logarithm of incident light intensity over transmitted light intensity, as the fraction of aggregated/denatured protein to quantitatively measure the formation of aggregation (**Figure 3.6.b-c**). In parallel, as a more sensitive tool for measuring thermostability, I performed a thermal inactivation assay using casein-BODIPY as a substrate (**Figure 3.6.d**). The changes in OD at 320 nm and activity as a function of temperature were fitted to the two-state model (**Figure 3.6.c**). I note that the transition from the folded to the aggregated state is irreversible, not representing the folding-unfolding equilibrium. Thus, the fitted transition temperature (T_m) represents the temperature corresponding to a half maximum of the total OD or activity change not informing of any thermodynamic meanings. Still, the T_m obtained in this way can be used as a reasonable quantitative measure of thermostability of the rhomboids.

Overall, the two thermal denaturation assays (aggregation vs activity) agreed with each other (**Figure 3.6.b** and **3.6.d**). Surprisingly, I found that *TpRh* and *PfuRh* from thermophilic archaea started to aggregate at temperatures much lower than their optimal growth temperatures in the dilapidated micellar environments (**Figure 3.6.c**). The T_m of *TpRh* was ~65 °C, 15 °C lower than the optimal growth temperature of *T. profundus* (~80 °C). For *PfuRh*, the T_m was also ~15 °C lower than the optimal growth temperature of *P. furiosus* (~90 °C). Interestingly, for *TmRh* the T_m was comparable to the optimal growth temperature (80 to 90 °C). In contrast, *EcGlpG* was highly thermostable with the T_m of 77 °C, much higher than the optimal living temperature of 37 °C. My T_m value for *EcGlpG* was around 5 °C higher than the literature value⁵¹.

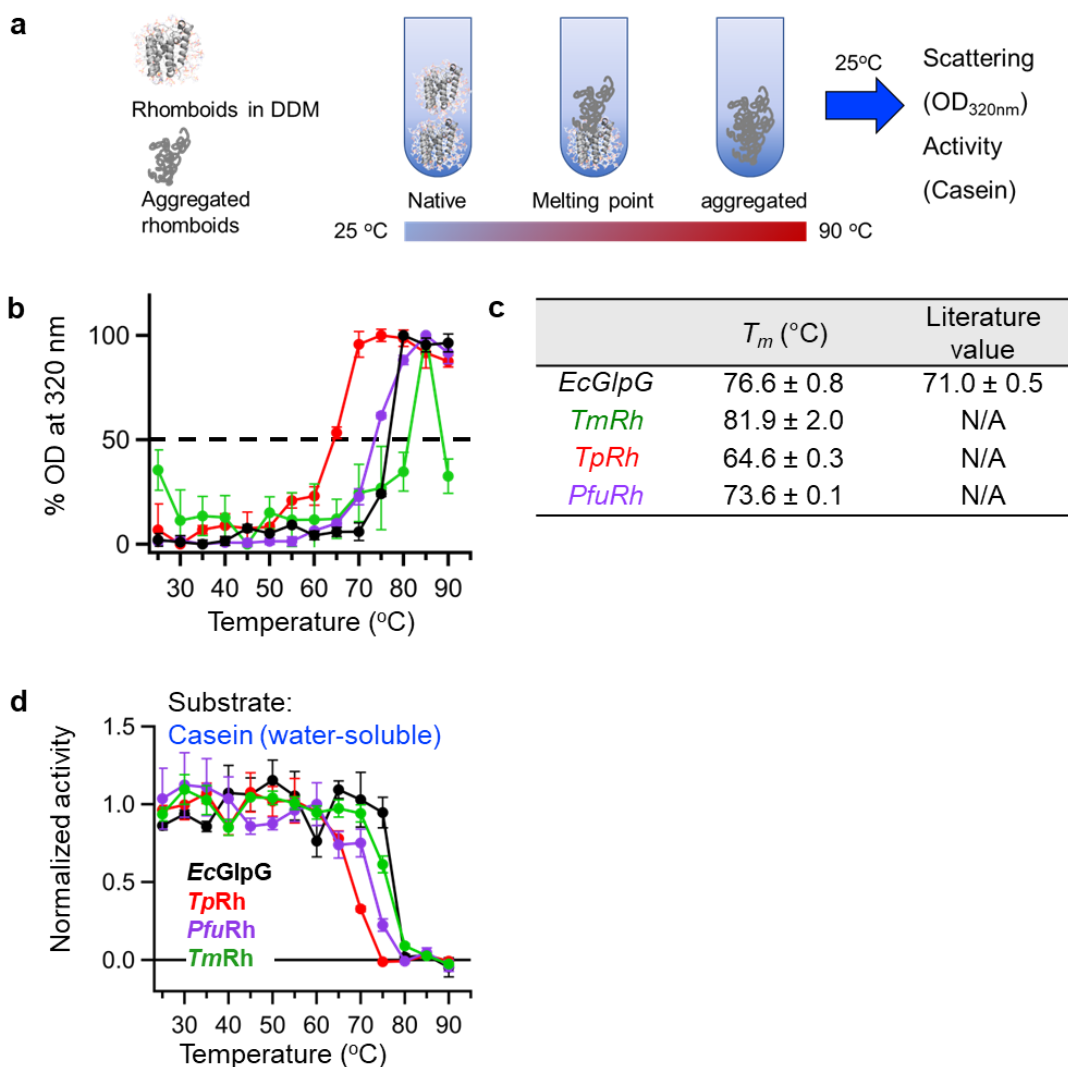


Figure 3.6. Thermostability of the rhomboids in DDM micelles. (a) Schemes for measuring heat-induced denaturation/aggregation and inactivation. **(b)** Thermal denaturation assay by measuring the temperature-dependent irreversible aggregation using light scattering at 320 nm. **(c)** The apparent melting temperatures, T_m (mean \pm s.d., $N = 3$) of the rhomboids with the literature value for *EcGlpG*⁵¹. **(d)** Thermal denaturation assay by measuring the temperature-dependent inactivation towards the water-soluble substrate casein-BODIPY (mean \pm s.d., $N = 3$).

3.4.4. In-silico structural analysis of thermophilic rhomboids

I further studied what is the molecular origin of various proteolytic activities of thermophilic rhomboids. Although there are currently no experimental structures of rhomboid proteases from thermophiles, I obtained the structural models using AlphFold²⁶⁷ and IntFold²⁶⁸ (**Figure 3.7.a**). A multiple sequence alignment (**Figure 3.7.b**) predicts the catalytic dyads for *TmRh* (Ser135/His198), *TpRh* (S128/His174) and *PfuRh* (Ser117/His163). The predicted structures show that these Ser-His pairs are buried under the membrane plane in the periplasmic side similar to the rhomboid proteases with known structures^{165,269}. Structural studies on *EcGlpG* suggest the critical role of TM5 helix as the lateral gate for the entrance of TM substrates^{165,173}. 7-residue long L4 (connecting TM4 and TM5 in the cytosolic side) and 8-residue long L5 (connecting TM5 and TM6 in the periplasmic side) are likely to be responsible for the gating movement of TM5. In *EcGlpG*, L5 is also known to play a critical role in capping the catalytic dyad such that the opening of L5 allows the access of the scissile bond in a substrate to the dyad^{175,176}. L5 harbors two bulky methionine residues (Met247 and Met249) above active site. In contrast, the predicted *TmRh* structure suggested a longer L4 loop (22-residue long) and a shorter L5 cap (7-residues long) compared to those of *EcGlpG*. The predicted structures of the other two thermophilic rhomboids both suggested shorter L4 and L5 (**Figure 3.7.a**).

Taking together, we hypothesize that the lengths of L4 and L5 are critical to the proteolytic activity of rhomboid proteases. Longer L4 and L5 would allow the TM5 lateral gate to open wider and increase its flexibility, facilitating the access of both TM and water-soluble substrates to the catalytic dyad. This is what we observe for *TmRh* with longer TM4 and TM5 than those of *EcGlpG*, leading to an increase in activity for both LYTM2 and casein-BODIPY (**Figure 3.5.**). In contrast, shorter L4 and L5 would allow the TM5 gate to open narrower and decrease its flexibility. The tightening of TM5 will hinder the lateral access of TM substrates to the catalytic dyad, but shorter L5 would shrink the cap size, thus exposing a larger portion of the catalytic dyad for water-soluble substrates. This is what we observe for *TpRh* and *PfuRh* with the shorter TM4 and TM5 (**Figure 3.5.**), leading to a decrease in activity for LYTM2 and an increase in activity for casein-BODIPY.

3.4.5. The effect of loop lengths on the activity of rhomboids towards a TM substrate

My hypothesis is that the lengths of the flanking loops of the gating helix TM5 control the activity of rhomboid proteases. Using *EcGlpG* as a template for testing the hypothesis, I first investigated whether the residues in L4 and L5 of *EcGlpG* are engaged in any tertiary interactions with their surrounding regions and if any, whether those interactions affect proteolytic activity. Accordingly, I engineered most of the residues in L4 and L5 into Gly to disrupt potential tertiary contacts while maintaining the original loop lengths. Then, I tested the role of substituted residues in activity since these Gly substitutions are expected to increase the flexibility of the loops, affecting activity. Next, after establishing the constructs with all possible WT residues in L4 and L5 replaced by Gly, I changed the lengths of L4 and L5 by inserting or deleting multiple Gly residues and measured proteolytic activities of these loop variants towards TM and water-soluble substrates.

The impacts of the series of mutations on *EcGlpG* proteolytic activities were first evaluated for the TM substrate LYTM2 that laterally access the catalytic dyad (**Figure 3.8.**). The activities of the variants were measured in both DDM micelles and DMPC:CHAPS bicelles. Overall, *EcGlpG* and variants were more active in bicelles than in micelle.

L4 is the cytosolic loop in the *trans* side of the catalytic dyad (residues 220-227). The replacement of the N-terminal Gln220 with Gly in L4 (L4_1G) moderately increased the activity by ~2 fold in micelles without a noticeable change in bicelles. The subsequent replacements of the N-terminal WT residues with Gly's did not change the activity up to L4_4G (the four WT residues in the N-terminal side of L4 were replaced by four Gly's) in both micelles and bicelles. However, the proteolytic activity is compromised when either of Ile223 or Leu225 (either or both of them) was mutated to glycine (see L4_4G1I and L4_4G1L in **Figure 3.8.a**). The crystal structure of *EcGlpG* indicates that Ile223 and Leu225 in L4 point inward forming tertiary contacts with Arg168 on TM helix 2 and Val211 on TM helix 4 while the other residues are facing outward without contacting with other TM helices. Thus, the tertiary contacts involving Ile223 and Leu225 are critical to the activity towards a TM substrate.

When three additional Gly residues were inserted into the N-terminal side of L4 (L4_3G+5G111L in **Figure 3.8.a**), a significant reduction in proteolytic activity was observed in micelles but not in bicelles. When three additional Gly residues were inserted into the C-terminal side of L4 (L4_5G111L+3G in **Figure 3.8.a**), activity was restored to the WT level in both micelles and bicelles. The addition of six extra glycine residues to L4 (that is, three Gly residues to each end of L4, L4_3G+5G111L+3G in **Figure 3.8.a**) reduces activity to a substantially lower level than WT *EcGlpG*.

Taken together, this result indicates that a few tertiary contacts between L4 and the rest of the protein (i.e., Ile223 and Leu225 with TM2 and TM4) are critical to the maintenance of activity. That is, the opening and closing of the TM5 gate mediated by the L4 loop are not random motions. This result agrees to our previous result showing that the cavity creating mutations on the other side of the TM2-TM5 substrate binding site significantly reduce activity¹⁴⁸. However, the flexibility of L4 still seems to affect activity. The insertion of additional Gly residues to the C-terminal side of L4 substantially increases activity in bicelles when the critical Ile223 and Leu225 are kept in their positions (compare the activities of L4-5G111L and L4-5G111L+3G).

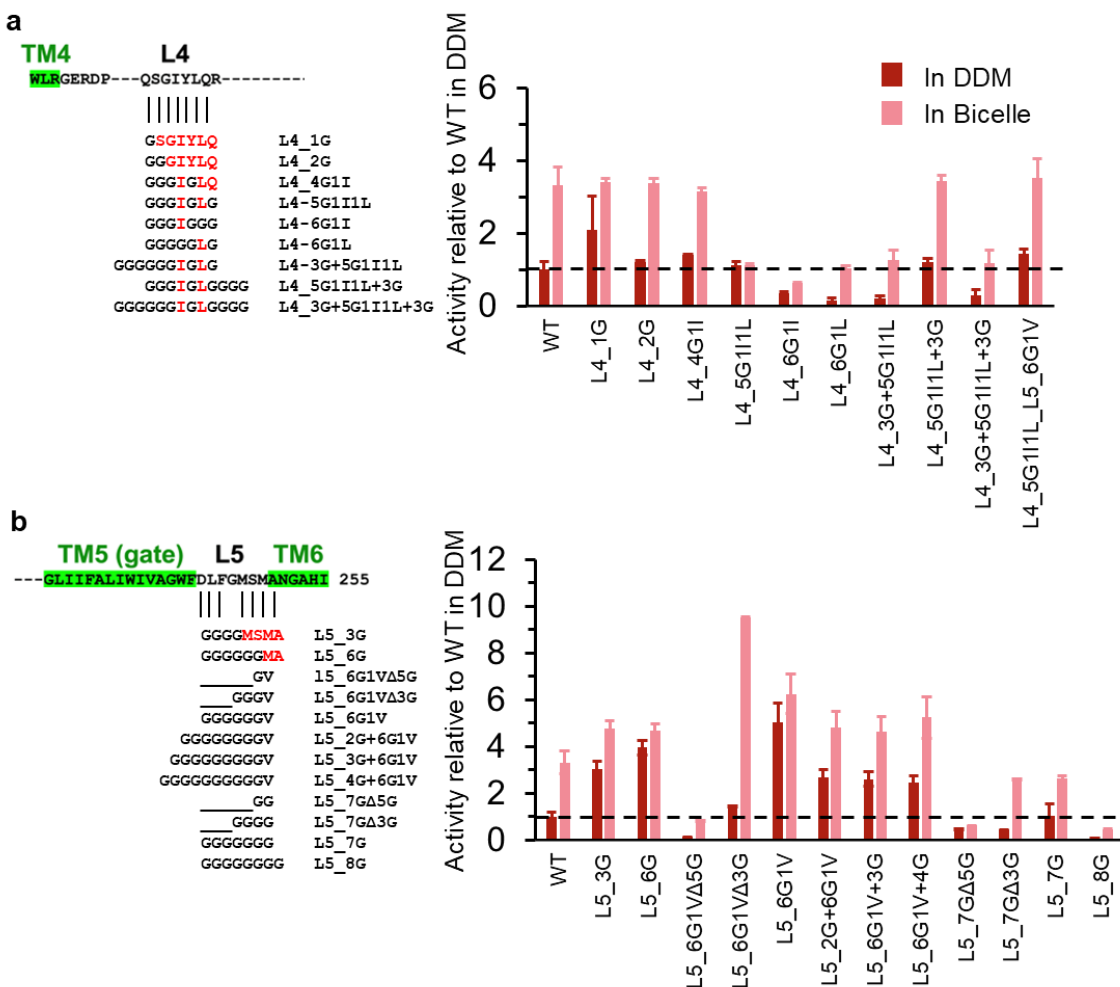


Figure 3.8. Proteolytic activities of the variants with modified L4 or L5 of *Ec*GlpG toward the TM substrate SN-LYTM2 in DDM micelles and DMPC:CHAPS bicelles (1 w/v-%, $q = 2$). (a) The effects of modifying L4 on activity. The modification schemes and notations of the variants (left) and the proteolytic activities of the variants (mean \pm s.d., $N = 3$) (right) are shown. (b) The effects of modifying L5 on activity. The modification schemes and notations of the variants (left) and the proteolytic activities of the variants (mean \pm s.d., $N = 3$) (right) are shown.

L5 is the periplasmic loop (residues 243-249) capping the catalytic dyad. When the residues in the N-terminal side of L5 were consecutively replaced by Gly with a constant loop length (L5_3G to L5_6G in **Figure 3.8.b**), the proteolytic activity towards LYTM2 was consistently enhanced in both micelles and bicelles. However, when Met249 and Ala250 were replaced with Gly (L5_7G in **Figure 3.8.b**), the WT activity

level was restored. This result aligns well with the Urban group's result showing that replacing WT L5 with all-Gly L5 moderately reduces the reaction towards another TM substrate TatA¹⁷⁶. They interpreted this result in the way that the detailed sequence of L5 is not important for the opening of the L5 cap. On the other hand, our result demonstrates that the disruption of tertiary contacts mediated by several WT residues substantially improves activity (L5_3G to L5_6G in **Figure 3.8.b**), indicating an importance of WT tertiary interactions between L5 and the rest of the protein to activity. Interestingly, we noticed when the upstream six residues were mutated into Gly and Met249 was switched to Val (L5_6G1V in **Figure 3.8.b**), proteolytic activity further improved by 4-5 fold compared to WT in micelles, and 1.2-1.5 fold in bicelles. The deletion of three to five Gly residues from the N-terminal side of the L5 cap significantly inactivated the protein relative to the hyperactive variant L5_6G1V (L5_6G1VΔ5G, L5_7GΔ5G and L5_7GΔ3G in **Figure 3.8.b**). However, the smaller deletion 6G1VΔ3G was only inactivating in micelles but highly activating in bicelles, which was difficult to interpret. Interestingly, the installation of two to four Gly residues that increased the length of L5 (L5_2G+6G1V, L5_3G+6G1V, and L5_4G+6G1V in **Figure 3.8.b**) showed an activity level comparable to the hyperactive L5_6G1V.

In summary, the disruption of the WT tertiary contacts involving the 6 residues in the N-terminal side of L5 greatly improved activity, suggesting the importance of this part in modulating the activity of *EcGlpG*. The tightening of the L5 cap by deletion of several residues overall reduced the activity towards the TM substrate LYTM2 while the further loosening of the cap maintained the hyperactivity. Overall, these results on L5 still support our hypothesis that the length and flexibility of the L5 cap controls the activity of rhomboid proteases.

3.4.6. The effect of loop lengths on the activity of rhomboids towards a water-soluble substrate

Next, I tested proteolytic activity of the loop-modified variants towards the water-soluble substrates that directly access the catalytic dyad through the L5 cap. I used two distinct substrates which have different requirements for their proteolysis: 1) the large generic substrate casein-BODIPY (20-25 kD) that would require a large opening of L5 and 2) a small self-quenched peptide substrate MMPS-024 whose susceptibility to proteolysis

has been optimized. Although the water-soluble substrate casein-BODIPY can effectively probe activity⁴⁸, the fluorescent signals of casein-BODIPY in bicelles were too poor to be analyzed. Thus, for casein-BODIPY, we only measured activity in micelles while for MMPS-024, in both micelles and bicelles.

When the multiple residues in the N-terminal side of L4 were replaced by Gly with the loop length, Ile223, and Leu225 kept intact (the variants L4_1G, L4_2G, L4_4G1I, and L4_5G1I1L), the activity towards casein-BODIPY was greatly improved by 2- to 4-fold relative to WT *EcGlpG* (**Figure 3.9.a**). This is a distinct feature from the trend for the TM substrate LYTM2, where those substitutions maintained the WT-level activity. Consistent with the trend for the TM substrate, the replacement of either Ile223 or Leu225 (participating in the tertiary contacts with TM2 and TM4) with Gly in the background of L4_5G1I1L substantially reduced activity from the hyperactive level to the WT-level. The addition of three additional Gly residues to the N- or C-terminal side of L4 (the variants L4_3G+5G1I1L) further reduced activity by 40%-70% compared to WT. It is interesting that the modification of L4 which is on the opposite side to the catalytic dyad, dramatically changes the activity towards the water-soluble substrate casein, whose access to the catalytic dyad. Except for Ile223 or Leu225 that are deeply inserted into the core of the GlpG structure, the replacements of other WT residues with Gly in L4 are expected to disrupt potential interactions with the protein itself or with the cytosolic interfacial region of the membrane. This disruption is likely to increase the mobility of the TM5 gate whose effect is transmitted to the periplasmic loop L5, leading to the enhancement of activity.

In the L5 cap region, the trend in activity variation for the water-soluble substrate agreed with the trend for the TM substrate (**Figure 3.9.b** vs **Figure 3.8.b**). That is, the replacement of the residues in the N-terminal side of L5 maintaining the constant loop length (L5_3G to L5_6G in **Figure 3.9.b**) enhanced the proteolytic activity towards casein-BODIPY by 6- to 7-fold. Both the deletion (i.e., the L5 shortening) and addition (i.e., the L5 extending) of Gly residues increased activity by 3- to 4-fold. However, the dramatic shortening of L5 by five residues (L5_6G1V Δ 5G with only 2 residues in L5) almost abolished activity probably due to the distortion of the active site structure. The shortening of the L5 cap (L5_6G1V Δ 3G) will lead to an incomplete shielding of the

catalytic dyad towards the aqueous phase and the extending L5 will increase the gating motions. Both of these cases can lead to an enhancement in activity.

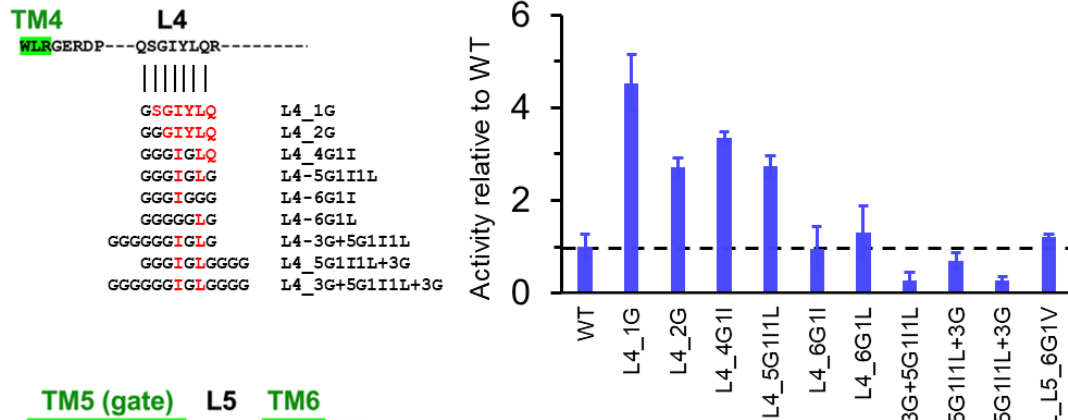
Finally, we probed the activity of the loop-modified variants towards the small water-soluble substrate MMPS-0024, a 10 amino acid peptide with the fluorophore 7-methoxycoumarin on the amino-terminus and the quencher dinitrophenol on the carboxyl side. The sequence of this peptide is highly optimized rendering it highly susceptible to proteolysis by *EcGlpG*²²¹. Due to the small size, MMPS-0024 would easily penetrate deep into the catalytic dyad without a large opening of the L5 cap. Thus, the activity result is expected to reflect the “intactness” of the active site conformation for proteolysis, not adequate for probing the movements of the gating helix TM5 and the flanking loops L4 and L5. Indeed, the variations of activities towards MMPS-0024 were overall smaller (varying from 50% to 300% relative to WT activity) than those towards the larger casein-BODIPY (varying from 30% to 800% relative to WT activity). Also, the variants displayed larger activity in bicelles than in micelles probably due to the enhancement of stability and cooperativity by the lipid environment¹⁴⁷.

For the modifications in L4, the trend in activity for MMPS-0024 was very similar to that for casein-BODIPY (**Figure 3.10.a** vs **Figure 3.9.a**), indicating that the activity variation observed for casein-BODIPY reflects an allosteric effect of the modifications in L4 on the catalytic dyad. However, we note that the activity changes within a narrower range (50% to 200% relative to WT), indicating that these modifications did not dramatically disrupt the intactness of the active site. In contrast, for the modifications in L5, the trend in activity for MMPS-0024 was different from that for casein-BODIPY (**Figure 3.10.b** vs **Figure 3.9.b**) in that the activity towards MMPS-0024 varied in a much narrower range than that towards casein-BODIPY. This result indicates that the perturbations on the intact active site were not severe for most of the loop modifications on L5 and it is reasonable to interpret the activity changes towards the casein-BODIPY based on the changes in the gating motions and loop dynamics.

In summary, I obtained a preliminary understanding about the correlation between the loop lengths and the activity of rhomboid proteases: 1) The native interactions in L4 and L5 are critical for overall proteolytic activities; 2) The disruption of the WT residue interactions in the flanking loops while maintaining a few structurally

important WT residues overall substantially enhances activity, indicating the importance of the loop flexibility to activity; 3) Increasing or decreasing the loop lengths can increase or decrease activity depending on the type of substrate, overall supporting our hypothesis (i.e., the control of rhomboid activity by the lengths of the flanking loops of the gating helix). Nonetheless, I acknowledge that a further activity analysis is required to confirm this hypothesis.

a



b

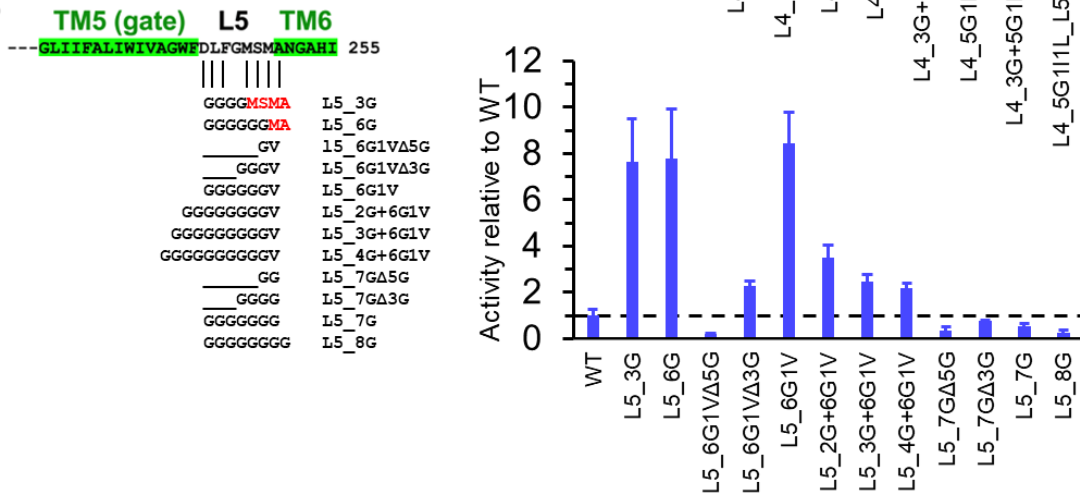


Figure 3.9. Proteolytic activities of the variants with modified L4 or L5 of *EcGlpG* toward the water-soluble substrate casein-BODIPY in DDM micelles. (a) The effects of modifying L4 on activity. The modification schemes and notations of the variants (left) and the proteolytic activities of the variants (mean \pm s.d., $N = 3$) (right) are shown. **(b)** The effects of modifying L5 on activity. The modification schemes and notations of the variants (left) and the proteolytic activities of the variants (mean \pm s.d., $N = 3$) (right) are shown.

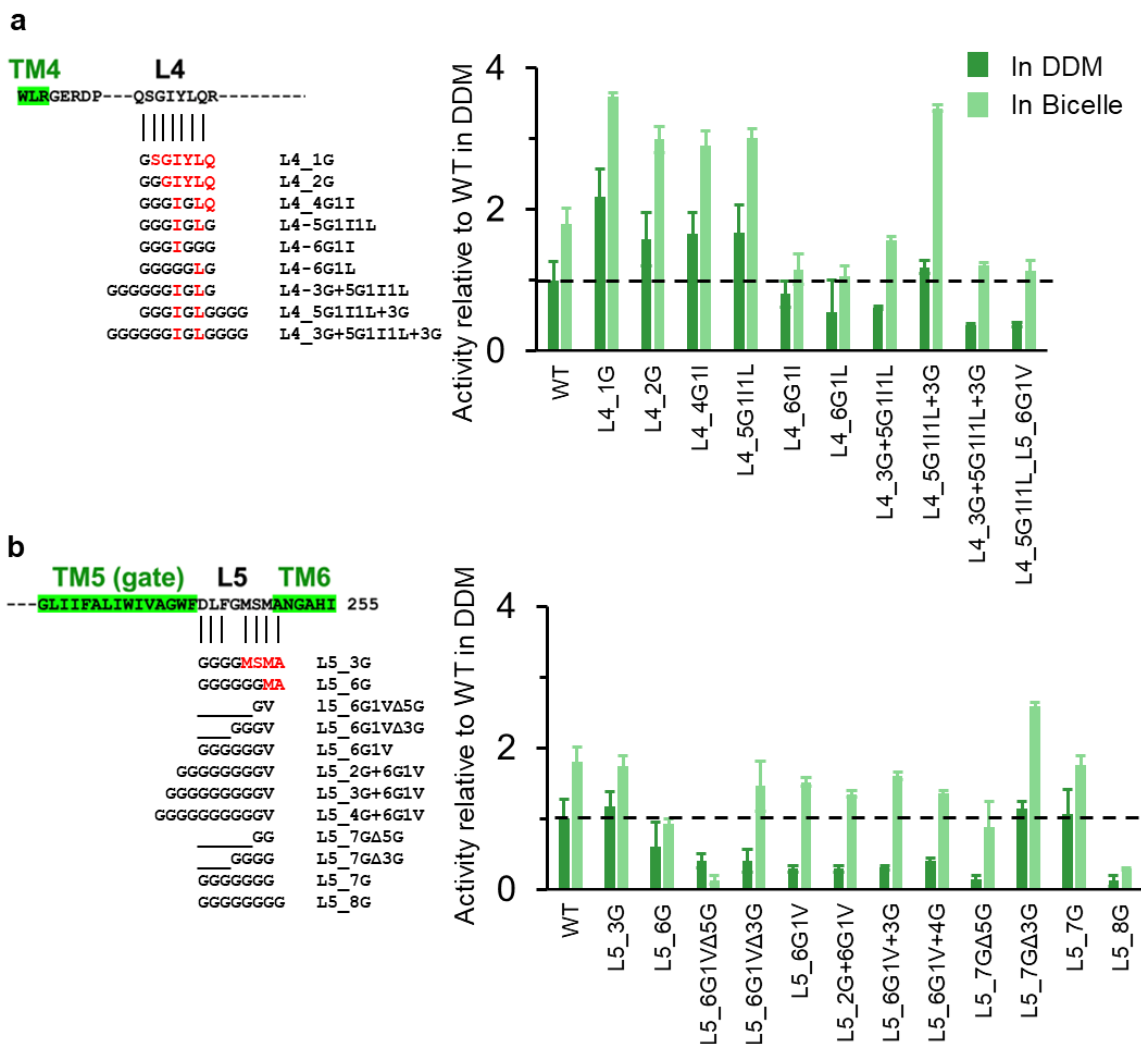


Figure 3.10. Proteolytic activities of the variants with modified L4 or L5 of *EcGlpG* toward the water-soluble substrate MMPS-0024 in DDM micelles and DMPC:CHAPS bicelles (1 w/v-%, $q = 2$). (a) The effects of modifying L4 on activity. The modification schemes and notations of the variants (left) and the proteolytic activities of the variants (mean \pm s.d., $N = 3$) (right) are shown. (b) The effects of modifying L5 on activity. The modification schemes and notations of the variants (left) and the proteolytic activities of the variants (mean \pm s.d., $N = 3$) (right) are shown.

3.5. Discussion

My results provide insights into how helical membrane protein in thermophiles acquire thermostability in thermophiles and how the motions of the gating helix TM5 and its flanking loops modulate the proteolytic activity of rhomboid proteases. I found that thermophilic rhomboids are not as thermostable as expected in a delipidated micellar environment. Surprisingly, thermophilic rhomboids exhibit even lower stability than their mesophilic homologue. The large differences between the measured inactivation transition temperatures of thermophilic rhomboids and the optimal growth temperatures of their native organisms imply the potential role of native lipids in stabilizing thermophilic membrane proteins.

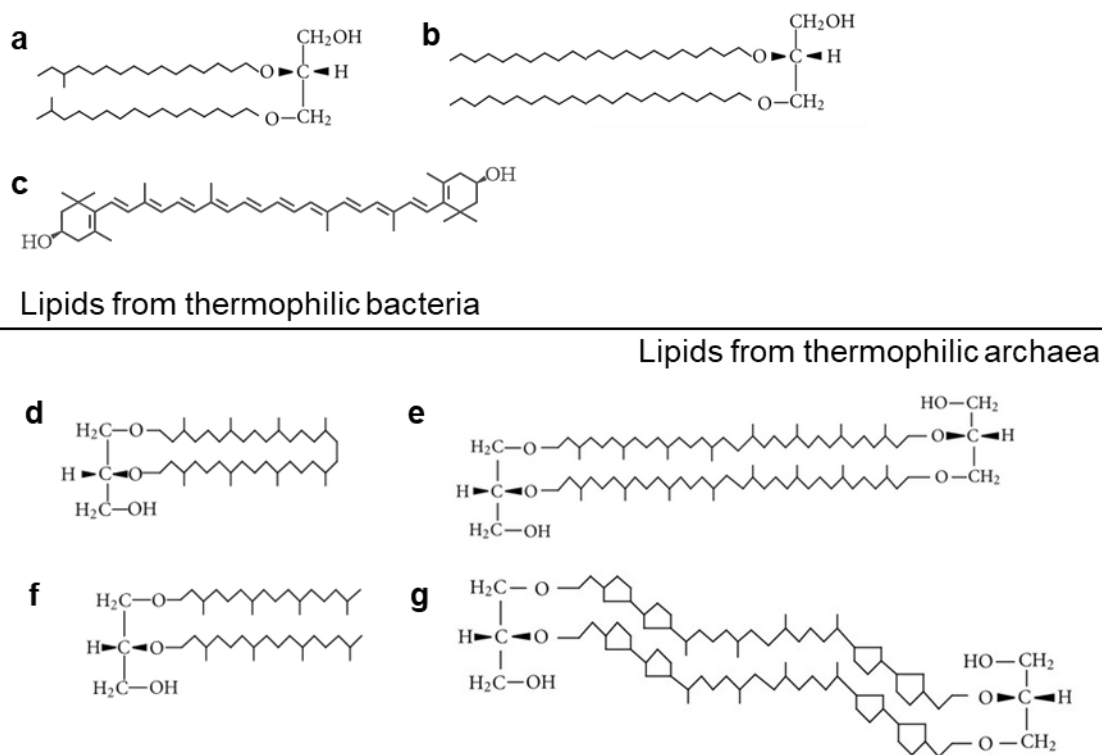


Figure 3.11. Lipids in thermophilic bacteria and archaea^{250,270}. (a) Bacterial dither lipid with iso-branched fatty acid chain. (b) Bacterial dither lipid with long, saturated fatty acid chains. (c) Polar carotenoid. (d) Macrocyclic archaeol. (e) Archaeal tetraether lipid (caldarchaeol). (f) Diphytanylglycerol. (g) Cyclopentane-containing caldarchaeol. Reprint permission from Springer Nature (license number: 6043801195643).

Lipids in thermophilic bacteria or archaea typically contain an ether backbone, which is linked to various types of fatty acid chains (**Figure. 3.11.**). These types of lipids enhance thermal adaptation by enhanced lipid-lipid packing and reduced permeability of the membrane. As the delipidation of the thermophilic rhomboids leads to lower thermostability than that of the mesophilic rhomboid *EcGlpG*, it is expected that their native lipid bilayer environments would be required for the acquisition of their thermostability. Interestingly, a study of our group has shown that the enhancement of lipid-lipid packing interaction improves the thermodynamic stability of *EcGlpG*. Therefore, testing the effect of thermophile-derived lipids on membrane protein stability will be important to the fundamental understanding of membrane protein stability. However, the interactions between membrane protein and thermophilic lipids have received insufficient attention. In addition to the global membrane properties of the lipid membranes such as the lipid-lipid packing and lateral pressure profile, measuring the binding properties of thermophile lipids to the protein including binding enthalpy, entropy, and free energy will also be a valuable effort to the understanding of thermostability of membrane proteins. Excitingly, Laganowsky and coworkers established the tools for studying how phospholipid binding modulates membrane protein stability and membrane protein-protein interactions using nano-electrospray ionization ion-mobility/mass spectrometry (nESI-IM-MS) ^{138–140}. Therefore, the nESI-IM-MS incorporating temperature-control apparatus would serve as a suitable platform for future investigations.

Previously, Baker and Urban measured the T_m of four different mesophilic rhomboids with low sequence identity (i.e., the rhomboids from *Escherichia coli*, *Haemophilus influenzae*, *Vibrio cholerae*, and *Providencia stuartii*)⁵¹. Their T_m vary widely by ~25 °C with the highest *EcGlpG* and lowest. The authors hypothesized that the origin of this discrepancy in thermostability stems from the variations in nonconservative residues. They further tested the hypothesis by incorporating one of the stabilizing residues Leu207 from *EcGlpG* into the corresponding residue Val122 in *HiGlpG*, leading to an elevation of T_m ⁵¹. The sequence alignments and predicted structures of thermophilic rhomboids indicate that the “hypothetically stabilizing” residue corresponding to Leu207 in *EcGlpG* are not conserved. Testing such potentially

stabilizing residues in thermophilic rhomboids under delipidated conditions would be an interesting future task.

In the Urban group's comprehensive stability and activity study of 151 mutants in *EcGlpG*, L4 is believed to be neither functional nor structurally important since the Ala mutations on some of the L4 residues have almost no change on thermostability and proteolytic activity⁵¹. In contrast, our systematic study demonstrates that a series of Gly mutations without disturbing the local loop-helix interactions enhance activity. Furthermore, I found that the changes in the length of L4 can also modulate activity. We speculate the length of L4 could be further increased to observe a salient impact.

A more recent structural and functional study on *EcGlpG* suggests that rather than being a flexible cap exposing the catalytic dyad to the substrate, it is more likely that L5 serves as a navigator guiding a substrate to the catalytic dyad¹⁷⁶. The high-resolution, time-resolved snapshots of the *EcGlpG*-substrate complex show a critical role of Met249 on L5 by interacting with the upstream residues from the scissile bond (i.e., P2, P3 and P4)¹⁷⁶. Nevertheless, our result shows that increasing the mobility of the residues in the upstream of Met249 and substituting it into Val249 further enhance the proteolytic activity, reemphasizing the "cap" role of L5. On the other hand, I found that the shortening of L5 significantly compromise activity, which contradict to the "cap" property. It is possible that the decrease in L5 length induces the distortion of the contacts between the gating helix TM5 and the rest of the protein as well as the shrinkage of the hydrophilic cavity harboring the active site. Thus, it is also necessary to test how the activity of *EcGlpG* is affected when the lengths of L4 and L5 are varied at the same time. Interestingly, my result shows that the simultaneous insertion of multiple Gly residues to L4 and L5 (i.e., L4_5G1I1L_L5_6G1V in **Figure 3.8.a** and **Figure 3.9.a**) restores activity to the WT level. Therefore, the L4_5G1I1L_L5_6G1V mutant can serve as the starting template.

CHAPTER 4: Concluding remark and outlook

In this research, I aimed to elucidate two important problems regarding the stability of helical membrane proteins with the universally conserved rhomboid proteases as a study model: 1) What are the energetic consequences of burying ionizable residue pairs in the core of membrane protein? 2) What is the origin of the thermostability of membrane proteins in thermophilic organisms?

In chapter 2, I addressed the first question by quantitatively evaluating the effects of internal ionizable residues on the kinetic and thermodynamic stability of *E. coli* GlpG in detergent micelles and lipid bicelles. Overall, the burial of ionizable residues destabilize GlpG to a varying extent. Double mutant cycle analysis shows that favorable interactions are formed between the acidic (e.g., Glu) and basic (e.g., Lys) ionizable residues that can form a salt bridge. Thermodynamic stability measurements indicate that the lipid environment provided by disc-shape like bicelles enhances the accommodation of ionizable residues in the protein interior compared to spherical detergent micelles ($\Delta\Delta G^{\circ}_{N-D, WT-Mut} = 1.5$ to 3.5 kcal/mol in micelles vs 1.2 to 1.9 kcal/mol in bicelles) while the interactions between the acidic and basic residues were more favorable in micelles. On the other hand, the kinetic stability measurements in general show no significant stabilizing effect from lipids for most of the buried ionizable residues. Substituting the acidic residues (Glu) with a non-ionizable proxy (Gln) yields a similar degree of destabilization, which suggests that the ionizable residues exist as a neutral form and their acidic-basic residue pairs are likely engaged with each other by H-bonds rather than a salt-bridge.

Although our current results cover multiple aspects regarding how an introduction of ionizable residues into hydrophobic internal cavities impact the stability of helical membrane proteins, there are remaining questions that can be further explored in the future: what would be the energetic consequences of polarity reversal compared to the current ionizable pairs (i.e., M100K/T178E to M100E/T178K and L207K/V165E to L207E/V165K)? Hwang and Warshal's computational study on aspartate aminotransferase revealed that switching the native positively charged Arg292 into the negatively charged Asp in the active site causes drastic destabilization of 6.8 kcal/mol for substrate binding²⁷¹. They suggested that the protein microenvironment around the ionizable residues is preorganized for accommodating a certain type of ionizable

residue while switching polarity is highly destabilizing and unfavorable²⁷¹. What if the microenvironment is not predefined for the ionizable residues, in other words, will the artificially buried ionizable residue pair be reversed without major energetic constraints? Garcia-Moreno and coworkers addressed this question with the study model of water-soluble protein staphylococcal nuclease (SNase). They studied the thermodynamic stability, ionization states, and structures of the artificial charge-reversal pairs V23E/L36K (EK) and V23K/L36E(KE)²¹⁷. They found that the charge-neutral KE pair is less stable than the charged EK pair and a structural reorganization occurs to accommodate the highly destabilizing KE pair^{215,217}. Such rearrangement includes the penetration of water molecules into the ionizable residue pair. In a membrane protein, I speculate that the structural reorganization for accommodating such polarity reversal will be even limited because the bilayer environment shielding the protein may not be able to facilitate a conformational rearrangement through solvation by water molecules. Instead, a distortion of the native helical packing may occur to accommodate the unfavorable polarity reversal.

Another important future direction is to obtain the atomic level investigation of how the buried ionizable residues inside the low dielectric, insulated environment impacts the behavior of the helical membrane protein. Our group's extensive thermodynamic studies in different environments will provide useful guidelines for computational studies such as Molecular Dynamic (MD) simulations. Deng and Cui's simulation work on SNase^{215,217} shows that with the nonpolarized CHARMM36m force field, the simulated interaction free energy of an ion-pair is extremely destabilizing (i.e., $\Delta\Delta G^{\circ}_{U,WT-Mut} > 55$ kcal/mol vs. ~ 9 kcal/mol from experiment)²⁷². Interestingly, the use of the electronic polarization including 2019 CHARMM-Drude force field, which include greatly improves the prediction, implying an essential role of polarization of the low-dielectric constant environment in stabilizing a buried ion pair. Their simulation results further suggest a higher degree of water mobility and penetration to the internal charged pair compared to those to a charge-neutral pair²⁷².

While MD simulation is a powerful tool for obtaining the information on dynamical properties of water molecules, it is much harder to investigate water dynamics experimentally. One of the few strategies is *in situ* H₂O¹⁶/H₂O¹⁸ exchange Fourier

transform infrared spectroscopy (FTIR) that has been used to monitor the H-bonding of water molecules in the helical membrane protein bacteriorhodopsin²⁷³. The advantage of FTIR spectroscopy is that the measurement is performed at room temperature rather than under cryogenic conditions. In future, incorporating both experimental and *in silico* methods would further strengthen our understanding of the behaviors of buried ionizable pairs in helical membrane proteins.

In chapter 3, I tackled the second question by measuring the heat-induced denaturation and the proteolytic activity of various rhomboid proteases from thermophilic bacteria and archaea. Surprisingly, in detergent micelles where the native lipids from *E. coli* are extensively removed, the thermophilic rhomboids were similarly or less thermostable than *EcGlpG*⁵¹. The thermophilic rhomboids are denatured and inactivated at temperatures (i.e., 60-70°C) lower than their intrinsic living temperature (80-100°C). Thus, unlike water-soluble thermophilic proteins, the amino acid sequences of the thermophilic membrane proteins may not necessarily have evolved to achieve thermostability by their intraprotein interactions. Rather, it is highly possible that the lipid environment plays a key role in their thermostability.

Our preliminary activity assays demonstrate that thermophilic rhomboids that are believed to be an ancient form of rhomboid family proteins, can cleave four widely used TM substrates with divergent sequences (i.e., Spitz, LYTM2, Gurken and TatA)²³⁵ that have been used to test the activity of mesophilic rhomboids. Notably, *TmRh* exhibits high proteolytic activities towards all four substrates in micelles and might be able to cleavage TatA from multiple sites. Although previous studies indicate that the rhomboids originated from evolutionarily distant bacteria and eukarya can recognize the same substrate motif²³⁵, the sequence specificity of thermophilic rhomboids are still elusive. Moreover, it would be informative to screen more divergent substrates other than naturally derived substrates. Therefore, a synthetic peptide library and a high throughput quantitative analytic platform would achieve this goal. The O'Donoghue group successfully combined tandem mass tags with an established peptide cleavage assay to yield a quantitative Multiplex Substrate Profiling by Mass Spectrometry (qMSP-MS)²⁷⁴. Their qMSP-MS platform was validated with well-characterized water-soluble cysteine proteases against 275 unique peptide bonds in parallel²⁷⁴. Later they assayed

the sequence specificity of the rhomboid proteases, *Haemophilus influenzae* GlpG and *Providencia stuartii* AarA against 228 peptide substrates. They showed that *HiGlpG* exhibits a narrower sequence preference but higher catalytic activity than *PsAarA*²⁷⁴. With the wide substrate library and high throughput tool, we can further characterize the sequence specificity and catalytic efficiency of thermophilic rhomboids.

Although our current results under delipidated micellar conditions imply that lipids from thermophiles may play a role in stabilizing thermophilic membrane proteins, the potential role of stabilizing residue motifs cannot be excluded. To discover such stabilizing motifs and evaluate the strength of interaction, a “motif exchange” strategy can be developed. That is, a potential stabilizing motif such as aromatic clusters, disulfide bonds, and salt bridge networks that are unique in either thermophilic or mesophilic structure could be implanted into the other structure. For example, previous sequence and structural analyses of thermophilic proteins have demonstrated the abundance of aromatic residues (especially, Phe) and the presence of their clusters in 17 out of 24 thermophilic enzyme families^{199,201}. These aromatic clusters are usually found in relatively rigid regions on the protein surface. These aromatic clusters occupy the positions of Leu, Ser, or Ile in their mesophilic counterparts¹⁹⁹. It is possible aromatic clusters increase the rigidity of secondary structure enhancing thermostability at higher temperatures. Interestingly, the predicted structures of our thermophilic rhomboids suggest the existence of inter- or intra-helical aromatic clusters (**Fig. 4.1.**). Thus, I hypothesize that the aromatic clusters would contribute to the thermostability of membrane proteins. In the next step, the hypothetical “aromatic cluster motif” from thermophilic rhomboids could be engineered into the corresponding regions of mesophilic rhomboids, which would further stabilize them. The impacts of the exchanged motifs on stability and cooperativity can be measured with steric trapping in different hydrophobic medium (i.e., neutral detergent micelles and lipid bicelles). Double mutant cycle analysis can be used to further confirm if there are favorable interactions between the aromatic residues.

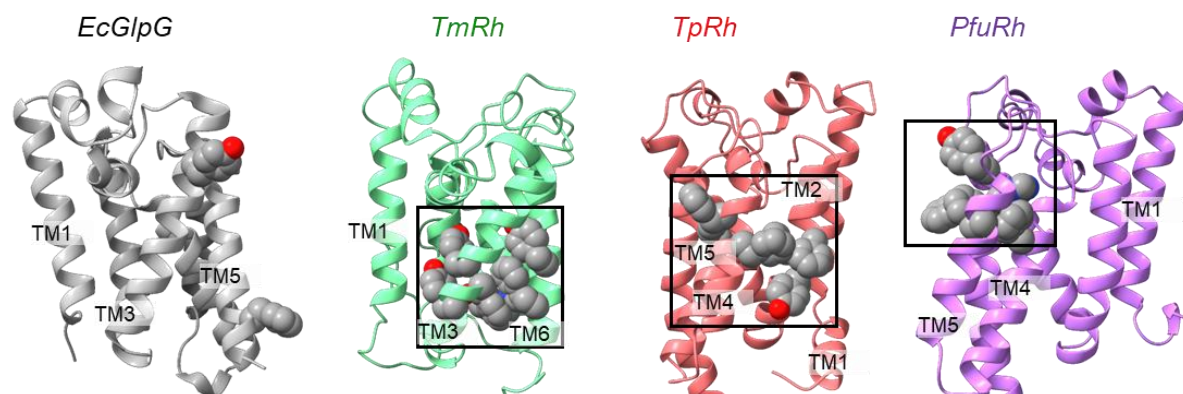


Figure 4.1. Predicted structures of mesophilic and thermophilic rhomboids.

EcGlpG with known structure (PDB ID: 3B45, resolution 1.9 Å) in which Important TM helices are labeled. The *TmRh* structure was predicted by IntFold²⁶⁸; The *TpRh* and *PfuRh* structures were predicted by Alpha fold²⁶⁷. The surface aromatic residues in *EcGlpG* are shown in spheres. The residues involved in the aromatic clusters in thermophilic rhomboids are shown as grey spheres. In *TmRh*: Tyr105, Tyr109, Tyr145, Phe108, Tyr204, and Trp208; In *TpRh*: Phe81, Phe82, Tyr85, Phe132, Phe158, and Phe160; in *PfuRh*: Phe121, Phe149, Tyr160, His163, and Phe164.

REFERENCES

1. Prusiner SB. Prions. *Proceedings of the National Academy of Sciences*. 1998;95(23):13363-13383. doi:10.1073/pnas.95.23.13363
2. Calabresi P, Mechelli A, Natale G, Volpicelli-Daley L, Di Lazzaro G, Ghiglieri V. Alpha-synuclein in Parkinson's disease and other synucleinopathies: from overt neurodegeneration back to early synaptic dysfunction. *Cell Death Dis*. 2023;14(3):176. doi:10.1038/s41419-023-05672-9
3. Trepte P, Strempel N, Wanker EE. Spontaneous self-assembly of pathogenic huntingtin exon 1 protein into amyloid structures. *Essays Biochem*. 2014;56:167-180. doi:10.1042/bse0560167
4. Mackenzie IRA, Bigio EH, Ince PG, et al. Pathological TDP-43 distinguishes sporadic amyotrophic lateral sclerosis from amyotrophic lateral sclerosis with SOD1 mutations. *Ann Neurol*. 2007;61(5):427-434. doi:10.1002/ana.21147
5. Soto C. Unfolding the role of protein misfolding in neurodegenerative diseases. *Nat Rev Neurosci*. 2003;4(1):49-60. doi:10.1038/nrn1007
6. Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC. A Three-Dimensional Model of the Myoglobin Molecule Obtained by X-Ray Analysis. *Nature*. 1958;181(4610):662-666. doi:10.1038/181662a0
7. Dill KA, MacCallum JL. The Protein-Folding Problem, 50 Years On. *Science* (1979). 2012;338(6110):1042-1046. doi:10.1126/science.1219021
8. Watson HC, Kendrew JC. The stereochemistry of the protein myoglobin. Published online May 19, 1976. doi:10.2210/pdb1mbn/pdb
9. Levinthal C. How to Fold Graciously. In: DeBrunner JTP, Munck E, eds. *Mossbauer Spectroscopy in Biological Systems: Proceedings of a Meeting Held at Allerton House*. University of Illinois Press; 1969:22-24.
10. Anfinsen CB, Haber E, Sela M, White FH. The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci U S A*. 1961;47(9):1309-1314. doi:10.1073/pnas.47.9.1309
11. Dill KA, Chan HS. From Levinthal to pathways to funnels. *Nat Struct Mol Biol*. 1997;4(1):10-19. doi:10.1038/nsb0197-10
12. Romero P, Obradovic Z, Li X, Garner EC, Brown CJ, Dunker AK. Sequence complexity of disordered protein. *Proteins*. 2001;42(1):38-48. doi:10.1002/1097-0134(20010101)42:1<38::aid-prot50>3.0.co;2-3

13. Riback JA, Bowman MA, Zmyslowski AM, et al. Innovative scattering analysis shows that hydrophobic disordered proteins are expanded in water. *Science* (1979). 2017;358(6360):238-241. doi:10.1126/science.aan5774
14. Oldfield CJ, Cheng Y, Cortese MS, Romero P, Uversky VN, Dunker AK. Coupled Folding and Binding with α -Helix-Forming Molecular Recognition Elements. *Biochemistry*. 2005;44(37):12454-12470. doi:10.1021/bi050736e
15. Oldfield CJ, Cheng Y, Cortese MS, Brown CJ, Uversky VN, Dunker AK. Comparing and Combining Predictors of Mostly Disordered Proteins. *Biochemistry*. 2005;44(6):1989-2000. doi:10.1021/bi047993o
16. Kannan S, Lane DP, Verma CS. Long range recognition and selection in IDPs: the interactions of the C-terminus of p53. *Sci Rep*. 2016;6(1):23750. doi:10.1038/srep23750
17. Wiener MC, White SH. Structure of a fluid dioleoylphosphatidylcholine bilayer determined by joint refinement of x-ray and neutron diffraction data. III. Complete structure. *Biophys J*. 1992;61(2):434-447. doi:10.1016/S0006-3495(92)81849-0
18. Petrache HI, Feller SE, Nagle JF. Determination of component volumes of lipid bilayers from simulations. *Biophys J*. 1997;72(5):2237-2242. doi:10.1016/S0006-3495(97)78867-2
19. White SH, Wimley WC. MEMBRANE PROTEIN FOLDING AND STABILITY: Physical Principles. *Annu Rev Biophys Biomol Struct*. 1999;28(1):319-365. doi:10.1146/annurev.biophys.28.1.319
20. Marsh D. Thermodynamics of phospholipid self-assembly. *Biophys J*. 2012;102(5):1079-1087. doi:10.1016/j.bpj.2012.01.049
21. Chandler D. Interfaces and the driving force of hydrophobic assembly. *Nature*. 2005;437(7059):640-647. doi:10.1038/nature04162
22. Beltrán-Heredia E, Tsai FC, Salinas-Almaguer S, Cao FJ, Bassereau P, Monroy F. Membrane curvature induces cardiolipin sorting. *Commun Biol*. 2019;2:225. doi:10.1038/s42003-019-0471-x
23. Wallin E, von Heijne G. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci*. 1998;7(4):1029-1038. doi:10.1002/pro.5560070420
24. Lindén M, Sens P, Phillips R. Entropic Tension in Crowded Membranes. *PLoS Comput Biol*. 2012;8(3):e1002431. doi:10.1371/journal.pcbi.1002431
25. Bowie JU. Helix packing in membrane proteins. *J Mol Biol*. 1997;272(5):780-789. doi:10.1006/JMBI.1997.1279

26. Wimley WC. Toward genomic identification of β -barrel membrane proteins: Composition and architecture of known structures. *Protein Science*. 2002;11(2):301-312. doi:10.1110/ps.29402
27. Bogdanov M, Dowhan W, Vitrac H. Lipids and topological rules governing membrane protein assembly. *Biochim Biophys Acta*. 2014;1843(8):1475-1488. doi:10.1016/j.bbamcr.2013.12.007
28. Heijne G. The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *EMBO J*. 1986;5(11):3021-3027. doi:10.1002/j.1460-2075.1986.tb04601.x
29. Kim H, Paul S, Jennity J, Inouye M. Reversible topology of a bifunctional transmembrane protein depends upon the charge balance around its transmembrane domain. *Mol Microbiol*. 1994;11(5):819-831. doi:10.1111/j.1365-2958.1994.tb00360.x
30. Harley CA, Holt JA, Turner R, Tipper DJ. Transmembrane Protein Insertion Orientation in Yeast Depends on the Charge Difference across Transmembrane Segments, Their Total Hydrophobicity, and Its Distribution. *Journal of Biological Chemistry*. 1998;273(38):24963-24971. doi:10.1074/jbc.273.38.24963
31. Almén MS, Nordström KJ V, Fredriksson R, Schiöth HB. Mapping the human membrane proteome: a majority of the human membrane proteins can be classified according to function and evolutionary origin. *BMC Biol*. 2009;7:50. doi:10.1186/1741-7007-7-50
32. Sakakura M, Hadziselimovic A, Wang Z, Schey KL, Sanders CR. Structural basis for the Trembler-J phenotype of Charcot-Marie-Tooth disease. *Structure*. 2011;19(8):1160-1169. doi:10.1016/j.str.2011.05.009
33. Huang H, Kuenze G, Smith JA, et al. Mechanisms of KCNQ1 channel dysfunction in long QT syndrome involving voltage sensor domain mutations. *Sci Adv*. 2018;4(3). doi:10.1126/sciadv.aar2631
34. Terragni B, Scalmani P, Franceschetti S, Cestè S, Mantegazza M. Post-translational dysfunctions in channelopathies of the nervous system. *Neuropharmacology*. 2018;132:31-42. doi:10.1016/j.neuropharm.2017.05.028
35. Niforou K, Cheimonidou C, Trougakos IP. Molecular chaperones and proteostasis regulation during redox imbalance. *Redox Biol*. 2014;2:323-332. doi:10.1016/j.redox.2014.01.017
36. Kučerka N, Liu Y, Chu N, Petrache HI, Tristram-Nagle S, Nagle JF. 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*. 2005;88(4):2626-2637. doi:10.1529/biophysj.104.056606

37. Plotkin SS, Onuchic JN. Understanding protein folding with energy landscape theory Part I: Basic concepts. *Q Rev Biophys.* 2002;35(2):111-167. doi:10.1017/S0033583502003761
38. Oliveberg M, Wolynes PG. The experimental survey of protein-folding energy landscapes. *Q Rev Biophys.* 2005;38(3):245-288. doi:10.1017/S0033583506004185
39. Jacob J, Krantz B, Dothager RS, Thiyagarajan P, Sosnick TR. Early Collapse is not an Obligate Step in Protein Folding. *J Mol Biol.* 2004;338(2):369-382. doi:10.1016/J.JMB.2004.02.065
40. Walters BT, Mayne L, Hinshaw JR, Sosnick TR, Englander SW. Folding of a large protein at high structural resolution. *Proceedings of the National Academy of Sciences.* 2013;110(47):18898-18903. doi:10.1073/pnas.1319482110
41. Fersht AR. Transition-state structure as a unifying basis in protein-folding mechanisms: Contact order, chain topology, stability, and the extended nucleus mechanism. *Proceedings of the National Academy of Sciences.* 2000;97(4):1525-1529. doi:10.1073/pnas.97.4.1525
42. Cecconi C, Shank EA, Bustamante C, Marqusee S. Direct Observation of the Three-State Folding of a Single Protein Molecule. *Science* (1979). 2005;309(5743):2057-2060. doi:10.1126/science.1116702
43. Muñoz V. Conformational Dynamics and Ensembles in Protein Folding. *Annu Rev Biophys Biomol Struct.* 2007;36(1):395-412. doi:10.1146/annurev.biophys.36.040306.132608
44. Bai Y, Sosnick TR, Mayne L, Englander SW. Protein Folding Intermediates: Native-State Hydrogen Exchange. *Science* (1979). 1995;269(5221):192-197. doi:10.1126/science.7618079
45. Moon CP, Fleming KG. Side-chain hydrophobicity scale derived from transmembrane protein folding into lipid bilayers. *Proceedings of the National Academy of Sciences.* 2011;108(25):10174-10177. doi:10.1073/pnas.1103979108
46. Huysmans GHM, Baldwin SA, Brockwell DJ, Radford SE. The transition state for folding of an outer membrane protein. *Proceedings of the National Academy of Sciences.* 2010;107(9):4099-4104. doi:10.1073/pnas.0911904107
47. Hong H, Tamm LK. Elastic coupling of integral membrane protein stability to lipid bilayer forces. *Proceedings of the National Academy of Sciences.* 2004;101(12):4065-4070. doi:10.1073/pnas.0400358101
48. Sanders MR, Findlay HE, Booth PJ. Lipid bilayer composition modulates the unfolding free energy of a knotted α -helical membrane protein. *Proc Natl Acad Sci U S A.* 2018;115(8):E1709-E1808. doi:10.1073/pnas.1714668115

49. Panigrahi R, Arutyunova E, Panwar P, Gimpl K, Keller S, Lemieux MJ. Reversible Unfolding of Rhomboid Intramembrane Proteases. *Biophys J*. 2016;110(6):1379-1390. doi:10.1016/J.BPJ.2016.01.032
50. Findlay HE, Rutherford NG, Henderson PJF, Booth PJ. Unfolding free energy of a two-domain transmembrane sugar transport protein. *Proceedings of the National Academy of Sciences*. 2010;107(43):18451-18456. doi:10.1073/pnas.1005729107
51. Baker RP, Urban S. Architectural and thermodynamic principles underlying intramembrane protease function. *Nat Chem Biol*. 2012;8(9):759-768. doi:10.1038/nchembio.1021
52. Haltia T, Freire E. Forces and factors that contribute to the structural stability of membrane proteins. *Biochimica et Biophysica Acta (BBA) - Bioenergetics*. 1995;1228(1):1-27. doi:10.1016/0005-2728(94)00161-W
53. Reading E, Hall Z, Martens C, et al. Interrogating Membrane Protein Conformational Dynamics within Native Lipid Compositions. *Angewandte Chemie International Edition*. 2017;56(49):15654-15657. doi:10.1002/anie.201709657
54. Stelzer W, Poschner BC, Stalz H, Heck AJ, Langosch D. Sequence-Specific Conformational Flexibility of SNARE Transmembrane Helices Probed by Hydrogen/Deuterium Exchange. *Biophys J*. 2008;95(3):1326-1335. doi:10.1529/BIOPHYSJ.108.132928
55. Xiao P, Bolton D, Munro RA, Brown LS, Ladizhansky V. Solid-state NMR spectroscopy based atomistic view of a membrane protein unfolding pathway. *Nat Commun*. 2019;10(1):3867. doi:10.1038/s41467-019-11849-8
56. Deisenhofer J, Epp O, Miki K, Huber R, Michel H. Structure of the protein subunits in the photosynthetic reaction centre of *Rhodospseudomonas viridis* at 3 Å resolution. *Nature*. 1985;318(6047):618-624. doi:10.1038/318618a0
57. Lau FW, Bowie JU. A Method for Assessing the Stability of a Membrane Protein. *Biochemistry*. 1997;36(19):5884-5892. doi:10.1021/bi963095j
58. Booth PJ, Flitsch SL, Stern LJ, Greenhalgh DA, Kim PS, Khorana HG. Intermediates in the folding of the membrane protein bacteriorhodopsin. *Nat Struct Mol Biol*. 1995;2(2):139-143. doi:10.1038/nsb0295-139
59. Popot JL, Engelman DM. Membrane protein folding and oligomerization: the two-stage model. *Biochemistry*. 1990;29(17):4031-4037. doi:10.1021/bi00469a001
60. Hong H, Choi HK, Yoon TY. Untangling the complexity of membrane protein folding. *Curr Opin Struct Biol*. 2022;72:237-247. doi:10.1016/j.sbi.2021.11.013

61. Vinothkumar KR, Strisovsky K, Andreeva A, Christova Y, Verhelst S, Freeman M. The structural basis for catalysis and substrate specificity of a rhomboid protease. *EMBO J.* 2010;29(22):3797-3809. doi:10.1038/emboj.2010.243
62. Huang KS, Bayley H, Liao MJ, London E, Khorana HG. Refolding of an integral membrane protein. Denaturation, renaturation, and reconstitution of intact bacteriorhodopsin and two proteolytic fragments. *J Biol Chem.* 1981;256(8):3802-3809.
63. Cao Z, Hutchison JM, Sanders CR, Bowie JU. Backbone Hydrogen Bond Strengths Can Vary Widely in Transmembrane Helices. *J Am Chem Soc.* 2017;139(31):10742-10749. doi:10.1021/jacs.7b04819
64. Xie K, Dalbey RE. Inserting proteins into the bacterial cytoplasmic membrane using the Sec and YidC translocases. *Nat Rev Microbiol.* 2008;6(3):234-244. doi:10.1038/nrmicro3595
65. Shurtleff MJ, Itzhak DN, Hussmann JA, et al. The ER membrane protein complex interacts cotranslationally to enable biogenesis of multipass membrane proteins. *Elife.* 2018;7. doi:10.7554/eLife.37018
66. Baño-Polo M, Baeza-Delgado C, Tamborero S, et al. Transmembrane but not soluble helices fold inside the ribosome tunnel. *Nat Commun.* 2018;9(1):5246. doi:10.1038/s41467-018-07554-7
67. Akopian D, Shen K, Zhang X, Shan S ou. Signal Recognition Particle: An Essential Protein-Targeting Machine. *Annu Rev Biochem.* 2013;82(1):693-721. doi:10.1146/annurev-biochem-072711-164732
68. Voigts-Hoffmann F, Schmitz N, Shen K, Shan S ou, Ataide SF, Ban N. The Structural Basis of FtsY Recruitment and GTPase Activation by SRP RNA. *Mol Cell.* 2013;52(5):643-654. doi:10.1016/j.molcel.2013.10.005
69. Du Plessis DJF, Nouwen N, Driessen AJM. The Sec translocase. *Biochimica et Biophysica Acta (BBA) - Biomembranes.* 2011;1808(3):851-865. doi:10.1016/J.BBAMEM.2010.08.016
70. Park E, Ménétret JF, Gumbart JC, et al. Structure of the SecY channel during initiation of protein translocation. *Nature.* 2014;506(7486):102-106. doi:10.1038/nature12720
71. Hegde RS, Keenan RJ. The mechanisms of integral membrane protein biogenesis. *Nat Rev Mol Cell Biol.* 2022;23(2):107-124. doi:10.1038/s41580-021-00413-2
72. Smalinskaitė L, Hegde RS. The Biogenesis of Multipass Membrane Proteins. *Cold Spring Harb Perspect Biol.* 2023;15(4):a041251. doi:10.1101/cshperspect.a041251

73. Smalinskaitė L, Kim MK, Lewis AJO, Keenan RJ, Hegde RS. Mechanism of an intramembrane chaperone for multipass membrane proteins. *Nature*. 2022;611(7934):161-166. doi:10.1038/s41586-022-05336-2
74. Sundaram A, Yamsek M, Zhong F, Hooda Y, Hegde RS, Keenan RJ. Substrate-driven assembly of a translocon for multipass membrane proteins. *Nature*. 2022;611(7934):167-172. doi:10.1038/s41586-022-05330-8
75. McGilvray PT, Anghel SA, Sundaram A, et al. An ER translocon for multi-pass membrane protein biogenesis. *Elife*. 2020;9. doi:10.7554/eLife.56889
76. Hessa T, Kim H, Bihlmaier K, et al. Recognition of transmembrane helices by the endoplasmic reticulum translocon. *Nature*. 2005;433(7024):377-381. doi:10.1038/nature03216
77. Yao J, Hong H. Steric trapping strategy for studying the folding of helical membrane proteins. *Methods*. 2024;225:1-12. doi:10.1016/j.ymeth.2024.02.007
78. O'Donnell JP, Phillips BP, Yagita Y, et al. The architecture of EMC reveals a path for membrane protein insertion. *Elife*. 2020;9. doi:10.7554/eLife.57887
79. Engelman DM, Chen Y, Chin CN, et al. Membrane protein folding: beyond the two stage model. *FEBS Lett*. 2003;555(1):122-125. doi:10.1016/S0014-5793(03)01106-2
80. Gaffney KA, Guo R, Bridges MD, et al. Lipid bilayer induces contraction of the denatured state ensemble of a helical-bundle membrane protein. *Proceedings of the National Academy of Sciences*. 2022;119(1). doi:10.1073/pnas.2109169119
81. Lu W, Schafer NP, Wolynes PG. Energy landscape underlying spontaneous insertion and folding of an alpha-helical transmembrane protein into a bilayer. *Nat Commun*. 2018;9(1):4949. doi:10.1038/s41467-018-07320-9
82. Southall NT, Dill KA, Haymet ADJ. A View of the Hydrophobic Effect. *J Phys Chem B*. 2002;106(3):521-533. doi:10.1021/jp015514e
83. Stevens TJ, Arkin IT. Are membrane proteins “inside-out” proteins? *Proteins*. 1999;36(1):135-143. doi:10.1002/(sici)1097-0134(19990701)36:1<135::aid-prot11>3.0.co;2-i
84. Adamian L, Nanda V, DeGrado WF, Liang J. Empirical lipid propensities of amino acid residues in multispans alpha helical membrane proteins. *Proteins: Structure, Function, and Bioinformatics*. 2005;59(3):496-509. doi:10.1002/prot.20456
85. Wimley WC, Creamer TP, White SH. Solvation Energies of Amino Acid Side Chains and Backbone in a Family of Host–Guest Pentapeptides. *Biochemistry*. 1996;35(16):5109-5124. doi:10.1021/bi9600153

86. Wimley WC, White SH. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat Struct Mol Biol.* 1996;3(10):842-848. doi:10.1038/nsb1096-842
87. Hessa T, Meindl-Beinker NM, Bernsel A, et al. Molecular code for transmembrane-helix recognition by the Sec61 translocon. *Nature.* 2007;450(7172):1026-1030. doi:10.1038/nature06387
88. Hunt KLC, Bohr JE. Effects of van der Waals interactions on molecular dipole moments: The role of field-induced fluctuation correlations. *J Chem Phys.* 1985;83(10):5198-5202. doi:10.1063/1.449732
89. Mravic M, Thomaston JL, Tucker M, Solomon PE, Liu L, DeGrado WF. Packing of apolar side chains enables accurate design of highly stable membrane proteins. *Science* (1979). 2019;363(6434):1418-1423. doi:10.1126/science.aav7541
90. Joh NH, Oberai A, Yang D, Whitelegge JP, Bowie JU. Similar Energetic Contributions of Packing in the Core of Membrane and Water-Soluble Proteins. *J Am Chem Soc.* 2009;131(31):10846-10847. doi:10.1021/ja904711k
91. MacKenzie KR, Engelman DM. Structure-based prediction of the stability of transmembrane helix-helix interactions: The sequence dependence of glycoporphin A dimerization. *Proceedings of the National Academy of Sciences.* 1998;95(7):3583-3590. doi:10.1073/pnas.95.7.3583
92. Lemmon MA, Flanagan JM, Hunt JF, et al. Glycophorin A dimerization is driven by specific interactions between transmembrane alpha-helices. *Journal of Biological Chemistry.* 1992;267(11):7683-7689. doi:10.1016/S0021-9258(18)42569-0
93. Smith SO, Song D, Shekar S, Groesbeek M, Ziliox M, Aimoto S. Structure of the transmembrane dimer interface of glycophorin A in membrane bilayers. *Biochemistry.* 2001;40(22):6553-6558. doi:10.1021/bi010357v
94. MacKenzie KR, Prestegard JH, Engelman DM. A transmembrane helix dimer: structure and implications. *Science.* 1997;276(5309):131-133. doi:10.1126/science.276.5309.131
95. Lemmon MA, Flanagan JM, Hunt JF, et al. Glycophorin A dimerization is driven by specific interactions between transmembrane alpha-helices. *J Biol Chem.* 1992;267(11):7683-7689.
96. Kim T, Lee J, Im W. Molecular dynamics studies on structure and dynamics of phospholamban monomer and pentamer in membranes. *Proteins.* 2009;76(1):86-98. doi:10.1002/prot.22322
97. Ben-Tal N, Sitkoff D, Topol IA, Yang AS, Burt SK, Honig B. Free Energy of Amide Hydrogen Bond Formation in Vacuum, in Water, and in Liquid Alkane Solution. *J Phys Chem B.* 1997;101(3):450-457. doi:10.1021/jp961825r

98. Lear JD, DeGrado WF, Choma C, Gratkowski H. Asparagine-mediated self-association of a model transmembrane helix. *Nat Struct Biol.* 2000;7(2):161-166. doi:10.1038/72440
99. Engelman DM, Xiao Zhou F, Cocco MJ, Russ WP, Brunger AT. Interhelical hydrogen bonding drives strong interactions in membrane proteins. *Nat Struct Biol.* 2000;7(2):154-160. doi:10.1038/72430
100. Faham S, Yang D, Bare E, Yohannan S, Whitelegge JP, Bowie JU. Side-chain Contributions to Membrane Protein Structure and Stability. *J Mol Biol.* 2004;335(1):297-305. doi:10.1016/j.jmb.2003.10.041
101. Bowie JU. Membrane protein folding: how important are hydrogen bonds? *Curr Opin Struct Biol.* 2011;21(1):42-49. doi:10.1016/J.SBI.2010.10.003
102. Barlow DJ, Thornton JM. Ion-pairs in proteins. *J Mol Biol.* 1983;168(4):867-885. doi:10.1016/S0022-2836(83)80079-5
103. Kumar S, Nussinov R. Salt bridge stability in monomeric proteins. *J Mol Biol.* 1999;293(5):1241-1255. doi:10.1006/JMBI.1999.3218
104. Waldburger CD, Schildbach JF, Sauer RT. Are buried salt bridges important for protein stability and conformational specificity? *Nat Struct Mol Biol.* 1995;2(2):122-128. doi:10.1038/nsb0295-122
105. Hong H, Szabo G, Tamm LK. Electrostatic couplings in OmpA ion-channel gating suggest a mechanism for pore opening. *Nat Chem Biol.* 2006;2(11):627-635. doi:10.1038/nchembio827
106. Senes A, Chadi DC, Law PB, Walters RFS, Nanda V, DeGrado WF. Ez, a Depth-dependent Potential for Assessing the Energies of Insertion of Amino Acid Side-chains into Membranes: Derivation and Applications to Determining the Orientation of Transmembrane and Interfacial Helices. *J Mol Biol.* 2007;366(2):436-448. doi:10.1016/J.JMB.2006.09.020
107. Schobert B, Cupp-Vickery J, Hornak V, Smith SO, Lanyi JK. Crystallographic Structure of the K Intermediate of Bacteriorhodopsin: Conservation of Free Energy after Photoisomerization of the Retinal. *J Mol Biol.* 2002;321(4):715-726. doi:10.1016/S0022-2836(02)00681-2
108. Baxa MC, Haddadian EJ, Jumper JM, Freed KF, Sosnick TR. Loss of conformational entropy in protein folding calculated using realistic ensembles and its implications for NMR-based calculations. *Proceedings of the National Academy of Sciences.* 2014;111(43):15396-15401. doi:10.1073/pnas.1407768111
109. Frey L, Lakomek N, Riek R, Bibow S. Micelles, Bicelles, and Nanodiscs: Comparing the Impact of Membrane Mimetics on Membrane Protein Backbone

- Dynamics. *Angewandte Chemie International Edition*. 2017;56(1):380-383. doi:10.1002/anie.201608246
110. Van Eps N, Caro LN, Morizumi T, et al. Conformational equilibria of light-activated rhodopsin in nanodiscs. *Proceedings of the National Academy of Sciences*. 2017;114(16). doi:10.1073/pnas.1620405114
 111. O'Brien ES, Fuglestad B, Lessen HJ, et al. Membrane Proteins Have Distinct Fast Internal Motion and Residual Conformational Entropy. *Angewandte Chemie International Edition*. 2020;59(27):11108-11114. doi:10.1002/anie.202003527
 112. Curnow P, Booth PJ. The transition state for integral membrane protein folding. *Proceedings of the National Academy of Sciences*. 2009;106(3):773-778. doi:10.1073/pnas.0806953106
 113. Otzen DE. Folding of DsbB in Mixed Micelles: A Kinetic Analysis of the Stability of a Bacterial Membrane Protein. *J Mol Biol*. 2003;330(4):641-649. doi:10.1016/S0022-2836(03)00624-7
 114. Lau FW, Bowie JU. A Method for Assessing the Stability of a Membrane Protein. *Biochemistry*. 1997;36(19):5884-5892. doi:10.1021/bi963095j
 115. Guo R, Gaffney K, Yang Z, et al. Steric trapping reveals a cooperativity network in the intramembrane protease GlpG. *Nat Chem Biol*. 2016;12(5):353-360. doi:10.1038/nchembio.2048
 116. Chang YC, Bowie JU. Measuring membrane protein stability under native conditions. *Proceedings of the National Academy of Sciences*. 2014;111(1):219-224. doi:10.1073/pnas.1318576111
 117. Dutta A, Kim TY, Moeller M, Wu J, Alexiev U, Klein-Seetharaman J. Characterization of Membrane Protein Non-native States. 2. The SDS-Unfolded States of Rhodopsin. *Biochemistry*. 2010;49(30):6329-6340. doi:10.1021/bi100339x
 118. Anosov AA, Smirnova EYu, Korepanova EA, Shogenov IM. The effects of SDS at subsolubilizing concentrations on the planar lipid bilayer permeability: Two kinds of current fluctuations. *Chem Phys Lipids*. 2019;218:10-15. doi:10.1016/j.chemphyslip.2018.11.005
 119. Klimov DK, Straub JE, Thirumalai D. Aqueous urea solution destabilizes A β 16–22 oligomers. *Proceedings of the National Academy of Sciences*. 2004;101(41):14760-14765. doi:10.1073/pnas.0404570101
 120. Hua L, Zhou R, Thirumalai D, Berne BJ. Urea denaturation by stronger dispersion interactions with proteins than water implies a 2-stage unfolding. *Proceedings of the National Academy of Sciences*. 2008;105(44):16928-16933. doi:10.1073/pnas.0808427105

121. Das A, Mukhopadhyay C. Urea-Mediated Protein Denaturation: A Consensus View. *J Phys Chem B*. 2009;113(38):12816-12824. doi:10.1021/jp906350s
122. Finer EG, Franks F, Tait MJ. Nuclear magnetic resonance studies of aqueous urea solutions. *J Am Chem Soc*. 1972;94(13):4424-4429. doi:10.1021/ja00768a004
123. Hoccart X, Turrell G. Raman spectroscopic investigation of the dynamics of urea–water complexes. *J Chem Phys*. 1993;99(11):8498-8503. doi:10.1063/1.465626
124. Chen GQ, Gouaux E. Probing the folding and unfolding of wild-type and mutant forms of bacteriorhodopsin in micellar solutions: evaluation of reversible unfolding conditions. *Biochemistry*. 1999;38(46):15380-15387. doi:10.1021/bi9909039
125. Panigrahi R, Arutyunova E, Panwar P, Gimpl K, Keller S, Lemieux MJ. Reversible Unfolding of Rhomboid Intramembrane Proteases. *Biophys J*. 2016;110(6):1379-1390. doi:10.1016/j.bpj.2016.01.032
126. Nakatogawa H, Murakami A, Ito K. Control of SecA and SecM translation by protein secretion. *Curr Opin Microbiol*. 2004;7(2):145-150. doi:10.1016/j.mib.2004.01.001
127. Ito K, Chiba S. Arrest peptides: cis-acting modulators of translation. *Annu Rev Biochem*. 2013;82:171-202. doi:10.1146/annurev-biochem-080211-105026
128. Cymer F, Hedman R, Ismail N, Von Heijne G. Exploration of the Arrest Peptide Sequence Space Reveals Arrest-enhanced Variants. *Journal of Biological Chemistry*. 2015;290(16):10208-10215. doi:10.1074/JBC.M115.641555
129. Nicolaus F, Metola A, Mermans D, et al. Residue-by-residue analysis of cotranslational membrane protein integration in vivo. *Elife*. 2021;10:1-16. doi:10.7554/eLife.64302
130. Yu H, Siewny MGW, Edwards DT, Sanders AW, Perkins TT. Hidden dynamics in the unfolding of individual bacteriorhodopsin proteins. *Science* (1979). 2017;355(6328):945-950. doi:10.1126/science.aah7124
131. Choi HK, Min D, Kang H, et al. PROTEIN FOLDING Watching Helical Membrane Proteins Fold Reveals a Common N-to-C-Terminal Folding Pathway. <http://science.sciencemag.org/>
132. Min D, Jefferson RE, Qi Y, et al. Unfolding of a ClC chloride transporter retains memory of its evolutionary history. *Nat Chem Biol*. 2018;14(5):489-496. doi:10.1038/s41589-018-0025-4
133. Kim S, Lee D, Wijesinghe WB, Min D. Robust membrane protein tweezers reveal the folding speed limit of helical membrane proteins. *Elife*. 2023;12. doi:10.7554/eLife.85882

134. Choi HK, Kang H, Lee C, et al. Evolutionary balance between foldability and functionality of a glucose transporter. *Nat Chem Biol.* 2022;18(7):713-723. doi:10.1038/s41589-022-01002-w
135. Min D, Jefferson RE, Bowie JU, Yoon TY. Mapping the energy landscape for second-stage folding of a single membrane protein. *Nat Chem Biol.* 2015;11(12):981-987. doi:10.1038/nchembio.1939
136. von Heijne G. Membrane-protein topology. *Nat Rev Mol Cell Biol.* 2006;7(12):909-918. doi:10.1038/nrm2063
137. Gupta K, Donlan JAC, Hopper JTS, et al. The role of interfacial lipids in stabilizing membrane protein oligomers. *Nature.* 2017;541(7637):421-424. doi:10.1038/nature20820
138. Laganowsky A, Reading E, Allison TM, et al. Membrane proteins bind lipids selectively to modulate their structure and function. *Nature.* 2014;510(7503):172-175. doi:10.1038/nature13419
139. Cong X, Liu Y, Liu W, Liang X, Russell DH, Laganowsky A. Determining Membrane Protein–Lipid Binding Thermodynamics Using Native Mass Spectrometry. *J Am Chem Soc.* 2016;138(13):4346-4349. doi:10.1021/jacs.6b01771
140. Patrick JW, Boone CD, Liu W, et al. Allostery revealed within lipid binding events to membrane proteins. *Proceedings of the National Academy of Sciences.* 2018;115(12):2976-2981. doi:10.1073/pnas.1719813115
141. Hong H, Bowie JU. Dramatic Destabilization of Transmembrane Helix Interactions by Features of Natural Membrane Environments. *J Am Chem Soc.* 2011;133(29):11389-11398. doi:10.1021/ja204524c
142. Blois TM, Hong H, Kim TH, Bowie JU. Protein unfolding with a steric trap. *J Am Chem Soc.* 2009;131(39):13914-13915. doi:10.1021/ja905725n
143. Hong H, Blois TM, Cao Z, Bowie JU. Method to measure strong protein–protein interactions in lipid bilayers using a steric trap. *Proceedings of the National Academy of Sciences.* 2010;107(46):19802-19807. doi:10.1073/pnas.1010348107
144. Yang Y, Guo R, Gaffney K, et al. Folding-Degradation Relationship of a Membrane Protein Mediated by the Universally Conserved ATP-Dependent Protease FtsH. *J Am Chem Soc.* 2018;140(13):4656-4665. doi:10.1021/jacs.8b00832
145. Howarth M, Chinnapen DJF, Gerrow K, et al. A monovalent streptavidin with a single femtomolar biotin binding site. *Nat Methods.* 2006;3(4):267-273. doi:10.1038/nmeth861

146. Hong H, Chang YC, Bowie JU. Measuring Transmembrane Helix Interaction Strengths in Lipid Bilayers Using Steric Trapping. In: ; 2013:37-56. doi:10.1007/978-1-62703-583-5_3
147. Muhammednazaar S, Yao J, Necelis MR, et al. Lipid bilayer strengthens the cooperative network of membrane proteins. bioRxiv. Published online December 23, 2024. doi:10.1101/2023.05.30.542905
148. Guo R, Cang Z, Yao J, et al. Structural cavities are critical to balancing stability and activity of a membrane-integral enzyme. Proc Natl Acad Sci U S A. 2020;117(36). doi:10.1073/pnas.1917770117
149. Gaffney KA, Hong H. The rhomboid protease GlpG has weak interaction energies in its active site hydrogen bond network. Journal of General Physiology. 2019;151(3):282-291. doi:10.1085/jgp.201812047
150. Jefferson RE, Blois TM, Bowie JU. Membrane proteins can have high kinetic stability. J Am Chem Soc. 2013;135(40):15183-15190. doi:10.1021/ja407232b
151. Deng L, Kitova EN, Klassen JS. Dissociation Kinetics of the Streptavidin–Biotin Interaction Measured Using Direct Electrospray Ionization Mass Spectrometry Analysis. J Am Soc Mass Spectrom. 2013;24(1):49-56. doi:10.1007/s13361-012-0533-5
152. Kim S, Lee D, Wijesinghe WB, Min D. Robust membrane protein tweezers reveal the folding speed limit of helical membrane proteins. Elife. 2023;12. doi:10.7554/eLife.85882
153. Brown MS, Ye J, Rawson RB, Goldstein JL. Regulated Intramembrane Proteolysis. Cell. 2000;100(4):391-398. doi:10.1016/S0092-8674(00)80675-3
154. Beard HA, Barniol-Xicota M, Yang J, Verhelst SHL. Discovery of Cellular Roles of Intramembrane Proteases. ACS Chem Biol. 2019;14(11):2372-2388. doi:10.1021/acscchembio.9b00404
155. Lewis AP, Thomas PJ. A novel clan of zinc metallopeptidases with possible intramembrane cleavage properties. Protein Science. 1999;8(2):439-442. doi:10.1110/ps.8.2.439
156. Wolfe MS, Xia W, Ostaszewski BL, Diehl TS, Kimberly WT, Selkoe DJ. Two transmembrane aspartates in presenilin-1 required for presenilin endoproteolysis and γ -secretase activity. Nature. 1999;398(6727):513-517. doi:10.1038/19077
157. Ben-Shem A, Fass D, Bibi E. Structural basis for intramembrane proteolysis by rhomboid serine proteases. Proceedings of the National Academy of Sciences. 2007;104(2):462-466. doi:10.1073/pnas.0609773104

158. Wu Z, Yan N, Feng L, et al. Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry. *Nat Struct Mol Biol.* 2006;13(12):1084-1091. doi:10.1038/nsmb1179
159. Hampton SE, Dore TM, Schmidt WK. Rce1: mechanism and inhibition. *Crit Rev Biochem Mol Biol.* 2018;53(2):157-174. doi:10.1080/10409238.2018.1431606
160. Jürgens G, Wieschaus E, Nüsslein-Volhard C, Kluding H. Mutations affecting the pattern of the larval cuticle in *Drosophila melanogaster* : II. Zygotic loci on the third chromosome. *Wilehm Roux Arch Dev Biol.* 1984;193(5):283-295. doi:10.1007/BF00848157
161. Urban S, Dickey SW. The rhomboid protease family: a decade of progress on function and mechanism. *Genome Biol.* 2011;12(10):231. doi:10.1186/gb-2011-12-10-231
162. Urban S, Lee JR, Freeman M. *Drosophila* rhomboid-1 defines a family of putative intramembrane serine proteases. *Cell.* 2001;107(2):173-182. doi:10.1016/s0092-8674(01)00525-6
163. Rather PN, Ding X, Baca-DeLancey RR, Siddiqui S. *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.* 1999;181(23):7185-7191. doi:10.1128/JB.181.23.7185-7191.1999
164. Srinivasan P, Coppens I, Jacobs-Lorena M. Distinct Roles of Plasmodium Rhomboid 1 in Parasite Development and Malaria Pathogenesis. *PLoS Pathog.* 2009;5(1):e1000262. doi:10.1371/journal.ppat.1000262
165. Wang Y, Zhang Y, Ha Y. Crystal structure of a rhomboid family intramembrane protease. *Nature.* 2006;444(7116):179-180. doi:10.1038/nature05255
166. Lemieux MJ, Fischer SJ, Cherney MM, Bateman KS, James MNG. The crystal structure of the rhomboid peptidase from *Haemophilus influenzae* provides insight into intramembrane proteolysis. *Proceedings of the National Academy of Sciences.* 2007;104(3):750-754. doi:10.1073/pnas.0609981104
167. Ha Y, Akiyama Y, Xue Y. Structure and Mechanism of Rhomboid Protease. *Journal of Biological Chemistry.* 2013;288(22):15430-15436. doi:10.1074/JBC.R112.422378
168. Hubbard SJ. The structural aspects of limited proteolysis of native proteins. *Biochimica et Biophysica Acta (BBA) - Protein Structure and Molecular Enzymology.* 1998;1382(2):191-206. doi:10.1016/S0167-4838(97)00175-1
169. Urban S, Freeman M. Substrate Specificity of Rhomboid Intramembrane Proteases Is Governed by Helix-Breaking Residues in the Substrate Transmembrane Domain Physiologically (Bier et Al. Vol 11.; 2003.

170. Urban S, Schlieper D, Freeman M. Conservation of Intramembrane Proteolytic Activity and Substrate Specificity in Prokaryotic and Eukaryotic Rhomboids. *Current Biology*. 2002;12(17):1507-1512. doi:10.1016/S0960-9822(02)01092-8
171. Strisovsky K, Sharpe HJ, Freeman M. Sequence-Specific Intramembrane Proteolysis: Identification of a Recognition Motif in Rhomboid Substrates. *Mol Cell*. 2009;36(6):1048-1059. doi:10.1016/j.molcel.2009.11.006
172. Zoll S, Stanchev S, Began J, et al. Substrate binding and specificity of rhomboid intramembrane protease revealed by substrate–peptide complex structures. *EMBO J*. 2014;33(20):2408-2421. doi:10.15252/embj.201489367
173. Baker RP, Young K, Feng L, Shi Y, Urban S. Enzymatic analysis of a rhomboid intramembrane protease implicates transmembrane helix 5 as the lateral substrate gate. *Proceedings of the National Academy of Sciences*. 2007;104(20):8257-8262. doi:10.1073/pnas.0700814104
174. Xue Y, Ha Y. Large lateral movement of transmembrane helix S5 is not required for substrate access to the active site of rhomboid intramembrane protease. *Journal of Biological Chemistry*. 2013;288(23):16645-16654. doi:10.1074/jbc.M112.438127
175. Wang Y, Ha Y. Open-cap conformation of intramembrane protease GlpG. *Proceedings of the National Academy of Sciences*. 2007;104(7):2098-2102. doi:10.1073/pnas.0611080104
176. Cho S, Baker RP, Ji M, Urban S. Ten catalytic snapshots of rhomboid intramembrane proteolysis from gate opening to peptide release. *Nat Struct Mol Biol*. 2019;26(10):910-918. doi:10.1038/s41594-019-0296-9
177. Paslawski W, Lillielund OK, Kristensen JV, et al. Cooperative folding of a polytopic α -helical membrane protein involves a compact N-terminal nucleus and nonnative loops. *Proceedings of the National Academy of Sciences*. 2015;112(26):7978-7983. doi:10.1073/pnas.1424751112
178. Gaffney KA, Guo R, Bridges MD, et al. Lipid bilayer induces contraction of the denatured state ensemble of a helical-bundle membrane protein contributed new reagents/analytic tools; K. BIOPHYSICS AND COMPUTATIONAL BIOLOGY. doi:10.1073/pnas.2109169119/-/DCSupplemental
179. Pati S, Banerjee S, Sengupta A, et al. Adaptation strategies of thermophilic microbes. *Bacterial Survival in the Hostile Environment*. Published online January 1, 2023;231-249. doi:10.1016/B978-0-323-91806-0.00012-6
180. Brock TD. Life at High Temperatures. *Science* (1979). 1967;158(3804):1012-1019. doi:10.1126/science.158.3804.1012

181. Stetter KO. Hyperthermophiles in the history of life. In: Philosophical Transactions of the Royal Society B: Biological Sciences. Vol 361. Royal Society; 2006:1837-1843. doi:10.1098/rstb.2006.1907
182. Uemori T, Sato Y, Kato I, Doi H, Ishino Y. A novel DNA polymerase in the hyperthermophilic archaeon, *Pyrococcus furiosus* : gene cloning, expression, and characterization. *Genes to Cells*. 1997;2(8):499-512. doi:10.1046/j.1365-2443.1997.1380336.x
183. Liu Y, Yu P, Song X, Qu Y. Hydrogen production from cellulose by co-culture of *Clostridium thermocellum* JN4 and *Thermoanaerobacterium thermosaccharolyticum* GD17. *Int J Hydrogen Energy*. 2008;33(12):2927-2933. doi:10.1016/j.ijhydene.2008.04.004
184. Ivanova G, Rákhely G, Kovács KL. Thermophilic biohydrogen production from energy plants by *Caldicellulosiruptor saccharolyticus* and comparison with related studies. *Int J Hydrogen Energy*. 2009;34(9):3659-3670. doi:10.1016/j.ijhydene.2009.02.082
185. April TM, Foght JM, Currah RS. Hydrocarbon-degrading filamentous fungi isolated from flare pit soils in northern and western Canada. *Can J Microbiol*. 1999;46(1):38-49. doi:10.1139/w99-117
186. Chaalal O, Islam MR. Integrated management of radioactive strontium contamination in aqueous stream systems. *J Environ Manage*. 2001;61(1):51-59. doi:10.1006/jema.2000.0399
187. Liao WY, Shen CN, Lin LH, et al. Asperjinone, a Nor-Neolignan, and Terrein, a Suppressor of ABCG2-Expressing Breast Cancer Cells, from Thermophilic *Aspergillus terreus*. *J Nat Prod*. 2012;75(4):630-635. doi:10.1021/np200866z
188. Thongpat K, Milehman N, Rojanaverawong W, et al. Total Synthesis and Anti-inflammatory Activity of Asperjinone and Asperimide C. *J Nat Prod*. 2024;87(8):2045-2054. doi:10.1021/acs.jnatprod.4c00557
189. Zhang X, Liu Y, Zheng B, et al. Protein interface redesign facilitates the transformation of nanocage building blocks to 1D and 2D nanomaterials. *Nat Commun*. 2021;12(1). doi:10.1038/s41467-021-25199-x
190. Liang M, Fan K, Zhou M, et al. H-ferritin–nanocaged doxorubicin nanoparticles specifically target and kill tumors with a single-dose injection. *Proceedings of the National Academy of Sciences*. 2014;111(41):14900-14905. doi:10.1073/pnas.1407808111
191. Rothschild LJ, Mancinelli RL. Life in extreme environments. *Nature*. 2001;409(6823):1092-1101. doi:10.1038/35059215

192. Siliakus MF, van der Oost J, Kengen SWM. Adaptations of archaeal and bacterial membranes to variations in temperature, pH and pressure. *Extremophiles*. 2017;21(4):651-670. doi:10.1007/s00792-017-0939-x
193. Tourte M, Kuentz V, Schaeffer P, Grossi V, Cario A, Oger PM. Novel Intact Polar and Core Lipid Compositions in the *Pyrococcus* Model Species, *P. furiosus* and *P. yamanosii*, Reveal the Largest Lipid Diversity Amongst Thermococcales. *Biomolecules*. 2020;10(6):830. doi:10.3390/biom10060830
194. Sprott GD, Meloche M, Richards JC. Proportions of diether, macrocyclic diether, and tetraether lipids in *Methanococcus jannaschii* grown at different temperatures. *J Bacteriol*. 1991;173(12):3907-3910. doi:10.1128/jb.173.12.3907-3910.1991
195. MATSUNO Y, SUGAI A, HIGASHIBATA H, et al. Effect of Growth Temperature and Growth Phase on the Lipid Composition of the Archaeal Membrane from *Thermococcus kodakaraensis*. *Biosci Biotechnol Biochem*. 2009;73(1):104-108. doi:10.1271/bbb.80520
196. Lai D, Springstead JR, Monbouquette HG. Effect of growth temperature on ether lipid biochemistry in *Archaeoglobus fulgidus*. *Extremophiles*. 2008;12(2):271-278. doi:10.1007/s00792-007-0126-6
197. Kim YH, Leriche G, Diraviyam K, et al. Entropic effects enable life at extreme temperatures. *Sci Adv*. 2019;5(5). doi:10.1126/sciadv.aaw4783
198. Fukuchi S, Nishikawa K. Protein surface amino acid compositions distinctively differ between thermophilic and mesophilic bacteria. *J Mol Biol*. 2001;309(4):835-843. doi:10.1006/jmbi.2001.4718
199. Kannan N, Vishveshwara S. Aromatic clusters: a determinant of thermal stability of thermophilic proteins. *Protein Engineering, Design and Selection*. 2000;13(11):753-761. doi:10.1093/protein/13.11.753
200. Serrano L, Bycroft M, Fersht AR. Aromatic-aromatic interactions and protein stability. *J Mol Biol*. 1991;218(2):465-475. doi:10.1016/0022-2836(91)90725-L
201. Meruelo AD, Han SK, Kim S, Bowie JU. Structural differences between thermophilic and mesophilic membrane proteins. *Protein Science*. 2012;21(11):1746-1753. doi:10.1002/pro.2157
202. Islam ST, Lam JS. Topological mapping methods for α -helical bacterial membrane proteins - an update and a guide. *Microbiologyopen*. 2013;2(2):350-364. doi:10.1002/mbo3.72
203. Baradaran R, Berrisford JM, Minhas GS, Sazanov LA. Crystal structure of the entire respiratory complex I. *Nature*. 2013;494(7438):443-448. doi:10.1038/nature11871

204. Kim JM, Altenbach C, Kono M, Oprian DD, Hubbell WL, Gobind Khorana H. Structural Origins of Constitutive Activation in Rhodopsin: Role of the K296E113 Salt Bridge.; 2004. www.pnas.org/cgi/doi/10.1073/pnas.0404519101
205. Abramson J, Smirnova I, Kasho V, Verner G, Kaback HR, Iwata S. Structure and Mechanism of the Lactose Permease of Escherichia Coli. <https://www.science.org>
206. Lemoine F, Correia D, Lefort V, et al. NGPhylogeny.fr: new generation phylogenetic services for non-specialists. *Nucleic Acids Res.* 2019;47(W1):W260-W265. doi:10.1093/nar/gkz303
207. Mader SL, Lopez A, Lawatscheck J, et al. Conformational dynamics modulate the catalytic activity of the molecular chaperone Hsp90. *Nat Commun.* 2020;11(1):1410. doi:10.1038/s41467-020-15050-0
208. Wimley WC, White SH. Experimentally determined hydrophobicity scale for proteins at membrane interfaces. *Nat Struct Biol.* 1996;3(10):842-848. doi:10.1038/nsb1096-842
209. Wimley WC, Gawrisch K, Creamer TP, White SH. Direct measurement of salt-bridge solvation energies using a peptide model system: implications for protein stability. *Proceedings of the National Academy of Sciences.* 1996;93(7):2985-2990. doi:10.1073/pnas.93.7.2985
210. Ulmschneider MB, Sansom MSP, Di Nola A. Properties of integral membrane protein structures: Derivation of an implicit membrane potential. *Proteins: Structure, Function, and Bioinformatics.* 2005;59(2):252-265. doi:10.1002/prot.20334
211. vonHeijne G. Control of topology and mode of assembly of a polytopic membrane protein by positively charged residues. *Nature.* 1989;341(6241):456-458. doi:10.1038/341456a0
212. von Heijne G. The distribution of positively charged residues in bacterial inner membrane proteins correlates with the trans-membrane topology. *EMBO J.* 1986;5(11):3021-3027. doi:10.1002/j.1460-2075.1986.tb04601.x
213. Isom DG, Castañeda CA, Cannon BR, García-Moreno E. B. Large shifts in pKa values of lysine residues buried inside a protein. *Proceedings of the National Academy of Sciences.* 2011;108(13):5260-5265. doi:10.1073/pnas.1010750108
214. Harms MJ, Schlessman JL, Sue GR, García-Moreno E. B. Arginine residues at internal positions in a protein are always charged. *Proceedings of the National Academy of Sciences.* 2011;108(47):18954-18959. doi:10.1073/pnas.1104808108
215. Robinson AC, Castañeda CA, Schlessman JL, Bertrand García-Moreno E. Structural and thermodynamic consequences of burial of an artificial ion pair in

- the hydrophobic interior of a protein. *Proc Natl Acad Sci U S A*. 2014;111(32):11685-11690. doi:10.1073/pnas.1402900111
216. Isom DG, Cannon BR, Castañeda CA, Robinson A, García-Moreno E. B. High tolerance for ionizable residues in the hydrophobic interior of proteins. *Proceedings of the National Academy of Sciences*. 2008;105(46):17784-17788. doi:10.1073/pnas.0805113105
 217. Robinson AC, Schlessman JL, García-Moreno BE. Dielectric Properties of a Protein Probed by Reversal of a Buried Ion Pair. *Journal of Physical Chemistry B*. 2018;122(9):2516-2524. doi:10.1021/acs.jpcb.7b12121
 218. Hong H, Szabo G, Tamm LK. Electrostatic couplings in OmpA ion-channel gating suggest a mechanism for pore opening. *Nat Chem Biol*. 2006;2(11):627-635. doi:10.1038/nchembio827
 219. Islam ST, Lam JS. Topological mapping methods for α -helical bacterial membrane proteins – an update and a guide. *Microbiologyopen*. 2013;2(2):350-364. doi:10.1002/mbo3.72
 220. Nagase H, Fields CG, Fields GB. Design and characterization of a fluorogenic substrate selectively hydrolyzed by stromelysin 1 (matrix metalloproteinase-3). *J Biol Chem*. 1994;269(33):20952-20957.
 221. Arutyunova E, Jiang Z, Yang J, et al. An internally quenched peptide as a new model substrate for rhomboid intramembrane proteases. *Biol Chem*. 2018;399(12):1389-1397. doi:10.1515/hsz-2018-0255
 222. Schneider CA, Rasband WS, Eliceiri KW. NIH Image to ImageJ: 25 years of image analysis. *Nat Methods*. 2012;9(7):671-675. doi:10.1038/nmeth.2089
 223. Howarth M, Chinnapen DJF, Gerrow K, et al. A monovalent streptavidin with a single femtomolar biotin binding site. *Nat Methods*. 2006;3(4):267-273. doi:10.1038/nmeth861
 224. Horovitz A. Double-mutant cycles: a powerful tool for analyzing protein structure and function. *Fold Des*. 1996;1(6):R121-R126. doi:10.1016/S1359-0278(96)00056-9
 225. Darby NJ, Creighton TE. *Protein Structure: In Focus*. Oxford University Press; 1993.
 226. Tan KP, Varadarajan R, Madhusudhan MS. DEPTH: a web server to compute depth and predict small-molecule binding cavities in proteins. *Nucleic Acids Res*. 2011;39(suppl_2):W242-W248. doi:10.1093/nar/gkr356

227. Park C, Marqusee S. Pulse proteolysis: A simple method for quantitative determination of protein stability and ligand binding. *Nat Methods*. 2005;2(3):207-212. doi:10.1038/nmeth740
228. Na YR, Park C. Investigating protein unfolding kinetics by pulse proteolysis. *Protein Sci*. 2009;18(2):268-276. doi:10.1002/pro.29
229. Raschke TM, Marqusee S. The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. *Nat Struct Biol*. 1997;4(4):298-304. doi:10.1038/nsb0497-298
230. Larsen AN, Moe E, Helland R, Gjellesvik DR, Willassen NP. Characterization of a recombinantly expressed proteinase K-like enzyme from a psychrotrophic *Serratia* sp. *FEBS J*. 2006;273(1):47-60. doi:10.1111/j.1742-4658.2005.05044.x
231. Min D, Jefferson RE, Bowie JU, Yoon TY. Mapping the energy landscape for second-stage folding of a single membrane protein. *Nat Chem Biol*. 2015;11(12):981-987. doi:10.1038/nchembio.1939
232. Sivaraman T, Robertson AD. Kinetics of Conformational Fluctuations by EX1 Hydrogen Exchange in Native Proteins. In: *Protein Structure, Stability, and Folding*. Humana Press; :193-214. doi:10.1385/1-59259-193-0:193
233. Muhammed Nazaar FS. *Folding, Stability and Degradation of Membrane Protein in the Bilayer*. Michigan State University; 2022.
234. Maegawa S, Ito K, Akiyama Y. Proteolytic action of GlpG, a rhomboid protease in the *Escherichia coli* cytoplasmic membrane. *Biochemistry*. 2005;44(41):13543-13552. doi:10.1021/bi051363k
235. Strisovsky K, Sharpe HJ, Freeman M. Sequence-Specific Intramembrane Proteolysis: Identification of a Recognition Motif in Rhomboid Substrates. *Mol Cell*. 2009;36(6):1048-1059. doi:10.1016/j.molcel.2009.11.006
236. Akiyama Y, Maegawa S. Sequence features of substrates required for cleavage by GlpG, an *Escherichia coli* rhomboid protease. *Mol Microbiol*. 2007;64(4):1028-1037. doi:10.1111/j.1365-2958.2007.05715.x
237. Razvi A, Scholtz JM. Lessons in stability from thermophilic proteins. *Protein Science*. 2006;15(7):1569-1578. doi:10.1110/ps.062130306
238. Stetter KO. Hyperthermophiles in the history of life. In: *Philosophical Transactions of the Royal Society B: Biological Sciences*. Vol 361. Royal Society; 2006:1837-1843. doi:10.1098/rstb.2006.1907

239. Uemori T, Ishino Y, Toh H, Asada K, Kato I. Organization and nucleotide sequence of the DNA polymerase gene from the archaeon *Pyrococcus furiosus*. *Nucleic Acids Res.* 1993;21(2):259-265. doi:10.1093/nar/21.2.259
240. Ceci P, Forte E, Di Cecca G, Fornara M, Chiancone E. The characterization of *Thermotoga maritima* ferritin reveals an unusual subunit dissociation behavior and efficient DNA protection from iron-mediated oxidative stress. *Extremophiles.* 2011;15(3):431-439. doi:10.1007/s00792-011-0374-3
241. Becktel WJ, Schellman JA. Protein stability curves. *Biopolymers.* 1987;26(11):1859-1877. doi:10.1002/bip.360261104
242. Nojima H, Ikai A, Oshima T, Noda H. Reversible thermal unfolding of thermostable phosphoglycerate kinase. Thermostability associated with mean zero enthalpy change. *J Mol Biol.* 1977;116(3):429-442. doi:10.1016/0022-2836(77)90078-x
243. Li W tyng, Grayling RA, Sandman K, Edmondson S, Shriver JW, Reeve JN. Thermodynamic Stability of Archaeal Histones. *Biochemistry.* 1998;37(30):10563-10572. doi:10.1021/bi973006i
244. Szilágyi A, Závodszky P. Structural differences between mesophilic, moderately thermophilic and extremely thermophilic protein subunits: results of a comprehensive survey. *Structure.* 2000;8(5):493-504. doi:10.1016/S0969-2126(00)00133-7
245. Fukuchi S, Nishikawa K. Protein surface amino acid compositions distinctively differ between thermophilic and mesophilic bacteria. *J Mol Biol.* 2001;309(4):835-843. doi:10.1006/jmbi.2001.4718
246. Chan CH, Yu TH, Wong KB. Stabilizing Salt-Bridge Enhances Protein Thermostability by Reducing the Heat Capacity Change of Unfolding. *PLoS One.* 2011;6(6):e21624. doi:10.1371/journal.pone.0021624
247. Szilágyi A, Závodszky P. Structural Differences between Mesophilic, Moderately Thermophilic and Extremely Thermophilic Protein Subunits: Results of a Comprehensive Survey.; 2000.
248. Schneider D, Liu Y, Gerstein M, Engelman DM. Thermostability of membrane protein helix–helix interaction elucidated by statistical analysis. *FEBS Lett.* 2002;532(1-2):231-236. doi:10.1016/S0014-5793(02)03687-6
249. Sinensky M. Homeoviscous adaptation--a homeostatic process that regulates the viscosity of membrane lipids in *Escherichia coli*. *Proc Natl Acad Sci U S A.* 1974;71(2):522-525. doi:10.1073/pnas.71.2.522
250. Yokoyama A, Shizuri Y, Hoshino T, Sandmann G. Thermocryptoxanthins: novel intermediates in the carotenoid biosynthetic pathway of *Thermus thermophilus*. *Arch Microbiol.* 1996;165(5):342-345. doi:10.1007/s002030050336

251. Oshima M, Miyagawa A. Comparative studies on the fatty acid composition of moderately and extremely thermophilic bacteria. *Lipids*. 1974;9(7):476-480. doi:10.1007/BF02534274
252. Patel BKC, Skerratt JH, Nichols PD. The Phospholipid Ester-linked Fatty Acid Composition of Thermophilic Bacteria. *Syst Appl Microbiol*. 1991;14(4):311-316. doi:10.1016/S0723-2020(11)80304-8
253. Jung S, Zeikus JG, Hollingsworth RI. A new family of very long chain alpha,omega-dicarboxylic acids is a major structural fatty acyl component of the membrane lipids of *Thermoanaerobacter ethanolicus* 39E. *J Lipid Res*. 1994;35(6):1057-1065. doi:10.1016/S0022-2275(20)40101-4
254. Boucher Y. Lipids: Biosynthesis, Function, and Evolution. In: *Archaea*. ASM Press; 2014:341-353. doi:10.1128/9781555815516.ch15
255. Dannenmuller O, Arakawa K, Eguchi T, et al. Membrane properties of archaeal macrocyclic diether phospholipids. *Chemistry*. 2000;6(4):645-654. doi:10.1002/(sici)1521-3765(20000218)6:4<645::aid-chem645>3.0.co;2-a
256. Gliozzi A, Paoli G, De Rosa M, Gambacorta A. Effect of isoprenoid cyclization on the transition temperature of lipids in thermophilic archaeobacteria. *Biochimica et Biophysica Acta (BBA) - Biomembranes*. 1983;735(2):234-242. doi:10.1016/0005-2736(83)90298-5
257. Freeman M. The Rhomboid-Like Superfamily: Molecular Mechanisms and Biological Roles. *Annu Rev Cell Dev Biol*. 2014;30(1):235-254. doi:10.1146/annurev-cellbio-100913-012944
258. Koonin E V, Makarova KS, Rogozin IB, Davidovic L, Letellier MC, Pellegrini L. The Rhomboids: A Nearly Ubiquitous Family of Intramembrane Serine Proteases That Probably Evolved by Multiple Ancient Horizontal Gene Transfers.; 2003. <http://genomebiology.com/2002/3/11/preprint/0010>
259. Kobayashi T, Kwak YS, Akiba T, Kudo T, Horikoshi K. *Thermococcus profundus* sp. nov., A New Hyperthermophilic Archaeon Isolated from a Deep-sea Hydrothermal Vent. *Syst Appl Microbiol*. 1994;17(2):232-236. doi:10.1016/S0723-2020(11)80013-5
260. Fiala G, Stetter KO. *Pyrococcus furiosus* sp. nov. represents a novel genus of marine heterotrophic archaeobacteria growing optimally at 100 °C. *Arch Microbiol*. 1986;145(1):56-61. doi:10.1007/BF00413027
261. Huber R, Langworthy TA, König H, et al. *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90 °C. *Arch Microbiol*. 1986;144(4):324-333. doi:10.1007/BF00409880

262. Stevenson LG, Strisovsky K, Clemmer KM, Bhatt S, Freeman M, Rather PN. Rhomboid Protease AarA Mediates Quorum-Sensing in *Providencia Stuartii* by Activating TatA of the Twin-Arginine Translocase.; 2007. www.pnas.org/cgi/doi/10.1073/pnas.0608140104
263. Strisovsky K, Sharpe HJ, Freeman M. Sequence-Specific Intramembrane Proteolysis: Identification of a Recognition Motif in Rhomboid Substrates. *Mol Cell*. 2009;36(6):1048-1059. doi:10.1016/j.molcel.2009.11.006
264. Lemberg MK, Menendez J, Misik A, Garcia M, Koth CM, Freeman M. Mechanism of intramembrane proteolysis investigated with purified rhomboid proteases. *EMBO Journal*. 2005;24(3):464-472. doi:10.1038/sj.emboj.7600537
265. Lohi O, Urban S, Freeman M. Diverse Substrate Recognition Mechanisms for Rhomboids: Thrombomodulin Is Cleaved by Mammalian Rhomboids. *Current Biology*. 2004;14(3):236-241. doi:10.1016/J.CUB.2004.01.025
266. Käll L, Krogh A, Sonnhammer ELL. A Combined Transmembrane Topology and Signal Peptide Prediction Method. *J Mol Biol*. 2004;338(5):1027-1036. doi:10.1016/j.jmb.2004.03.016
267. Jumper J, Evans R, Pritzel A, et al. Highly accurate protein structure prediction with AlphaFold. *Nature*. 2021;596(7873):583-589. doi:10.1038/s41586-021-03819-2
268. Roche DB, Buenavista MT, Tetchner SJ, McGuffin LJ. The IntFOLD server: an integrated web resource for protein fold recognition, 3D model quality assessment, intrinsic disorder prediction, domain prediction and ligand binding site prediction. *Nucleic Acids Res*. 2011;39(suppl):W171-W176. doi:10.1093/nar/gkr184
269. Lemieux MJ, Fischer SJ, Cherney MM, Bateman KS, James MNG. The crystal structure of the rhomboid peptidase from *Haemophilus influenzae* provides insight into intramembrane proteolysis. *Proceedings of the National Academy of Sciences*. 2007;104(3):750-754. doi:10.1073/pnas.0609981104
270. Koga Y. Thermal adaptation of the archaeal and bacterial lipid membranes. *Archaea*. 2012;2012. doi:10.1155/2012/789652
271. Hwang JK, Warshel A. Why ion pair reversal by protein engineering is unlikely to succeed. *Nature*. 1988;334(6179):270-272. doi:10.1038/334270a0
272. Deng J, Cui Q. Electronic Polarization Is Essential for the Stabilization and Dynamics of Buried Ion Pairs in Staphylococcal Nuclease Mutants. *J Am Chem Soc*. 2022;144(10):4594-4610. doi:10.1021/jacs.2c00312

273. Garczarek F, Gerwert K. Functional waters in intraprotein proton transfer monitored by FTIR difference spectroscopy. *Nature*. 2006;439(7072):109-112. doi:10.1038/nature04231
274. Lapek JD, Jiang Z, Wozniak JM, et al. Quantitative Multiplex Substrate Profiling of Peptidases by Mass Spectrometry. *Molecular & Cellular Proteomics*. 2019;18(5):968a-9981. doi:10.1074/MCP.TIR118.001099