New Insights Into the Role of Membrane Interactions and Conformational Dynamics in Intramembrane Proteolysis by GlpG Rhomboid

By

Alexander Foo

A thesis submitted to the
Faculty of Graduate & Postdoctoral Studies
In partial fulfillment of the requirements for the
Doctorate in Philosophy degree in Chemistry

Department of Chemistry and Biomolecular Sciences
Faculty of Science
University of Ottawa

Candidate

Supervisor

_____

Alexander Foo

_____

Dr. Natalie K. Goto

Abstract

The rhomboid family of intramembrane serine proteases can catalyze proteolysis of substrates that are normally embedded in the cell membrane, making them key players in a diverse range of biological processes. While X-ray crystal structures provide detailed insights into the mechanism of intramembrane hydrolysis, questions remain concerning how transmembrane (TM) substrates are able to gain access to the rhomboid active site, and whether interactions with the membrane environment can influence its structure and function. In this thesis, these questions were investigated using the *E. coli* rhomboid *ec*GlpG.

In Chapter 3, the effect of hydrophobic mismatch between lipid and protein was investigated using families of amphiphiles with saturated alkyl chains. While *ec*GlpG displayed maximal activity against a water-soluble model substrate when solubilized in detergents containing 10-12 carbon atoms, shorter and longer chain detergents led to loss of activity. An even larger effect was observed when *ec*GlpG was reconstituted into phospholipid bicelles, with no proteolytic activity being detected in 14-carbon lipids. These results suggest that mismatch between the hydrophobic regions of the catalytic TM domain (TMD) and the local membrane environment is detrimental to proteolysis.

To obtain further insight into the structure and dynamics of *ec*GlpG, sample conditions were identified in Chapter 4 that enabled, for the first time, the acquisition of NMR spectra showing signals from the *ec*GlpG TMD. While significant peak broadening prevented chemical shift assignment, the sensitivity and resolution of peaks corresponding to the tryptophan indole NH group allowed their use as structural probes. These were employed in Chapter 5 to characterize the open conformation of *ec*GlpG that is postulated to facilitate substrate entry. These spectra showed evidence of an open conformation in which the intact

α5 is laterally displaced. Interactions with a substrate-derived peptide also appeared to stimulate gate opening; however, activity assays suggested that formation of the open state could compromise catalytic activity against water-soluble substrates, and that interactions with TM substrates could counter this effect. Taken together, these results provide new insight into the role of both the local membrane environment and α5-conformational dynamics on intramembrane proteolysis, and suggest a mechanism to prevent cleavage of off-target rhomboid substrates *in vivo*.

Acknowledgements

First and foremost, I would like to thank Dr. Natalie Goto, my doctoral supervisor, for accepting me into her lab and providing me with many years of guidance, mentorship and, most importantly, patience during the conduct of my doctoral research project and concurrent development as a professional researcher. I would also like to thank her for the countless hours she has spent helping to prepare presentations and written documents; in particular her feedback has helped develop and refine this thesis for your reading pleasure, while her keen eyes have no doubt prevented the *ingress* of numerous careless errors and typoes. I am also grateful for her support for my passion for science education and outreach, providing me with opportunities to become involved in the teaching process for her undergraduate chemistry classes including the occasional classroom demo, not to mention putting up with all the strange (and occasionally pyrotechnic) science demos and experiments I have brewed up in her lab throughout the years. It is clear from the past five years that she always looks out for her students, and I feel honoured to be one of them.

I am also grateful for the direct contributions from my thesis committee Drs. Yves Aubin, Roberto Chica, Corrie daCosta, Joanne Lemieux, and John Pezacki for all the help and advice they have provided me throughout the long and winding road of this research project. Additionally, I would also like to acknowledge the contributions from Jeff Metz, Brandon Harvey, Carolina Uribe, Paulina Przeracki, Eric Burge, Shawna Rieux, and Rebecca Brown; undergraduate students who assisted with the design and implementation of the various experiments throughout my research. All of them have contributed beyond what was expected, while contributing to a fun and engaging work environment.

I would also like to acknowledge the indirect contributions from Drs. Jeffery Keillor and Tom Moon who provided our lab generous use of their instruments and, along with Dr. Adam Schuhendler have followed my research and posed excellent questions and insights throughout my research process. Glen Facey, the NMR lab coordinator, was instrumental in enabling the NMR experiments presented throughout this thesis, and was always willing to accommodate our requests with what appeared to be an endless degree of patience, while Dr. Kvido Strisovsky (Academy of Sciences of the Czech Republic), kindly provided the plasmid for the PKS187 TatA construct. Other members of the Goto lab, especially (soon to be Dr.) Saud Ayed and Adam Cloutier, along with our adoptive student Adam Damry have all helped with various aspects of the research process while sharing the trials and tribulations of graduate student life. This research was also aided by funding provided through the Ontario Graduate Scholarship program (2013, 2015)

Finally, no acknowledgement page would be complete without the deepest thanks to my family. My parents Simon and Elise, along with Brian, my twin brother, who have been with me quite literally since the very beginning, and have provided me with endless amounts of love and support in all of life's endeavours, regardless of where they may have lead. Lastly, I would like to thank my dearest friend Lianna Teeter who has been with me for almost half my life, and kept me going through thick and thin.

Table of Contents

List of Abbreviations

3D	3-dimensional
Aβ	Amyloid Beta Peptide
AarA	*Providencia stuartii* rhomboid protease
AMS	4-acetamido-4'-maleimidylstilbene-2,2'-disulfonic Acid
APP	Amyloid Precursor Protein
ATP	Adenosine Triphosphate
β-NG	β-nonylglucoside
β-OG	β-octylglucoside
β2AR	Beta 2 Adrenergic Receptor
BACE1	Beta-secretase 1
Bcl2	B-cell Lymphoma 2
BIK	Bcl2-interacting Killer
BODIPY	Boron Dipyrromethene
bR	Bacteriorhodopsin
$C_{12}E_8$	Octaethylene Glycol Monododecyl Ether
CD	Circular Dichroism
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
CMC	Critical Micelle Concentration
CPMG	Carr-Purcell-Meiboom-Gill
CytD	Cytoplasmic Domain
DCI	3,4-Dichloroisocoumarin
***dm*Rho4**	*Drosophila melanogaster* Rhomboid 4
DFP	Diisopropylfluorophosphate
DMSO	Dimethyl Sulfoxide
DSS	4,4-dimethyl-4-silapentane-1-sulfonic Acid
***ec*GlpG**	*Escherichia coli* rhomboid protease
EDTA	2,2',2",2"'-(Ethane-1,2-diyldinitrilo)tetraacetic Acid
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
ER	Endoplasmic Reticulum

ERAD	Endoplasmic Reticulum-Associated Degradation
ExPEC	Extraintestinal Pathogenic *E. coli*
FITC	Fluorescein Iso-thiocyanate
Fos-8	Octylphosphocholine
Fos-9	Nonylphosphoholine
Fos-10	Decylphosphocholine
Fos-12	Dodecylphosphocholine
Fos-14	Tetradecylphosphocholine
Fos-16	Hexadecylphosphocholine
FPLC	Fast Protein Liquid Chromatography
FRET	Förster Resonance Energy Transfer
GPCR	G-protein Coupled Receptor
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
***hi*GlpG**	*Haemophilus influenzae* Rhomboid
HSQC	Heteronuclear Single Quantum Coherence
iCcp1	Intermediate of Cytochrome C Peroxidase
IMP	Integral Membrane Protease
IPTG	Isopropyl B-D-Thiogalactopyranoside
iRhom	Inactive Rhomboid
K_D	Dissociation Constant
kDa	Kilodalton
K_M	Michaelis Constant
K_i	Inhibition Constant
KscA	*Streptomyces lividans* Potassium Ion Channel
l-Mgm-1	Mitochondrial Genome Maintenance 1, Long Isoform
L1	Loop 1 Region
L5	Loop 5 Region
LDAO	N,N-Dimethyldodecylamine N-oxide
LY2	LacY Transmembrane Helix 2
Lyso-Gly-16	Lysopalmitoylphosphoglycerol
Lyso-Phos-12	Lysolaurylphosphocholine

Lyso-Phos-14	Lysomyristoylphosphocholine
M2M	1,2-ethanediyl Bismethanethiosulfonate
Mal-10	n-decyl-β-D-maltoside
Mal-12	n-dodecyl-β-D-maltoside
Mal-14	n-tetradecyl-β-D-maltoside
MD	Molecular Dynamics Simulations
Mgm-1	Mitochondrial Genome Maintenance 1
MHz	Megahertz
MPP	Mitochondrial Processing Peptidase
MRE	Mean Residue Ellipticity
MW	Molecular Weight
NG-10	Decyl Maltose Neopentyl Glycol
NG-12	Lauryl Maltose Neopentyl Glycol
n_H	Hill Coefficient
Ni-NTA	Nickel-Nitrilotriacetic Acid
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
OD	Optical Density
OmpA	Bacterial Outer Membrane Protein A
OmpX	Bacterial Outer Membrane Protein X
***pa*GlpG**	*Pseudomonas aeruginosa* Rhomboid
PagP	PhoPQ-activated Gene P
PARL	Presenilin-associated Rhomboid-like Protease
PC-7	Diheptanoylphosphatidylcholine
PC-12	Dilaurylphosphatidylcholine
PC-14	Dimystroylphosphatidylcholine
PDB	Protein Database (RCSB)
PDC	Protein-Detergent Complex
***pf*Rom1**	*P. falciparum Rhomboid 1*
***pf*Rom4**	*P. falciparum Rhomboid 4*
POPG	1-palmitoyl-2-oleoyl-phosphatidylglycerol

PRE	Paramagnetic Relaxation Enhancement
RHBDL2	Human Rhomboid Related Protein 2
RHBDL4	Human Rhomboid Related Protein 4
s-Mgm-1	Mitochondrial Genome Maintenance 1, Short Isoform
SAXS	Small Angle X-ray Scattering
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SMR	Small Multidrug Resistance Transporter
Tat	Twin Arginine Translocase
TatA	Twin Arginine Translocase, Component A
***tg*ROM4**	Rhomboid 4 from *Toxoplasma gondii*
TM	Transmembrane
TMD	Transmembrane Domain
TMD60	*ec*GlpG Transmembrane Domain Residues 60-276
TMD81	*ec*GlpG Transmembrane Domain Residues 81-276
TMD91	*ec*GlpG Transmembrane Domain Residues 91-276
TRIS	2-Amino-2-(hydroxymethyl)propane-1,3-diol
TROSY	Transverse Relaxation Optimized Spectroscopy
UB	Ubiquitin
UIM	Ubiquitin Interacting Motif
VDAC- 1	Human Mitochondrial Voltage-dependent Anion Channel 1
VDAC- 2	Human Mitochondrial Voltage-dependent Anion Channel 2
WT	Wild Type

List of Figures

List of Tables

List of Equations

Chapter 1: Introduction to Rhomboid Proteases

Preamble

The cell membrane can be considered the single most important component of a cell, acting as a barrier between the interior and exterior environment. Approximately 20-30% of all coding sequences in both prokaryotic and eukaryotic genomes encode for proteins that are associated with the membrane where they carry out functions critical for life including ion conduction, receptor-mediated cell signalling, and biocatalytic processes such as hydrolysis.[1,2] While the hydrophobic nature of the membrane environment poses a unique challenge to the latter,[3] there are many families of intramembrane enzymes which catalyze hydrolytic reactions including phospholipases and intramembrane proteases.[4,5]

Intramembrane proteases (IMPs) are enzymes that catalyze the cleavage of peptide bonds. The active site is typically formed by a multi-spanning transmembrane domain (TMD) that can cleave both solvent exposed and buried regions of transmembrane (TM) sequences, the latter of which are not normally accessible to other water-soluble proteases. This gives rise to a diverse range of biological roles such as propagation of cell signalling pathways, and the processing and degradation of other integral membrane proteins. Several families of IMPs have been described in the literature, the most recent of which is the rhomboid serine protease. The first rhomboid protease was discovered through a mutation in *Drosophila* which gave rise to pointy or "rhomboid" shaped heads during embryonic development, leading to its designation as Rhomboid-1.[6,7] Subsequent genetic and functional studies have identified homologous proteases in numerous prokaryotic and eukaryotic species including humans and *E. coli*, the latter of which will be the subject of this thesis. [8,9]

While the sequence identity within the rhomboid family is relatively low (around 10-15%), all members are predicted to share a similar topology consisting of 6 or 7 TM alpha-

helices containing a conserved Ser-His catalytic dyad located within conserved GxSx and AHxxGxxxG motifs positioned on TM helices α4 and α6 respectively.[10] Other sequence features include conserved histidine and asparagine residues within α2, which form part of the oxyanion hole required for proteolysis discussed in 1.2.1.[10–12] Structural and kinetic studies over the last decade have provided a detailed understanding of the rhomboid catalytic cycle. However, questions remain concerning how rhomboids and other IMPs are able to gain access to TM substrates while minimizing exposure of the hydrophilic active site to the hydrophobic membrane environment, and whether interactions with the lipid phase can influence protease structure and function. This thesis aims to address both of these questions through the use of enzyme kinetics coupled with structural techniques such as solution-state nuclear-magnetic resonance (NMR) and circular dichroism (CD) spectroscopy.

1.1: Biological Function of Rhomboids

1.1.1: Cell Signalling

The first biological activity to be characterized for a rhomboid was the cleavage of Spitz by the *Drosophila* Rhomboid-1 to release and secrete a soluble Epidermal Growth Factor (EGF), which activates the corresponding EGF receptor (EGFR) in neighboring cells to regulate cell growth and development.[8,9] Three other rhomboid homologues have since been characterized in *Drosophila* (Rhomboids 2-4), all of which also participate in EGF signalling pathways related to different aspects of cell growth and development.[13,14]

Mammalian rhomboids have also been associated with EGFR signalling pathways, as exemplified by the cleavage of EGF and proTGFα EGF precursors by the human RHBDL2 and RHBDL4 rhomboids respectively.[15,16] The latter has also been associated with the

regulation of apoptosis through the proteolytic activation of Bcl-2 Interacting Killer (BIK) and TSAP6, making it difficult to identify a unified role for these enzymes in mammalian cell signalling.[17,18]

The prokaryotic rhomboid AarA from *Providencia stuartii* also contributes to cell signalling, though the mechanism through which this occurs is distinct from its eukaryotic counterparts. Here, AarA is responsible for the removal of a short N-terminal inhibitory sequence from the twin arginine translocase component A (TatA), which subsequently oligomerizes to form a membrane pore (Fig. 1.1A) which, along with TatB and TatC, comprises the complete twin-arginine translocation complex.[19–22] This complex facilitates the export of yet to be identified virulence factors that facilitate quorum sensing and biofilm formation. This mode of action appears to be unique to prokaryotic rhomboids, and makes them a potential target for the development of antimicrobial compounds.

1.1.2: Protein Degradation

In addition to its role in EGF-mediated cell signalling, the human rhomboid RHBDL4 has been shown to contribute to the degradation of misfolded proteins through the Endoplasmic Reticulum-Associated Degradation (ERAD) pathway.[23,24] Here, the soluble C-terminal domain of RHBDL4 is able to bind ubiquitinated TM proteins, facilitating their cleavage by the rhomboid catalytic domain as shown in Fig. 1.1B.[23,24] Functional studies suggest that RHBDL4 interacts with p97 to facilitate the excision and proteasomal degradation of the substrate fragments following cleavage, though the structural details of this interaction have yet to be determined.[23,24]

A recent report found that RHBDL4 could also cleave the amyloid precursor protein (APP) in an ERAD/ubiquitin independent manner.[25] Normally intact APP is exported from the ER to the plasma membrane, where it is cleaved by the BACE1 and γ-secretase IMPs to generate the amyloidogenetic Aβ peptide.[26] RHBDL4 is able to intercept APP prior to its transport to the plasma membrane, cleaving it to produce a 20-25 kDa TM fragment and a 70-75 kDa soluble ectofragment.[25] While the final fate of these fragments is not known, the action of RHBDL4 has been shown to significantly reduce the amount of APP which can be transported to the cell surface, potentially limiting the formation of amyloid plaques and the onset of their associated disease states.[25]

1.1.3: Mitochondrial Rhomboids and Apoptosis

The yeast *S. cerevisiae* contains two rhomboids: RBD2p and RBD1p, the latter of which was found to localize to the mitochondrial inner membrane where it is responsible for the processing of cytochrome c peroxidase (iCcp1) to produce the mature enzyme that catalyzes the oxidation and subsequent release of cardiolipin from the mitochondrial membrane, triggering apoptosis through the activation of caspase-9.[27–30] Conversely, RBD1p also plays a role in maintaining mitochondrial structure and DNA retention through its processing of Mgm1 GTPase.[31,32] The Mgm1 precursor contains two hydrophobic TM segments, as shown in Fig. 1.1C. Insertion of the first segment into the mitochondrial inner membrane enables cleavage by the Mitochondrial Processing Peptidase (MPP), resulting in the production of the long isoform (l-Mgm1) anchored in the mitochondrial inner membrane.[31] Alternatively, Mgm1 can be pulled further into the membrane in an ATP dependant process, resulting in the insertion of the second TM segment.[33] Under these

conditions Mgm1 is cleaved by the RBD1p rhomboid to liberate the soluble short (s-Mgm1) isoform from the inner membrane into the intermembrane space.[33] Both isoforms, along with the associated RBD1p rhomboid processing machinery, are required for the maintenance of mitochondrial morphology and function, though the exact role of Mgm1 itself has yet to be determined.[31,34] The mitochondrial rhomboid PARL has also been shown to carry out a similar function in mammalian systems, cleaving the kinase PINK1 to ensure its proper subcellular localization.[35–37] Dysregulation of these processes can lead to a loss of mitochondrial function and ultimately cell death through apoptosis, with implications for neurodegenerative diseases such as Parkinson's.[37,35,38,39]

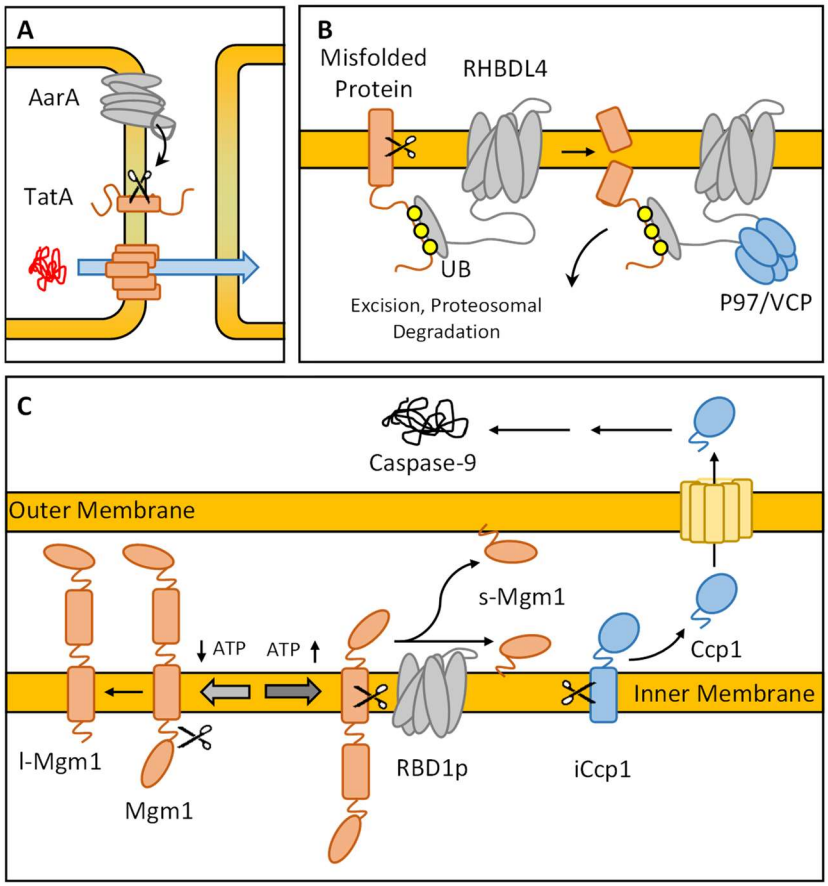


Figure 1.1 (Previous Page): Schematic diagrams illustrating the biological roles of various rhomboid proteases. A) Cleavage of the intramembrane TatA precursor protein (orange) by the *Providencia* rhomboid AarA (grey) enables the formation of membrane pores, facilitating secretion of a signalling molecule (red) critical for quorum sensing. B) The soluble N-terminal domain from the human endoplasmic reticulum rhomboid RHBDL4 (grey) contains Ubiquitin Interacting Motifs (UIM) that are able to recognize and bind to ubiquitinated, misfolded membrane proteins. Cleavage of these substrates and the subsequent recruitment of P97/VCP facilitates their excision and degradation via the ERAD pathway. C) Processing of Ccp1 (blue) and Mgm1 (orange) by the mitochondrial inner-membrane rhomboid RBD1p (grey). Cleavage of iCcp1 (right) results in the release of the mature Ccp1 from the membrane, which can then exit the mitochondria to stimulate apoptosis via the caspase signalling pathway. Processing of Mgm1 by RBD1p (left) yields either the long (l-Mgm1) or short (s-Mgm1) isoform depending on the rate of ATP-dependant translocation of the Mgm1 precursor.

1.1.4: Rhomboids and Pathogenicity

Rhomboids play a role in the pathogenicity of a range of organisms including apicomplexan parasites (e.g. *Plasmodium falciparum,* the causative agent of malaria), and bacteria (*e.g. P. stuartii* and *E. coli*), contributing to a range of human disease states. *P. falciparum* and other apicomplexan parasites express cell surface proteins known as adhesins, which bind to receptors on the host cell to initiate the internalization process.[40] Remodeling of the actin cytoskeleton drives the receptor-bound adhesins to the opposite side of the parasite cell, resulting in the invagination of the host plasma membrane and ultimately the complete internalization of the pathogen.[41] Cleavage of these adhesion proteins by the *P. falciparum* *pf*Rom1 and *pf*Rom4 rhomboids is required to release the mature parasite into the host cell, completing the invasion process. [42] Rhomboids from related apicomplexan parasites such as *Toxoplasma gondii* and *Trichomonas vaginalis* have also been shown to facilitate release from the host plasma membrane albeit through the cleavage of different or, in the case of the latter,

yet to be identified TM substrates, making them potential targets for the development of therapeutic compounds.[43–46]

As seen in 1.1.1, the *Providencia* AarA rhomboid contributes to bacterial pathogenicity by facilitating quorum sensing and biofilm formation. In contrast, the role of its *E. coli* homologue (*ec*GlpG) is less well defined. The gene encoding *ec*GlpG is located in the glpEGR operon,[47] which is potentially involved in the regulation of glycerol metabolism due to the role of GlpR as a transcriptional repressor for the glycerol-3-phosphate assimilation operon.[47] However, *ec*GlpG activity is not required for the growth and replication of *E. coli* in liquid media, raising questions regarding its biological function.[48,49] Recently it was shown that knocking out *ec*GlpG in extraintestinal pathogenic strains of *E. coli* (ExPEC) resulted in a decrease in gut colonization and pathogenicity *in vivo*.[50] This correlated with a decrease in viability when grown *ex vivo* on mucous substrates or plates with oleate as the sole carbon source.[50] GlpR knock-outs produced a similar phenotype under some of the conditions tested, and the viability of both knockout strains could be recovered through the addition of glycerol-3-phosphate. Taken together, the data suggest that *ec*GlpG contributes to the enhanced pathogenicity of ExPEC through activation of glycerol metabolic pathways that facilitate host-tissue colonization, although the exact mechanism through which this regulation occurs and the endogenous substrates involved in this process remain unknown.

The studies presented in this section highlight the ability of rhomboids to participate in a diverse range of biological processes, all of which are dependent on its ability to carry out hydrolytic cleavage within the hydrophobic membrane environment. Accordingly, there is significant interest in elucidating the structural mechanisms by which this unique proteolytic event can occur.

1.2: Rhomboid Proteolysis

1.2.1: Mechanism of Rhomboid Proteolysis

Rhomboids employ a serine protease catalytic mechanism similar to that found in many soluble proteases. Here, a catalytic serine residue within the rhomboid active site acts as a nucleophile as shown in Fig. 1.2; attacking the peptide carbonyl to produce a negatively charged tetrahedral intermediate that is stabilized by a cationic "oxyanion hole". Collapse of this tetrahedral state generates the acyl-enzyme intermediate, which is subsequently hydrolyzed via a second tetrahedral intermediate to release the final peptide cleavage products and restore the catalytic serine side chain to its nucleophilic state, as shown in Fig. 1.2. The nucleophilicity of the catalytic hydroxyl group is enhanced by interactions with a nearby histidine indole base. While most serine proteases utilize a catalytic "triad" in which an aspartic acid residue is employed to polarize the activating histidine side chain and align it for serine nucleophile activation. However, rhomboid proteases lack this third residue and rely solely on the catalytic Ser-His dyad.[51–53] Interestingly, several members of the rhomboid family lack the complete catalytic dyad.[54–56] These inactive rhomboids (iRhoms) also contain a conserved proline residue at the first x position of the GxSx motif not generally found in their active protease counterparts, though the specific role of this residue has yet to be determined.[11] Functional assays confirm the catalytically inactive status of these proteins, but suggest a role in a range of important biological functions such as cell signalling and protein trafficking.[57–59]

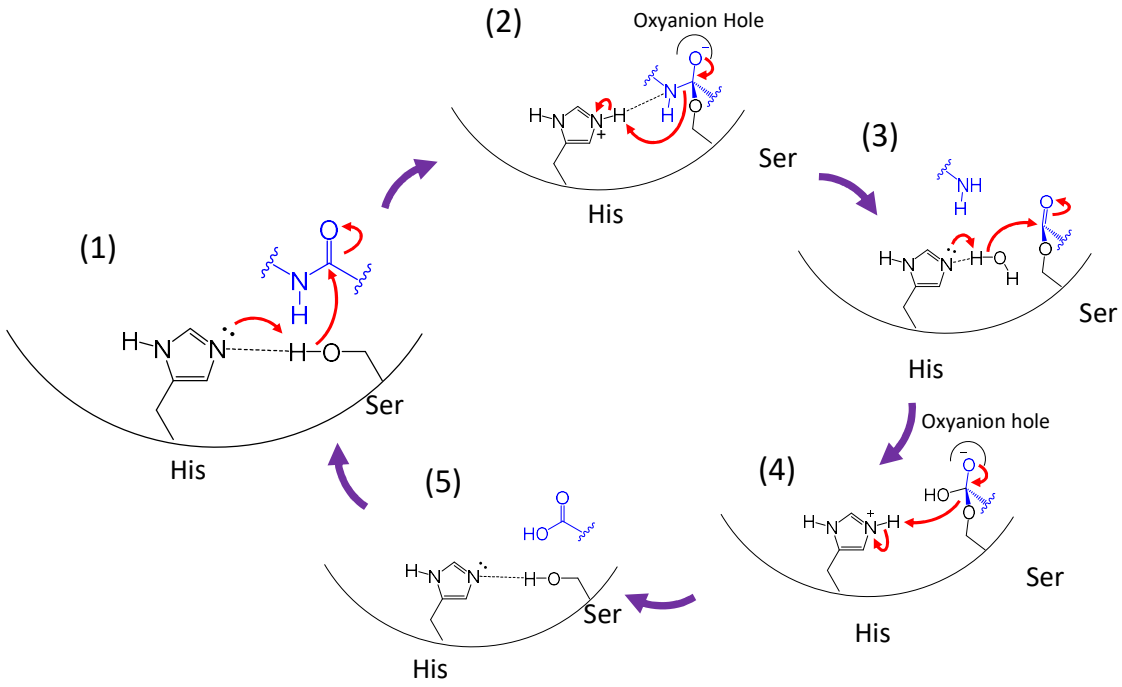


Figure 1.2: Mechanism of peptide hydrolysis catalyzed by the rhomboid serine protease catalytic dyad illustrating the initial nucleophilic attack (1), formation of the initial tetrahedral (2) and acyl-enzyme intermediates (3), and subsequent hydrolysis (4) to produce the final products and regenerate the rhomboid active site (5)

Most proteases cleave substrates at specific sequences that are often described using classic nomenclature where the cleavage occurs between the P1 and P1' positions, with Px or Px' being used to describe the residues of increasing distance N-terminal or C-terminal to the cut site (Fig. 1.3). The interaction between substrate Px/Px' side chains with protease substrate-binding pockets (denoted with the corresponding Sx/Sx') contributes to substrate recognition, though additional factors including substrate dynamics and exosite binding may also be involved (*vide infra*).

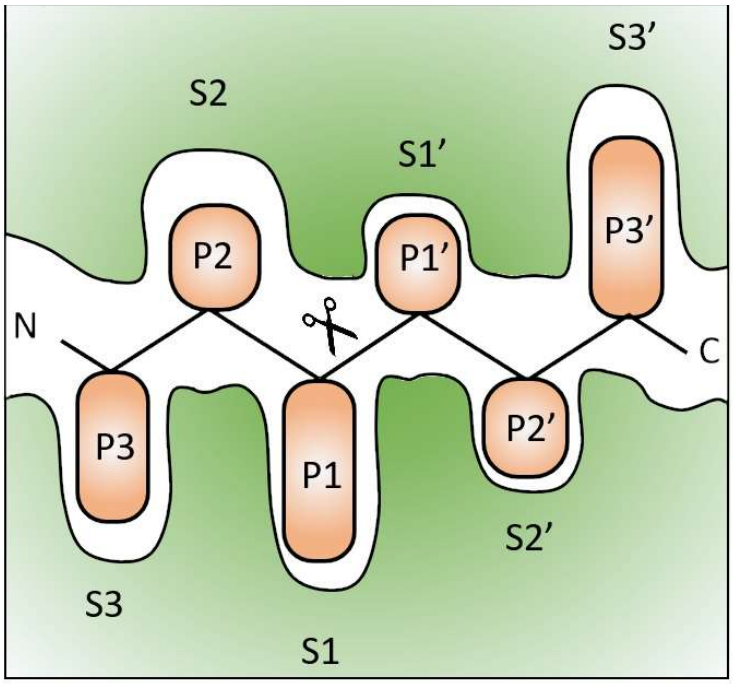


Figure 1.3: Schematic view of a protease active site showing the interaction between substrate side chains (P3-P3') and substrate binding pockets (S3-S3').

1.2.2: Protease Kinetics

The kinetics of many enzymatic processes, including rhomboid proteolysis, are approximated using the Michaelis-Menten model. Use of this model requires that the free enzyme (E) and substrate (S) are able to diffuse freely, the concentration of substrate is in excess (>10x K_M),[60] and product inhibition is negligible (therefore requiring initial reaction rates v_o to be used). According to the general kinetic scheme proposed for the Michaelis-Menten model (Equation 1.1), the binding of substrate to the enzyme to form the enzyme-substrate complex (ES) is approximated as an equilibrium between substrate-bound and free enzyme, with a dissociation constant K_D. Enzyme-catalyzed conversion of the enzyme-substrate complex into free enzyme and product can be described as a first order process with a rate constant of k_{cat}. Based on this model, the initial rate of proteolysis (v_o) is determined by the first order rate equation shown in Equation 1.2

(1.1) $E + S \underset{k_{1-}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_{cat}} E + P$ where $K_D = \frac{k_1}{k_{1-}}$ *(1.2)* $v_o = k_{cat}[ES]$

The Michaelis-Menten model assumes that the rate of ES conversion into products is so slow that the concentration of the ES complex can be considered to be approximately equal to equilibrium concentrations expected from substrate binding alone.[61] Based on this assumption, the initial rate of an enzyme catalyzed reaction can be represented by Equation 1.3:

$$v_o = \frac{k_{cat}[E]_T[S]}{K_M + [S]} \quad (1.3)$$

where K_M is known as the Michaelis constant, and is equal to the dissociation constant for the substrate-enzyme interaction.

A more general variation on this model was subsequently forwarded by Briggs and Haldane, where the concentration of the ES complex is considered to be approximately constant during the rate measurement. Using this steady-state approximation, the form of the rate equation remains the same as the Michaelis-Menten equation, however, the Michaelis constant K_M, represents a combination of k_1, k_{-1}, and k_{cat}, as shown in Equation 1.4.

$$K_M = \frac{k_{-1}+k_{cat}}{k_1} \approx \frac{k_{-1}}{k_1} \approx K_D \quad (1.4)$$

Since $k_{cat} << k_{-1}$ in most enzyme catalyzed reactions, K_M is often used as an approximation for K_D (Equation 1.4). According to the Michaelis-Menten equation, a plot of

v_o versus substrate concentration yields a hyperbolic curve as shown in Figure 1.4. The plateau of this graph represents the maximum velocity (V_{max}), where the enzyme is fully saturated with substrate (ie: $[ES] = [E]_t$), allowing calculation of k_{cat}. The substrate concentration required to reach the half-maximal rate represents K_M.

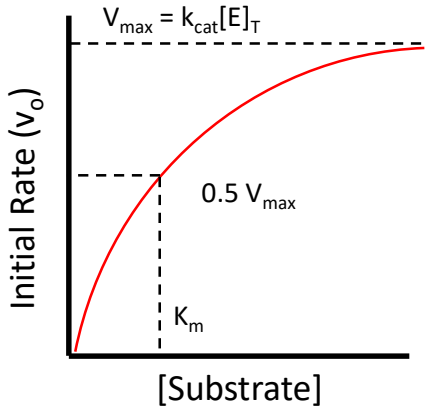


Figure 1.4: Schematic example of a Michaelis-Menten graph with corresponding kinetic constants.

This thesis examines the role of membrane interactions and conformational dynamics on the function of the *E. coli ec*GlpG rhomboid, and thus will rely heavily on kinetic analyses to quantify proteolytic activity. As discussed elsewhere in this thesis, the catalytic cycle of serine proteases, including rhomboids, involves multiple first-order processes, including the formation of an acyl intermediate (EI) and its subsequent hydrolysis as represented in the kinetic scheme shown in 1.5 where EI represents the acyl intermediate, k_2 represents its rate of formation, and k_3 represents its rate of subsequent hydrolysis:

(1.5) $E + S \underset{k_{1-}}{\overset{k_1}{\rightleftharpoons}} ES \xrightarrow{k_2} EI \xrightarrow{k_3} E + P$

(K_M = k_{1-}; k_{cat} = k_2, k_3)

While the kinetics of rhomboid proteolysis can be fit to the Michaelis-Menten equation, the physical interpretation of the kinetic constants will differ from that developed in the original model shown in Equation 1.1. For example, k_{cat} represents the aggregate rate constant for all the first order steps in the catalytic cycle, and thus the enzymatic turnover number for the overall proteolytic reaction. For kinetic scheme 1.5 k_{cat} will depend on the rate-limiting process under conditions where $k_2 << k_3$ or vice versa. Otherwise, k_{cat} can be found with the equation:

(1.6) $$\frac{1}{k_{cat}} = \frac{1}{k_2 + k_3}$$

Michaelis-Menten kinetics assumes that all substrate-binding sites behave independently. As discussed in subsequent sections, a subset of prokaryotic rhomboids have been shown to form dimers, and hence binding of substrate to one subunit could potentially alter the substrate affinity for the second subunit such that $K_D \neq K_D'$. This cooperative mode of substrate binding can be described using the Hill equation, as shown in Figure 1.5. Here, the Michaelis-Menten equation is modified by the Hill coefficient n_H, which is used to describe the degree of cross-communication between the active sites in the complex.

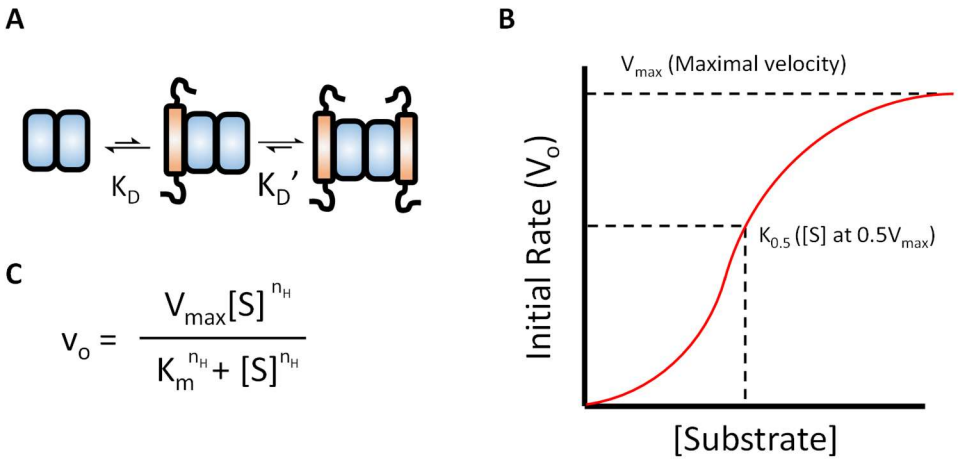


Figure 1.5: Principles of cooperative enzyme kinetics. A) Schematic showing one possible model for cooperative binding of substrate (orange) to an enzyme (blue). Note that $K_D \neq K_D'$. B) Schematic example of an enzyme activity graph for enzyme kinetics displaying positive cooperativity. C) Hill equation for cooperative kinetics.

1.2.3: Kinetic Measurements of Rhomboid Proteases

The proteolytic activity in *ec*GlpG was first shown when coexpressed in mammalian cells with potential TM substrates, the cleavage of which was monitored using Western blotting.[62] Numerous techniques have since been developed to quantify rhomboid function both *in vitro* and *in vivo*. One such technique employs mechanism-based inhibitors to monitor the active site integrity.[63] Here, a fluorophosphonate mechanism-based inhibitor coupled to a fluorophore was used to covalently label the rhomboid catalytic serine, allowing gel-based quantitation of rhomboid labeling and thus activity in crude membrane extracts without the need for further purification. The small size of the probe facilitates its entry into the active site, providing a convenient tool for the evaluation of loss/gain-of function mutations on *ec*GlpG catalytic ability.[63] Subsequent studies have used similar probes to explore the mechanism of rhomboid inactivation by isocoumarin inhibitors,[64] and verify the catalytic integrity of *ec*GlpG upon reconstitution into phospholipid vesicles and liposomes.[63,65]

Activity-based probes have also been used to monitor *ec*GlpG activity *in vivo*. In one such study, isocoumarin-based probes were incubated with live *E. coli* cells, and *in vivo* fluorescent labeling of *ec*GlpG was assessed using SDS-PAGE analysis of the resulting cell lysate.[64] While Michaelis-Menten kinetic parameters cannot generally be obtained using these methods, they have been successfully employed to determine the effect of point mutations on rhomboid activity, identify substrate sequence features required for rhomboid proteolysis, and to monitor the effectiveness of small-molecule and peptide-based inhibitors, often providing *in vivo* validation of more detailed results obtained from the *in vitro* studies described herein.[48,66–69]

The most common technique described in the literature to monitor the proteolytic activity of rhomboid *in vitro* employs a discontinuous assay in which samples of the purified rhomboid are incubated with a TM model substrate such as that derived from the second TM segment of lactose permease LacY (LY2) or, more commonly, physiological rhomboid substrates such as TatA (*Providencia*), Gurken (*Drosophia*), or Spitz (*Drosophila*), an example of which is shown in Fig. 1.6A. Samples are removed from the reaction mixture at regular time intervals, and analysed using western blotting, radiographic gel imaging or SDS page analysis to determine the amount of cleavage product produced over a given period of time and thus the rates of proteolysis. A plot of the initial rates of proteolysis against the substrate concentration can be fit to either the Michaelis-Menten model or Hill model in order to obtain the corresponding kinetic parameters as discussed in 1.2.2. [21,51,69–71]

Continuous, or “real time” assays have since been developed, the first of which employed a fluorescein iso-thiocyanate (FITC) fluorophore attached to the N-terminus of a TatA-based TM model substrate that could be quenched by interactions with the surrounding

bilayer environment. Cleavage of this substrate releases the fluorophore from the membrane surface, giving rise to an increase in fluorescence that provides a real-time readout of proteolytic activity (Fig. 1.6B).[53] However, since generation of a membrane-quenched substrate required both rhomboid and substrate to be in a lipid bilayer environment, it was necessary to reconstitute both proteins into liposomes. To limit proteolysis during the reconstitution process, a reduction in pH was used to keep the catalytic histidine in a protonated state, although studies carried out on *ec*GlpG, suggest a significant (~30%) level of rhomboid activity is likely retained at the pH values employed in this process.[52] Once the protease and substrate had been successfully reconstituted into the liposomes, maximal activity could be quickly restored by increasing the pH to 7.4 to initiate the reaction. Using this technique the authors were able to monitor proteolysis rates across a range of substrate concentrations to obtain the Michaelis-Menten kinetic parameters for a variety of rhomboid homologues. In an attempt to create a more versatile continuous assay that would not depend on the fluorescence-quenching properties of the lipid membrane, a FRET-based approach was designed in which fluorescent proteins, corresponding to the donor and acceptor fluorophores are fused to the N and C-terminal ends of a TatA model substrate, which can then be expressed and purified without a need for covalent labeling.[52] Cleavage of the TM substrate liberates the N-terminal fluorophore which can be detected as a decrease in FRET as illustrated in Fig. 1.6C, allowing for the measurement of AarA cleavage rates and the corresponding kinetic parameters in detergent micelles.[52] Curiously, neither *ec*GlpG nor *hi*GlpG were able to cleave the modified TatA model substrate, suggesting that the addition of fluorescent protein tags could potentially interfere with TatA substrate recognition and/or binding, and that further development is required before this method can be applied to other rhomboid systems.[52]

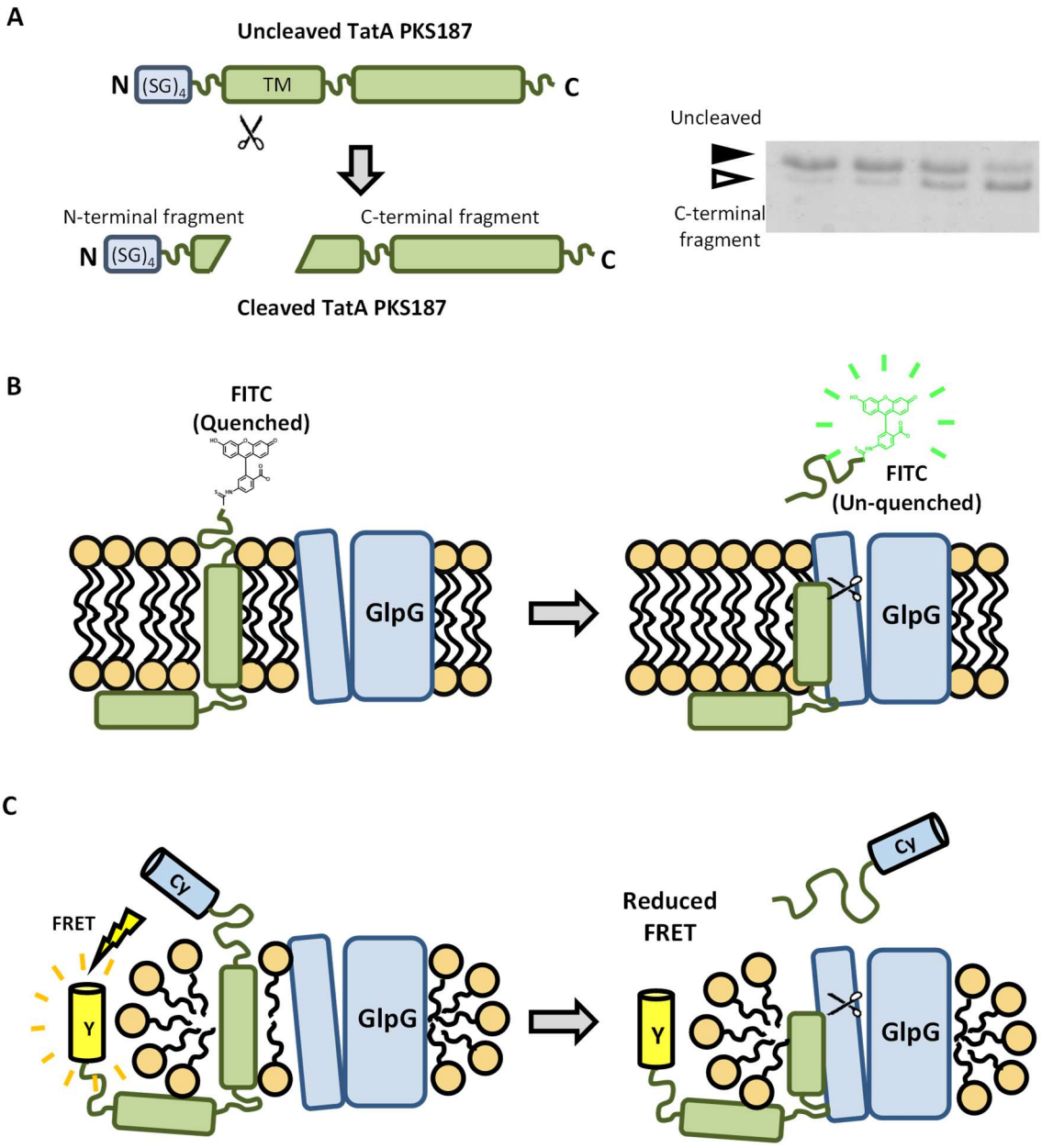


Figure 1.6: Assessing rhomboid activity against TM substrates. A) Schematic illustrating the structure of the PKS187 TM model substrate generated by extending the N-terminus of TatA by an $(SG)_4$ repeat, and subsequently employed in this thesis.[69] Cleavage of this substrate generates an N and C-terminal fragment, the latter of which can be separated and quantified using SDS-PAGE (right). B) Schematic illustrating the FITC-based TM model substrate developed to measure rhomboid activity in real time in phospholipid bilayers. C) Schematic illustrating the FRET-based TM model substrate developed to measure rhomboid activity in real time in detergent micelle systems.

In addition to the TM substrates discussed above, rhomboids such as *ec*GlpG are able to effectively cleave a fluorogenic casein protein. This commercially-available, amphipathic, water-soluble protein has been extensively labelled with a boron-dipyrromethene (BODIPY) fluorophore, whose fluorescence is reduced by quenching interactions with nearby fluorophores. Proteolytic cleavage relieves this quenching interaction, producing a fluorescent signal .[48,63] The β-casein used to produce this substrate adopts an extended, partially unstructured conformation in solution and displays a tendency to associate with both detergent micelles and phospholipid membranes, helping to bring it in close proximity to rhomboid to facilitate its cleavage. [72–77] While the β-casein sequence contains a number of potential *ec*GlpG cleavage sites (Fig. 1.7) according to the rhomboid protease consensus motif described in Chapter 1.3.2.1,[69] SDS-PAGE analysis resolves just one cleavage product suggesting that proteolysis occurs at one of the two overlapping cut sites in the middle of the sequence shown in Fig. 1.7.[78,79] While BODIPY-casein is not a native rhomboid substrate, it has been a useful reporter of rhomboid proteolytic activity, allowing detailed kinetic profiles to be acquired for a range of rhomboid homologues including *ec*GlpG.[52,79–81] In this thesis, the activity of *ec*GlpG was monitored using BODIPY-casein, along with the TM TatA-based substrate (PKS187) shown in Fig. 1.6A.

Sequence (β-casein – *Bos taurus*)

```
1    MKVLILACLVALALARELEELNVPGEIVESLSSSEESITRINKKIEKFQSEEQQQTEDE
60   LQDKIHPFAQTQSLVYPFPGPIHNSLPQNIPPLTQTPVVVPPFLQPEVMGVSKVKEAMAP
120  KHKEMPFPKYPVEPFTESQSLTLTDVENLHLPLPLLQSWMHQPHQPLPPTVMFPPQSVLS
180  LSQSKVLPVPQKAVPYPQRDMPIQAFLLYQEPVLGPVRGPFPIIV
```

Figure 1.7: Primary sequence of BODIPY-casein fluorogenic substrate showing potential *ec*GlpG cut sites.[69] The most probable cleavage site as identified by SDS-PAGE analysis is highlighted by the box.[78,79]

Although more challenging, it is also possible to measure rhomboid kinetics *in vivo* by co-expression of model substrates. Here, rhomboids are expressed within the cell at endogenous levels, while the substrate is overexpressed using an inducible operon and allowed to accumulate over the course of a fixed incubation period. Samples are then removed at fixed time intervals and the amount of N and/or C-terminal cleavage products produced, along with the remaining un-cleaved substrate can be isolated from the crude lysate and quantified using the Western blotting techniques discussed above. The amount of substrate cleaved per unit time can then be normalized against both the total amount (cleaved + un-cleaved) substrate produced, and the amount of enzyme present to obtain relative rates of proteolysis.[48,68,69,53] In order to obtain a more detailed understanding of rhomboid kinetics *in vivo*, techniques were developed in which expression of the TM substrate, in this case under the control of the pBAD plasmid, was induced with varying concentrations of arabinose, enabling *ec*GlpG proteolysis rates to be obtained over a range of substrate concentrations. The resulting data can be fit to either the Michaelis-Menten model or Hill model in order to obtain the corresponding kinetic parameters as discussed in 1.1.3.[53] The values for k_{cat} and K_M obtained using this method are similar to those measured for *ec*GlpG reconstituted into phospholipid bilayers, indicating that existing methods for assessing rhomboid activity *in vitro* provide a good representation of physiological conditions.[53]

1.3: Rhomboid Structure and Function

1.3.1: Rhomboid Structure

Since 2001, protocols have been available to purify rhomboids expressed using recombinant *E. coli* systems, allowing for the acquisition of 28 high-resolution X-ray crystal structures of the *ec*GlpG catalytic TMD, an example of which is shown in Figure 1.8 with a full list of the available structures provided in Table 1.1. While there are also two X-ray structures for the *H. influenzae* (*hi*GlpG) rhomboid,[82,83] *ec*GlpG is often used as a model to enhance our understanding of rhomboid function due to the wealth of structural information available for this enzyme and the highly conserved nature of the catalytic domain.

Table 1.1: Summary of rhomboid crystal structures reported in the literature

PDB ID	Membrane Mimetic†	Notes	Source
2IC8	β -NG	One of the first structures	Wang *et al*., 2006[78]
2NRF	β -NG	One of the first structures. Captured two antiparallel molecules within the asymmetric unit showing variation in the structure and position of α5, one in hypothetical open state	Wu *et al.,* 2006[84]
2IRV	LDAO/Mal-12	Two antiparallel molecules in asymmetric unit in which the intact α5 is rotated away from α2 relative to 2IC8. One molecule contains co-purified phospholipid bound in active site	Ben-Shem *et al*., 2007[85]
2O7L	β-NG	L5 loop region connecting α5 and α6 is disordered	Wang *et al*., 2007[86]
3B44	β-NG	W136A mutant	Wang *et al*., 2007[87]
3B45	β-NG	Wild-type control for 3B44	Wang *et al*., 2007[87]
2XOW	β-NG	In complex with covalent isocoumarin-based inhibitor.	Vinothkumar *et al*., 2010[88]
2XOV	β-NG	Native structure control for 2XOW	Vinothkumar *et al*., 2010[88]
2XTU	β-NG	Detergent-micelle control for 2XTV. S201T inactive mutant	Vinothkumar *et al*., 2011[89]
2XTV	PC-14/CHAPS (Bicelle)	First structure of *ec*GlpG in a phospholipid bilayer, uses S201T inactive mutant	Vinothkumar *et al*., 2011[89]
3TXT	β-NG	In complex with covalent diisopropylfluorophosphonate (DFP) inhibitor.	Xue *et al.,* 2012[90]
3UBB	β-NG	In complex with Cbz-Ala^{p}(O-*i*Pr)F (CAPF) phosphonofluoridate based covalent inhibitor, L5 loop region displaced upwards relative to apo structure	Xue *et al.,* 2012[91]
3ZEB	β-NG	In complex with covalent isocoumarin-based inhibitor	Voska *et al.,* 2013[64]
4H1D	β-NG	In complex with covalent diisopropylfluorophosphonate (DFP) inhibitor	Xue *et al.,* 2013[92]

Table 1.1 (Continued): Summary of rhomboid crystal structures reported in the literature

3ZMH	β-NG	In complex with covalent β-lactam inhibitor, L5 is displaced upwards relative to 2IC8	Vinothkumar *et al.,* 2013[93]
3ZMI	β-NG	In complex with covalent N-phenyl β-lactam inhibitor. L5 is unstructured	Vinothkumar *et al.,* 2013[93]
3ZMJ	β-NG	In complex with covalent N-isobutyl β-lactam inhibitor. L5 is partially unstructured, no electron density for residue F245	Vinothkumar *et al.,* 2013[93]
3ZOT	β-NG	In complex with covalent N-cyclopentyl β-lactam inhibitor. L5 is structured, in similar conformation to 3ZMH. Potential binding of second inhibitor at exosite located at the interface between α2 and α5	Vinothkumar *et al.,* 2013[93]
4NJN	β-NG	Structure obtained at pH 4.5	Dickey *et al.,* 2013[53]
4NJP	β-NG	pH 7.5 control for 4NJN	Dickey *et al.,* 2013[53]
4QNZ	β-NG	In complex with Ac-FATA-cmk peptide-based covalent inhibitor. F146I mutant	Zoll *et al.,* 2014[67]
4QO0	β-NG	In complex with Ac-FATA-cmk peptide-based covalent inhibitor. Wild-type control to 4QNZ	Zoll *et al.,* 2014[67]
4QO2	β-NG	In complex with Ac-IATA-cmk peptide-based covalent inhibitor	Zoll *et al.,* 2014[67]
5F5B	β-NG	In complex with Ac-VRMA-CHO peptide-based inhibitor. No shift in α5	Cho *et al*., 2016[10]
5F5D	PC-14/CHAPS (Bicelle)	Apo-enzyme showing ~5 A˚ shift in α5 away from α2. Y205A Mutant	Cho *et al*., 2016[10]
5F5G	PC-14/CHAPS (Bicelle)	In complex with Ac-RMA-CHO peptide-based inhibitor. ~5A˚ shift in α5 away from α2. Y205A Mutant	Cho *et al*., 2016[10]
5F5J	PC-14/CHAPS (Bicelle)	In complex with Ac-VRMA-CHO peptide-based inhibitor. ~5A˚ shift in α5 away from α2. Y205A mutant	Cho *et al*., 2016[10]
5F5K	PC-14/CHAPS (Bicelle)	In complex with Ac-RKVRMA-CHO peptide-based inhibitor. ~5A˚ shift in α5 away from α2. Y205A mutant	Cho *et al*., 2016[10]
2NR9	$C_{12}E_8$	First structure of *hi*GlpG. α5 less helical than in *ec*GlpG	Lemieux *et al.,* 2007[82]
2ODJ	$C_{12}E_8$	Structure of *hi*GlpG with disorder in α5 and adjacent loop regions L4 and L5	Brooks *et al.,* 2011[83]

* All crystals obtained from the wild-type catalytic TM core generated by proteolytic digestion of full-length *ec*GlpG beginning at residues 87 or 90 unless specified.

† All structures obtained in detergent micelles unless otherwise specified. Abbreviations are as follows. **β-NG**: Nonyl β-D-glucopyranoside. **LDAO:** N,N-Dimethyldodecylamine N-oxide. **Mal-12:** N-dodecyl-β-D-maltopyroniside. **PC-14:** 1,2-dimyristoyl-sn-glycero-3-phosphocholine. **CHAPS:** 3-[(3-Cholamidopropyl) dimethylammonio]-1-propanesulfonate hydrate. **$C_{12}E_8$:** octaethylene glycol monododecyl ether

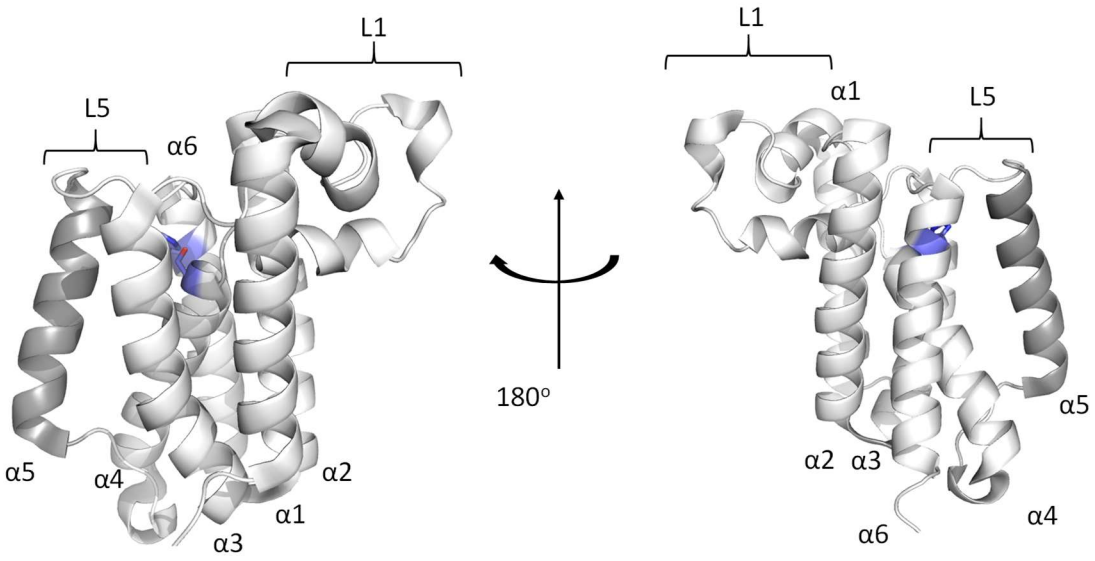


Figure 1.8: Overall structure of *ec*GlpG (PDB 2IC8) indicating the position of the L1 and L5 loop, and α5 gate region. The active site, as delineated by the catalytic dyad, is shown in blue.

Both *hi*GlpG and ec*GlpG* contain 6 TM helices, while sequence analysis suggests the presence of an additional helix in some members of the rhomboid family such as the *Providencia* AarA and *Drosophila* Rhomboid-1.[94] The catalytic serine is located on α4 while its activating histidine residue is located on α6. Both residues lie approximately 10 Å beneath the surface of the cell membrane within a water filled cavity delineated by α4, α5, and α6. In addition to housing the catalytic dyad, both the GxSx and AHxxGxxxG motifs play a role in active site integrity by maintaining the close packing of α4 and α6, ensuring the catalytic serine and histidine side chains remain within hydrogen bond distance.[62,78]

Other important residues within the *ec*GlpG active site include Y205, the loss of which significantly reduces proteolytic activity, potentially due to disrupted rhomboid-substrate interactions or altered positioning of the catalytic histidine and/or W236 on α5 (Fig. 1.9A), though studies carried out using activity-based probes suggest that this residue does not contribute to the nucleophilicity of the catalytic serine.[63,67,70,90,93] Initial structures obtained

using isocoumarin inhibitors suggest that the oxyanion hole is primarily formed by the backbone amide protons of residues L200 and S201, with hydrogen bonds to the side chains of H150 and N154.[90] However, subsequent structures obtained using β-lactam, chloromethylketone and, most recently, peptide-aldehyde inhibitors which better approximate the tetrahedral intermediate suggests that the oxyanion hole is formed by the backbone amide of S201 along with shorter (~3 Å vs ~3.5 Å) hydrogen bonds with H150 and N154 side chains as shown in Figure 1.9B. [10,88] The positioning of the oxyanion hole required that nucleophilic attack by the catalytic serine occurs at the *si*-face of the scissile peptide bond; an uncommon arrangement not usually found in other serine protease systems.[90,94]

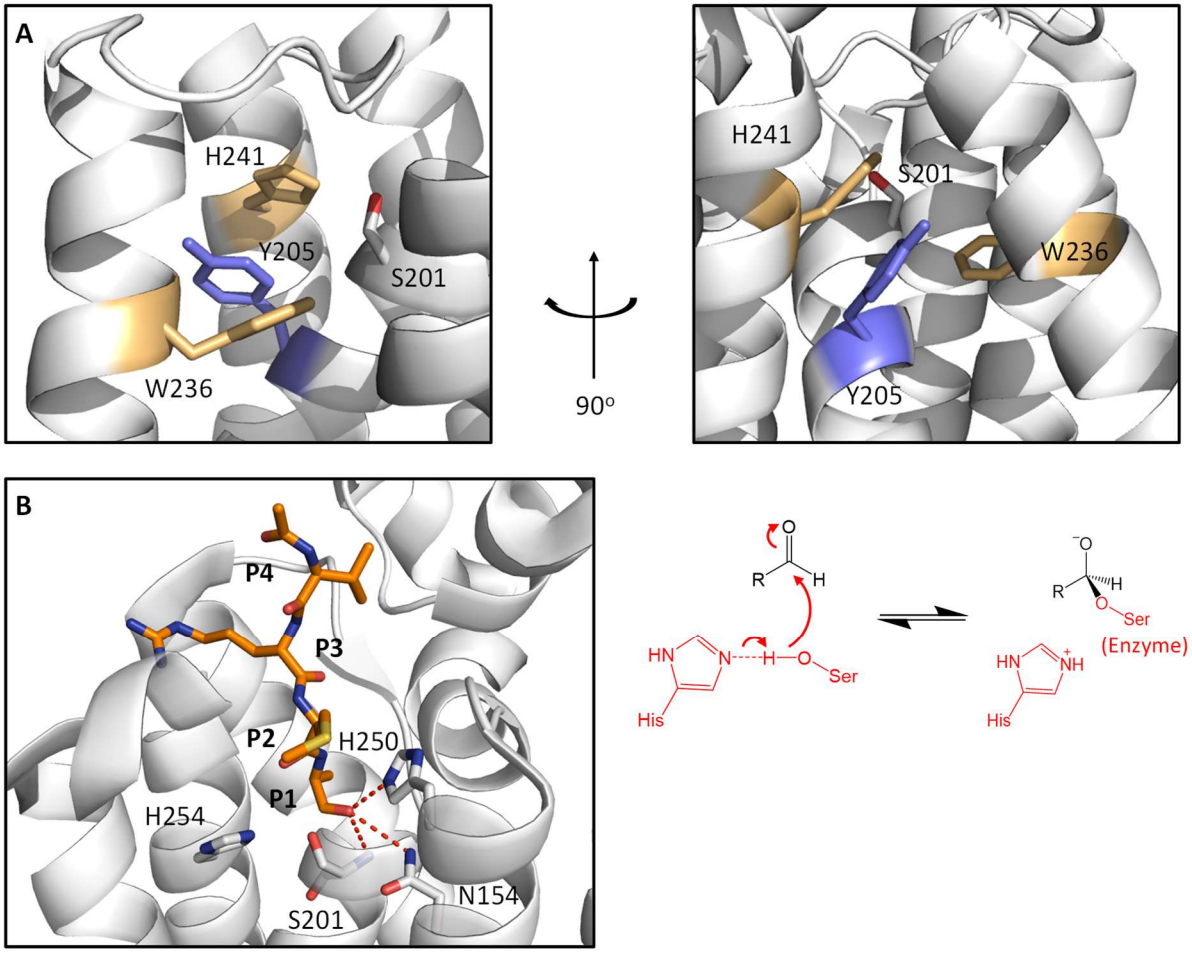


Figure 1.9 (Previous page): Detailed view of the *ec*GlpG active site A) Close-up of the *ec*GlpG active site (PDB 3B45) showing the position of Y205 (blue) relative to W236 and H251 (orange). B) X-ray structure of *ec*GlpG in complex with a peptide-aldehyde inhibitor illustrating the position of the oxyanion hole formed by hydrogen bonding interactions between the inhibitor oxyanion and the side chain protons from H250 and N154 with the backbone amide proton of Ser 201. (PDB 5F5K). The basic structure and mechanism of the aldehyde-based inhibitor employed in this study is shown at the top right, where R refers to an N-terminally acetylated peptide sequence described in Table 1.1. Note that the *si* face of the peptide bond is facing the viewer.

While mutations in helices 1-3 can influence the thermostability of *ec*GlpG, the corresponding effect on catalysis is disproportionately small, giving rise to the hypothesis that residues from this region contribute primarily to maintaining the overall rhomboid structure.[95] The L1 loop sequence between α1 and α2 forms a compact 3-helix structure that protrudes into the lipid environment, away from the core of the catalytic domain (Fig. 1.8). This structure is thought to be partially embedded in the membrane, and mutations to residues at the predicted lipid/water interface have the potential to significantly reduce proteolytic activity without significantly perturbing the overall structure,[87] suggesting a role in maintaining the correct positioning of the rhomboid TMD within its membrane environment.[95] More recent X-ray structures obtained for *ec*GlpG in complex with peptide-based inhibitors have also revealed an additional role for L1 in substrate binding and recognition (*vide infra*).

It should be noted that most rhomboids, including *ec*GlpG, include a soluble cytoplasmic domain (CytD). The cytoplasmic domains from *Drosophila* Rho4 and human RHBDL4 rhomboids contain specific ligand-binding motifs, and have been shown to influence both catalytic activity and facilitate substrate recognition.[23,80] In contrast, the functional role of the analogous structure on the *E. coli* rhomboid has yet to be determined;

while the *ec*GlpG CytD is able to form domain-swapped dimers, its removal does not appear to be detrimental to catalytic activity against TM and water-soluble model substrates, nor does it alter active site integrity when challenged using an activity-based probe.[63,79] However, the *ec*GlpG CytD can interact with some detergent micelles in a manner which disrupts its tertiary structure.[96] An analogous denaturing interaction was observed in the *Pseudomonas aeruginosa ps*GlpG rhomboid CytD, which correlated with a loss of activity under certain detergent conditions.[97] Since these denaturing interactions can complicate functional and structural studies of the catalytic domain, various truncations of *ec*GlpG were tested in this thesis to identify the optimal construct length to conserve activity and stability in the TM catalytic core.

One interesting observation to emerge from the X-ray crystal structures summarized in Table 1.1 was that the hydrophobic core of the rhomboid TMD is shorter than that of its native membrane environment by ~10 Å.[87,89,98] Given the plasticity of phospholipid bilayers relative to alpha-helical integral membrane proteins, it was proposed that the local membrane environment would deform in response to this hydrophobic mismatch, reducing its thickness in order to accommodate the shorter dimensions of rhomboid. This hypothesis is supported by the presence of a detergent or phospholipid "belt" in some X-ray structures, whose dimensions are more closely matched to the structure of the rhomboid TMD rather than the native bilayer or micellar state.[87,89] Molecular dynamics simulations performed for the TMD in a phospholipid bilayer also support this hypothesis, with deformations of the bilayer in the local vicinity of the rhomboid.[63] In 2007 the Akiyama group inserted a series of cysteine residues at various positions in the L1 loop region, of which only those positioned on the top face of L1 were reactive against the membrane-impermeable alkylating agent 4-acetamido-

4′-maleimidylstilbene-2,2′-disulfonic acid (AMS), consistent with the position of the detergent or phospholipid "belt" observed in the X-ray studies and MD simulations.[66] This provides valuable evidence that membrane deformation can occur under conditions which support proteolytic activity, potentially reflecting a physiologically relevant process. Other integral membrane proteins have been shown to introduce similar perturbations to the local membrane environment in order to minimize hydrophobic mismatch, although in some cases the structure and function of the membrane protein itself can be affected by the mismatch.[99–102]

1.3.2: Rhomboid Substrate Selectivity

1.3.2.1: Consensus Sequence for Rhomboid Proteolysis

Despite being a constitutively active enzyme when analyzed *in vitro*, the sequence selectivity of *ec*GlpG is exceptionally broad, consisting of a large hydrophobic residue at the P4 position, a small residue at the P1 position directly adjacent to the cleavage site, and a hydrophobic residue at the P2' position as illustrated in Figure 1.10.[10,69,103,104] X-ray structures of *ec*GlpG in complex with peptide-based inhibitors provide some insight into the structural basis underlying this sequence specificity.

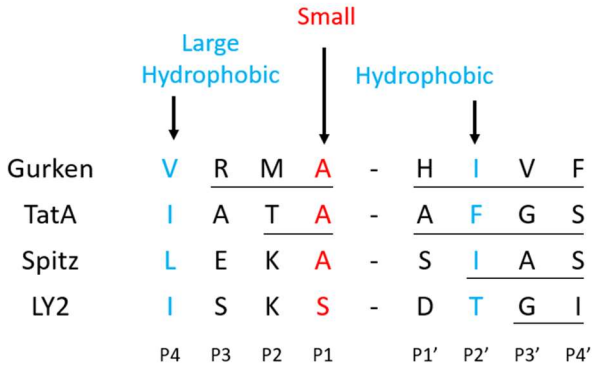


Figure 1.10: *ec*GlpG cleavage sites for various commonly employed TM substrates, illustrating the sequence features associated with *ec*GlpG rhomboid substrates, including the second TM region of lactose permease (LY2) employed in early rhomboid studies. Regions of the sequence predicted to be within the TM segment are underlined.[69,103,104]

The most recent of these structures depicts rhomboid in complex with a covalently bound tetrapeptide aldehyde with its sequence derived from the N-terminal cleavage product of the Gurken TM substrate (VRMA), shown in Figure 1.10. Although this sequence likely exists as a partly-submerged α-helix in the full-length substrate, the backbone of the bound tetrapeptide was found in an extended β-conformation, consistent with expected loss of helical structure in substrates prior to proteolysis discussed in the following section.

The P1 (Ala) side chain of the VRMA tetrapeptide points into a large cavity defined by residues H141, Q189, F197, G198, and S185. Of these residues, only the backbone of amide G198 was observed to interact directly with the P1 position, with its backbone amide forming a contact with the P1 carbonyl.[10] While such a large cavity is normally expected to accommodate large side chains at the P1 position, structures obtained of both the peptide-bound and apo forms of rhomboid show the presence of water within this cavity suggesting that it may function as a water retention site in addition to the S1 binding pocket, giving rise to the requirement for small side chains (A, S) at the P1 position.[10,105] This hypothesis is

supported by molecular dynamics studies, combined with functional studies in which the loss of water-retaining residues within this cavity impaired catalysis.[106]

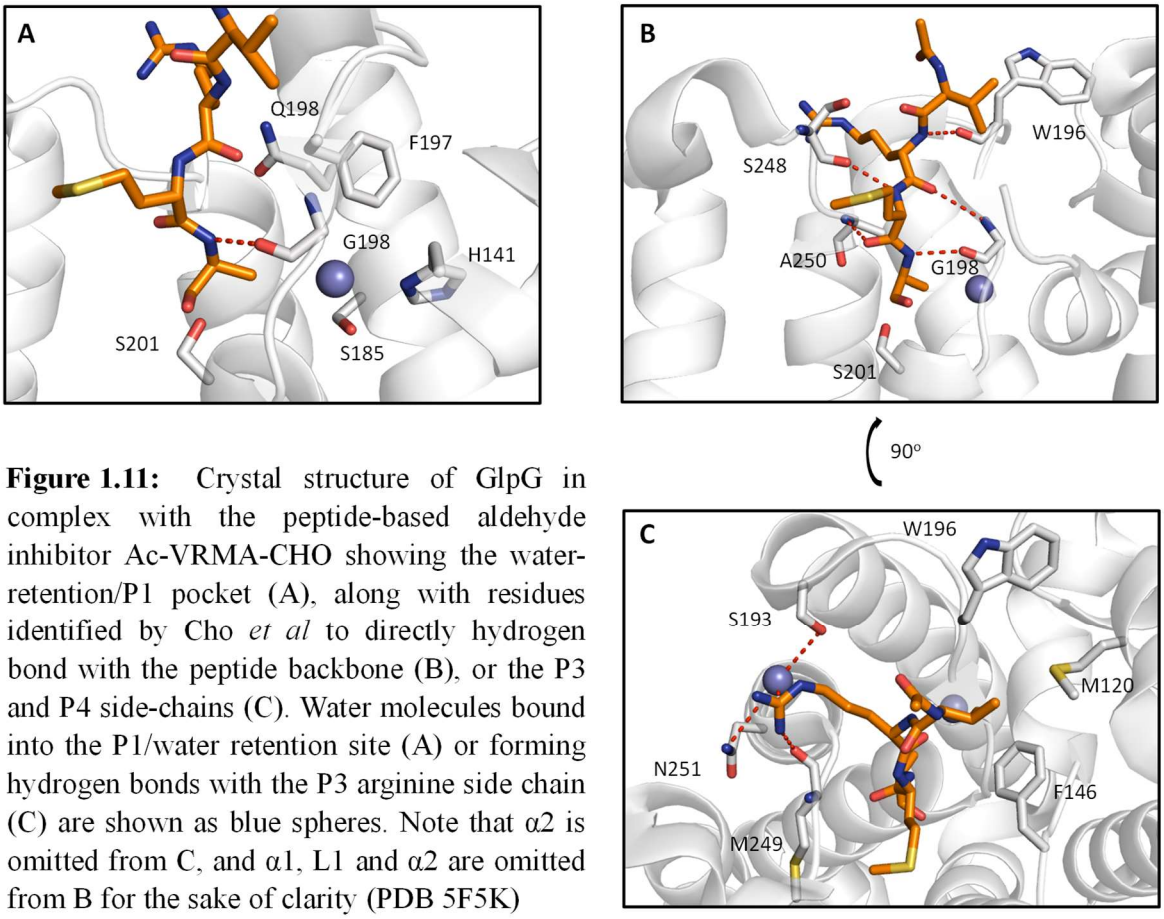


Figure 1.11: Crystal structure of GlpG in complex with the peptide-based aldehyde inhibitor Ac-VRMA-CHO showing the water-retention/P1 pocket (A), along with residues identified by Cho *et al* to directly hydrogen bond with the peptide backbone (B), or the P3 and P4 side-chains (C). Water molecules bound into the P1/water retention site (A) or forming hydrogen bonds with the P3 arginine side chain (C) are shown as blue spheres. Note that α2 is omitted from C, and α1, L1 and α2 are omitted from B for the sake of clarity (PDB 5F5K)

Further upstream, substrate residues P2 and P3 form a total of four additional backbone hydrogen bonds with residues A250 and S248 (P2), and W196 and G198 (P3) as shown in Fig. 1.11B, consistent with earlier structures depicting *ec*GlpG in complex with a TatA-derived peptidyl-chloromethylketone inhibitor.[67] The arginine side chain from the P3 residue of the bound peptide-inhibitor was also observed to make a direct hydrogen bond with the backbone of M249, while hydrogen bonds with a bridging water molecule allowed

additional interactions with side chains of S193, S248, and N251 (Fig. 1.11C). The importance of arginine at this position was confirmed by the approximately 10-fold decrease in binding affinity observed upon its mutation to an alanine. Since the P3 arginine is not found in other known TM substrates (Fig. 1.10), this interaction may be unique to the *E. coli* rhomboid. While the P4 peptide backbone does not make any direct hydrogen bonds with the rhomboid, its side chain fits into a shallow hydrophobic pocket formed by side chains from residues M120, F146, and W196, potentially accounting for the preference for large, hydrophobic residues at this position.[10,69] Longer peptide aldehyde inhibitors that include the Gurken P5 and P6 residues (RKVRMA) were even more effective inhibitors of *ec*GlpG than the shorter tetrapeptide inhibitors, suggesting the presence of additional rhomboid-substrate contacts at these positions, although no electron density could be resolved for residues beyond the P4 position in the corresponding X-ray crystal structure.

Other information on substrate interactions has been provided by X-ray structures of *ec*GlpG bound to isocoumarin, diisopropylfluorophosphonate (DFP), and monolactam inhibitors that delineate a potential S2' pocket formed by the side chains of V204, Y205, M208, W157, and W236, enabling it to accommodate the hydrophobic substrate residues commonly found at this position (Fig 1.12).[105] Unfortunately, no structures are currently available which depict a rhomboid in complex with a peptide-based inhibitor that extends beyond the P1 position to verify this hypothesis. Structural and functional studies indicate that residues beyond the P2' position do not form specific interactions with the *ec*GlpG active site, though they may still contribute to rhomboid selectivity by influencing substrate dynamics or exosite binding, as discussed below.[10,69,52,103]

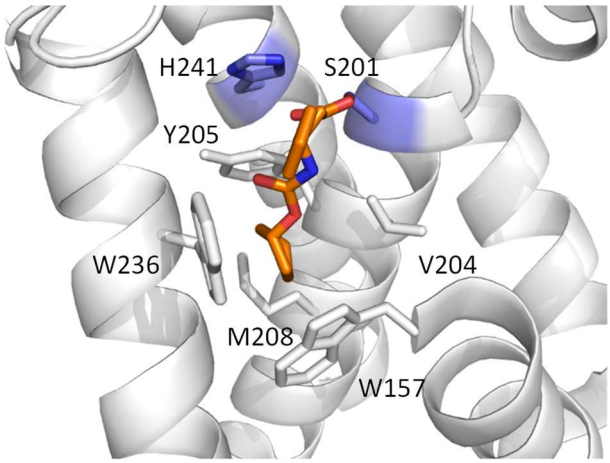


Fig 1.12: Crystal structure of Crystal structure of *ec*GlpG in complex with the monolactam L62 inhibitor (orange). Residues proposed to form the S2' pocket are indicated. The catalytic dyad is shown in blue. Note that the extracellular face of α2 (residues 146-158) has been omitted for the sake of clarity (PDB 3ZMH).

Most of the interactions observed in the rhomboid-inhibitor structures correlate well with functional studies examining the effect of *ec*GlpG mutations and substrate sequence on proteolytic activity.[10,95,105] However, the consensus sequence defined by these interactions as shown in Fig. 1.11 remains unusually broad, enabling the cleavage of a range of non-physiological substrates when assessed both *in vitro* and *in vivo*.[103] Given that rhomboids appear to be constitutively active, it is likely that there are additional factors which help to direct rhomboid selectivity beyond the interactions highlighted in this section.[103]

1.3.2.2: Role of Substrate Dynamics in Rhomboid Selectivity

In many rhomboid substrates, proteolytic cleavage occurs in regions predicted to be helical and potentially buried below the surface of the membrane.[69] The ability of rhomboids to alter the local membrane environment as discussed in 1.3.1 might help to expose a portion of the TM substrate to the aqueous environment as it approaches the protease, which would be expected to destabilize the helix in this region and facilitate its conversion into the extended

structure seen in crystal structures of *ec*GlpG in complex with the peptide-aldehyde inhibitors discussed above. Functional studies show that the presence of helix-destabilizing residues within the substrate TM segment facilitates proteolysis and the position of these residues can influence the site at which this cleavage occurs, suggesting that substrate dynamics and stability can provide an additional layer of selectivity.[69,53,103,104,107]

A 2013 study by the Urban group provided additional evidence supporting this hypothesis through a comparative analysis of the cleavage rates for various rhomboid homologues against several TM substrates. In all instances, the K_m values were significantly higher than the substrate concentrations that could be achieved in the membrane environment under normal physiological conditions.[53] This result was interpreted as an indication that rhomboid proteolysis *in vivo* is independent of binding affinity, with selectivity being determined exclusively by factors that affect k_{cat}, such as substrate gating and helix stability. However, a subsequent study by the Lemieux group examining the ability of different rhomboids to cleave the *Providencia* TatA rhomboid substrate contradicts this hypothesis, since physiologically relevant K_m values ranging from 7.6 μM to 120 μM were measured, with the lowest Km being observed when TatA is cleaved by its native AarA rhomboid.[52] Therefore, while rhomboids appear to be capable of cleaving a wide range of non-native substrates, sequence specificity appears to be important for determining activity against biologically relevant substrates.

1.3.2.3: Role of Cooperativity and Allostery in Rhomboid Selectivity

Curiously, the kinetic data obtained for both AarA and *hi*GlpG against TatA showed that substrate binding occurs in a cooperative fashion, with Hill coefficients ranging from 1.7-

2.1.[52] This suggests that these rhomboids self-associate, a feature that was independently confirmed by equilibrium sedimentation ultracentrifugation and size-exclusion chromatography, both of which showed a dimeric state for *ec*GlpG, *hi*GlpG, and AarA in n-dodecyl-β-D-maltoside (Mal-12) detergent micelles, while pull-down assays carried out on the *hi*GlpG demonstrate self-association occurs *in vivo*.[52,108] Conditions that disrupt dimer formation, such as solubilisation in n-decyl-β-D-maltoside (Mal-10) resulted in a loss of proteolytic activity for all three rhomboids, indicating that the dimer represents the physiologically-relevant functional state. However, none of the rhomboids tested displayed cooperative activity against the non-physiological casein model substrate, suggesting that cooperativity involves interactions with TM substrates that does not occur with casein.[52] It should be noted that the *E. coli ec*GlpG studied in this thesis was also found to exist as a dimer yet displayed only weak cooperativity, although none of the tested substrates are physiological targets for this rhomboid.

In the same study on rhomboid cooperativity, the TM TatA substrate was found to inhibit the cleavage of casein by AarA in a non-competitive fashion, suggesting the presence of a substrate binding site (exosite) outside the active site which can facilitate the recognition and binding of TM substrates, as shown in Fig.1.13. Similar results were reported with the peptide-aldehyde inhibitor depicted in Fig. 1.11, which was also a non-competitive inhibitor of TatA cleavage by both AarA and *ec*GlpG.[10,52]

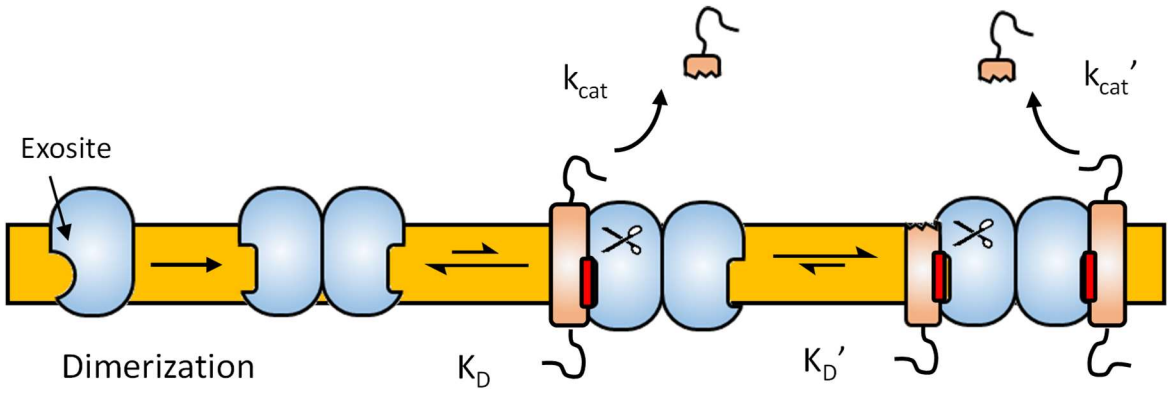


Figure 1.13: Exosite-mediated recognition of rhomboid substrates proposed by Arutynova *et al*.[52] Dimerization of rhomboid (blue) causes a conformation change in a membrane exposed exosite. Binding of substrates via an exosite recognition motif (red) can facilitate the binding of additional substrate molecules to the second dimer subunit to reduce K_D (K_D'), and/or enhance its catalytic turnover rate (k_{cat}').

Overall, these studies paint a complex picture of rhomboid selectivity, where the relatively broad sequence specificity discussed in 1.3.2.1 is modulated by a range of factors beyond the active site, such as substrate helix propensity, cooperativity and exosite binding. However, while these factors were found to enhance the activity and binding affinity of AarA against its physiological substrate, there does not appear to be any mechanism to hinder its ability to cleave non-physiological substrates such as casein, against which the physiological dimeric form displays similar values for K_M (2.5 μM vs 3.9 μM) and k_{cat} (0.37 min^{-1} vs 0.17 min^{-1}).[52] This suggests the presence of additional mechanisms by which TM substrates and non-target peptides can be differentiated *in vivo*, some of which will be explored during the course of this thesis.

1.3.3: Substrate Gating in Rhomboid Proteases

1.3.3.1: α5 as a Lateral Gate for Substrate Entry

The α-helical structure of the rhomboid TMD serves to isolate the aqueous active site from the hydrophobic membrane environment. While this limits the potential for unfavourable interactions between these two disparate environments, it raises the question as to how TM substrates are able to gain access to the active site. Early hypotheses identified the L1 region as a potential substrate gate due to its location between α1 and α3 where it blocked the V-shaped opening between these helices, along with the detrimental effect on activity when mutations were introduced into this region.[82] However subsequent X-ray structures, including one obtained for a loss-of-function L1 mutant suggest a high level of structural rigidity inconsistent with a role in substrate gating.[70,84,87,92] Instead, enzymatic analysis of mutants predicted to disrupt the interaction between α5 and α2 (Table 1.2, Fig. 1.14) showed enhanced proteolytic activity, leading to the suggestion that the movement of α5 facilitates substrate entry into the active site.[70,53,80,92,95] Curiously, these enhancements to proteolytic activity are primarily mediated by changes to k_{cat} rather than $K_{M,}$ suggesting that substrate gating may represent the rate-limiting step for the overall rhomboid catalytic process.[70,53,92] This is in contrast with a number of proteases with the rate-determining step involving hydrolysis of the acyl-enzyme intermediate, as indicated by relatively large solvent kinetic isotope effects (KIE), with values ≥2. Rhomboids display a relatively modest (1.26) KIE in deuterated solvents, consistent with the hypothesis that substrate gating rather than hydrolysis is rate-limiting.[53]

Table 1.2: Summary of rate enhancements by mutations around the substrate gate measured either in phospholipid liposomes or Mal-12 detergent micelles

Mutant	Fold Enhancement (Approximate)	Substrate	Medium	Paper
F153A/W157A	5	Spitz	Mal-12	Baker and Urban, 2012[95]
F153A/W236A	10	Spitz	Mal-12	Baker *et. al.,* 2007[70]
F153A/W236A	2	Casein	Mal-12	Sherratt *et al.,* 2012[63]
	None	Activity based probe	Mal-12	Sherratt *et al.,* 2012[63]
	10	Spitz	Mal-12	Xue and Ha, 2013[92]
	40	Gurken	Mal-12	Xue and Ha, 2013[92]
	3	TatA	Liposomes	Dickey *et al.,* 2013[53]
F153A/W236G	15	Spitz	Mal-12	Baker and Urban, 2012[95]
	10	Spitz/APP fusion[104]	Liposomes	Baker and Urban, 2015[80]
	None	Casein	Mal-12	Baker and Urban, 2012[95]
W157A	None	Spitz	Mal-12	Baker and Urban, 2012[95]
W157A/F232A	7	Spitz	Mal-12	Baker *et. al.,* 2007[70]
	5	Spitz	Mal-12	Xue and Ha, 2013[92]
	7	Gurken	Mal-12	Xue and Ha, 2013[92]
F232A/W236G	5	Spitz	Mal-12	Baker and Urban, 2012[95]
L234P	5	Spitz	Mal-12	Baker *et al.,* 2007[70]
W236G	5	Spitz	Mal-12	Baker and Urban, 2012[95]
	5	Spitz/APP fusion[104]	Mal-12	Baker and Urban, 2015[80]
	None	Casein	Mal-12	Baker and Urban, 2015[80]

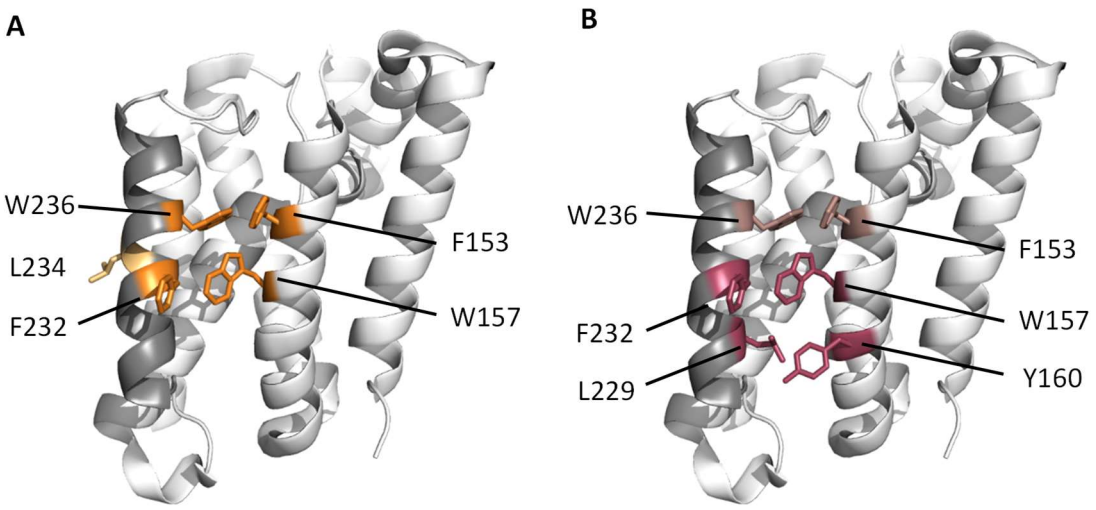


Figure 1.14: Investigations into the role of α5 dynamics in substrate gating. A) Structure of *ec*GlpG showing the location of residues which, when mutated to disrupt interactions around the putative substrate gate, enhances proteolytic activity (highlighted in orange). The L234, which when mutated to the helix-breaking proline residue also enhances activity, is shown in light orange. B) Structure of *ec*GlpG (PDB 3B45) showing the location of residues tested for cysteine cross-linking, with residues identified to be involved in substrate gating are shown in dark red. α5 is shown in grey for both structures.

Further evidence for a role of α5 dynamics in substrate gating was provided by electron paramagnetic resonance (EPR) experiments showing a correlation between the increased α5 dynamics in micelles relative to phospholipid bilayers, and increased proteolysis rates.[104] In 2006 the Shi group reported an X-ray structure which appeared to depict an alternate conformation of *ec*GlpG in which α5 appeared to be tilted away from α2. This was accompanied by an upward movement of the L5 loop and a slight distortion to the cytoplasmic end of α6 that exposed a gap through which substrates might enter the active site (Fig. 1.15A), providing a structural model for substrate gating.[70] However, of the 15 other apo-enzyme *ec*GlpG structures available in the Protein Data Bank (summarised in Table 1.1), none have reproduced the ~35° rotation of α5 observed in this model. Even the second molecule in the asymmetric unit of this same crystal structure adopts the canonical closed conformation.

Given that the α5 helix in this open structure is engaged in significant crystal packing interactions with other *ec*GlpG molecules in the crystal lattice, there is a strong possibility that the perturbation observed in α5 could represent an artifact of crystallization, rather than an accurate depiction of a physiologically relevant open state.

In line with this conclusion, subsequent structures of *ec*GlpG in complex with a range of covalent inhibitors, including isocoumarins, fluorophosphonates, monolactams, and chloromethylketones failed to reproduce the α5-tilted gate-open structure. Instead, some of these structures depict the entire α5 helix shifted laterally by ~1 Å,[90] hinting at a different mode of substrate gating. In 2016 the Urban group was able to provide more structural evidence for this model via a series of X-ray crystal structures in which α5 is displaced laterally away from α2 by ~5 Å, accompanied by smaller displacements of helices α6 and α2 relative to α4 (Figure 1.15B). While this structure could be obtained in the presence of peptide-aldehyde inhibitors derived from the Gurken TM substrate similar to those depicted in Figure 1.11, it could only be captured using a functionally-impaired Y205F construct solubilized in phospholipid bicelles; structures which did not meet *both* these requirements, such as those depicting WT *ec*GlpG in bicelles (PDB 2XTV) or the Y205 mutant in detergent micelles (5F5B), yielded only the closed conformation, even when bound to the peptide-aldehyde inhibitor.[10] This requirement for bicelles is particularly surprising, since previous studies had suggested that solubilisation in detergent micelles would enhance rhomboid gate dynamics and facilitate formation of the open state.[104] Based on these concerns, the physiological relevance of this gate-open conformation remains a matter of debate. Nonetheless, both these models, and the associated functional and structural studies put forward to support them, represent a growing body of evidence that α5 plays a key role in the

substrate gating and proteolysis, making it a potential point of control for the rhomboid catalytic cycle

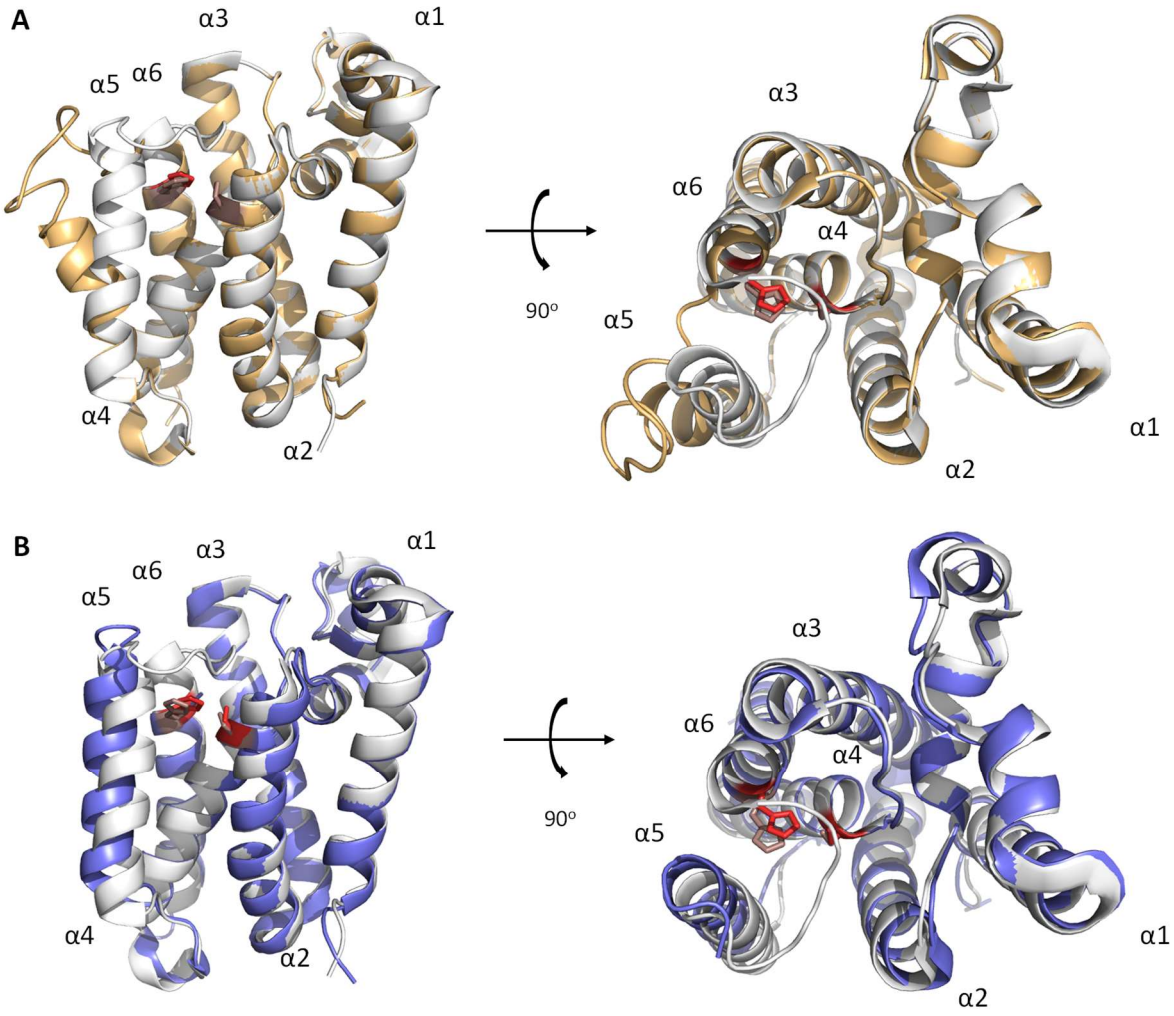


Figure 1.15: The gate-open conformation showing a5 in a bent conformation (A) or laterally displaced (B) shown in orange and blue respectively, superimposed on the closed conformation (white). Catalytic dyad is shown in red. (PDB 3B45, 2NRF, 5F5K)

1.3.3.2: Alternate Models of Substrate Gating

In 2007, the Ha group reported a crystal structure for *ec*GlpG in which no electron density could be resolved for the L5 loop region, suggesting a high degree of conformational flexibility. In its place were several water molecules, giving rise to the hypothesis that L5 can

act as a flexible loop to permit substrate access directly from the extracellular environment, potentially aided by the membrane deformation described in 1.3.1 and 1.3.2.2.[86,87,109] In line with this hypothesis, subsequent structures obtained for *ec*GlpG in complex with both small-molecule and peptide-based inhibitors also showed differences in the structure and position of the L5 loop but relatively little movement in α5 relative to the apo structure, suggesting that movement of the L5 loop alone may be sufficient to permit entry of peptide fragments and potentially full length substrates directly from the aqueous phase.[10,90] This was supported by cross-linking studies on a series of cysteine mutants introduced at the interface between α2 and α5 (Figure 1.14). The 1,2-ethanediyl bismethanethiosulfonate (M2M) cross-linking agent employed in one of these studies incorporates a 5.2 Å spacer that allows it to accommodate the 8.4 - 10.4 Å gap between the two helices without distorting the *ec*GlpG active site. Despite the restrictions this would have imposed on the movement of α5, cross-linking with M2M did not reduce the rate of proteolytic activity against a Spitz TM substrate whose cleavage site is predicted to be solvent exposed (Figure 1.10).[92] This appears to contradict an earlier study where direct cross-links between cysteine residues on α2 and α5 severely hindered *ec*GlpG activity against Gurken, whose cut site lies within the TM segment.[70] However, the shorter length of the disulfide bridges employed in this study could potentially introduce perturbations to the *ec*GlpG active site, making it difficult to directly compare results between the two studies. Nonetheless, these results suggest that movement of the L5 lid alone is sufficient for the cleavage of some substrates, though displacement of α5 may still be required for substrates such as Gurken or TatA, whose cut sites are contained within a buried TM helix. Overall, these studies are part of a growing body of evidence suggesting that dynamics in the α5/L5 region of *ec*GlpG play a critical role in substrate gating. However, the structural nature of

gating, and its role in rhomboid catalysis and selectivity remains a matter of debate and will be explored further in this thesis

1.4: Summary and Outlook

Since their initial discovery less than 20 years ago, significant progress has been made towards understanding the structure and function of rhomboid proteases as highlighted by the studies presented in this Chapter. However, several questions still remain concerning how helical TM substrates are able to gain access to the buried rhomboid active site, and the potential consequences of substrate gating on rhomboid structure and function.

The ability of rhomboids to manipulate the structure and dimensions of its membrane environment has been suggested by results from previous studies, and has given rise to the hypothesis that membrane deformation is required to facilitate substrate entry. However, the energetic cost of this interaction, and its potential to influence rhomboid structure and function, prior to the work presented in this thesis, had yet to be explored. Moreover, the mechanism through which substrates can gain access to the catalytic residues while minimizing exposure of the hydrophilic active site to the hydrophobic membrane environment was not also resolved. As highlighted in section 1.3.3 there is substantial evidence suggesting that conformation changes in the α5/L5 regions are involved. However, there remains some ambiguity concerning the structure of the physiologically-relevant "open" conformation, while the ability of some substrates to be cleaved under conditions that should restrict α5-opening raises questions concerning its contribution to the rhomboid catalytic cycle.

The three chapters in this thesis will explore these questions through the use of functional studies, such as those described in 1.2.3 coupled with structural techniques such as

solution-state NMR and CD spectroscopy, the theoretical bases of which will be introduced in the subsequent chapter. This will be followed by Chapter 3, which describes an investigation into the effect of hydrophobic mismatch on the structure and function of *ec*GlpG, providing new insight into the ability of the local environment to influence its catalytic activity. Chapter 4 describes the development and optimization of sample conditions for solution NMR studies of the *ec*GlpG TM catalytic domain, laying the groundwork for Chapter 5, which employs solution NMR to characterize the structure of the α5-open conformation. Functional assays carried out using both TM and water-soluble model substrates are used to demonstrate the role of both α5 dynamics and hydrophobic mismatch in the regulation of rhomboid activity and substrate selectivity, with potential implications for the function of *ec*GlpG in its native biological setting.

Chapter 2: Introduction to the Theory and Applications of Solution NMR and CD Spectroscopy

Preamble

This thesis employs both solution-state NMR and CD spectroscopy to provide structural information on rhomboid proteases. NMR is a technique that can be used to obtain information on the conformational dynamics of proteins, allowing it to complement the high-resolution structural information provided by X-ray crystallography. In this thesis, specific NMR signals were used to monitor the structure and dynamics of targeted regions of the *ec*GlpG structure, with a focus on the putative α5 substrate gate. CD spectroscopy is a complementary technique that can be used to monitor protein secondary structure content. In this thesis CD was used to assess the effect of point mutations and/or the local membrane environment on the thermostability and structure of *ec*GlpG, particularly under conditions that are not conducive to the acquisition of high-quality NMR spectra. This chapter aims to provide some introduction to theoretical and practical aspects of these two techniques, with a focus on features that are pertinent to the results presented in this thesis.

2.1: NMR Spectroscopy

2.1.1: Basic Principles of NMR Spectroscopy

NMR spectroscopy utilizes the magnetic properties of NMR-active nuclei to yield information on its local chemical environment. The nuclei typically used in protein NMR are ^{1}H, ^{15}N, and ^{13}C nuclei, all of which have intrinsic angular momentum, also referred to as spin, with a quantum number of ½. Application of an external magnetic field (B_o), as is provided by an NMR spectrometer, induces a dipole in these spins, which will orient in a parallel or antiparallel configuration relative to B_o. These states are separated by a small energy difference, and transitions between these two states can be induced by absorption of

energy in the radiofrequency range. This frequency is commonly referred to as the Larmor frequency (ω_o), which is defined by Equation 2.1;

2.1) $\omega_o = \gamma B_{eff}$

Where B_{eff} is the strength of the effective magnetic field experienced by the nucleus, and γ is its gyromagnetic ratio. The Larmor frequency is expressed in units of $rads^{-1}$ (ω_o), which can be converted into units of Hz (v_o) by dividing by 2π. The effective magnetic field experienced by a particular nucleus (B_{eff}) includes contributions from a number of factors, such as electron density, bonding environment and non-covalent interactions, all of which can introduce localized perturbations to the magnetic field. This is usually reported as a chemical shift (δ_{ppm}), which takes the difference in Larmor frequency of the spin (v_{sample}) from that of a reference compound (v_{ref}) and normalizes it against the Larmor frequency for the reference compound according to 2.2:

2.2) $\delta_{ppm} = \frac{v_{sample} - v_{ref}}{v_{ref}} * 10^6$

When carrying out ^{1}H NMR on proteins, 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) is employed as the reference compound. This is due to the fact that the protons in DSS methyl groups are more shielded from the external magnetic field than most other nuclei typically in a protein sample, allowing it to effectively demarcate one end of the chemical shift range likely to be encountered.

NMR spectroscopy has inherently low sensitivity due to the very small population difference between low and high-energy states. This small excess population of spins in the

low energy state gives rise to a net magnetization vector oriented along B_o (designated as the z-axis, by convention) that precesses about this axis with a frequency equal to ω_o. Application of electromagnetic radiation matching the Larmour frequency in a direction that is perpendicular to B_o will cause the bulk magnetization vector to rotate away from its equilibrium position along the z-axis. When the bulk magnetization vector is in the transverse plane, its precession in this plane can be detected as an oscillating field with a frequency of ω_o. Since the ability to detect this signal depends on the magnitude of the bulk magnetization vector, its relatively small size, even at large magnetic field strengths, makes NMR an inherently insensitive technique.

The sensitivity of an NMR experiment is also affected by spin relaxation, which is the return of excited spins back to their equilibrium populations, a phenomenon that also leads to loss of the NMR signal over time. This can be further broken down into two separate processes: spin-lattice relaxation, which is defined by the relaxation time T_1, and spin-spin relaxation, defined by T_2. Both of these forms of relaxation depend on fluctuations in the local magnetic field generated by the motion of nearby nuclei, primarily through rotational diffusion of molecules in the solution state. The rate at which these oscillations occur, and thus the rate of relaxation, can be broadly correlated to the rate of molecular tumbling which is described by the rotational correlation time (τ_c), with smaller molecules giving rise to shorter correlation times. Since relaxation is induced by oscillations which occur at certain frequencies, molecular motion is sometimes described using a spectral density function $J(\omega)$ which represents the power available from molecular motions to induce relaxation as a function of frequency, and can be described by a Lorenzian function shown in Equation 2.3.

2.3) $J(\omega) = \overline{B_{loc}^2} \frac{2(\tau_c)}{1+\omega^2\tau_c^2}$

Systems in which $\tau_c >> 1/\omega_o$ are classified as belonging to the slow motion regime, under which the overwhelming majority of power available is around zero frequency (ie: $J(0) >> J(\omega_o)$). In contrast to this, the spectral density function of systems which exist in the fast motion regime ($\tau_c >> 1/\omega_o$) is spread out across a wide range of frequencies as shown in Fig. 2.1.

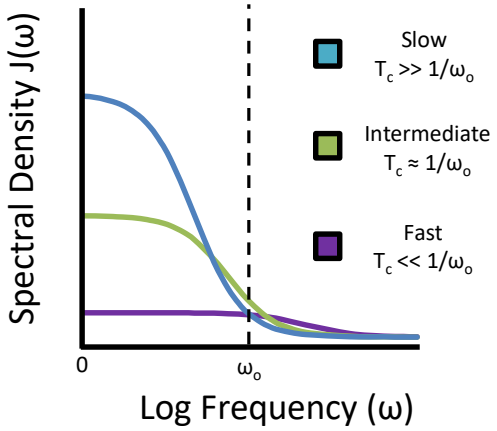


Figure 2.1: Spectral density function for systems undergoing slow ($\tau_c >> 1/\omega_o$), intermediate ($\tau_c \sim 1/\omega_o$), and fast ($\tau_c << 1/\omega_o$) molecular reorientation.

Spin-lattice, also known as longitudinal or T_1 relaxation, refers to the rate at which the normal equilibrium distribution of spins between high energy and low energy states is re-established. Due to the low sensitivity of NMR, multiple scans are usually required to generate spectra with useful signal-to-noise levels. Shorter T_1 times allow the delay between NMR scans required for the re-establishment of equilibrium populations to be reduced, thereby increasing the number that can be obtained over a given period of time. T_1 relaxation results from dipole-dipole interactions, and chemical shift anisotropy. In the case of the former,

dipole-dipole interactions with adjacent nuclei produce local magnetic fields (B_{loc}) that vary, depending on the relative position of these nuclei with respect to the external magnetic field. Chemical shift anisotropy also produces oscillating fields, since the electronic environment surrounding a given nucleus can be asymmetric, giving rise to small variations in the effective magnetic field depending on the orientation of the molecule relative to B_o. Local magnetic fields oscillating at 0, ω_o or $2\omega_o$ can stimulate transitions between high energy and low energy states to restore equilibrium populations, contributing to the rate of T_1 relaxation (T_1^{-1}) as shown in Equation 2.4: [110].

2.4) $T_1^{-1} = K^2[J(0) + 3J(\omega_o) + 6J(2\omega_o)]$ where $K = \frac{\gamma^2\hbar}{2r^3}$

where K represents a combination of constants including the gyromagnetic constant of the nucleus in question (γ), the Plank constant ($\hbar$), and the distance between the two interacting magnetic fields (r).

Since proteins are large molecules with τ_c values >> $1/\omega_o$ they fall into the slow motional regime in which the majority of the spectral density centered is concentrated in the low frequency range around J(0), reducing the power available at the higher frequencies (ω_o, $2\omega_o$) which promote effective T_1 relaxation. This is in contrast to small molecules with extremely short correlation times approaching the fast motion limit described in which the spectral density is spread across a large range of frequencies, again decreasing the power available at the specific frequencies required for effective T_1 relaxation.

Spin-spin, or transverse (T_2) relaxation refers to the loss of coherence in the magnetization in the transverse plane and plays a significant role in determining NMR

sensitivity, particularly in proteins and other large molecules with τ_c values in the slow motional regime. Recall that the bulk magnetization vector arises from the sum of the individual magnetic dipoles that all precess at a rate that depends on the effective magnetic field. Small variations in the effective magnetic field induced by molecular tumbling will cause changes in the precession frequencies of individual spins, resulting in the loss of coherence or dephasing of the bulk magnetization vector over time. In addition, the same processes which give rise to longitudinal (T_1) relaxation will also contribute to transverse relaxation, giving rise to the equation depicted in 2.5:[110]

2.5) $T_2^{-1} = K^2\left[\frac{5}{2}J(0) + \frac{9}{2}J(\omega_o) + 3J(2\omega_o)\right]$ where $K = \frac{\gamma^2\hbar}{2r^3}$

Here, J(0) is able to induce effective relaxation, becoming the predominant determinant of T_2 for proteins and other large molecules in the slow tumbling regime. As shown in Figure 2.1 and Equation 2.3, increasing τ_c will result in a further increase in J(0), increasing the rate of T_2 relaxation. The resolution of an NMR spectrum, as defined by peak width at half-height, is limited by the transverse relaxation time as shown in Equation 2.6.

2.6) $v_{1/2} \geq \frac{1}{\pi T_2^*}$

Where T_2^* represents the effective or observed relaxation time, which represents the intrinsic T_2 spin relaxation time resulting from both dipolar coupling and chemical shift anisotropy, but also includes contributions arising from inhomogeneities in the magnetic field. According to this equation, the shorter T_2 relaxation times of large molecules such as proteins will

contribute to peak broadening, reducing NMR peak resolution and sensitivity. Additionally, many types of NMR experiments, including those commonly employed in the study of proteins, involve the transfer of magnetization between nuclei through covalent bonds via J-coupling interactions. T_2 relaxation reduces the efficiency of these transfers, resulting in additional loss of signal intensity. All these factors contribute to the low sensitivity of NMR experiments performed on proteins, and therefore NMR spectra typically require the use of relatively high sample concentrations (~50-400 μM). Additionally, these samples are often isotopically-labeled with NMR-active ^{15}N and/or ^{13}C nuclei to enable the acquisition of multidimensional heteronuclear correlation spectra discussed in the subsequent section.

In some NMR applications, rapid relaxation induced by an unpaired electron in a paramagnetic species is sometimes used as a structural probe. Paramagnetic species can be introduced into specific sites by covalent linkages such as cysteine-mediated disulphide bonds. The paramagnetic relaxation enhancement (PRE) broadens peaks for spins that are in proximity to the paramagnetic species, providing information that can be converted into distance restraints to facilitate the determination of protein structure,[111] or the identification of specific regions involved in protein interactions.[112] Lipophilic PRE probes such as the 5-DOXYL stearic acid free radical employed in this thesis (Fig. 2.2) can also be used to provide structural information on membrane proteins by selectively broadening peaks from residues that face the lipid environment close to the micelle or membrane surface, while its 16-DOXYL counterpart reports on residues which are exposed to the hydrophobic micelle or membrane interior.[113]

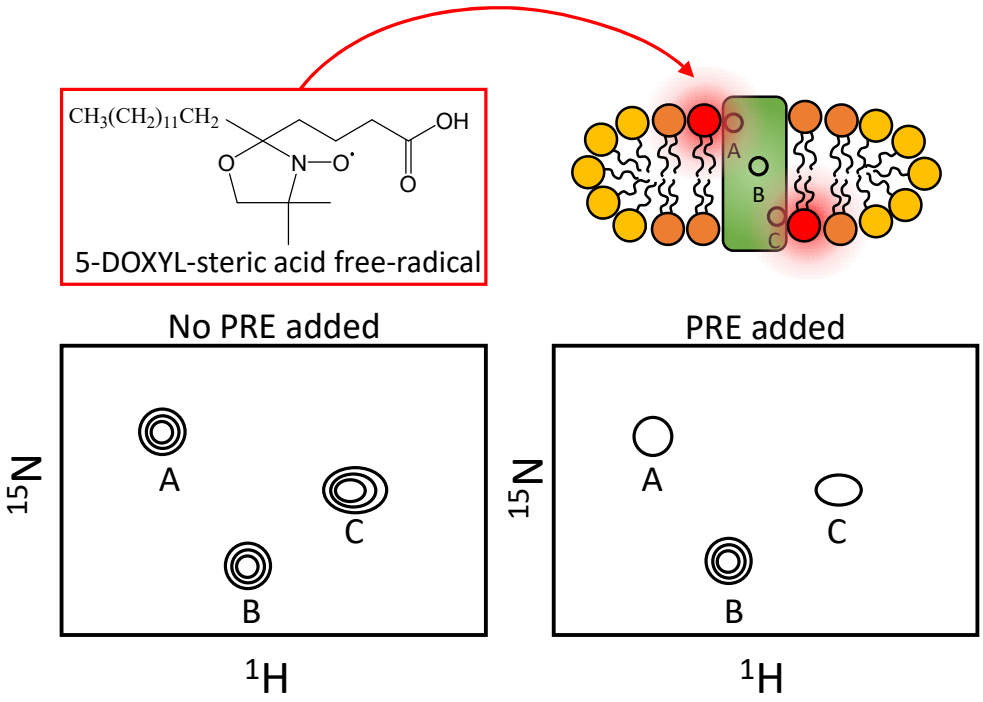


Figure 2.2: Use of the lipophilic paramagnetic relaxation enhancement (PRE) agent 5-DOXYL-steric acid (pictured) to probe the structure of membrane proteins. The PRE agent (red) is incorporated into the lipid or detergent micelle environment, selectively broadening NMR signals originating from residues in proximity to the membrane surface (A, C), while leaving those originating from the shielded core region of the protein (B) unaltered.

2.1.2: Heteronuclear Single Quantum Coherence (HSQC) and Transverse Relaxation-optimized Spectroscopy (TROSY) Spectra

In a 1D ^{1}H NMR experiment, each proton will display a characteristic chemical shift that reflects its local chemical environment. Large molecules such as proteins contain numerous proton resonances, resulting in densely populated, overlapping spectra from which useful structural information cannot be obtained. To circumvent this problem, multidimensional NMR experiments are employed. These experiments provide correlations between nuclei that are coupled either through bonds or through space, yielding valuable structural information on the sample while simultaneously reducing spectral overlap. The

most widely used multidimensional NMR experiment employed in the field of protein NMR is the 2D ^{1}H-^{15}N HSQC.

Peaks in the 2D ^{1}H-^{15}N HSQC spectrum correlate the chemical shift of a proton with that of its directly-bonded ^{15}N nucleus. Therefore each non-proline residue should give rise to a single peak corresponding to the backbone peptide amide proton. The spectrum can also contain peaks arising from the ^{1}H-^{15}N groups of Trp, Asn, Gln, Lys, and Arg side chains. Since each protein gives rise to a unique pattern of peaks, the ^{1}H-^{15}N HSQC spectrum is widely regarded as a "fingerprint" spectrum. This spectrum can also be used to assess structural integrity since spectra from folded proteins will contain peaks over a wide range of proton chemical shifts, while those from unfolded proteins will have a narrower proton chemical shift dispersion. As such, the 2D ^{1}H-^{15}N HSQC spectrum is often the first experiment performed on a protein sample to confirm that it is stably folded and to assess the feasibility of further NMR studies; if the HSQC spectrum shows the expected number of peaks over a wide range of proton chemical shifts, then triple-resonance experiments to assign chemical shifts are likely to be feasible. In standard applications, proteins are usually uniformly ^{15}N and/or ^{13}C-labeled, although selective labeling can be used to obtain site-specific information for larger, more challenging systems.[115]

As molecular weights increase, slow tumbling rates gives rise to rapid T_2 relaxation through both dipolar-coupling interactions and chemical shift anisotropy, limiting the sensitivity and resolution of the ^{1}H-^{15}N-HSQC spectrum. One way in which sensitivity can be improved for large species is through the use of a ^{1}H-^{15}N TROSY experiment, which takes advantage of differential relaxation properties of peaks in a decoupled ^{1}H-^{15}N HSQC experiment. In the absence of the decoupling sequences that are normally employed in an

HSQC experiment, the one-bond coupling constant between the amide proton and nitrogen will become apparent, giving rise to a doublet of peaks in both dimensions for a single amide NH group (Fig. 2.3). In the TROSY experiment, only the component of the decoupled spectrum that undergoes the slowest rate of T_2 relaxation time is preserved in the final spectrum. This peak is narrower than the others due to destructive interference between dipolar coupling and chemical shift anisotropy relaxation pathways. Although some signal can be lost by selecting just one component of the quartet, the corresponding spectrum will have higher resolution relative to the HSQC correlation spectrum, giving rise to a higher quality spectrum for large proteins.[115]

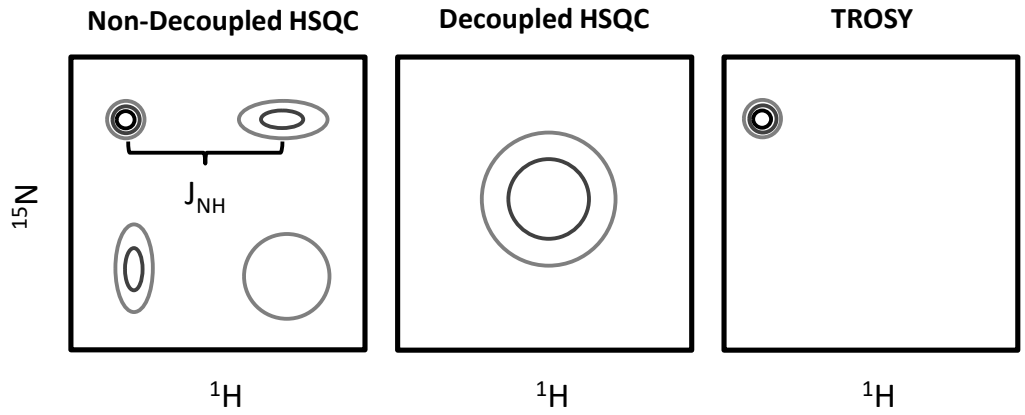


Figure 2.3: Schematic illustrating the pattern of peaks obtained for a single ^{1}H-^{15}N group using an HSQC experiment in the absence and presence of decoupling, and a TROSY experiment.

The resolution-enhancing ability of the TROSY experiment is dependent on the ability of the dipole-dipole and chemical-shift anisotropy relaxation mechanisms to cancel out for the slowly relaxing component. While relaxation caused by dipole-dipole interactions becomes more pronounced as the molecular weight increases, relaxation through chemical shift anisotropy increases with the strength of the applied magnetic field B_o. Consequently, enhancements in peak resolution and signal sensitivity is generally seen for proteins with

molecular masses greater than 20 kDa, and with a spectrometer field strength of at least 700 MHz. However, the level of enhancement that can be achieved depends on the specific relaxation properties of the sample being examined, with systems that undergo ms to μs exchange showing a loss of sensitivity relative to those that are rigid over this timescale.[115]

2.1.3: NMR Experiments to Monitor Conformational Exchange

NMR spectroscopy can be also used to provide information on dynamic processes such as conformational exchange. Residues that undergo a change in conformation are likely to experience a change in local chemical environment, potentially giving rise to a unique set of peaks for each discrete conformation. As illustrated for a two-state (e.g. open-closed) system in Figure 2.4, if the exchange rate (k_{ex}) is smaller than the frequency difference between these states ($|\Delta v|$), peaks from both states will be individually resolved in a spectrum, with the relative intensity reflecting the equilibrium distribution of the two states. Conversely if the system undergoes fast exchange (ie: $k_{ex} >> |\Delta v|$) the resulting NMR spectrum will show a single sharp peak representing a population-weighted average of the signal of both states. Intermediate exchange processes where $k_{ex} \approx |\Delta v|$ will give rise to a broad peak of low intensity centered on the weighted average chemical shift of the two species.[116] Generally, slow exchange processes as defined above have k_{ex} values $< 1\ s^{-1}$, while fast exchange processes occur on the μs timescale and faster. However, since $|\Delta v|$ is directly related to B_o while k_{ex} is independent of this variable, the effect of exchange on the quality of an NMR spectrum can be sometimes be affected by using a spectrometer of different field strength.

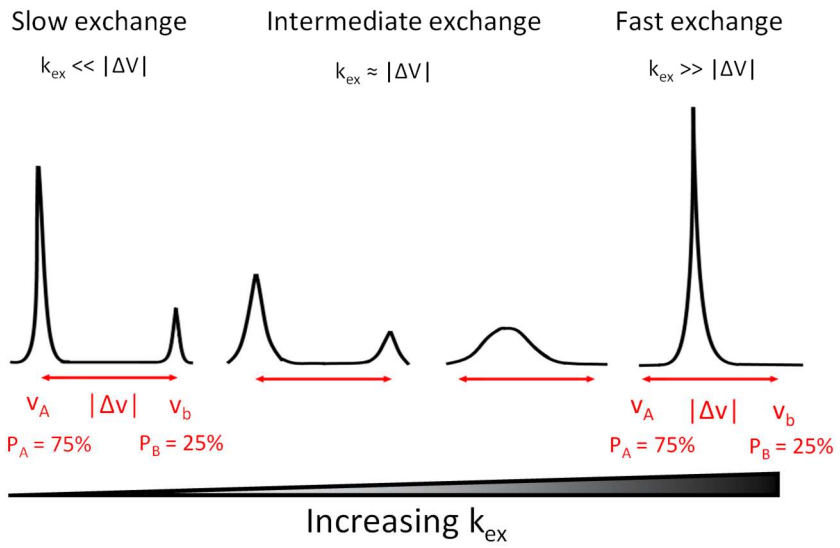


Figure 2.4: Schematic diagram showing the effect of exchange on NMR line shapes and chemical shifts. v_A and v_B represent the Larmour frequencies for the two exchanging species in units of Hz, while |Δv| represents the difference between V_A and V_B

A number of variations on the ^{1}H-^{15}N HSQC spectrum have been developed to measure rates of conformational exchange. For example, for slow exchange (millisecond to second time scales) between two states that have populations that are large enough to be seen as distinct peaks in the spectrum, the longitudinal (ZZ) exchange experiment produces cross-peaks between exchanging species, the intensity of which depends on a variable mixing time that can be used to determine exchange rates,[117] while faster processes (millisecond to microsecond) can be measured and quantified using Carr-Purcell-Meiboom-Gill (CPMG) experiments which monitor the rate of T_2-independent peak broadening resulting from chemical exchange occurring on the same timescale as the NMR pulse sequence.[118–120] However, in order for these experiments to be successful, it is essential that the quality of the HSQC spectrum be high, with intense, well-resolved peaks that can be detected within a reasonably short experiment time (e.g. less than a few hours).

2.1.4: NMR of Membrane Proteins

In order to be studied using solution-state NMR, membrane proteins such as rhomboid proteases require solubilisation in a membrane mimetic system that can maintain a folded, functional conformation in the liquid state. However, the addition of a membrane mimetic increases the apparent molecular weight of the protein complex, decreasing the T_2 relaxation time and reducing spectral sensitivity and resolution. Therefore, optimization of the sample conditions is often required to minimize the size of the protein complex while maintaining its structural integrity and function.

2.1.4.1 Micelles

The most common membrane mimetic system employed in the field of solution-state NMR is the detergent micelle due to its relatively small size and ease of handling.[121] Micelles arise from the spontaneous assembly of amphipathic molecules, usually detergents, into approximately spherical structures with the non-polar alkyl chains directed inwards to form a hydrophobic interior separated from the aqueous solvent by the polar detergent headgroups. Integral membrane proteins can be inserted into the micelle, allowing TM regions to be buried within the hydrophobic micelle interior (Figure 2.5A).

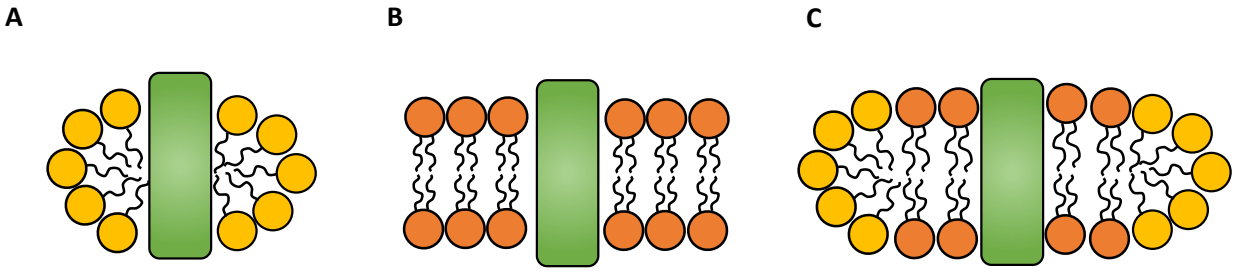


Figure 2.5: Schematic illustrating the solubilisation of integral membrane proteins using A) detergent micelles, B) phospholipid bilayers and C) phospholipid bicelles. Phospholipid molecules are shown in orange, and detergent molecules are shown in yellow.

While these protein-detergent complex (PDC) structures mimic the general features of a bilayer environment, there are some significant differences. Unlike phospholipids, many commonly used detergents for solution NMR contain a single hydrophobic alkyl chain. This gives rise to micelles with a high degree of surface curvature that is usually not representative of the native membrane environment. These structures also contain a higher level of entropy than the equivalent phospholipid bilayer, giving rise to packing defects that enable the penetration of water into the hydrophobic core while simultaneously reducing the lateral pressure exerted on the solubilized membrane protein.[70,101,104,122,123]. In addition, individual detergent molecules can freely associate and dissociate from the micellar phase, the dynamics of which may also potentially influence the structure or function of the solubilized protein.[124,125] Despite these issues, detergent micelles have successfully been used to characterize a range of integral membrane proteins by solution-state NMR, and detergents still remain the default system for many applications, especially if functional assays or other structural techniques are available to verify that native folding and dynamics is preserved.

The equilibrium between micellar and monomeric states for a detergent can be described using the critical micelle concentration (CMC), which defines the concentration above which micelles begin to form; any additional amphiphile that is added to the system will partition directly into the micellar phase. The CMC also reflects the concentration of monomeric detergent that undergoes dynamic exchange with micelles in equilibrium between the two states. Since the rate constant of micellar association (k_{on}) for most detergents is considered to be diffusion-limited, this equilibrium concentration can be directly correlated to exchange kinetics between micellar and monomeric detergent states.[124] The aggregation number *N*, is the average number of detergent molecules per individual micelle which,

together with the CMC, can be used to calculate [M], the concentration of micelles in a solution containing detergent concentration [D] as shown in Equation 2.7.

$$[M] = \frac{([D]-CMC)}{N} \quad \textbf{2.7)}$$

It should be noted that detergent micelle systems can have high levels of polydispersity, with some studies reporting up to a 100% variation in aggregation number within a single detergent solution.[126] This represents a potential challenge for solution-state NMR, and is a factor that must be taken into account during the optimization process, including the one presented in this thesis (Chapter 3).

For detergents of a given headgroup, decreasing the length of the alkyl chain gives rise to smaller, more disordered micelles with higher CMC values, more intense surface curvature, and smaller aggregation numbers. The chemistry of the polar headgroup can also influence these properties by altering the packing density of detergent molecules. There is also variation in the shape of the resulting micelle, with different systems giving rise to spherical, prolate, or oblate geometries depending on the length of the alkyl chain and the need to minimize the electrostatic repulsion between charged headgroups.[127] The effect of headgroup chemistry and alkyl chain length on various micelle properties of three example detergent systems can be found in Table 2.1 and Figure 2.6. The physical properties of mixed micelles can be approximated as a weighted average of equivalent physical properties for the component detergents, which allows mixtures to be chosen with specific CMC and *N* values that could potentially be tailored for solution-state NMR of specific membrane protein systems.[128]

Table 2.1: Physical properties of micelles formed from phosphocholine (Fos), lysophosphatidylcholine (Lyso-Fos), and maltoside (Mal) detergents. Data from Anatrace unless otherwise specified

Headgroup	Alkyl chain length	CMC (mM)	*N*	Shape
Fos	10	11	~24	Prolate
Fos	12	1.5	~54	Prolate
Fos	14	0.12	~108	Prolate
Lyso-Fos	10	6.06	~40[129]	N/a
Lyso-Fos	12	0.57	~82[129]	N/a
Lyso-Fos	14	0.04	~126[129,130]	N/a
Mal	10	1.8	~69	Oblate
Mal	12	0.17	78-140	Oblate
Mal	14	0.01	80-130[131]	Oblate

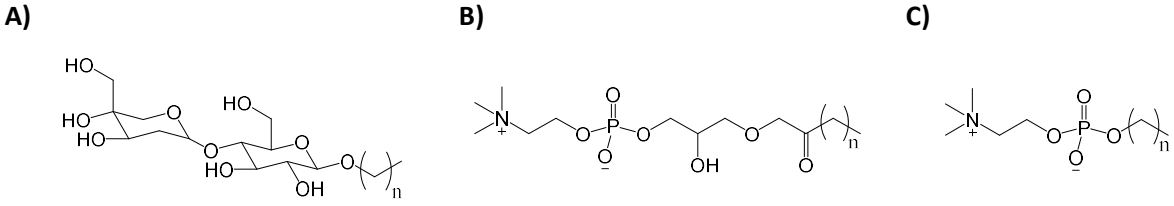


Figure 2.6: General chemical structure for A) maltoside, B) lysophosphatidylcholine , and c) phosphocholine (Fos) detergents.

Detergents with shorter alkyl chains and charged head groups generally give rise to smaller micelles that can help to reduce peak broadening in solubilized membrane protein samples. However, the small size can compromise the ability of these detergents to maintain a membrane protein in a functionally folded state.[132] Conversely, larger non-ionic detergents tend to be less denaturing, but at the cost of increased PDC size and peak broadening.[132] It should be noted however, that these are general trends, and that there are no rules for determining which detergent micelle system will produce the best result for a given protein. For example, the Sanders lab found that the anionic detergent dodecylphosphocholine (Fos-12) gave rise to the best combination of functional activity and NMR spectral quality for diacylglycerol kinase (DAGK),[133] while the same detergent micelle system was unable to

maintain the bacterial multidrug resistance pump Smr in its native conformation. In the case of the latter, the non-polar Mal-10 detergent was optimal for NMR.[121] Lysophospholipids such as lysomyristoylphosphocholine (Lyso-Fos-14) retain the same phosphocholine headgroup as Fos-12 but incorporate a glycerol spacer between this headgroup and the alkyl chain, providing a more gradual transition between the aqueous and hydrophobic phases of the micelle (Fig. 2.6). This allows them to better preserve the structure and function of some integral membrane proteins, facilitating the acquisition of high-quality NMR spectra of the human V2 receptor and other GPCR's. [134,135]

Conversely, SDS, which is widely considered to be a strongly denaturing detergent, is able to maintain the KcsA ion channel in a folded, active conformation, and was used for solution-state NMR studies of this protein.[136,137] It is also possible for the detergent to affect the function of a solubilized integral membrane while preserving its structure, as was seen for the PagP palmitoyl-transferase; while both β-octylglucoside (β-OG) and Fos-12 detergents are able to maintain its native structure, the latter was also able to bind into the PagP active site, interfering with attempts to study its function in the solution state.[138,139] Taken together, these examples highlight the importance of detergent selection in the study of membrane proteins using solution-state NMR, and the need for screening to identify the optimum detergent for the system at hand.

2.1.4.2 Phospholipid Bicelles

In contrast to detergents, phospholipid molecules are immiscible in water, forming insoluble lipid sheets in an aqueous environment (Fig. 2.5B). However, disruption of these sheets through physical agitation can produce lipid vesicles that are small enough to remain

suspended in the aqueous phase. While vesicles provide a more native-like environment for structural studies, their large size (25-100 nm)[140] makes them impractical for solution-state NMR studies of integral membrane proteins.[141] One way to overcome this issue is through the use of a mixed phospholipid/detergent bicelle in which the solvent-exposed hydrophobic edges of an otherwise insoluble bilayer disk are "capped" with short chain phospholipids or amphipathic detergents (Fig. 2.5C). Bicelle size can be controlled by varying the molar ratio of the long chain lipid and short chain lipid/detergent (q); bicelles which are small enough to tumble isotropically in solution can generally be produced with q values of 0.5-1, depending on the lipid/detergent system employed.[123,142,143] The presence of a phospholipid bilayer structure within the bicelle can help preserve the native structure and function of integral membrane proteins, albeit at the cost of slower tumbling rates and larger particle sizes compared to the corresponding detergent micelle system.[143–145]

2.2: Circular Dichroism Spectroscopy

Circular dichroism (CD) spectroscopy measures the differential absorption of left and right-handed circularly polarized light by chiral molecules such as proteins and amino acids, and is measured in terms of the ellipticity acquired by circularly polarized light as it passes through the sample. The absorption properties of proteins and polypeptides in the far-UV range (250-150 nm) commonly used in CD studies is determined primarily by n-π^* and π -π^* electronic transitions within the backbone amide group.[146,147] As a result, values for CD are usually expressed in mean residue ellipticity (MRE), which normalizes the CD signal for the total concentration of peptide bonds within the protein sample. CD spectroscopy in this region of the spectrum reports on backbone ψ and φ angles and thus protein secondary structure

content,[147] enabling its use as a tool to monitor changes in protein secondary structure in response to external stimuli such as ligand binding or thermal denaturation, as seen in the 2012 study by the Urban group examining the effect of various point mutations on *ec*GlpG rhomboid stability.[95]

CD spectroscopy can also be used to provide quantitative information on secondary structure, since α-helix, β-sheet and random coil configurations give rise to characteristic CD spectra. The experimental CD spectrum can be approximated as linear combination of these characteristic spectra (Fig. 2.7),[148] prompting the development of algorithms to deconvolute these spectra in order to provide a quantitative estimate the protein secondary structure content. Most of these algorithms accomplish this by comparing the sample spectrum to a set of reference spectra from proteins of known structure. Early methods employed relatively small reference sets which attempted to define a single characteristic spectrum for each secondary structure type using data obtained on model polypeptides known to exist predominantly in a single secondary structure type.[149,150] Recognizing that the structure of globular proteins can deviate from those of the model polypeptides, subsequent methods, including the CONTIN algorithm employed in this thesis,[150] utilize reference sets derived from a database of proteins representing a more realistic range of secondary structure content.[150] Initial reference sets only used soluble, globular proteins, which limited their utility when applied to the analysis of membrane proteins.[151] Inclusion of membrane bound proteins into the soluble protein reference sets improved performance against both types of proteins, and has since become standard practice.[148] The reference set employed in this thesis is the SMP56 reference, which contains the CD spectra from 56 soluble and membrane bound proteins of known structure.[148,150]

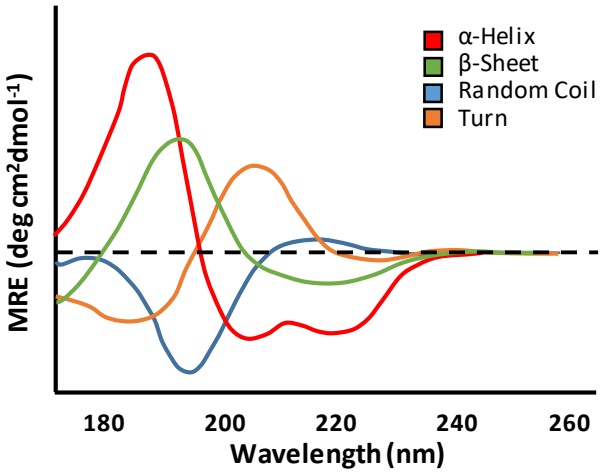


Figure 2.7: CD spectra characteristic of various protein secondary structure elements. Note that the signal is normalized against both polypeptide length and sample concentration to give mean residue ellipticity.

While CD spectroscopy has proven to be a powerful tool for secondary structure analysis, there are a number of limitations that should be noted. Initial experiments in the 1960's demonstrated that CD spectra of proteins will depend not only on the total secondary structure content, but also on the length of the individual secondary structure elements within the protein itself.[146] Additionally, while secondary structure is the primary determinant of CD spectral properties for proteins, other factors such as tertiary structure, variations in the local chemical environment or the presence of non-peptide chromophores (e.g. disulphide bonds) can also contribute to the measured ellipticity.[150] This results in a degree of uncertainty in the secondary structure predictions produced using these methods, with error rates of up to 5-10% for the CONTIN algorithm and SMP56 reference sets used in this thesis.[148,152] Nonetheless, these methods can provide a useful estimate of secondary structure content, particularly when comparing the structure of closely related samples such as those presented in this thesis.

Chapter 3: Influence of Hydrophobic Mismatch on the Catalytic Cycle of the *E. coli* GlpG Rhomboid

Contributions of collaborators:

Results from this chapter have been published, and are reproduced with permission from **Foo, A. C. Y.**, Harvey, B. G. R., Metz, J. J. & Goto, N. K. Influence of Hydrophobic Mismatch on the Catalytic Activity of *E. coli* GlpG Rhomboid Protease. *Protein Sci.* **24,** 464–473 (2015). Licence number 4137441215096 June 28, 2017.

Jeff J. Metz (University of Ottawa) purified and assayed TMD81 in Mal-12 and NG-10. Brandon G.R. Harvey (University of Ottawa) purified, assayed, and obtained CD spectra and thermal denaturation curves for TMD81 in Fos-12, Fos-14, and Mal-12. Additionally Brandon assisted in the reconstitution of TMD81 into PC-14 phospholipid bicelles

3.1: Introduction

It has been suggested that the phospholipid bilayer can be deformed in response to the hydrophobic mismatch between *ec*GlpG and its membrane environment, minimizing unfavourable interactions between the hydrophobic membrane core and the polar regions of *ec*GlpG (Figure 3.1). This thinning of the bilayer in the local vicinity of *ec*GlpG has been proposed to expose portions of α-helical TM substrates to the aqueous phase, destabilizing its secondary structure and facilitating its entry into the active site. However, the energetic cost and altered phospholipid packing density resulting from this deformation may also have an influence on rhomboid structure and function, as has been shown for various helical and β-barrel integral membrane proteins including Gramicidin, KcsA, and OmpA.[99,100,153–157] In the case of OmpA, hydrophobic mismatch was found to perturb protein folding, leading to the formation of a partly folded intermediate state, while other systems experienced more subtle perturbations such as altered helix packing, reduced stability, or changes in oligomeric state.[99,100,153–158] However, to date there have been no reported examples of intramembrane proteases whose structure and function was influenced by hydrophobic mismatch prior to the work presented in this chapter.

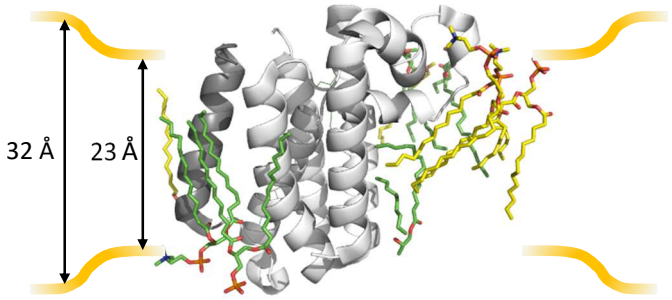


Figure 3.1: X-ray crystal structure of *ec*GlpG (PDB 2XTV) showing the position of co-crystalized phospholipid (PC-14) molecules from the same subunit (green) or symmetry-related subunit (yellow). The relative size of the 23 Å *ec*GlpG hydrophobic region and the 32 Å interior hydrophobic dimensions of the native bilayer environment are shown to scale to the left, illustrating the potential for mismatch. The α5 gate region is highlighted in grey.

In this chapter the impact of hydrophobic mismatch on rhomboid activity was examined using, an *E. coli ec*GlpG construct corresponding to the isolated catalytic TM domain (TMD). This construct was purified into phosphocholine (Fos), lysophosphocholine (Lyso-Fos) and maltoside (Mal) detergents of varying alkyl chain lengths, giving rise to micelles of varying hydrophobic dimensions. [126,127] Maximal activity was observed in detergents containing a saturated alkyl chain of 10-12 carbons regardless of headgroup. Detergents which fall outside this range significantly impaired catalytic activity, with a ~two-fold reduction in k_{cat} observed in systems containing 8 or 14 carbon alkyl chains. Similar results were obtained in mixed Fos-12/Fos-14 micelles, and phospholipid bicelles. A small but reproducible difference was observed between the CD spectra obtained of TMD81 solubilized in short- and long-chain detergents, indicating that subtle structural changes may be responsible for the loss in activity. These results suggest that the energetic cost of hydrophobic mismatch can impair *ec*GlpG-mediated cleavage of water-soluble substrates, potentially contributing to the regulation of rhomboid activity in vivo.

3.2: Results

3.2.1: Identification of TMD81 as the Optimal Construct for Exploring *ec*GlpG Function

Three *ec*GlpG constructs of varying lengths were tested to identify the optimal construct length to conserve activity and stability in the TM catalytic core while omitting the N-terminal cytoplasmic domain (CytD), thereby eliminating the possibility of denaturing interactions between this region and the local micelle environment which might interfere with proteolytic activity. The constructs examined consisted of TMD60, which begins at residue

60, and is missing only the structured CytD, TMD81 that also omits the unstructured linker region connecting it to the catalytic TM core, and TMD91 which omits all residues that are not resolved in *ec*GlpG X-ray structures (Fig. 3.2A). All truncation constructs, along with full-length *ec*GlpG were purified in the n-dodecyl-β-D-maltoside (Mal-12) detergent micelle system most commonly used in the study of *ec*GlpG. SDS-PAGE of the resulting protein fractions reveal a single band corresponding to the predicted MW for all constructs tested (Fig. 3.2B). Since the local membrane environment can influence *ec*GlpG-catalyzed proteolysis by altering the physical properties of its TM substrates,[104] all kinetics measurements were performed using a water-soluble BODIPY casein model substrate as described previously.[63] Initial rates of proteolysis were plotted as a function of substrate concentration, yielding a hyperbolic curve that could be fit to the Michaelis-Menten equation (Section 1.1.3) to allow determination of k_{cat} and (Fig. 3.3C, 3.3D). As shown previously, loss of the first 59 (TMD60) or 80 (TMD81) residues did not produce a significant change in either catalytic parameter, while a significant decrease in k_{cat} was observed in TMD91 (Fig. 3.2E)[79,81]

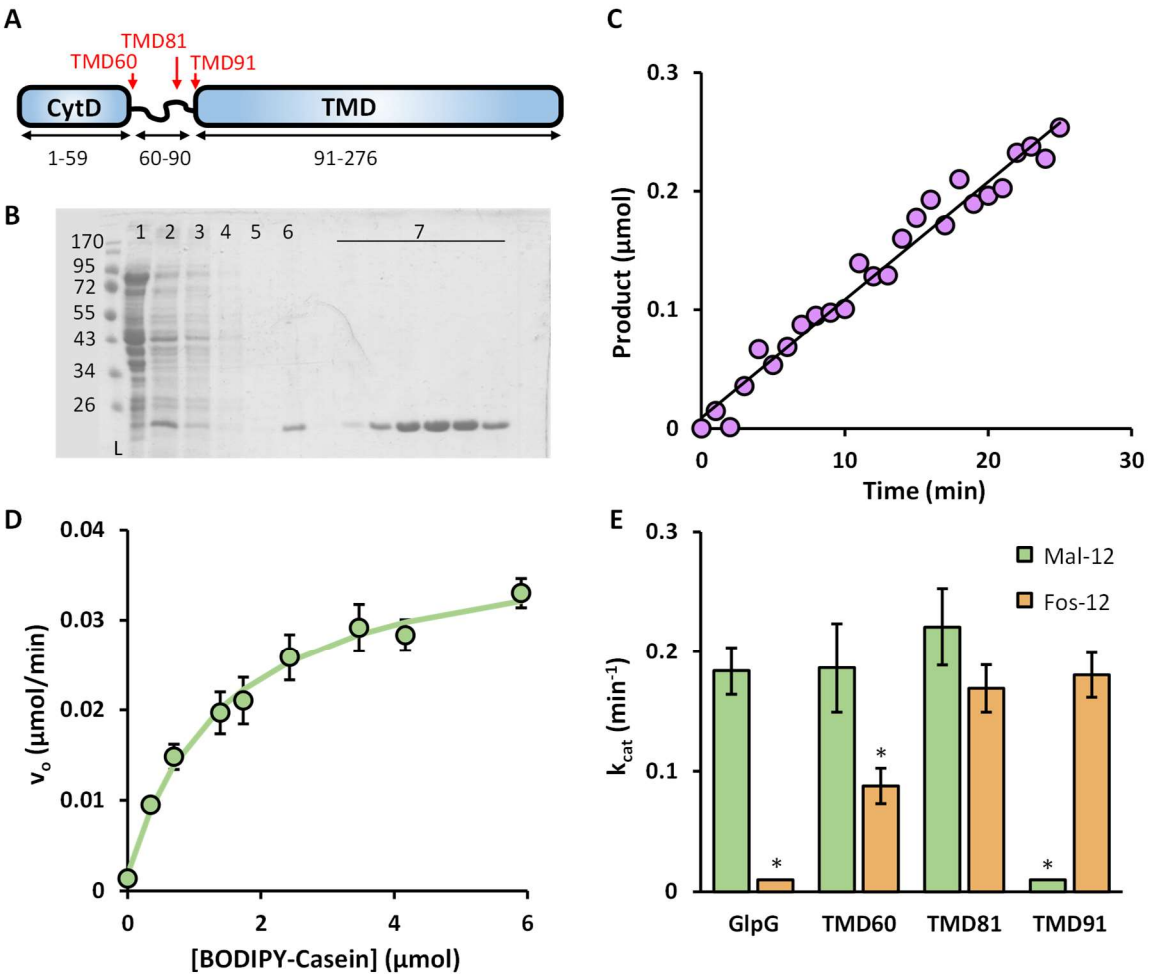


Figure 3.2: Optimization of *ec*GlpG truncation construct. A) Schematic representation of *ec*GlpG showing the relative size and position of the N-terminal cytoplasmic domain (CytD), the flexible linker region and catalytic TM domain (TMD). B) Representative SDS-PAGE gel of fractions taken during various stages of the purification of TMD81. L: Molecular weight ladder in kDa. 1: Soluble protein fraction. 2: Mal-12 solubilized membrane protein fraction. 3: Ni-NTA column flow through. 4 and 5: Non-specifically bound contaminants removed during Ni-NTA wash step. 6: Purified TMD81 eluted using 250 mM imidazole. 7: Representative fractions obtained after size-exclusion chromatography containing the final, purified TMD81 in 0.1% w/v Mal-2. C) Representative data showing the formation of the BODIPY-casein cleavage product generated by *ec*GlpG proteolysis over time. The slope of this linear relationship provided the initial rate of proteolysis. D) Example of data used to obtain kinetic parameters for *ec*GlpG cleavage of BODIPY-casein. Initial proteolysis rates were fit to the Michaelis-Menten equation (solid line) to obtain values of K_M and V_{max}.

Figure 3.2 (Con't): E) Catalytic turnover numbers for the various N-terminal truncation constructs solubilised in Mal-12 (green) or Fos-12 (orange) detergent micelles against a water-soluble casein substrate. Reported error values for this assay and all subsequent data reported in this thesis represent the standard deviation obtained from at least three independent trials unless otherwise noted. Values which are significantly different ($p < 0.05$) from those of TMD81 in their respective detergents are indicated with *, while the near-zero values with no error bars denote samples that show no detectable activity under the conditions employed (three trials). Statistical analysis for these values and all subsequent data was carried out using a student T-test assuming unequal variances.

Previous studies show that the CytD-micelle interaction of both *E. coli* and *P. aeruginosa* rhomboids depend on the detergent headgroup, with both homologues displaying a high affinity for phosphocholine based systems.[96,97] To assess the potential effect of this interaction on *ec*GlpG function, the activity of the N-terminally truncated constructs was also measured in micelles composed of Fos-12. The alkyl chains of both Fos-12 and Mal-12 contain the same number of carbons (12), and are expected to give rise to micelles with very similar internal hydrophobic dimensions.[126,127] Despite this, inclusion of the N-terminal cytoplasmic domain and associated linker region resulted in a significant decrease in k_{cat} in this detergent micelle system (Fig 3.2E), suggesting that CytD-micelle interactions can effect proteolytic activity, mirroring behavior previously observed in the *Pseudomonas* homologue.[97] Interestingly, full activity was observed for TMD91, in contrast with the results for this construct obtained in Mal-12, highlighting the sensitivity of the catalytic domain to the local membrane environment independent of hydrophobic mismatch. However, since TMD81 showed maximal activity under both conditions, it was chosen as the most robust TMD construct for our study of the effect of hydrophobic mismatch on *ec*GlpG catalytic function. For the remainder of this thesis, all work was carried out on the *ec*GlpG TMD81 construct unless otherwise specified.

3.2.2: Hydrophobic Mismatch Disrupts TMD81 Activity in Detergent Micelles

To assess the effect of hydrophobic mismatch on *ec*GlpG activity, samples of TMD81 were solubilised in a series of Fos detergents. These detergents share the same headgroup as Fos-12 (Fig. 2.6A), but differ in the number of carbons in their alkyl chains, giving rise to micelles with varying hydrophobic dimensions. While detergents with alkyl chains as short as 10 carbons (Fos-10) were able to support close to maximal turnover numbers, a significant decrease in k_{cat} was observed when the length of these alkyl chains exceeded 12 carbons, with up to a three-fold reduction observed in Fos-14 (Figure 3.3A). This was not accompanied by a significant change in K_M, suggesting that the loss in activity is not a result of competitive binding of the longer chain detergent molecules into the active site, but instead reflects a change in the rhomboid catalytic status.

Samples of TMD81 were also prepared in mixed micelles comprised of varying ratios of Fos-12 and its 14-carbon alkyl chain counterpart (Fos-14). This will change the dimensions of the hydrophobic phase of the micelle, as it has been shown that there is a linear correlation between micelle composition and hydrophobic phase size.[129] The size of each PDC was assessed using size-exclusion chromatography and the initial rate of TMD81 proteolysis monitored using a fixed concentration of casein substrate (Figure 3.3B). Both were found to vary linearly with micelle composition, with larger PDCs supporting lower levels of activity. The same correlation was obtained, regardless of whether TMD81 was purified into Fos-12 and diluted with Fos-14 or purified into Fos-14 and diluted with Fos-12, indicating that the observed change in activity is not the result of an irreversible denaturation. Despite the presence of a lipid-like glycerol linker as discussed in Chapter 2 (Fig. 2.6B), a similar loss in

activity was also observed in the 14-carbon Lyso-Fos detergent relative to its 12-carbon counterpart.

To verify that this loss in activity was not specific to the phosphocholine headgroup shared by both Fos and Lyso-Fos, TMD81 proteolytic activity was assessed in a series of Mal detergents (Fig. 2.6A). The 12-carbon alkyl chain Mal-12 has been extensively used to characterize the structure and function of *ec*GlpG and other rhomboid homologues in vitro, making it a good reference detergent. Additionally, the lower CMCs of these detergents relative to the Fos and Lyso-Fos series made it feasible to purify TMD81 into detergents with shorter alkyl chains, such as Mal-9 and Mal-8, allowing the lower micelle size limit for TMD81 proteolytic activity to be explored. A significant decrease in k_{cat} was observed in Mal detergents with alkyl chains containing more than 12 carbon atoms, confirming that this value represents the upper limit for optimal rhomboid proteolysis (Figure 3.3A). A lower limit of 9 carbon atom-long alkyl chains was observed for the Mal series of detergents, with a ~1.5-fold decrease in k_{cat} in Mal-8.

Size-exclusion chromatography was used to assess the PDC size for TMD81 in all of the detergent micelle systems tested and, while a general correlation was observed between alkyl chain length and PDC size within each series, no overall trend that encompassed all tested systems was apparent (Figure 3.3C). Conversely, a significant (~2x) variation in TMD PDC size was observed between the detergent micelle systems which gave rise to maximal activity (Fig. 3.D), providing additional support to the idea that the loss in k_{cat} is a result of hydrophobic mismatch between TMD81 and the micelle interior, rather than differences in overall micelle size or morphology.

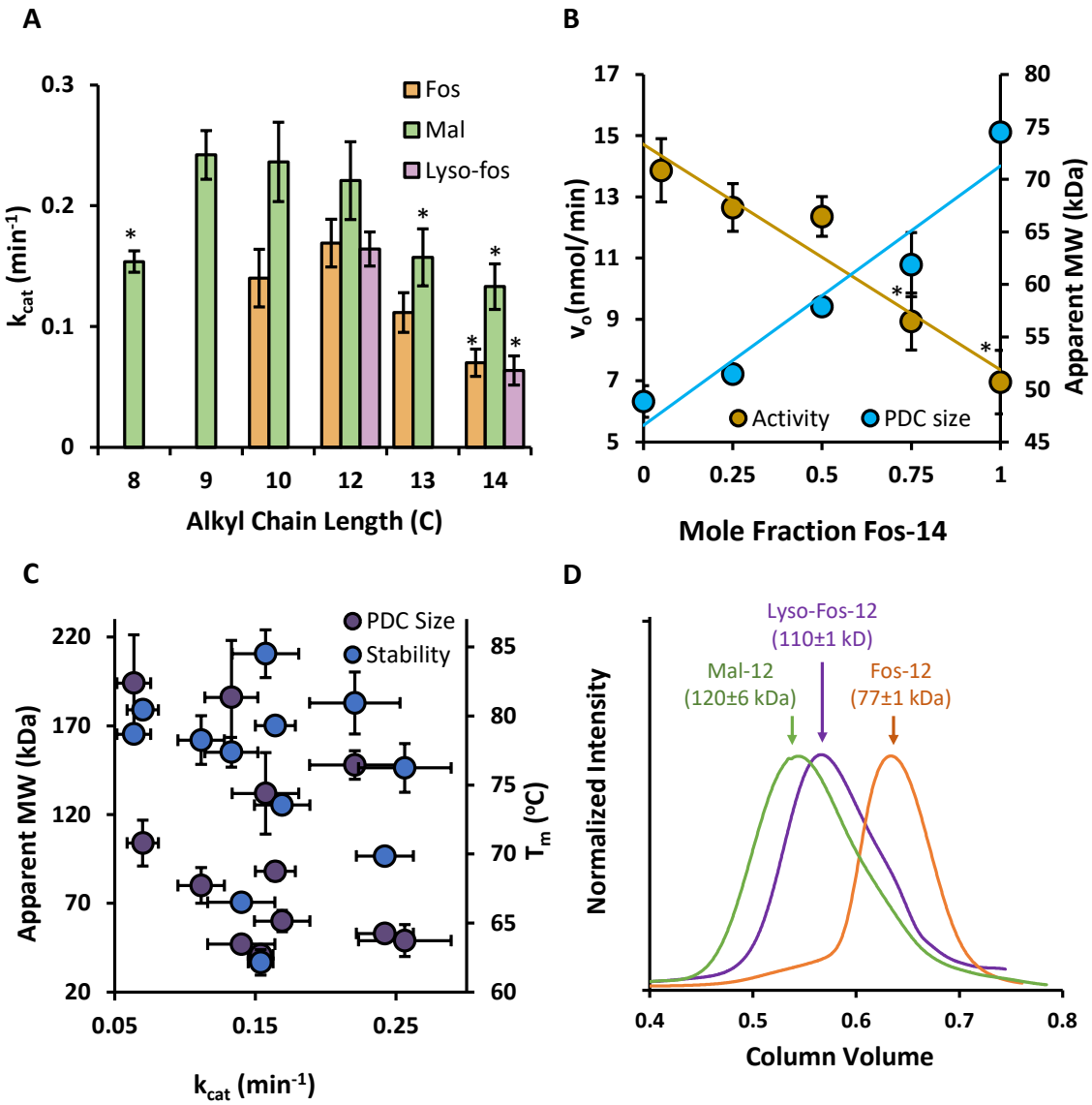


Figure 3.3: Effect of detergent micelle environment on TMD81. A) Catalytic turnover number (k_{cat}) values for samples of TMD81 solubilised in maltoside (Mal - green), phosphocholine (Fos - orange) or lysophosphocholine (Lyso-Fos - purple) detergents of varying alkyl chain lengths. B) Proteolysis rates (orange) of TMD81 assessed using a fixed concentration of casein substrate in mixed-micelles containing varying ratios of Fos-12 and Fos-14. The size of the resulting protein-detergent complexes (PDC's) as estimated from size exclusion chromatography is shown in blue. Values which are significantly different ($p < 0.05$) from those obtained in the corresponding 12-carbon Fos, Lyso-Fos, or Mal detergent are indicated with * in both A) and B). C) k_{cat} values of TMD81 solubilised in the various detergents tested in this chapter plotted against the size (purple) and melting temperature (blue) of the resulting protein-detergent complexes. D) Representative FPLC elution profiles and estimated PDC molecular weights of TMD81 purified into Lyso-Fos-12, Mal-12, and Fos-12 detergents which display maximal activity.

Previous studies showed that *ec*GlpG is exists primarily in the dimeric state in Mal-12 micelles, whereas it exists as a monomer in Mal-10.[52] However, both these detergents support high levels of catalytic activity against BODIPY-casein,[52] suggesting that the reduction in k_{cat} cannot be attributed to changes in the rhomboid oligomeric state. In line with this conclusion, the linear correlation observed between the composition of the Fos mixed micelles and PDC size (Fig 3.3B) suggests that the oligomeric state of TMD81 remains constant between the Fos-12 and Fos-14 micelles employed in this thesis, and is not responsible for the decrease in activity and k_{cat} observed in the latter. In addition, a plot of the apparent PDC molecular weight versus the predicted aggregation number for each detergent also yields a linear relationship (Fig. 3.4), providing further evidence that the oligomeric state of TMD81 likely remains unchanged throughout the range of Fos detergents tested.

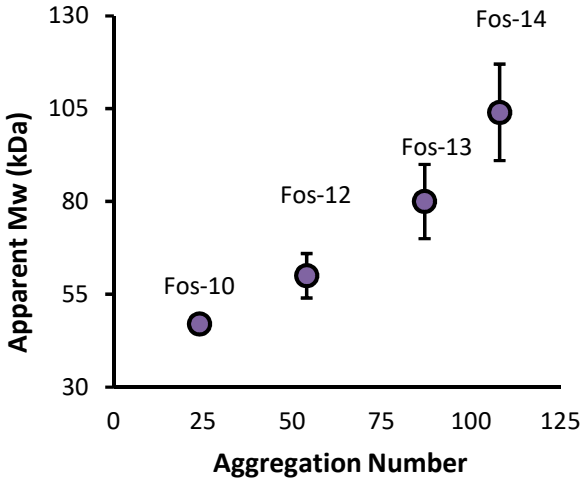


Figure 3.4: Plot of the apparent molecular weight of various TMD81-Fos complexes against the predicted aggregation number. The PDC sizes obtained for Fos-10 and Fos-14 were found to be significantly different ($p < 0.05$) from those of Fos-12.

While detergent micelles such as Mal-12 are able to preserve the structure and function of *ec*GlpG in vitro,[89] there are several limitations to these systems, a full discussion of which

was presented in Chapter 2. To evaluate whether hydrophobic mismatch could influence rhomboid activity in a more native-like bilayer environment, samples of TMD81 were reconstituted into bicelles comprised of phosphatidylocholine (PC) phospholipids and CHAPS detergent. In order to determine the ideal conditions for kinetic studies, samples of TMD81 were prepared in PC-12 bicelles of varying concentrations and phospholipid:bicellar detergent ratios (q_{eff}). Little change in proteolytic activity was observed upon incubation with a fixed concentration (2 μM) of BODIPY-casein (Table 3.1), indicating that variations in bicelle conditions are unlikely to significantly affect the observed proteolysis rates under the conditions employed in this thesis. While the 12-carbon phospholipid (PC-12) yielded a k_{cat} value of 0.01 min^{-1}, the proteolytic activity of TMD81 in bicelles made from the 14-carbon phospholipid PC-14 was found to be undetectable across the range of substrate concentrations tested (Figure 3.5). Therefore, hydrophobic mismatch also affects the ability of *ec*GlpG to catalyze peptide cleavage when reconstituted into bilayers.

Table 3.1: Proteolysis rates measured for TMD81 purified into PC-12/CHAPS bicelles.*

q_{eff}	Bicelle Concentration (%w/v)	v_o (nmol/min)
1	1.5	2.3±0.3**
0.5	4.7	3.2
0.40	4.4	2.8
0.33	4.2	2.4
0.2	3.8	2.5
0.2	2.2	2.4
0.14	1.5	2.4

* Data obtained from a single trial unless otherwise indicated.
** Average and standard deviation for two independent protein preparations.

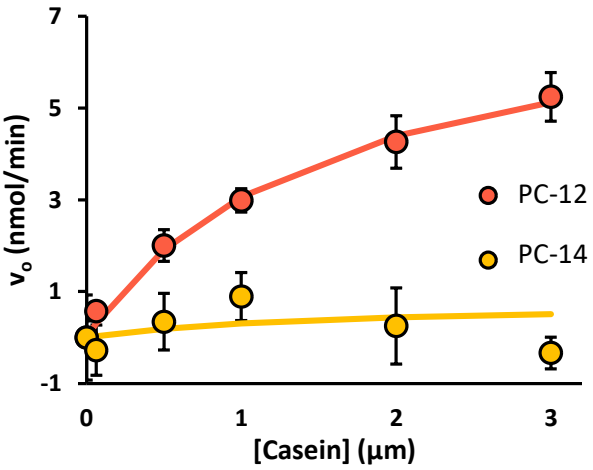


Figure 3.5: Michaelis-Menten curves obtained for TMD81 solubilised in PC-12/CHAPS (orange) or PC-14/CHAPS (yellow) phospholipid bicelles.

3.2.3: Effect of Hydrophobic Mismatch on TMD81 Structure and Stability

In some systems such as the *E. coli* OmpA, the energetic cost of membrane deformation in response to increased levels of hydrophobic mismatch can result in a reduction of thermodynamic stability and loss of tertiary structure.[157,159] To determine whether the loss of TMD81 activity under mismatch conditions could be attributed to a similar mechanism, thermal denaturation studies were carried out using the various detergent micelle systems tested. All samples yielded a sigmoidal denaturation curve indicative of cooperative unfolding (Fig. 3.6A), with melting temperatures (T_M) above 60ºC mirroring previous results reported for the full-length *ec*GlpG.[95] While purification into shorter alkyl-chain length detergents produced a slight decrease in thermostability, no overall correlation with k_{cat} was observed (Figure 3.3D). Therefore, the detergent-dependent loss in activity is not the direct result of a corresponding decrease in TMD81 structural stability.

CD spectroscopy was also employed to explore the possibility of more subtle structural rearrangements Spectra obtained for TMD81 in Fos-10, Fos-12, and Fos-14 are

consistent with the largely α-helical structure expected for a properly folded *ec*GlpG, with characteristic minima at 208 and 220 nm. However, a small but reproducible ($n \geq 3$) difference in the relative intensity of these two minima was observed in Fos-14, as shown in Figure 3.6B. Secondary structure prediction algorithms indicate that this shift is consistent with a slight loss in *ec*GlpG α-helical content in response to the increased hydrophobic mismatch present in Fos-14 detergent micelles (83±5%) relative to Fos-12 (91±1%), potentially reflecting a change in the structure of TMD81 that is detrimental to k_{cat}.

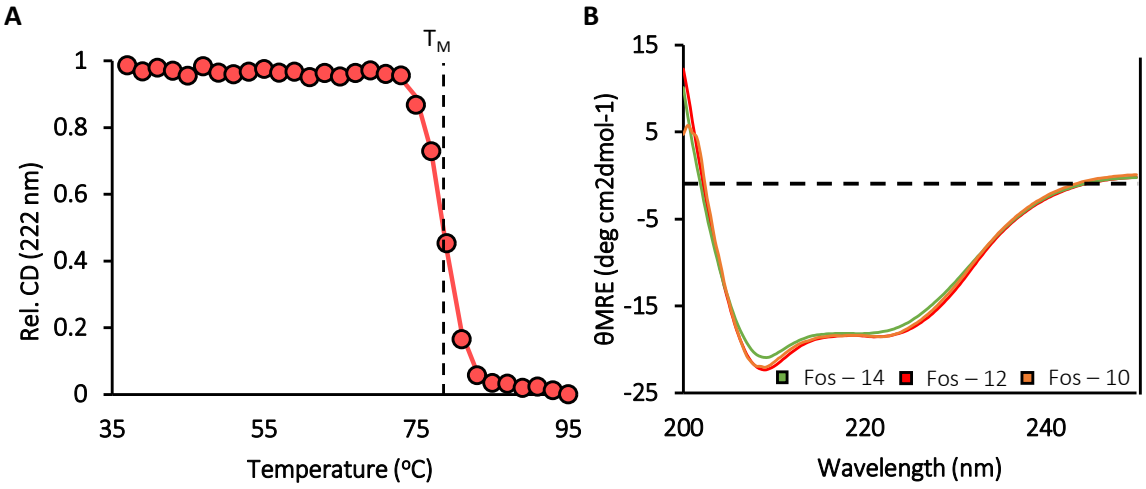


Figure 3.6: Characterization of TMD81 using CD. A) Example thermal denaturation curve for TMD81 in Mal-12 obtained through monitoring the loss of CD signal intensity at 222 nm. A linear baseline correction has been applied to facilitate interpolation of the sample melting temperature (T_M). B) Representative CD spectra obtained for 5 μM TMD81 solubilised in 0.1% w/v Fos-14 (green), Fos-12 (Red) or Fos-10 (orange) at 37°C

3.3: Discussion

3.3.1: New Insight into *ec*GlpG-Membrane Interactions

The results presented here show that rhomboids display maximal activity when solubilised in amphiphiles with alkyl chains between 9 and 12 carbon atoms in length, corresponding to a core hydrophobic thickness of approximately 22-33 Å.[126,127] The hydrophobic length of *ec*GlpG is generally smaller than this, being approximately 20-23 Å, suggesting that rhomboids have some capacity to accommodate hydrophobic mismatch through thinning the local membrane environment. While previous MD studies provide evidence for such an interaction in phospholipid bilayers,[98] the only functional evidence for this membrane deformation comes from a solvent accessibility study carried out using the hydrophilic alkylating agent AMS. However, this study only examined the ability of AMS to modify a small subset of residues, all of which were localized to the L1 region (Fig. 3.7).[87] Furthermore, these experiments were carried out in live cells, raising the possibility that the AMS alkylation was influenced by external factors such as other membrane proteins that could introduce additional distortion to the local bilayer environment, or variations in the local membrane composition. The evidence presented in this chapter provides functional evidence that the isolated rhomboid TMD is capable of altering the local membrane environment under controlled conditions *in vitro*. This is consistent with the proposed role of membrane thinning in facilitating rhomboid proteolysis by destabilizing α-helical TM substrates.[86,109] In contrast, the lower limit of membrane hydrophobic thickness (~22 Å, or 9 carbons) corresponds to the minimum hydrophobic thickness of the *ec*GlpG TM region, suggesting a reduced ability to accommodate membrane environments with shorter hydrophobic dimensions. The reduced stability afforded by these micelles could give rise to structural rearrangements in TMD81,

though the high CMC of the 8-carbon maltoside prevented the acquisition of CD spectra required to evaluate this hypothesis.

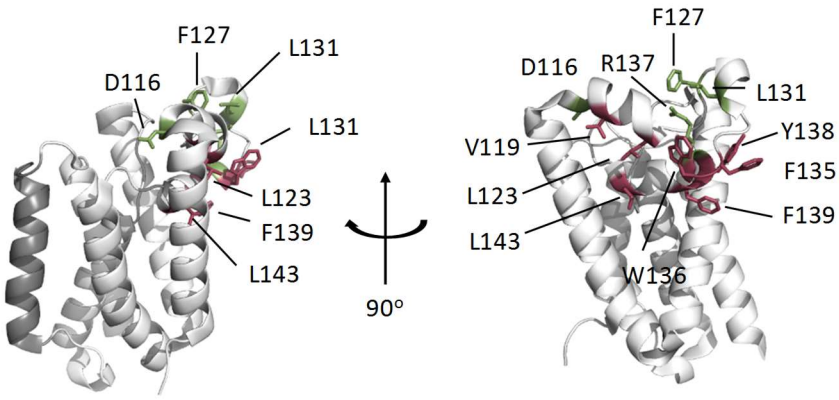


Figure 3.7: X-ray crystal structure of *ec*GlpG showing the location of all residues assessed for solvent exposure by Wang et al (2007).[87] Residues which were found to be solvent exposed are coloured in green, while those which were protected from alkylation are shown in red (PDB 2XTU)

The values of k_{cat} obtained against the soluble casein substrate in PC-12 bilayers (0.01 min^{-1}) represents a >10-fold reduction relative to the value reported for its Fos-12 detergent counterpart (0.17 min^{-1}). This is consistent with previous studies carried out on the *Pseudomonas* homologue,[97] and potentially reflects the lower entropy of the bilayer environment, which may restrict movement of the α5 lateral gate.[104] Curiously, the kinetic parameters against a model TM segment substrate reported for *ec*GlpG in phospholipid bicelles versus Mal-12 micelles yielded a significantly smaller (~5-fold) decrease in $k_{cat,}$ while activity of AarA against the same substrate was constant across Mal-12 detergents and phospholipid bicelles, suggesting the presence of additional factors which facilitate proteolysis of TM substrates in a bilayer environment.[53,52] The more ordered nature of the bilayer environment could also make it more resistant to deformation, reducing the ability of rhomboids to compensate for hydrophobic mismatch, giving rise to the loss of activity

observed in PC-14 despite sharing the same hydrophobic dimensions as the Mal-12 and Fos-12 micelles which support maximal activity. It should be noted that the bacterial membrane environment has similar hydrophobic dimensions to PC-14,[160] suggesting that interactions with the physiological substrates are able to overcome the negative effects of reduced bilayer entropy and hydrophobic mismatch present in the native membrane environment, providing a mechanism through which rhomboid activity and selectivity can be regulated *in vivo*.

3.3.2: Mechanistic Basis for Mismatch-Dependent Decrease in TMD81 k_{cat}.

Some integral membrane proteins undergo significant structural rearrangements in response to changes in the local membrane environment, altering TM helix tilt angles (eg: KscA ion channel)[153] or packing interactions (eg: Ca^{2+} ATPase)[154] to minimize exposure of hydrophobic residues to the aqueous environment. Currently, there are several X-ray structures available for *ec*GlpG solubilized in the 9-carbon nonyl-glucoside (β-NG) detergent, which is expected to produce micelles whose interior hydrophobic dimensions correspond to those which support maximal TMD81 activity. The structure of *ec*GlpG has also been solved in PC-14/CHAPS detergent bicelles; conditions which have been demonstrated in this chapter to be detrimental to catalytic activity due to hydrophobic mismatch. Despite this, no significant structural changes are observed between the two environments, indicating that *ec*GlpG does not undergo the same level of rearrangement as seen in other membrane proteins affected by hydrophobic mismatch.

Given the absence of obvious structural differences in β-NG versus bicelle-solubilized *ec*GlpG structures, it is possible that the observed loss in proteolytic activity instead reports on a change in the equilibrium distribution between different conformational states.[78,86,89] A previous study found that the mobility of an EPR probe attached to α5 was substantially

reduced in a phospholipid bilayer environment relative to detergent micelles.[53] Hydrophobic mismatch may similarly prevent opening of the α5 substrate gate by perturbing the equilibrium between the open and closed states without directly altering the structure of either state. If one of these states had less helix structure, then this equilibrium shift would be expected to manifest itself as a decrease in overall α-helical content consistent with the subtle differences in CD spectra observed between Fos-14 and Fos-12/Fos-10-solubalized TMD81.

The available X-ray structures and MD simulations reveal that the thickness of the *ec*GlpG hydrophobic belt is not uniform, being thinnest (~20 Å) in the vicinity of the α5 gate helix.[89,98] This results in an asymmetrical thinning of the membrane environment, with the most severe deformation occurring in the α5 gate region,[98] potentially promoting the entry of detergent or lipid molecules into the rhomboid active site. The observation that hydrophobic mismatch primarily effects k_{cat} rather than K_M suggests that these bound detergent or lipid molecules do not inhibit substrate binding in a competitive manner. However, entry of detergent or lipid into the active site could alter k_{cat} by disrupting substrate gating or perturbing the active site structure. The feasibility of such an interaction is supported by X-ray crystal structures showing a phospholipid molecule within the *ec*GlpG active site,[85] producing a slight rearrangement in the extracellular side of active site helices α4, α5 and α6, coupled with a shift in the position of S201 and to a lesser extent H254 (Fig. 3.8A). A subsequent structure that depicted a detergent molecule bound in the active site of a S201T *ec*GlpG mutant showed a more extensive perturbation in H254 which was shifted by ~0.6Å, potentially disrupting its ability to hydrogen bond with S201 as it does in the WT *ec*GlpG (Fig. 3.9B) and giving rise to the reduced catalytic activity observed under mismatch conditions.

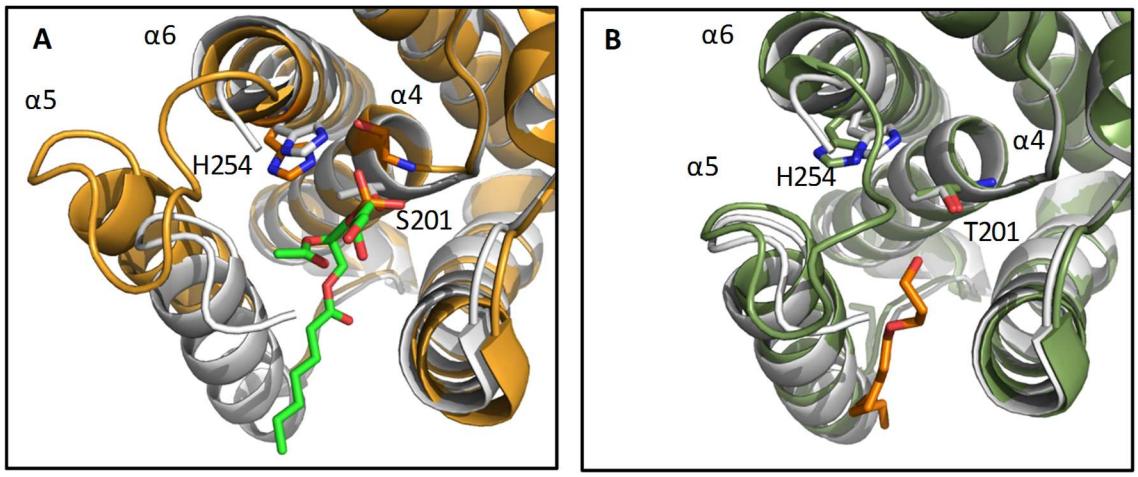


Figure 3.8: Lipid and detergent binding to the *ec*GlpG active Site A) Structure of *ec*GlpG (PDB 2IRV) depicting a POPG phospholipid bound into the rhomboid active site (orange) compared to that of the apo-structure obtained from the second molecule of the same asymmetric unit (white). The bound phospholipid is coloured in green. B) Structure of *ec*GlpG S201T (PDB 2XTU) depicting a β-nonyglucoside detergent (orange) bound into the active site (green). The structure of the apo-form of the same mutant obtained in a phospholipid bilayer (PDB 2XTV) is shown in white.

3.3.3: New Avenues for the Regulation of Intramembrane Proteolysis

The studies presented in this chapter demonstrate that the ability of hydrophobic mismatch to alter *ec*GlpG activity, activity, highlighting the complex interplay between rhomboid protease structure and function, and the local membrane environment. The physical properties of this membrane environment can be varied *in vivo* through changes in the local membrane composition, or the localization of rhomboid proteases to lipid rafts or microdomains enriched in lipids with specific headgroups or alkyl chain lengths, potentially providing a mechanism for the spatial and temporal regulation of rhomboid proteolysis required for various biological processes such as parasitic host-cell invasion by *Toxoplasma gondii*. Here, the *Tg*ROM4 rhomboid is responsible for cleaving the MIC cell-surface adhesins that initially bind to the host cell, and are slowly displaced towards the posterior pole of the parasite as the invasion process progresses (Fig. 3.9). While *Tg*ROM4 is uniformly

distributed across the cell, adhesin cleavage and shedding only occurs when the moving junction reaches the posterior region.[161,162] A similar phenomenon was also observed in the *Entamoeba histolytica* rhomboid *eh*ROM1 which is responsible for cleaving lectins on the amoeba cell surface. While *eh*ROM1 is initially distributed throughout the plasma membrane, it undergoes several changes in subcellular localization throughout the course of the invasion process. Here, catalytic activity is only observed at certain stages of the invasion process,[163] potentially reflecting the effect of different membrane compositions at particular subcellular locations. These findings may also be applied to other IMPs such as γ-secretase, the catalytic activity of which has been shown to be dependent on the local membrane composition, and whose localization to lipid rafts and cholesterol-rich microdomains has been associated with APP cleavage and Alzheimer's disease.[164–167]

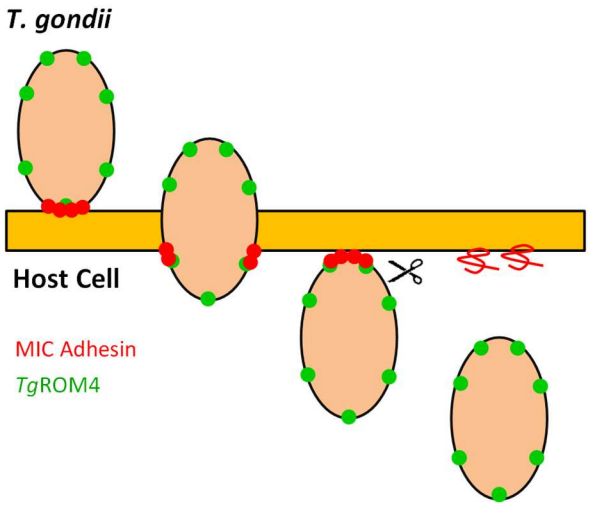


Figure 3.9: MIC adhesin cleavage during *T. gondii* parasite invasion. Note that while *Tg*ROM4 (green) is distributed throughout the plasma membrane, cleavage of MIC (red) occurs only at the posterior end of the cell

While rhomboids are constitutively active when assessed *in vitro*, the low levels of proteolysis observed in bilayer systems whose hydrophobic dimensions match those predicted

for the native membrane environment suggests that additional factors are required to maintain maximal proteolytic activity *in vivo*, providing additional avenues for the regulation of rhomboid activity and selectivity, some of which will be explored in greater detail in subsequent chapters of this thesis.

3.4: Materials and Methods

3.4.1: Protein Expression and Purification:

Wild-type and truncation mutant *ec*GlpG constructs were overexpressed in C43(DE3) *E. coli* cells and purified using previously published protocols.[63,81] In brief, bacterial cell pellets from 1 L of culture grown in M9 minimal media were resuspended in 20 mM HEPES buffer (pH 7.4) containing 100 mM NaCl, 5 mM $MgCl_2$, 10% glycerol, and 1X protease inhibitor cocktail (1mg/mL 4-(2-aminoethyl) benzenesulfonyl fluoride hydrochloride (AEBSF), 2 μg/mL pepstatin, 1.2 μg/mL E-64, 1 μg/mL bestatin, 1.6 μg/mL phosphoramidon). Cells were lysed by sonication (Fisher Sonic Dismembrator Model 500) at 50% intensity, using 60 one-second pulses separated by an equal delay time for a total of 1 minute. Alternatively, lysis was carried out using an Emulsiflex-B15 Homogenizer, undergoing a total of two passes at 80 PSI. The lysate was then centrifuged at 50,000xg for 1 h to isolate the insoluble membrane fraction, which was then solubilised for 1 hour in 50 mM HEPES buffer (pH 7.4) containing 200 mM NaCl, 5 mM imidazole, 10% w/v glycerol, 1x cocktail inhibitor and 1% w/v n-dodecyl β-D-maltoside (Mal-12). The supernatant containing the detergent-solubilised protein was applied to a nickel-affinity chromatography column. Unbound protein was removed through two sequential wash steps using 50 mM HEPES buffer containing either 500 mM or 200 mM NaCl and 0.1% w/v of the desired detergent. The bound

rhomboid was eluted using 250 mM imidazole and applied to a Superdex-200 10/300 GL size exclusion column using an AKTA FPLC (GE Healthcare) equilibrated in GlpG buffer containing 50 mM Tris pH 7.4, 150 mM NaCl, 100 μM EDTA, and 0.1% w/v detergent. In some cases, it was necessary to employ higher concentrations of detergents in the Ni-NTA and size-exclusion chromatography steps in order to exceed the critical-micelle concentration. In particular, 0.2%, 0.35%, and 1.3% (w/v) of Mal-10, Mal-9, and Mal-8 were used for the Ni-NTA, whereas 0.3% (w/v) was used for purification into Fos-10. Fractions of the main peak were collected, and the resulting protein concentrations and purity were assessed using a BCA assay (Pierce) and SDS-PAGE analysis respectively.

3.4.2: Analytical Size-exclusion Chromatography

Molecular weight standards (Amersham, Sigma- Aldrich) comprised of albumin (67 kDa), ovalbumin (45 kDa), chymotrypsin (25 kDa), and lysozyme (14.3 kDa) were loaded onto a Superdex 200 10/30 size exclusion chromatography column and eluted using an AKTA FPLC in GlpG buffer. The log of the molecular weight of the various standards employed was plotted against the observed elution volume (v_e) normalized against the void volume (v_o), as defined by the elution volume of blue dextran (~2000 kDa) to obtain a linear calibration curve.[168] Elution volumes from TMD81 samples in the various detergent-micelle systems tested were compared to the calibration curve to calculate an apparent molecular weight for the protein detergent complex.

3.4.3: CD Spectroscopy:

CD spectra were collected on 5 μM purified rhomboid solubilised in GlpG buffer containing the detergent of interest at 37ºC using a Jasco J-815 spectrometer as described previously.[81] Eight accumulations were collected for each sample using a data pitch of 0.2 nm and a scan rate of 20 nm/min, and averaged to produce a single spectrum. At least three spectra were acquired for each construct from at least two independent purifications unless otherwise noted. Secondary structure calculations were carried out with the CDPro software package (Jasco) using the CONTIN method and the SMP56 reference set.[152] To assess the effect of the local membrane environment on thermostabilty, samples of TMD81 in the various detergent-micelle systems were heated from 35ºC to 97ºC at a rate of 0.3ºC/min. The elipticity at 222 nm was monitored and fit to a two-state Boltzmann curve to determine the melting temperature (T_m).[81,95]

3.4.4: Reconstitution into Phospholipid Bicelles:

Samples of TMD81 were first reconstituted into phospholipid vesicles which were subsequently converted into bicelles by solubilisation in detergent, similar to previously established protocols.[123,169] PC-12 or PC-14 phospholipids (Avanti Polar Lipids) were deposited onto the surface of a glass vessel as previously described to produce a thin lipid film,[71] which was subsequently hydrated in detergent-free GlpG buffer to produce a 20 mg/mL proteoliposome suspension. This was then extruded through a 0.1 μm filter using an Avanti Mini-Extruder, and the resulting unilamellar vesicles were solubilised using 0.12 mM 12-Mal in GlpG buffer for 30 min. Purified TMD81 in 0.1% (w/v) Mal-12 was added to produce a final protein concentration of 0.1 mg/mL in a 1 mg/mL lipid solution, and incubated

at room temperature for an additional hour. Mal-12 was removed by overnight dialysis followed by the addition of two 30 mg aliquots of Amberlite beads (Sigma-Aldrich) per mg Mal-12, with an incubation time of 1–2 hours per aliquot to allow for liposome reformation and TMD81 incorporation. The resulting proteoliposomes were collected via ultracentrifugation, re-suspended in GlpG buffer, and re-extruded through a 0.1 μm filter as described above. Density gradient centrifugation in a 0-50% sucrose gradient (200,000g for 90 min), followed by SDS-PAGE analysis of the separate liposome and pellet fractions was used to confirm the successful reconstitution of TMD81 into the bilayer envelope. CHAPS detergent (Bioshop) was added to the proteoliposome solution to achieve a molar phospholipid:bicellar detergent ratio (q_{eff}) of 1 and a final total (phospholipid + CHAPS detergent) concentration of 1.5% (w/v). The mixture was then subjected to multiple cycles of heating to 42°C and cooling to room temperature to yield small isotropic bicelles similar to those employed in previous rhomboid studies.[97]

3.4.5: *ec*GlpG Kinetics Measurements:

The EnzChek Protease Assay Kit (Invitrogen) was used to measure k_{cat} against a soluble casein substrate as described previously.[81] Here, 0.25 μM TMD81 in GlpG buffer was incubated with increasing concentrations of the BODIPY-casein substrate in a black 96-well plate (Corning) at 37°C. Initial rates of hydrolysis were measured for 30-60 minutes using a Spectramax Gemini XS microplate reader and fit to the Michaelis-Menten equation to determine values of K_m and V_{max} (Fig. 3.2). Fractions taken at several time points during a reaction were subsequently analysed using SDS-PAGE to convert the fluorescence intensities reported by the microplate reader into molar concentrations of product as previously

described,[79] giving rise to k_{cat} values that were comparable to those previously reported in the literature for both the full length *ec*GlpG and isolated TM domain using similar substrates in Mal-12.[52,79] Activity values in Fos-12/Fos-14 mixed micelle systems were obtained by diluting samples of TMD81 purified in Fos-12 with GlpG buffer containing 0.1% Fos-14 to obtain the desired ratio of the two detergents. 0.25 μM of TMD81 in the resulting mixed micelles was incubated with a fixed concentration (1.5 μM) of BODIPY-casein substrate and the resulting rate of proteolysis recorded as describe above. The experiment was repeated using sample of TMD81 purified into Fos-14, and subsequently diluted with Fos-12 to confirm that the loss of activity observed in Fos-14 was reversible. The activity of TMD81 in phospholipid bicelles was measured using 0.75 μM of purified rhomboid incubated with increasing concentrations of BODIPY-casein. Samples were removed at regular time intervals and analyzed using SDS-PAGE to quantify the amount of product produced.

Chapter 4: The Road to Solution-NMR Studies of *ec*GlpG Rhomboid

Contributions of collaborators:

Jason K.C. Kwok and Adam Damry (University of Ottawa) provided guidance for the NMR data analysis.

4.1: Introduction

The numerous high-resolution X-ray crystal structures of *ec*GlpG and *hi*GlpG described in Table 1.1 provide valuable insight into the structure of rhomboid proteases. However, their static nature limits our ability to study the effect of conformational dynamics on rhomboid function. In this chapter, protocols were developed to study the rhomboid protease TM domain construct TMD81 using solution-state NMR, since this has the potential to report on the structure and dynamics of rhomboid in a way that compliments the static structures produced by X-ray crystallography. In particular, this chapter focuses on the identification of an optimal membrane-mimetic system which is able to preserve rhomboid structural integrity and catalytic function, while producing an overall protein-detergent complex (PDC) size that would minimize signal broadening due to transverse relaxation.

The results obtained in the preceding chapter demonstrated that phosphocholine (Fos), lysophosphocholine (Lyso-Fos) and maltoside (Mal)-based amphiphiles can be used to isolate TMD81 in a fully-folded, stable conformation with varying levels of catalytic activity. Some of the detergents tested, including Fos-12, Mal-12, Mal-10, and Lyso-Fos-14, have been successfully used in NMR studies of other integral membrane proteins, making them promising candidates for NMR studies of *ec*GlpG. However, other detergents are also available that could be appropriate for solution NMR of TMD81. For example, a novel family of glucose and maltose neopentyl glycol (NG) based detergents have been developed that contain two non-polar alkyl chains, potentially providing a better mimic of the physical properties of a phospholipid bilayer environment (Fig 4.1A) and contributing to the enhancements in thermostability and/or catalytic activity observed for a range of membrane proteins.[124,170–173] A member of these detergents (NG-10) was used in ^{19}F-NMR studies of the

β_2AR GPCR where it was found to reduce the level of conformational heterogeneity previously observed for this system in Mal-12 micelles, allowing researchers to capture the physiologically relevant, functional state.[124] The short-chain phospholipid diheptanoylphosphatidylcholine (PC-7, Fig. 4.1B) provides another promising alternative, since it also forms micelles in solution that would be more similar to natural membrane environments. PC-7 micelles have been used in NMR studies of integral membrane proteins such as OmpX and rhodopsin;[141,174,175] in the case of the latter, the choice of PC-7 as the detergent system was critical, as it was the only detergent of the 15 that were tested to yield NMR spectra of sufficient quality to allow its structure determination.[141,175] While the ability of lysophosphocholine detergents to sustain TMD81 in a stable confirmation has been assessed in the previous chapters, other lysophospholipids such as the anionic lysophosphoglycerol are also promising candidates, with the 16-carbon version Lyso-Gly-16 (lysopalmitoylphosphoglycerol) being identified from a screen of 27 other candidates including members of the Fos, Mal, and Lyso-Fos families as being the optimum detergent micelle system for NMR studies on five different membrane proteins including the SMR bacterial transporter and the OF4 ATP synthase.[176]

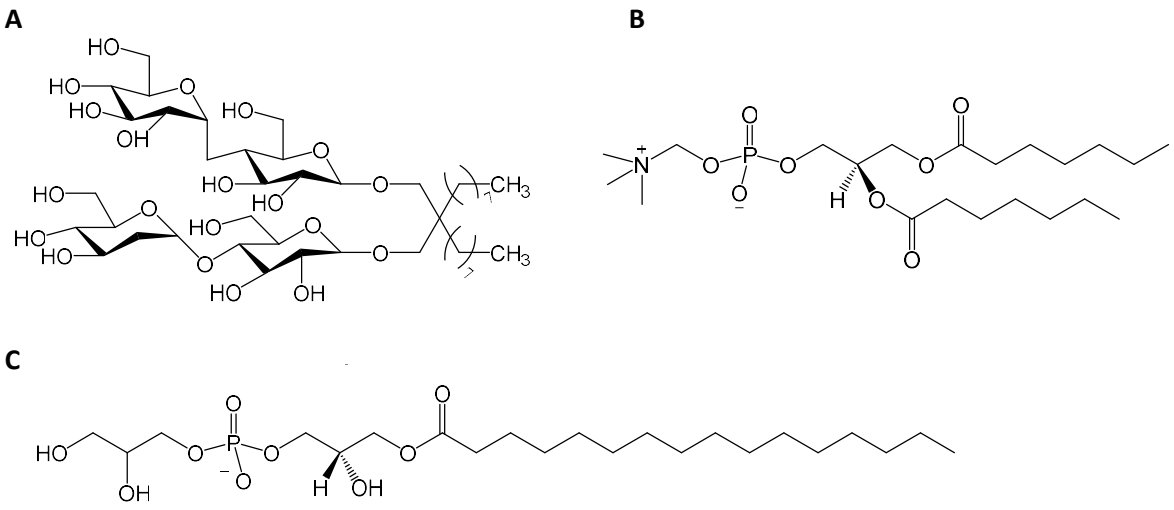


Figure 4.1: Structure of NG-10 (A), PC-7 (B) and Lyso-Gly-16 (C) amphiphiles employed in this chapter.

It also is possible that binding of small molecule inhibitors could enhance the quality of NMR spectra of a membrane protein through increasing its stability and/or altering its conformational dynamics. For example, the structure of *ec*GlpG bound to a 7-amino-4-chloro-3-methoxy isocoumarin inhibitor showed covalent modification of both active site histidine and serine side chains that essentially cross-links α4 and α6 (Fig 4.2). However, this was also accompanied by a rotation of the W236 side chain away from the substrate binding pocket while a smaller rotation around Y205 pulls its side chain away from the catalytic H254 enabling to form a new hydrogen bond with the repositioned W236 backbone (Fig. 4.2). This rearrangement within the active site was also accompanied by a slight change in the position of TM5 and TM6, with an average displacement of 1.25Å. Formation of a complex with isocoumarin-based inhibitors could potentially stabilize the structure of rhomboid, enhancing NMR spectral quality. However, this may also be accompanied by structural perturbations whose impact on NMR spectral quality is difficult to predict.

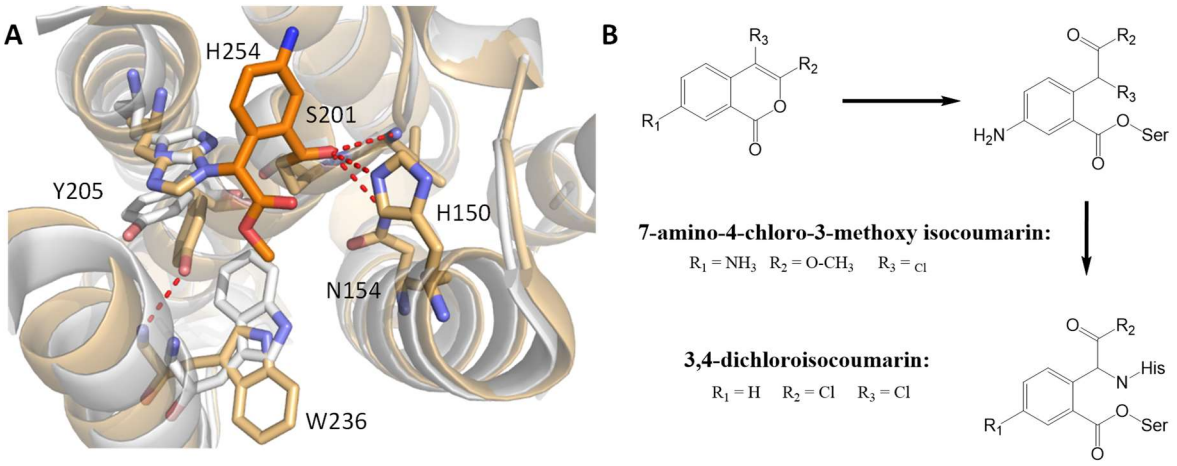


Figure 4.2: Binding of isocoumarin inhibitors to *ec*GlpG. A) Structure of *ec*GlpG bound to 7-amino-4-chloro-3-methoxy isocoumarin in orange (PDB 2XOW) overlaid upon the apo-form in white (PDB 2XTV). B) Structure and mechanism of isocoumarin based inhibitors

In this chapter, samples of the *ec*GlpG TMD81 construct were prepared in the detergent micelle systems described above. The resulting PDC's were assessed for their ability to maintain protein structure and catalytic activity using functional assays and CD spectroscopy, while complex size and homogeneity was assessed using size-exclusion chromatography. Promising candidates which produced small PDCs while sustaining high levels of catalytic activity and thermostability were identified and their spectral properties evaluated using ^{1}H-^{15}N HSQC spectra. Of all the detergents tested, Fos-12 was found to be optimal for solution-state NMR of TMD81 although these NMR spectra still lacked sufficient resolution to enable assignment of the backbone amide peaks using standard triple-resonance experiments, and attempts to enhance spectral quality through the addition of an isocoumarin-based inhibitor failed to produce any improvements. Nonetheless, these results represent the first time the rhomboid TMD has been observed using this technique, and provide the foundation for the structural characterization of the TMD81 α5-open conformation that is presented in the subsequent chapter.

4.2: Results

4.2.1: TMD81 Purification and Activity

Previous studies, including those outlined in the preceding chapter of this thesis, describe a method for the purification of both full-length and TMD81 rhomboid constructs into a range of Mal, Fos, and Lyso-Fos detergents, with yields of around 0.4-1 mg per 1 L of cell culture grown in M9 minimal media. This same protocol was used to purify TMD81 into maltoseneopentyl-glycol (NG), lysophospho-glycerol (Lyso-Gly), and short-chain phosphatidylcholine (PC) detergent micelle systems, from which similar levels of protein yield and purity were obtained (Fig 4.3). All samples retained significant levels of catalytic activity against BODIPY-casein (Fig 4.4A) and produced CD spectra broadly similar to those obtained in previous chapters (Fig 4.4B), suggesting that TMD81 structural integrity was preserved in all of the detergents tested.

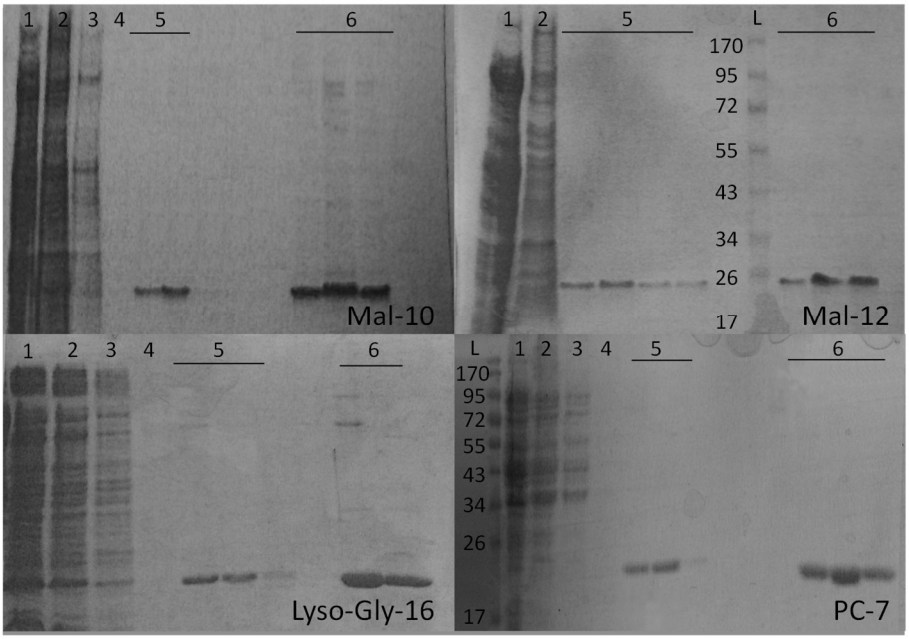


Figure 4.3: Representative SDS-PAGE gel of fractions taken during various stages of the purification of TMD81. L: Molecular weight ladder in kDa. 1: Soluble protein fraction. 2: Ni-NTA column flow through. 3 and 4: Non-specifically bound contaminants removed during Ni-NTA wash step. 5: Purified TMD81 eluted using 250 mM imidazole. 6: Representative fractions obtained after size-exclusion chromatography containing the final, purified TMD81

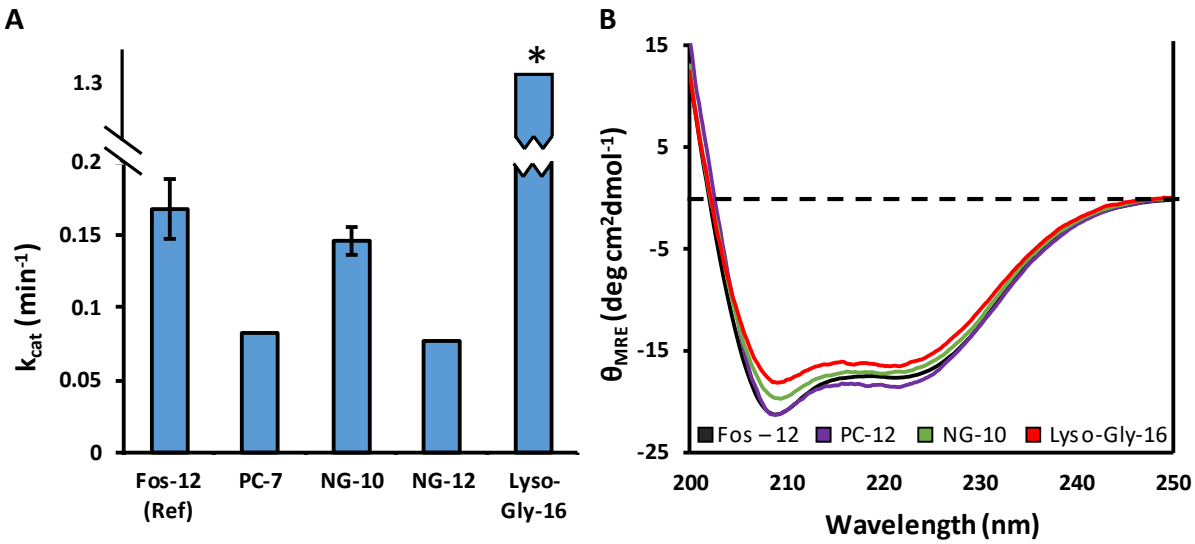


Figure 4.4: Effect of various detergents on rhomboid structure and activity. A) k_{cat} values for TMD81 solubilised in micelles composed of PC-7, NG-10, NG-12 and Lyso-Gly-16. Values lacking error bars represent data obtained from only a single trial. Samples solubilized in Lyso-Gly-16 (*) displayed abnormally high values of K_M, which may contribute to the elevated values of k_{cat} predicted in this detergent B) Representative CD spectra of TMD81 purified into PC-7, Lyso-Gly-16 and NG-12. Note that k_{cat} values and CD spectra for TMD81 in Fos-12 were presented in Chapter 3, and are included here for reference.

4.2.2: PDC Size and Thermostability

Size exclusion chromatography was used to determine the apparent molecular weight of the PDCs obtained using the various detergent micelle systems. In general, it was found that the size of the PDCs is dependent on the length of the alkyl chain, consistent with previous studies carried out on empty detergent micelles (Figure 4.5A).[126,127] Curiously, Lyso-Gly gave rise to a broader size-exclusion chromatography elution profiles (Fig. 4.5B), similar to the broad profile observed for Lyso-Fos and Mal-detergents (Fig. 3.3D). The broad elution peaks suggest that these detergents give rise to a heterogeneous combination of two or more species, potentially resulting from a monomer-dimer equilibrium. The apparent molecular weight measured for the Mal-12 PDC (120±6 kDa) was similar to that of the dimeric full-length *ec*GlpG (~140 kDa) observed previously in Mal-12 micelles using sedimentation equilibrium

analysis.[108] This is also consistent with analytical ultracentrifugation data demonstrating that the full-length *ec*GlpG and the homologue from *H. influenza* that does not contain a cytoplasmic domain are both able to form dimers. To further verify the dimeric state of TMD81 under NMR conditions, exclusion chromatography measurements were obtained for theTMD81 Mal-12 PDC across a range of detergent concentrations (0.05%-0.15%). No significant variation (p=0.1-0.2) in apparent PDC size was observed, suggesting it was the detergent type, and not detergent-to-*ec*GlpG ratio that is important for determining oligomerization state.

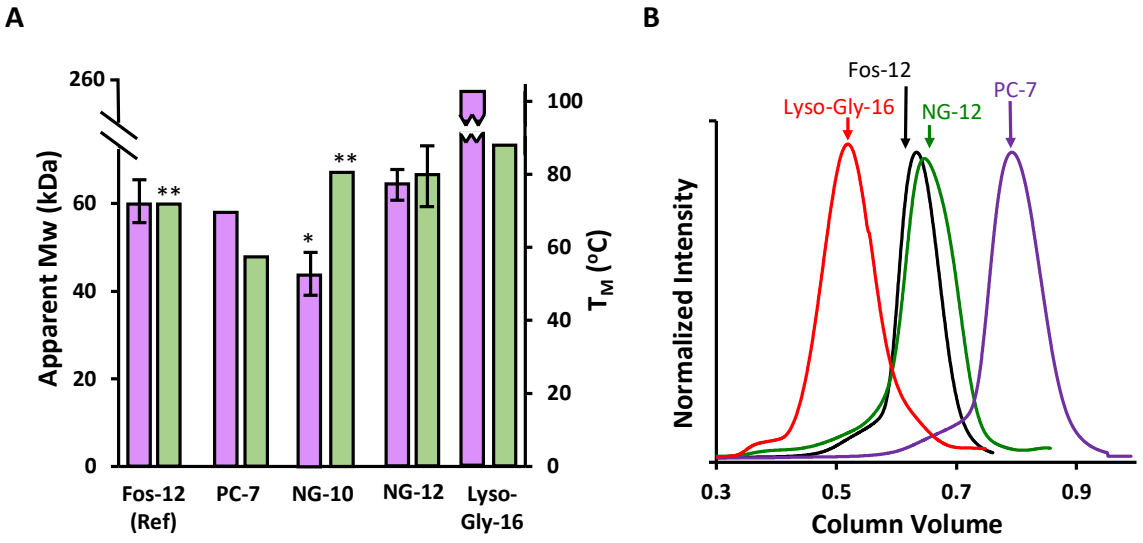


Figure 4.5: Physical properties of candidate TMD81 PDCs for solution-NMR studies. A) Apparent molecular weight (purple) and thermostability (green) values of TMD81 solubilised in PC-7 NG-10, NG-12 and Lyso-Gly-16. Values lacking error bars represent data obtained from a single trial except for those indicated by **, whose error bars are too small (±0.7°C) to display. Values that were significantly different from those obtained in Fos-12 ($p < 0.05$) are indicated by *. B) Representative FPLC elution profiles of TMD81 purified into PC-7, Lyso-Gly-16 and NG-12. All data for TMD81 in Fos-12 were collected in the previous chapter, and are presented here as a reference.

PDC stability was tested using thermal denaturation monitored by CD spectroscopy, and showed that PC-7 solubilized TMD81 had the lowest melting point, in line with the

general correlation between alkyl chain length and thermostability. In contrast, NG-10 yielded a 10-15^{o}C increase in T_m relative to the equivalent Fos or Mal detergents, potentially due to the stabilizing effect of its twin alkyl chains. However, increasing the length of the alkyl chain to 12 carbons did not further increase the thermostability (Fig. 4.5A). Lyso-Gly-16 also did not appear to provide further enhancements in thermostability over the other detergents tested, despite the longer alkyl chain.

4.2.3: NMR Spectroscopy of Detergent-solubilized TMD81

To test the utility of each amphiphile in solution NMR, ^{1}H-^{15}N HSQC TROSY spectra were acquired for TMD81 in detergents that supported high levels of proteolytic activity while maintaining a small PDC size. NG-10 was one of the first to be tested due to the favourable characteristics of its PDC. However, as shown in Figure 4.6A, TMD81 solubilised in NG-10 yielded poor-quality spectra, with only 23 of the 191 expected backbone amide peaks (Table 4.1). Extensive precipitation was observed in NMR samples removed following acquisition of the NMR experiment (~6h at 35^{o}C), suggesting that this detergent was deficient in maintaining the long-term stability of TMD81 at the elevated concentrations and temperatures required for solution-NMR

As described in Chapter 3, Lyso-Fos-14 PDCs also displayed high levels of thermostability, and so HSQC spectra were acquired for TMD81 in this micelle system. Although this detergent prevented precipitation during the course of a 12-hour NMR experiment at 45^{o}C, significant peak broadening was observed in the resulting spectra (Fig. 4.6B), potentially due to the larger size and increased polydispersity of the PDC. Reducing PDC size through the use of Lyso-Fos-12 produced a noticeable enhancement in spectral quality despite the corresponding loss in thermostability (Fig. 4.6C). Spectra were also

acquired in Mal-12, since this is the detergent most widely used for in vitro assays of *ec*GlpG activity. However, a similar level of peak broadening was observed (Fig 4.6D).

Given the robust activity and good thermostability supported by Fos-12, this detergent was also tested for its suitability for solution NMR studies of TMD81. As shown in Figure 4.6E there was a substantial increase in peak number and dispersion, in addition to enhanced signal intensity compared to all the other detergents tested. Fos-14 was also tested, since increasing the length of the alkyl chain by 2-carbons increases its thermostability (Chapter 3). However, this comes at a cost of increased PDC size and reduced molecular tumbling rates, resulting in peak broadening (Fig. 4.6F). Based on these results, Fos-12 was identified as the most promising candidate for the study of TMD81 using solution-NMR.

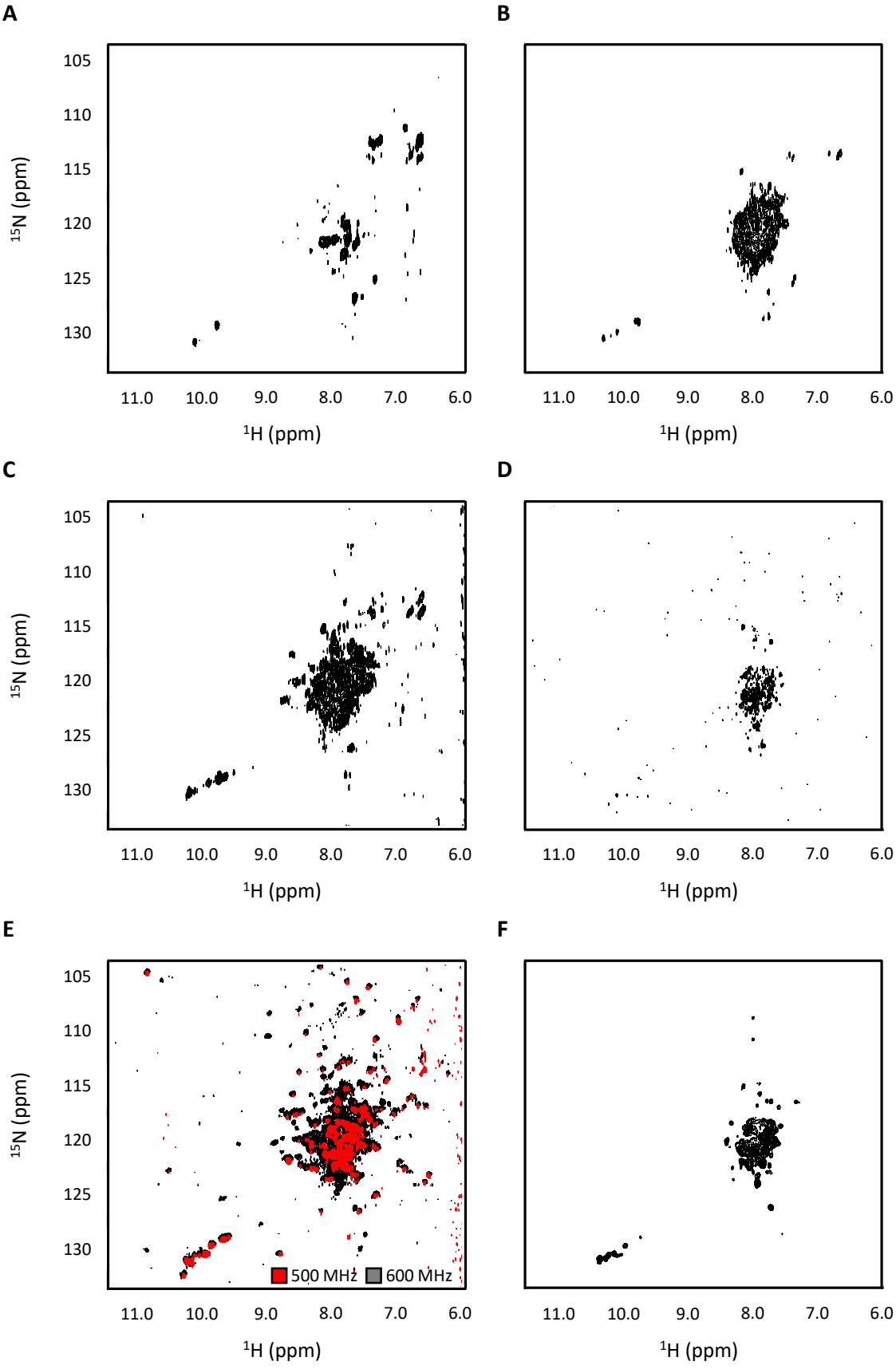

A
B
C
D
E
F
105
110
115
120
125
130
^{15}N (ppm)
11.0
10.0
9.0
8.0
7.0
6.0
^{1}H (ppm)
500 MHz
600 MHz

Figure 4.6 (Previous Page): ^{1}H-^{15}N TROSY NMR spectra of TMD81 in A) NG-10, B), Lyso-Fos-14, C) Lyso-Fos-12, D) Mal-12 E) Fos-12 and F) Fos-14. All spectra have been normalized for sample concentration, and were acquired using a 500 MHz spectrometer with the exception of (D) and (F), which were collected using a 600 MHz spectrometer with cryoprobe. (E) Includes spectra obtained from both instruments for the sake of comparison. All samples were analyzed at 45ºC with the exception of NG-10 (A), which was examined at 35ºC.

Table 4.1: Number of backbone and tryptophan side chain peaks observed in TMD81 ^{1}H-^{15}N TROSY spectra acquired at 45ºC, using a 500 MHz spectrometer at 45ºC unless otherwise specified.

Detergent	Backbone Peaks	Trp Side Chain Peaks	Notes
NG-10	23*	2	Precipitation observed
Lyso-Fos-12	25*	7	
Lyso-Fos-14	21*	9	
SDS	14*	6	Acquired at 35ºC
Fos-12	89	10	
Fos-12	128	13	600 MHz with cryoprobe
Fos-14	63*	11	600 MHz with cryoprobe
Mal-12	47*	2	600 MHz with cryoprobe

* Peak numbers represent estimates.

To identify the experimental conditions that give rise to maximum spectral quality using this Fos-12 detergent system, ^{1}H-^{15}N HSQC and TROSY spectra were collected across a range of temperatures and detergent concentrations. Increasing the temperature from 25ºC to 45ºC increased the number and resolution of peaks in the spectrum while preserving sample stability (Fig. 4.7A). No significant changes were observed in spectra obtained across a range of Fos-12 concentrations as shown in Figure 4.7B, in line with previous studies suggesting that micelle shape and morphology is independent of detergent concentration once the CMC has been reached.[126,177]

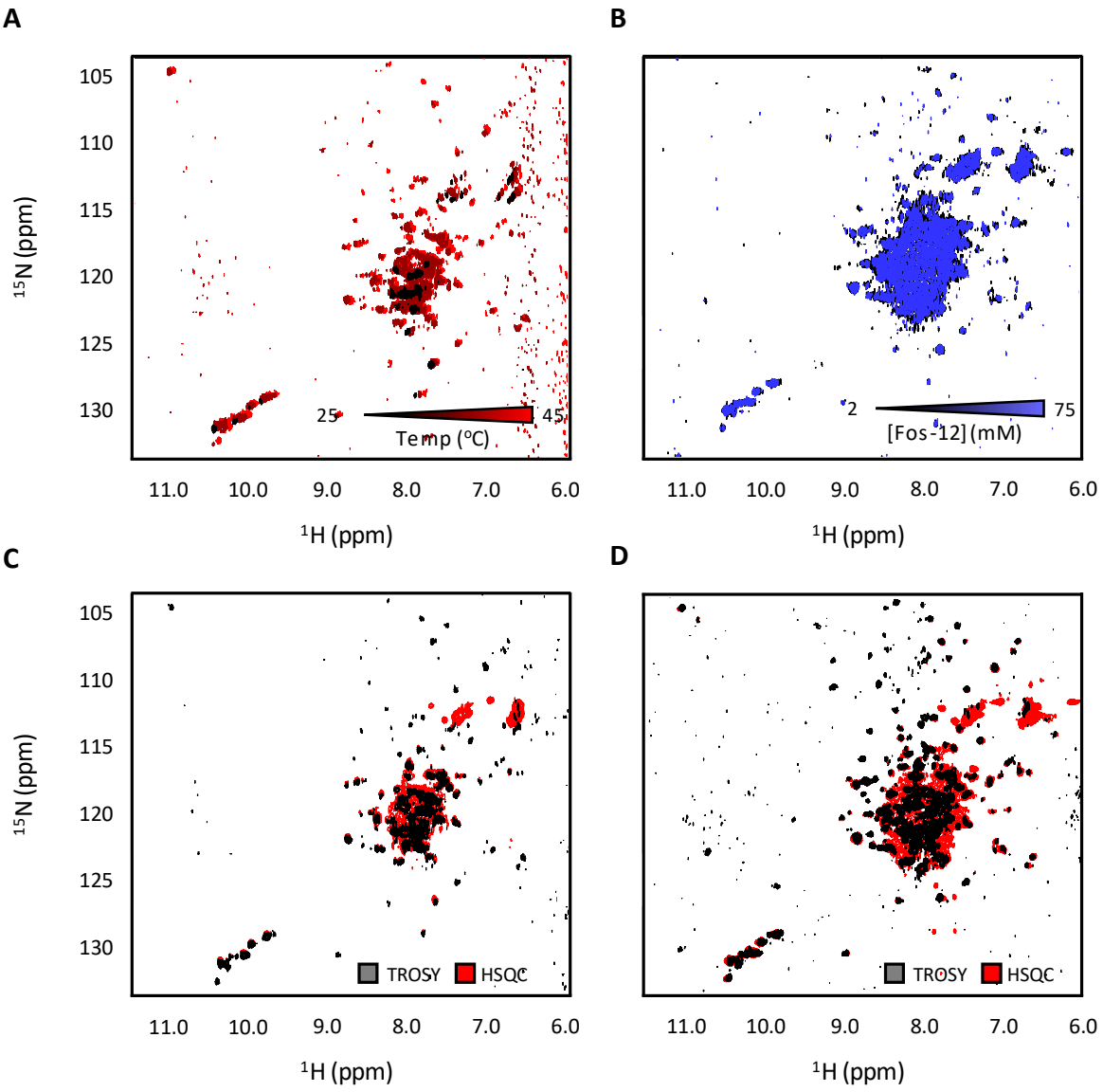


Figure 4.7: Optimization of NMR conditions for TMD81 in Fos-12 micelles. A) ^{1}H-^{15}N TROSY spectra of TMD81 in 2 mM DPC at 25ºC, 35ºC, or 45ºC. B) ^{1}H-^{15}N HSQC spectra of TMD81 in 2 mM (0.1%) and 75 mM DPC. Comparison of ^{1}H-^{15}N HSQC and TROSY spectra obtained for TMD81 using either a 500 MHz spectrometer with room temperature probe (C) or 600 MHz spectrometer with cryoprobe (D) in 0.1% Fos-12 at 45ºC. Note that the TROSY spectra in C and D were shifted in both dimensions by ½ $^{1}J_{NH}$ to facilitate comparison with their HSQC counterparts. All other spectra were obtained using ~200 µM TMD81 with the 500 MHz instrument unless otherwise specified.

Spectral quality was also assessed for TMD81 in Fos-12 at 45ºC for TROSY versus non-TROSY variants of the ^{1}H-^{15}N HSQC. There was a 2-fold reduction in overall signal intensity with the use of the TROSY experiment for backbone amide peaks in the central region of the spectrum. However, there was also an increase in the number of backbone ^{1}H-^{15}N correlations coupled with an enhancement of peak resolution in TROSY spectra obtained on both the 500 MHz and 600 MHz instruments, making this the preferred experiment for the study of TMD81 in Fos-12 (Fig. 4.7C, 4.7D). Under these optimized conditions (TROSY experiment at 45ºC, 0.1% Fos-12 pH 7.0 or 7.4), 89 backbone amide peaks could be identified using the 500 MHz instrument, representing just under 50% of the expected peaks based on the number of non-proline residues in the TMD81 construct. Use of the 600 MHz spectrometer with cryoprobe increased the number of backbone peaks to 128, representing 67% of the expected correlations. In contrast, samples solubilized in Mal-12 yielded only 47 backbone peaks under similar conditions while samples in Fos-14 yielded approximately 65 backbone peaks, or 34% sequence coverage (Table 4.1).

As discussed in Chapter 2, phospholipid bicelles provide a more accurate approximation of the native bilayer environment, preserving the native structure and function of any embedded membrane proteins and potentially enhancing NMR spectral quality. To explore this possibility, samples of TMD81 were prepared in the PC-12/CHAPS bicelle systems which supported proteolytic activity in Chapter 3. However, extensive precipitation was observed at the protein concentrations required for NMR studies (~200 μM). Subsequent SDS-PAGE analysis of the resulting sample indicated that the majority of the TMD81 had precipitated as part of the insoluble fraction, with little protein remaining in the supernatant. Subsequent attempts to prevent precipitation by reducing the protein concentration and the q-

ratio failed, while NMR spectra obtained under all conditions tested were unable to resolve any peaks corresponding to TMD81.

4.2.4: Small Molecule Inhibitor Binding

Conformational exchange on a slow to intermediate NMR timescale likely contributes to the poor spectral quality obtained for TMD81. To test the ability of a covalent inhibitor to alter these dynamics and potentially improve the quality of the spectrum, spectra were acquired for TMD81 in Fos-12 in the presence of 800 μM 3,4-dichloroisocoumarin (DCI), a small-molecule covalent rhomboid inhibitor. However, as shown in Figure 4.8, there was no significant change in NMR spectral quality, even though the concentration exceeded the IC_{50} value by ~40-fold.[178] The activity of this inhibitor was confirmed in an assay of TMD81 activity performed under NMR conditions, which showed complete inhibition. A modest (24±8%) recovery in catalytic activity was observed in samples analyzed following acquisition of NMR spectra, which reflects the loss of the enzyme-inhibitor complex during the course of the 6-hour experiment due to hydrolysis. This resulted in a heterogeneous, incompletely inhibited sample and may partly explain the absence of an improvement in TMD81 NMR spectral quality when inhibited by DCI.

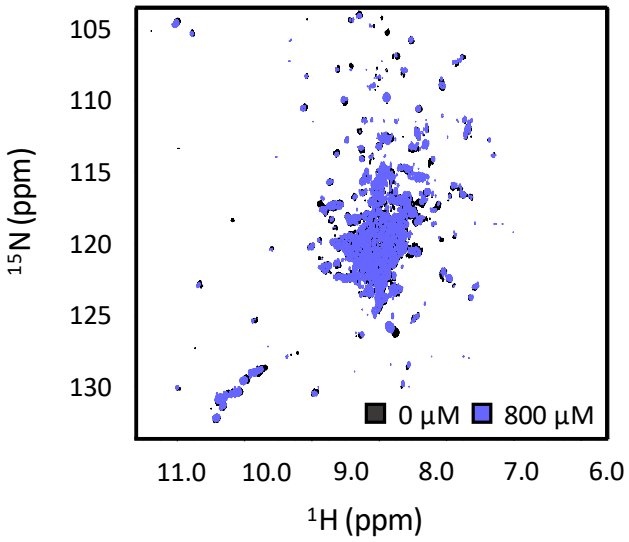


Figure 4.8: ^{1}H-^{15}N TROSY NMR spectra of TMD81 in Fos-12 micelles in the presence of 800 µM DCI protease inhibitor (blue) overlaid upon a control spectra collected in the absence of inhibitor (black). Both spectra obtained using the 600 MHz spectrometer with cryoprobe.

4.3: Discussion

4.3.1: Assessment of Membrane Mimetic Systems for Solution-NMR Studies of *ec*GlpG

The results presented in this chapter demonstrate that *ec*GlpG TMD81 can be purified into a wide range of detergent micelle systems in an active state with good thermostability. However, most of the detergent systems tested gave rise to poor NMR spectra for TMD81, with extensive peak broadening that, in most cases (i.e. Mal-12, Lyso-Fos-12, Lyso-Fos-14), could be attributed to the large size of the PDC. The ability of the isolated *ec*GlpG TMD to form dimers when purified into Mal-12 detergent micelles has been demonstrated in previous studies and is consistent with the results presented in this chapter.[52,108] While the oligomeric state of TMD81 in other detergent systems was not directly assessed in this thesis, the size the PDCs obtained in Lyso-Fos detergents significantly exceed the combined molecular weight of the free detergent micelle and the monomeric rhomboid species, potentially reflecting the presence of a dimeric or even a tetrameric state. Exchange between these oligomeric states

could give rise to the broad elution profiles obtained for TMD81 in the Mal-12, Lyso-Fos-12, and Lyso-Fos-14 detergents, representing a significant barrier to further characterization via NMR.[52]

Table 4.2: Molecular weights of TMD81 PDCs, free micelles and theoretical PDCs.

Detergent	TMD81 PDC Mw[#] (kDa)	Predicted Mw of Free Micelle* (kDa)	Predicted Mw of Free Micelle + Monomeric TMD81 (kDa)
NG-10	59±1	N/A	N/A
Lyso-Fos-12	110±1	34-36[129]	50
Lyso-Fos-14	223	42-59[129,130]	80
Fos-12	77±1	15-28[126,127,177†]	42
Fos-14	123±3	33-44[126,127,†]	65

Measured by size exclusion chromatography. All data was acquired in Chapter 3, except for NG-10, which was a new measurement.
* Calculated by multiplying aggregation numbers by molecular weight of detergent monomer. Aggregation numbers for NG-10 are not available at the time of writing.
† Aggregation numbers obtained from Anatrace.

While both Fos detergents also produced PDCs that are compatible with a dimeric form of TMD81, the dimensions are significantly smaller than their Mal and Lyso-Fos equivalents. Moreover, both detergents produced much narrower elution profiles when analyzed using size-exclusion chromatography, suggesting the presence of a homogenous sample composed primarily of a single oligomeric state and contributing to the significant enhancement in Fos-12 spectral quality compared to the other candidates.

The NG-10 detergent tested in this chapter also produced small monodisperse micelles which supported high levels of thermostability, potentially reflecting the extremely low CMC (0.036 mM) of this detergent system. The high thermodynamic stability of the NG-10 micelle suggests a slow rate of exchange between the monomeric and micellar states, a feature that has been shown to enhance thermostability in other membrane proteins systems such as the β_2AR receptor.[124] However the lack of long term stability for the TMD81 PDC with NG-10,

as evidence by the formation of precipitate during the course of a single 6-hour NMR experiment, prevents the acquisition of high-quality spectra. In contrast to this, while the T_m of TMD81 in Fos-12 was lower than that observed for its NG, Mal and, Lyso-Fos counterparts, sufficient thermostability remained to allow for the collection of multiple NMR spectra at elevated temperatures (45ºC) for upwards of 48 hours with little loss of spectral quality.

A previous study examining the β_2AR receptor showed a correlation between the CMC of the detergent micelle system and conformational dynamics, with detergents such as NG-10 giving rise to exchange on the slow timescale, allowing individual states to be separately resolved using solution NMR. In that study, exchange between the functional states of β_2AR was accompanied by rearrangements in the local micelle environment; a process which can be limited by the slow rate of monomer/micellar exchange in low CMC detergents such as NG-10.[124] In contrast, spectra obtained for β_2AR in Mal-12, a detergent with a significantly higher CMC (150 μM vs 36 μM) showed only a single set of broad peaks, indicative of intermediate-to-fast exchange kinetics.[124] The high (1.5 mM) CMC of Fos-12 relative to even Mal-12 could potentially have the opposite effect on TMD81, increasing the rate of conformational exchange towards a timescale more favourable for solution-NMR studies. A survey of the protein structure databank showed that Fos-12 is the most frequently used detergent for structure determination of integral membrane proteins using solution-NMR, suggesting that TMD81 is not the only system to benefit from this characteristic.[179] While the use of Fos-12 significantly enhances spectral quality relative to the other detergents tested, it is not yet at a level that would enable chemical shift assignment. Other types of media that might yield further improvements include small isotropic bicelles due to their

ability to provide a more accurate mimic of the native bilayer environment. While attempts to produce samples at the concentrations required for NMR studies were unsuccessful, further optimization of could identify conditions from which high-quality spectra can be obtained. However, one potential limitation of bicelle systems for the study of integral membrane protein is the requirement for high concentrations of short-chain lipids or detergents; for example, the protocol employed in this thesis requires >10 mM of CHAPS, which can distort or denature the native protein structure. These high concentrations are partly due to the high CMCs of the detergents employed, giving rise to a high rate of exchange between micellar and monomeric detergent species that could potentially perturb structure of encapsulated bilayer and reduce thermodynamic stability of overall system.

4.3.2: Alternate Membrane Mimetic Systems for Solution-NMR Studies of *ec*GlpG

As an alternative to micelles or bicelles, there are a range of emerging techniques to solubilise membrane proteins that might also prove useful for solution NMR of TMD81, including lipid-protein nanodisks and amphipols.[180,181] Nanodiscs are fundamentally similar to phospholipid bicelles, except the phospholipid bilayer surrounding the embedded membrane protein is capped by an alpha-helical membrane protein scaffold, thereby avoiding the need for detergents. Under current protocols, thermodynamically stable nanodiscs with diameters as low as 7.3 nm and molecular weights of 70 kDa can be produced, enabling them to rival the size of small isotropic bicelles (~5-8 nm, 100 kDa), and even some of the detergent micelle systems presented in this chapter.[142,143,145,182,183] While nanodiscs have been successfully used in NMR studies of a range of integral membrane proteins such as VDAC-1 and VDAC-2,[184,185] OmpA,[186] YgaP,[187] and even cytochrome P450,[186] the most convincing

application was reported by the Wagner group in 2013, where they used small 70 kDa nanodiscs to study both the 7-TM helix protein bacteriorhodopsin (bR), and the bacterial outer membrane protein OmpX.[188] While the use of nanodiscs gave rise to longer correlation times compared to Mal-12 micelles (41 ns vs 60 ns), the spectral quality observed in the former was sufficient to produce a solution-NMR structure of OmpX, illustrating the potential for this membrane mimetic system for the structural characterization of *ec*GlpG and other integral membrane proteins.[188]

Amphipols, or amphipathic polymers, provide another avenue by which membrane proteins can be solubilized for solution NMR studies. These are molecules composed of an amphipathic carbon backbone to which long hydrophobic side chains are attached (Fig. 4.9A), which interact with the hydrophobic TM segments of integral membrane proteins to form a protective "ring" insulating it from the aqueous solvent (Fig. 4.9B). The extensive surface area involved in this interaction, coupled with the hydrophobic nature of the amphipol chains gives rise to an extremely high affinity, ensuring a low concentration of free amphipol in solution and slow exchange kinetics between the bound and unbound states.[189] While several integral membrane proteins have been successfully characterized in amphipol systems,[189–191] the size of the resulting complexes is usually significantly larger than those generated using detergent micelles.[191,192] Additionally, most commercial amphipols are only soluble under basic conditions unfavourable for the detection of amide protons using solution-NMR and upscaling the amphipol reconstitution process to produce the concentrations required for solution-NMR has been reported to increase the size and heterogeneity of the resulting complexes, although the cause of this change currently remains unknown.[190] Nonetheless, high-quality NMR spectra for bacteriorhodopsin and OmpX have been reported using

recently-developed pH-insensitive sulfonated and non-ionic amphipols,[192–194] demonstrating their viability as a platform for solution-NMR studies on *ec*GlpG

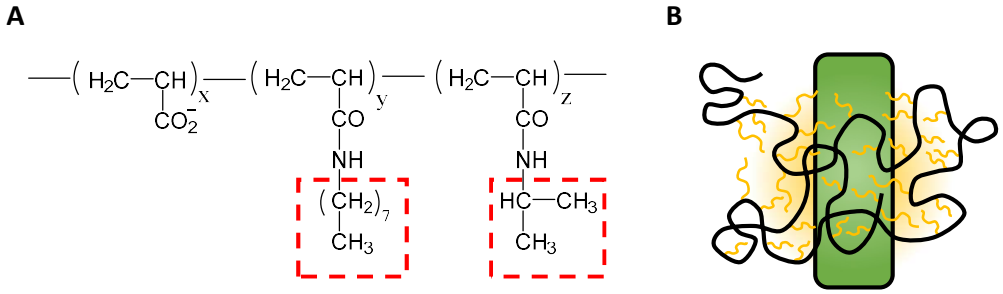


Figure 4.9: Amphipols as membrane mimetic systems for membrane proteins. A) Structure of A8-35, one of the most widely used and best characterized amphipols to date.[194] Hydrophobic side chains are outlined in red, which are randomly distributed along the length of the chain with molar percentages corresponding to x, y, and z. This particular example consists of ~70 repeats, with x, y, and z values of approximately 24 (35%), 18(25%), and 28 (40%) respectively B) Schematic illustrating the solubilisation of a membrane protein with amphipols. The hydrophobic side chains that interact with the TM segments are shown in yellow.

Despite their shortcomings, the results obtained using detergent micelles represent the first time that peaks from a rhomboid protease TMD have been observed using solution-state NMR. While binding of DCI to the rhomboid active site failed to produce any noticeable enhancement to NMR spectral quality, it is possible that other inhibitors that interact with a larger area of the rhomboid such as the 7-amino-4-chloro-3-methoxy isocoumarin or peptidyl-chloromethylketones employed in previous X-ray studies of *ec*GlpG could have a greater effect on *ec*GlpG dynamics and NMR spectral quality, while their ability to promote the formation of a slightly-open state in these studies could be used to explore structural nature of substrate gating using solution-NMR. While the quality of the spectra obtained using Fos-12 did not allow for the assignment of the backbone amide peaks through triple resonance experiments, their ability to resolve ~2/3 of the expected backbone amide demonstrates the

potential of solution-NMR as a tool to gain insight into the structure and dynamics of TMD81. Additionally, the good resolution observed for correlations originating from ^{15}N-containing side chains such as Trp also opens the door for their use as structural probes, which are employed in the subsequent chapter to explore the impact of α5-opening on TMD81 structure.

4.4: Materials and Methods

4.4.1: Protein Expression, Purification, and *In Vitro* Analysis.

Wild-type and mutant TMD81 constructs were overexpressed in C43(DE3) chemically-competent *E. coli* cells, purified into GlpG buffer and characterized using analytical size-exclusion chromatography, CD spectroscopy and enzymatic analysis against a water-soluble BODIPY-casein substrate as described in the preceding chapter. Size-exclusion chromatography was used to determine whether TMD81 was completely exchanged into the desired detergent system. Samples were prepared in PC-7 with a final detergent concentration of 0.3% w/v in order to accommodate its higher CMC. ^{15}N-labeled samples for solution-state NMR were grown in M9 media enriched with $^{15}NH_4Cl$ as the sole nitrogen source, and the resulting protein concentrated to 60-300 μM using a 30,000 kDa MWCO centrifugal filter unit (Amicon). Bicelle-solubilized samples were prepared as described in Chapter 3 with q-values of 0.33 and 0.16.

4.4.2: NMR Spectroscopy

^{15}N-labeled TMD81 was analysed using either a Varian 500 MHz Inova spectrometer equipped with a room temperature triple resonance probe, or a 600 MHz Bruker Avance III spectrometer with a triple resonance cryoprobe, both at the University of Ottawa NMR

Facility. In the case of bicelle-solubilized TMD81, spectra were acquired by Dr. Simon Sauvé using the 600 MHz Avance spectrometer with cryoprobe located at the Centre for Biologics Evaluation, Health Products and Food Branch at Health Canada. Samples collected using the Avance 600 were purified into a modified GlpG buffer containing 50 mM Tris-HCl pH 7.0, 50 mM NaCl, 100 μM EDTA, and 0.1% w/v of the desired detergent. This change in buffer composition was done to improve sensitivity in the cryoprobe, and was confirmed to preserve thermostability and catalytic activity. Standard ^{1}H-^{15}N HSQC and TROSY experiments were acquired at 25-45ºC with a spectral width of 30 ppm spread across 128 increments in the indirect dimension, and the resulting data was processed using NMRPipe and NMRViewJ,[195,196] with signal intensities normalized for the number of scans (typical range 128-256), receiver gain (typical range 57-203) and total protein concentration (typically 80-250 μM). To carry out the small-molecule binding assays, a 10 mM stock of DCI was prepared in DMSO, and added to the NMR samples to produce the desired concentration, to a maximum DMSO concentration of 8%. Control spectra were obtained for WT TMD81 in the presence of DMSO alone, and showed no significant perturbations across the range of concentrations employed in the DCI binding assay.

Chapter 5: Solution-NMR Investigation of Substrate Gating in the *E. coli* GlpG Rhomboid

Contributions of collaborators:

Carolina Uribe (University of Ottawa) assisted in the creation of the F153A, W157F, W236G, and W241A TMD81 mutants, along with its subsequent purification and enzymatic analysis in Fos-12 against BODIPY-casein. Eric R. Burge (University of Ottawa) purified TMD81 and W236G into Mal-12 and Fos-12, and assessed their catalytic activity against the TatA-based TM model substrate. Rebecca Brown and Shawna Rieux (University of Ottawa) assisted in the creation of the W122F, W125F, and W136Y TMD81 mutant constructs. Kvido Strisovsky (Academy of Sciences of the Czech Republic) provided the plasmid for the TatA model substrate

5.1: Introduction:

The ability of rhomboids to cleave TM segments within the phospholipid membrane contributes to the wide range of biological functions attributed to this family. However, the mechanism through which these TM substrates are able to transition from this hydrophobic environment into the buried, hydrophilic rhomboid active site remains unknown. As discussed in Chapter 1, results from enzyme kinetics and EPR experiments suggest that α5 plays a role in the substrate gating process. This is supported by X-ray structures reported in 2006 and 2016 that depict an "open" conformation in which the C-terminal half of α5 is rotated 35^{o} away from α2, or the entire α5 helix is laterally displaced from α2 respectively. However, it is not clear whether these structures accurately represent physiologically relevant states related to substrate gating. While it would be ideal to study the structure of the enzyme-substrate complex to gain high-resolution insights into substrate gating, the hydrophobic nature of the substrate TM segment has presented challenges that have limited progress towards this goal. As a result, the structural nature of the α5-open conformation and its relevance to substrate gating and the rhomboid catalytic cycle remain a subject of debate.

Enzyme kinetics-based evidence for the role of α5 in substrate gating has been obtained primarily through the use of what have been termed "gate-open" mutations located at the interface of α5 and α2, (summarized in Table 1.2, Fig. 5.1). While these give rise to up to 20-fold enhancement in proteolytic activity, no significant change in labelling efficiency by an activity based probe was detected, suggesting that the increase in proteolytic activity is not due to an increase in the reactivity of active site residues.[63] While this supports the hypothesis that substrate gating represents the rate-determining step in the catalytic cycle, there has yet to be a demonstration that these mutants are able to facilitate the displacement

of α5, and that this increased propensity to adopt a gate-open conformation causes an increase in the rate of proteolysis by *ec*GlpG.

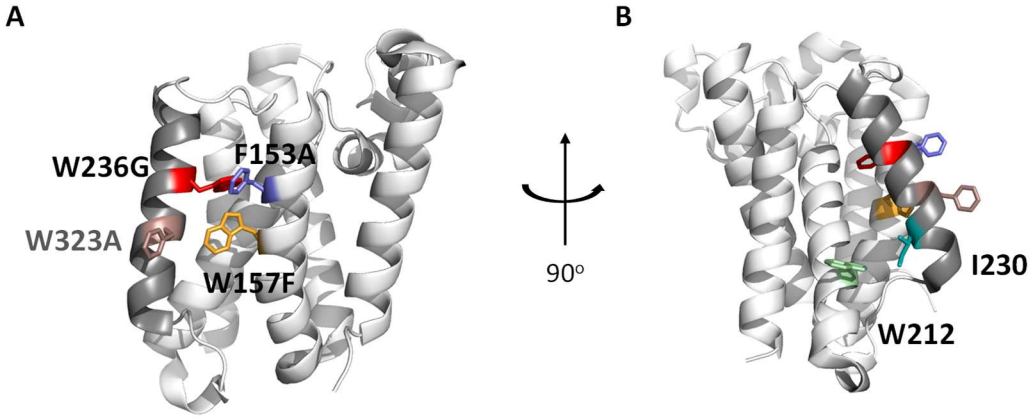


Figure 5.1: Structure of *ec*GlpG (PDB 3B45) showing A) the position of all gate-open mutants examined in this chapter. Gate-open mutants characterized in previous studies but not assessed in this chapter (F232A) are coloured mauve. B) Rotated view to show location of W212 (green) and I230 (teal). Colour coding of gate-open mutants consistent with A. In both structures α5 is shown in grey.

Due to the distribution of tryptophan residues throughout the *ec*GlpG structure and the reasonably well-resolved NMR spectrum that could be obtained for indole protons using the conditions identified in Chapter 4, mutagenesis was performed to assign tryptophan indole chemical shifts. This provided a series of structural probes which were subsequently employed to study the effect of predicted gate-open mutations on the structure of TMD81. These mutations were found to produce a conformational change consistent with an open structure involving the lateral displacement of the intact α5 helix. Binding of a peptide fragment containing sequence elements from a native rhomboid substrate promoted formation of a α5-shifted WT TMD81 species when monitored using solution-NMR. Functional assays carried out using water-soluble substrates indicate that, while α5-gating facilitates proteolysis, opening of the lateral gate can compromise catalytic activity potentially by increasing its

exposure to the hydrophobic membrane environment. Use of a TM substrate counteracted these negative effects, suggesting a mechanism to promote the preferential cleavage of physiological substrates *in vivo*

5.2: Results

5.2.1: Tryptophan Side Chain Peak Assignments

Tryptophan indole N-H groups have distinctive ^{1}H chemical shifts in the 9-11 ppm range, enabling them to be easily identified and resolved from the main cluster of backbone and side-chain amide peaks. In the NMR spectra obtained in Chapter 4, these peaks are individually well resolved, and can be used to monitor changes to TMD81 conformation. To assist this process series of tryptophan knockout mutations was created at all 11 tryptophan positions (Figure 5.2A). ^{1}H-^{15}N TROSY spectra were then collected for the resulting constructs and compared to that of the wild-type TMD81 in order to assign the tryptophan side chain peaks to specific positions within the *ec*GlpG sequence (representative example shown in Figure 5.2B). Based on this comparison, it was possible to assign 10 out of 11 Trp residues (Fig. 5.2C). This was complicated by changes in chemical shifts induced by some of the mutations and the fact that the number of peaks in the tryptophan indole region (13) exceeds the expected number based on the wild-type TMD81 sequence. This suggests that there is some conformational heterogeneity, where a subset of tryptophan residues are undergoing slow exchange on the NMR timescale.

Further validation of these peak assignments was obtained by NMR spectra of WT TMD81 acquired in the presence of the spin labeled amphiphile 5-DOXYL-steric acid, which was compared to the spectrum acquired with no spin label (Fig 5.2D). Selective broadening

was observed for peaks corresponding to W157, W159, W212, W236, and W241, all of which are expected to be exposed to the hydrophobic phase of the detergent micelle that is in proximity to the headgroup region and thus the 5-DOXYL-steric acid PRE probe. While the effect of the PRE probe on W125 is difficult to interpret due to overlap with an unassigned peak, no significant peak broadening was observed for the two other L1 Trp residues at positions 122 and 136 despite their location near the membrane surface predicted by both MD and X-ray crystallography studies.[66,98] This suggests that this region could potentially form a dimer interface which insulates the L1 residues from the local membrane environment. This hypothesis is consistent with a dimeric state for TMD81 in Fos-12 micelles observed in both Chapter 3 and 4.

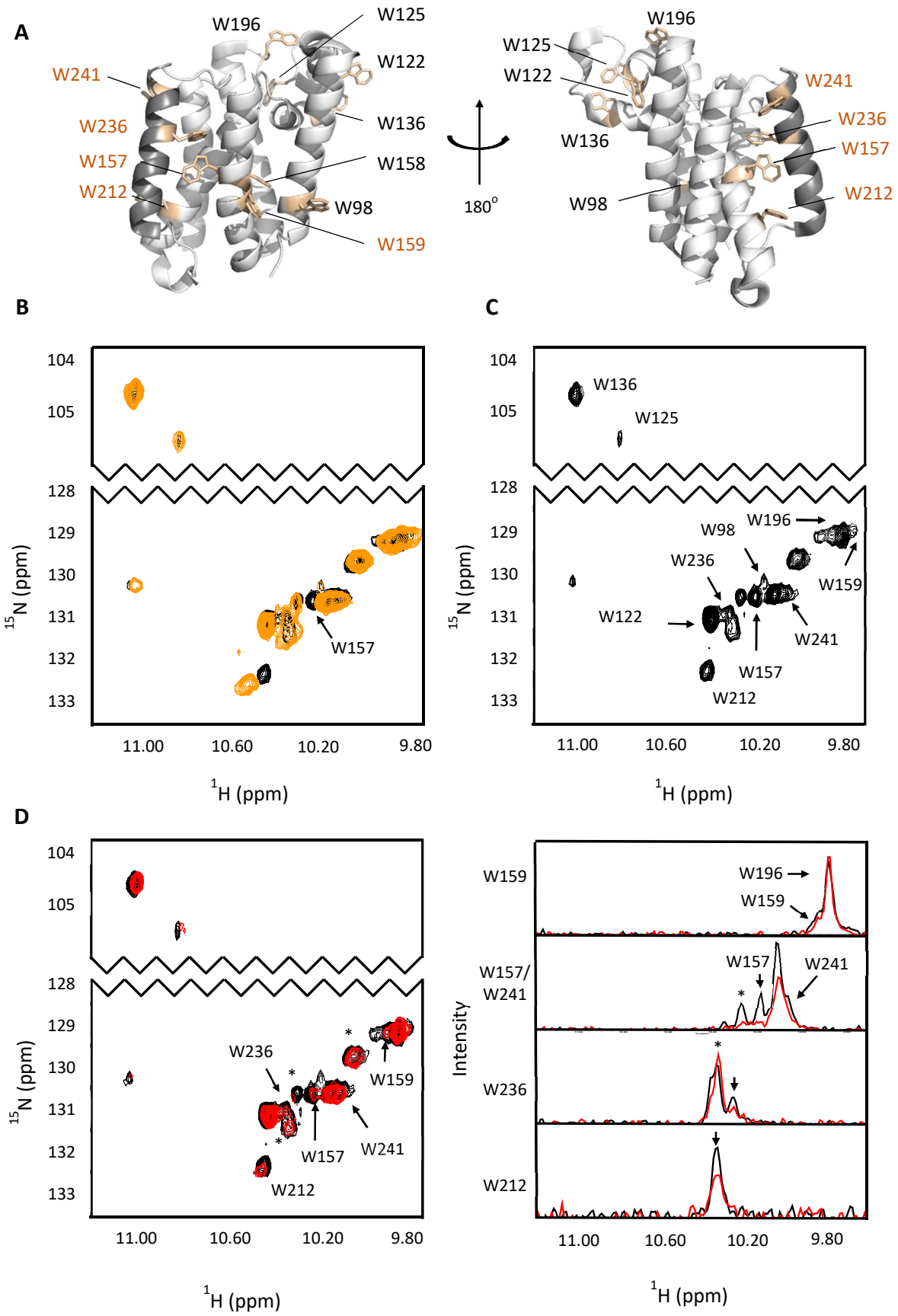

A
W196
W125
W241
W122
W236
W136
W157
W158
W212
W98
W159
180°
W196
W125
W122
W241
W236
W157
W136
W98
W212
B
C
104
105
128
129
130
131
132
133
^{15}N (ppm)
W157
W136
W125
W98
W196
W236
W159
W122
W241
W157
W212
11.00
10.60
10.20
9.80
^{1}H (ppm)
D
W236
W159
W157
W241
W212
W159
W196
W159
W157/
W241
W157
W241
W236
W212
Intensity

Figure 5.2 (Previous Page): Trp peak assignments for TMD81 in Fos-12 detergent micelles. Figure A) Location of all Trp residues on the WT *ec*GlpG. Peaks which are selectively broadened by the lipophilic 5-DOXYL steric acid spin label are labeled in red. B) Example spectra illustrating the assignment process. The tryptophan side chain region from the TMD81 W157F mutant ^{1}H-^{15}N TROSY spectrum (orange) is overlaid upon that of the wild type (black). Spectra illustrating the assignment of the remaining 10 tryptophan residues are shown as supplemental figures at the end of this chapter. C) Peak assignments for TMD81 tryptophan. D) ^{1}H-^{15}N TROSY spectra of TMD81 in the presence of 0 (black) or 1:1 mole-PRE probe: protein ratio (red). Peaks that show significant broadening are labeled, and shown as a 1D slice in the inset to the right. Peaks for which no assignments are available are indicated with *.

Enzyme kinetics measured using the soluble BODOPY-casein showed that all mutants retained significant levels of proteolytic activity (Table 5.1) and gave rise to CD spectra and thermostabilities that were similar to that of the WT TMD, indicating that these mutations did not introduce significant structural changes.

Table 5.1: Kinetic parameters (BODIPY-casein), melting temperatures, and helix content of TMD81 WT and Trp mutants in 0.1% Fos-12.

Mutant	k_{cat} (min^{-1})	K_m (uM)	T_m (^oC)	% α-helix
TMD81[#]	0.16±0.02	2.0±0.4	73.5±0.4	92±2
W98F	0.18±0.05	1.7±0.8	79±1*	94±1
W122F	0.22±0.04	1.4±0.4	-	-
W125F†	0.19±0.04*	2.8±0.2*	80.0±0.9*	89±4
W136Y	0.14±0.01	2.4±0.6	80.3±0.7*	91±2
W157F	0.09±0.02*	1.3±0.8	77±2	90±2
W158F†	0.17±0.04	4.9±0.2*	75±1	89±1
W159F	0.33±0.02*	1.1±0.3*	78.3±0.2	79±1*
W196F†	0.20±0.03	1.1±0.3*	82±1*	81±3*
W212F	0.145±0.006	1.7±0.8	77±3	88±5
W236G	0.17±0.02	1.9±0.6	74.2±0.4	92±3
W241A	0.23±0.01*	2.9±0.5*	83±3*	81±5*

Data acquired in Chapter 3.
† Replicates were performed from samples produced in a single purification.
* Values that are significantly different from TMD81 ($p < 0.05$)

5.2.2: Solution-state NMR Studies of Gate-open Rhomboid Constructs

In addition to assignment information, NMR spectra of the tryptophan substitution mutants could potentially provide some information into the gate-open state of TMD81. In particular, W236 on α5 interacts with F153 from α2, and its substitution to glycine or alanine was previously found to significantly enhance proteolytic activity against both soluble and TM model substrates in Mal-12 detergent micelles and phospholipid liposomes, potentially by promoting opening of the lateral gate (Table 1.2).[63,70,53,104] Similarly, the α2 residue W157 interacts closely with F232 in α5 and, while mutations to this residue have not been extensively studied in isolation, the W157A and F232A double mutant shows an increase in activity in Mal-12 against both Spitz and Gurken TM substrates suggesting a role for both these residues in regulating α5 gating.[70] Both the W236G and W157F "gate-open" mutations used in the tryptophan peak assignment process gave rise to a chemical shift perturbation for the W212 indole peak relative to WT, (Fig.5.3A). In contrast, other mutations to the gate region that were not expected to perturb α2/α5 interaction (i.e. W158F, W159F and W241A) failed to produce a similar effect (Appendix) suggesting that the W212 shift reports on changes to the TMD81 structure associated with gate-opening.

In an attempt to provide further structural information on the open conformation, the effect of an additional gate-open mutant (F153A) on the Trp indole spectrum was examined. This mutation was introduced into both wild-type and W236G TMD81, and purified into Fos-12. CD spectra for both constructs showed characteristic minima at 208 and 222 nm, with melting temperatures above 70°C, suggesting that overall rhomboid structural integrity and thermostability was not significantly altered (Table 5.2, Fig. 5.4). The overall appearance of the ^{1}H-^{15}N TROSY spectrum of F153A was similar to that of WT TMD81, mirroring results

obtained for W236G and W157F. Curiously, no major change in chemical shifts were observed for the peak corresponding to W236 despite its location on the proposed gate helix. However, the tryptophan side-chain peak corresponding to W212 appeared to be split to yield multiple separate peaks potentially corresponding to different conformations undergoing slow exchange, suggesting a change to the conformational dynamics of the gate region (Fig. 5.3B).

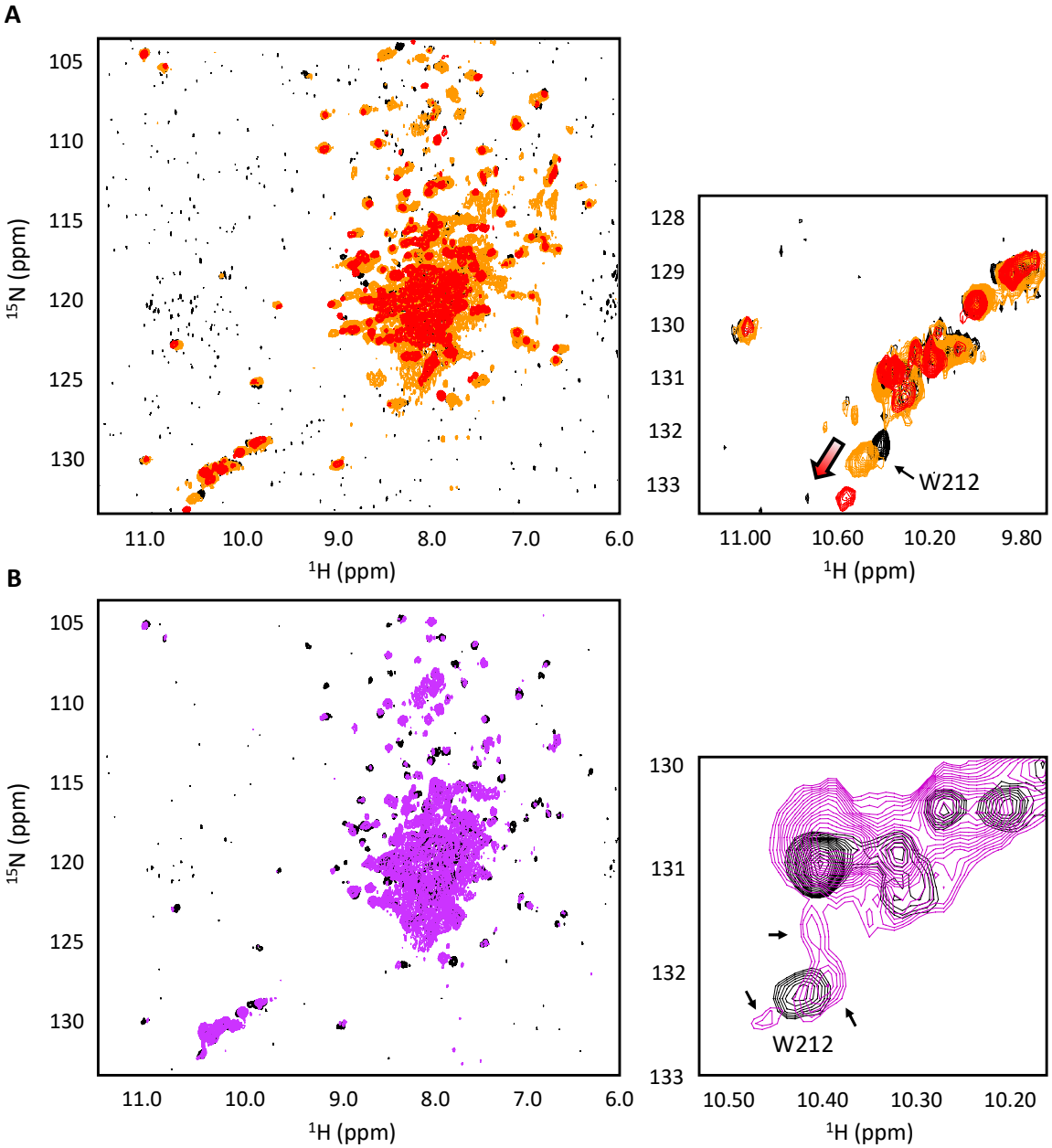


Figure 5.3 (Previous Page): Structural characterization of gate-open TMD81 mutants by solution-state NMR: A) ^{1}H-^{15}N TROSY spectra for wild-type TMD81 (black) and gate-open mutants W236G (red) and W157F (orange).B) ^{1}H-^{15}N TROSY correlation spectra for wild-type TMD81 (black), and the gate-open mutant F153A (purple). A close-up view of the W212 side-chain peak is shown to the top right. The three species observed in the F153A mutant are indicated with the black arrows

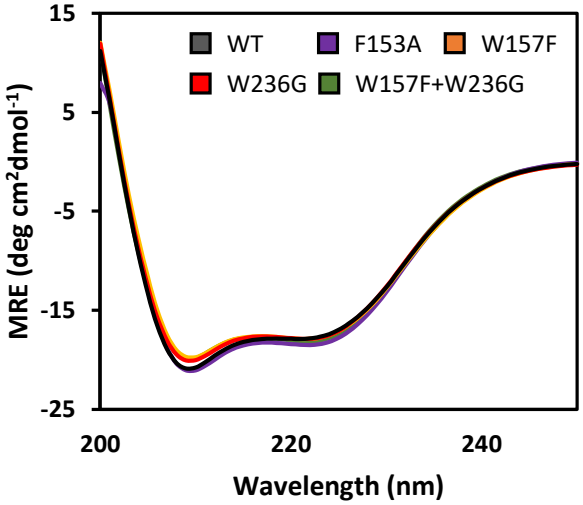


Figure 5.4: Representative CD spectra collected for gate-open mutants solubilized in 0.1% Fos-12 at 35ºC showing a small, but consistent shift in secondary structure. Note that the F153A+W236G CD spectrum overlaps those obtained from the F153A and WT rhomboid.

While W212 is located on α4, it lies in close proximity to α5 residues such as I230 (Fig. 5.5), suggesting that the W212 shift in mutants W236G and W157F might be reporting on a change in the position of the gate helix. A spectrum was also acquired on the I230 mutant to provide additional confirmation that this peak corresponds to W212. This α5 residue packs against W212 in the crystal structure of the closed state, and so its mutation to alanine would be expected to change the local chemical environment of W212 giving rise to a corresponding change in chemical shift. Such a shift is observed in Figure 5.5, providing additional confirmation that the peak that shows changes in response to gate-opening mutations corresponds to W212.

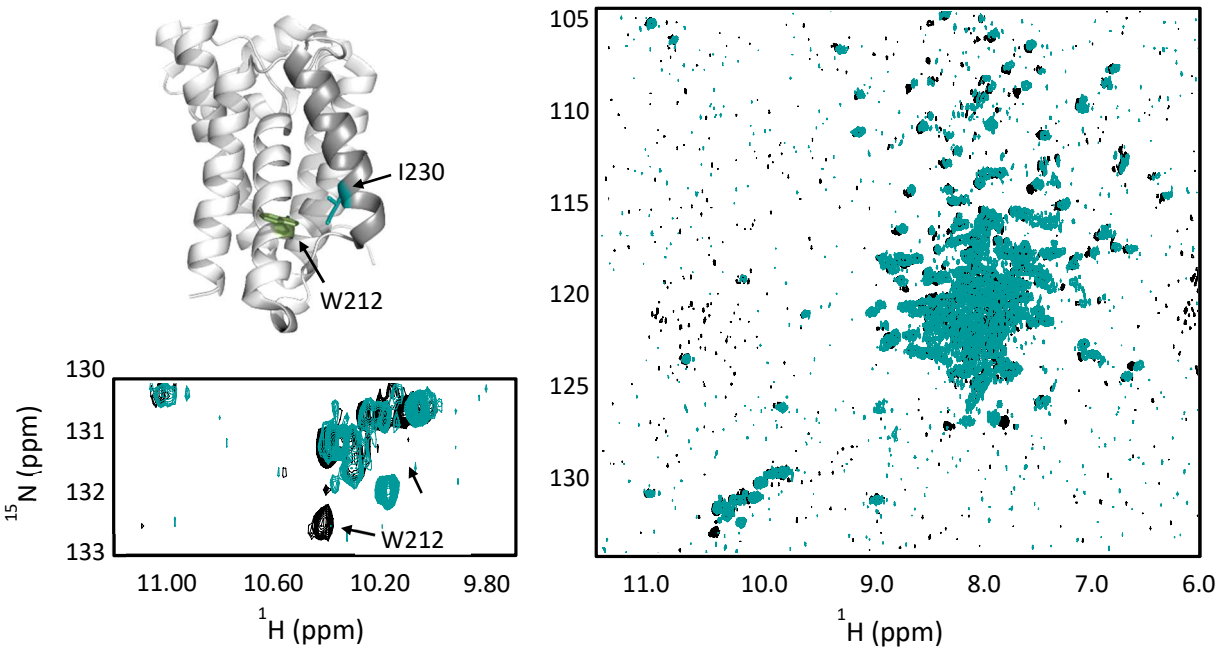


Figure 5.5: ^{1}H-^{15}N TROSY correlation spectra for WT (black) and I230A (teal) TMD81. X-ray structure (PDB 3B45) showing the location of W212 relative to α5-residue I230 is shown in the top left.

The chemical shift change in W212 observed for gate-open mutants of TMD81 suggests that opening of the lateral gate alters the local chemical environment around this residue, potentially via its interactions with I230 discussed above. While the initial structural model for lateral gating does not show any significant change in the environment around either residue (Fig. 5.6A), the alternate open state in which the entire α5 is laterally shifted away from α2 gives rise to altered packing interactions between W212 and the α5 residue I230 (Fig. 5.6B). To determine whether the change in W212 indole NH chemical shift observed in the gate-open mutants corresponds to the lateral shift in α5 towards α4 and away from α2 proposed in this model, NMR spectra of W236G were collected in the presence of increasing concentrations of 5-DOXYL-steric acid PRE probe and compared to the equivalent data obtained for the WT TMD81. As shown previously, the W212 indole NH peak in the WT

TMD81 NMR spectrum is selectively broadened by the lipophilic PRE probe (Fig. 5.2B), resulting in a decrease in peak intensity with increasing probe concentrations (Fig. 5.6C). However, this loss of intensity is reduced in the gate-open W236G mutant, consistent with the proposed change in helix packing interactions in which W212 is brought closer to I230 by ~0.8 Å upon adoption of the open conformation, reducing its exposure to the micelle environment. This effect was specific for W212, since the intensity of other Trp residues that are far from the lateral gate did not show a significant difference in broadening, as shown for W241 in Fig. 5.6D.

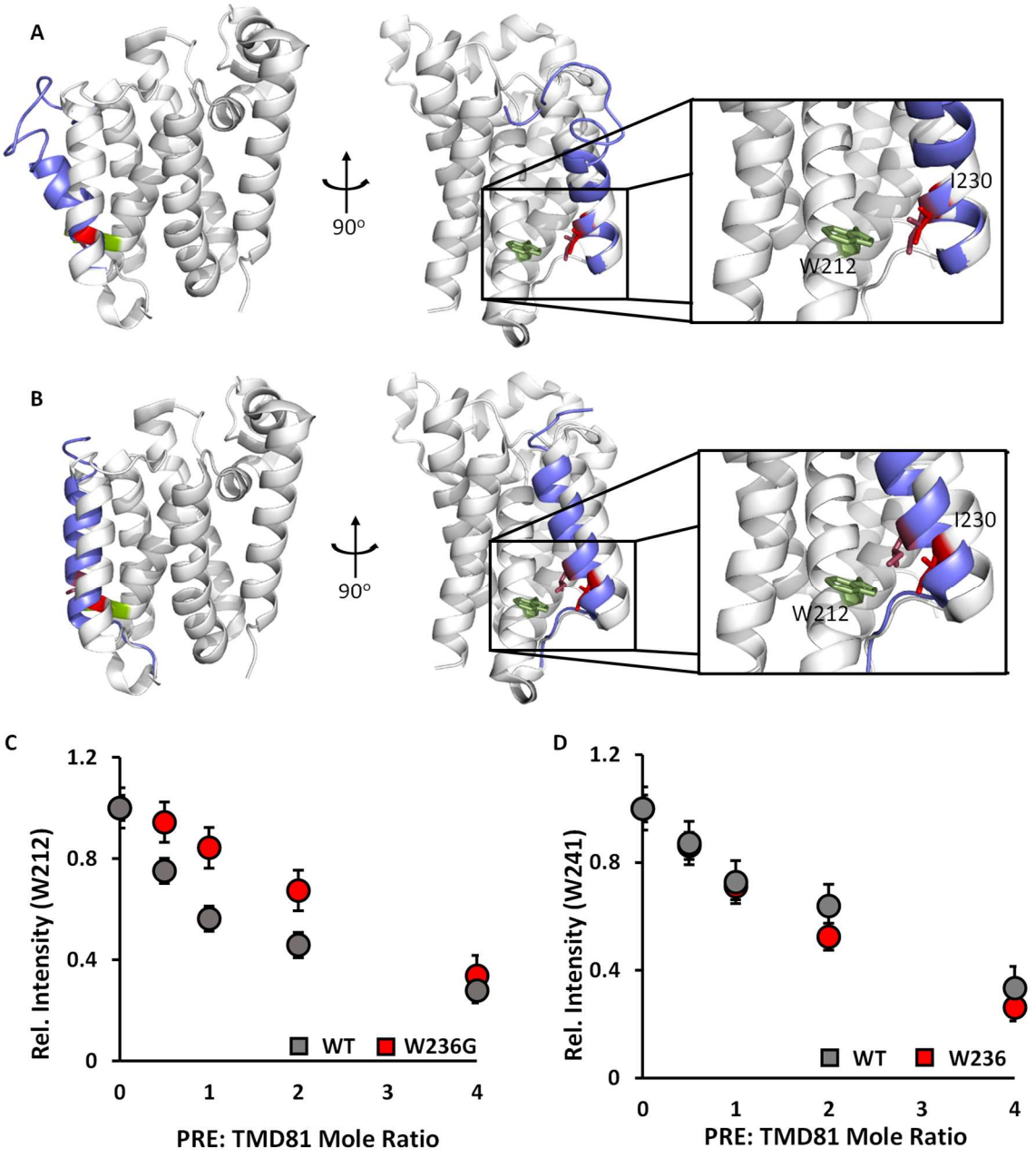


Figure 5.6: Effect of gate-opening on the local environment of W212 A) Comparison of the closed structure (white PDB 3B45) and an open structure with the C-terminal side of a5 (blue) pulled away from the main body of the catalytic domain (PDB 2NRF). B) X-ray structure of *ec*GlpG in the closed (white) and alternate open (PDB 5F5K) structure showing a lateral displacement of a5 (blue). Movement of I230 on α5 results in a change in the local chemical environment around W212 (right) C) Decrease in W212 side chain peak intensity in the W236G (red) upon addition of 5-DOXYL steric acid PRE probe. D) As in (C) measured for W241 side chain peak intensity. Data from both C and D obtained using only one trial. Error bars for both graphs represent the standard variation in peak intensity measured for a non-membrane exposed Trp (W196) side-chain peak across all PRE concentrations tested.

5.2.3: Elucidating the Role of α5 Dynamics in Substrate Gating

To explore whether substrate binding influences *ec*GlpG gating, solution-state NMR was used to monitor TMD81 interactions with a peptide derived from the Gurken TM substrate. This fragment shares the same sequence as a peptide-aldehyde inhibitor that is one of the most potent and specific inhibitor of *ec*GlpG to date.[10] ^{1}H-^{15}N TROSY spectra obtained in the presence of this peptide show splitting of the W212 peak to yield a new species consistent with an α5-shifted open conformation similar to that observed in the gate-open mutants (Fig. 5.7A). When the relative population of the W212-shifted open conformation was plotted as a function of peptide concentration, a hyperbolic curve was obtained, with a K_D value similar to that previously reported for similar peptide fragments (Figure 5.7B).[10] As this peptide is expected to provide a reasonable approximation the enzyme-substrate complex involving the N-terminal half of the substrate,[10] these results suggest that substrate binding may stimulate gate opening. These spectra also show an additional peak that appears over the course of the titration, the volume of which could be fit to a hyperbolic curve to yield a K_D value similar to that obtained using the W212 peak (~660 μM), showing that some other part of TMD81 also senses the peptide interaction, although it is not possible to assign this new peak to a specific region of the protein.

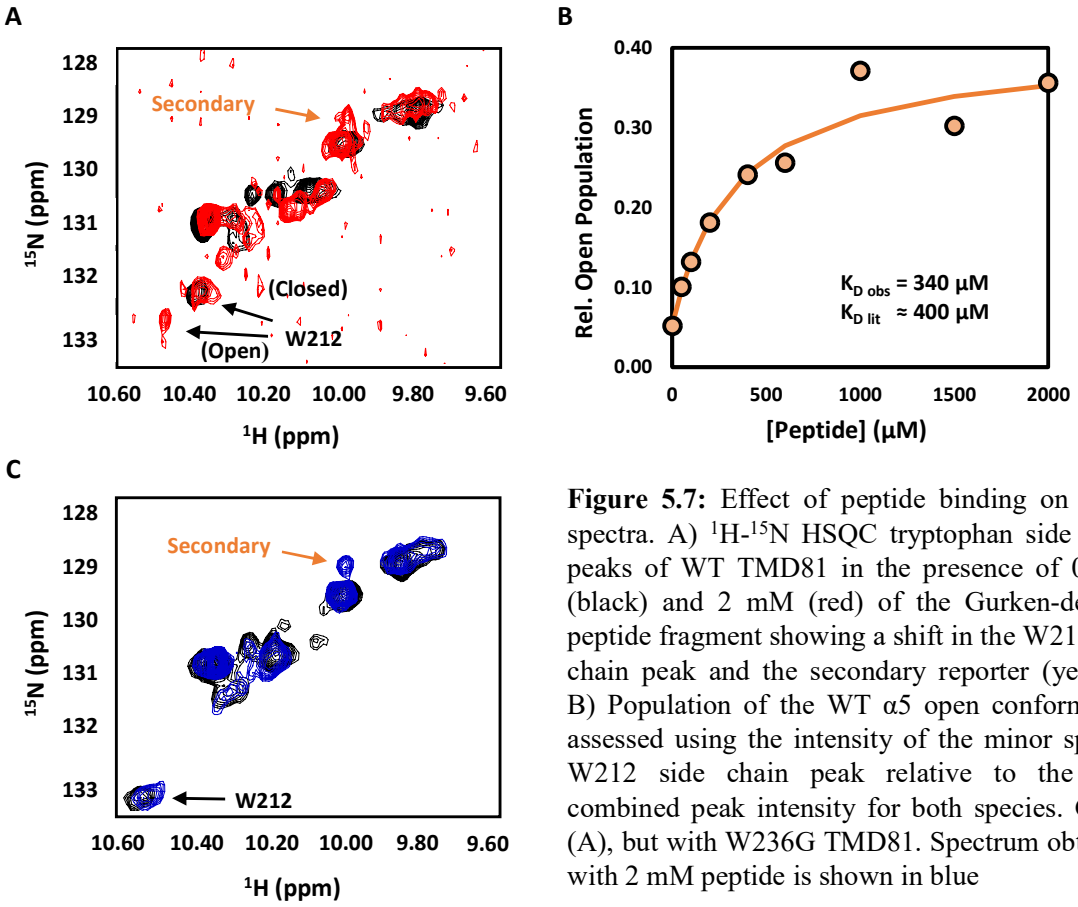


Figure 5.7: Effect of peptide binding on GlpG spectra. A) ^{1}H-^{15}N HSQC tryptophan side chain peaks of WT TMD81 in the presence of 0 mM (black) and 2 mM (red) of the Gurken-derived peptide fragment showing a shift in the W212 side chain peak and the secondary reporter (yellow). B) Population of the WT α5 open conformation assessed using the intensity of the minor species W212 side chain peak relative to the total combined peak intensity for both species. C), as (A), but with W236G TMD81. Spectrum obtained with 2 mM peptide is shown in blue

The effect of gate-open mutations on α5 conformation was verified using a similar titration experiment carried out using the W236G mutant. It was possible to use the unassigned "secondary" reporter peak of peptide binding to obtain K_D values that were comparable to WT values (~760 µM). However, this binding event failed to diminish the intensity of the W212 side chain peak or give rise to any new peaks in this region (Fig. 5.7C), indicating that W236G successfully promotes formation of the α5-open state even in the absence of peptide or substrate binding.

5.2.4: Exploring the Role of α5-opening in Rhomboid Proteolysis

To determine the functional implications of mutations around the α5-gate, catalytic activity of the various gate-open mutants was assessed under conditions similar to those employed in the solution-NMR studies (ie: 0.1% Fos-12) using the BODIPY-tagged casein substrate employed in Chapter 3.[81] The F153A and F153A+W236G mutants displayed a 3-fold enhancement in activity over WT TMD81 against this substrate in Fos-12 detergent micelles (Fig. 5.8A, Table 5.2), indicating that the movement of α5 plays a role in the proteolysis of water-soluble substrates despite previous speculation that such substrates are able to bypass the lateral gate and enter the *ec*GlpG active site directly from the aqueous environment.[80,92] Curiously, W236G failed to enhance proteolytic activity against this substrate, while W157F displayed a 2-fold decrease in turnover numbers despite the ability of both mutants to promote formation of the α5-open conformation as suggested by the solution-NMR spectra. Similar results were obtained in Mal-12 detergent micelles (Table 5.3), suggesting that this loss in activity is independent of the membrane mimetic system employed. Additionally, the I230A mutant, which could potentially facilitate α5-opening by promoting closer packing with α4, yielded a similar 2x decrease in activity. These results suggest that opening of the lateral gate may have a deleterious effect on proteolysis of water-soluble substrates, potentially resulting from increased exposure of the hydrophilic active site to the hydrophobic membrane environment which could introduce structural perturbations that compromise catalytic activity. To assess this hypothesis, CD spectra were obtained for the gate-open mutants which revealed a small but reproducible change in secondary structure suggesting a loss in α-helical content in the gate-open mutants (Fig. 5.4).

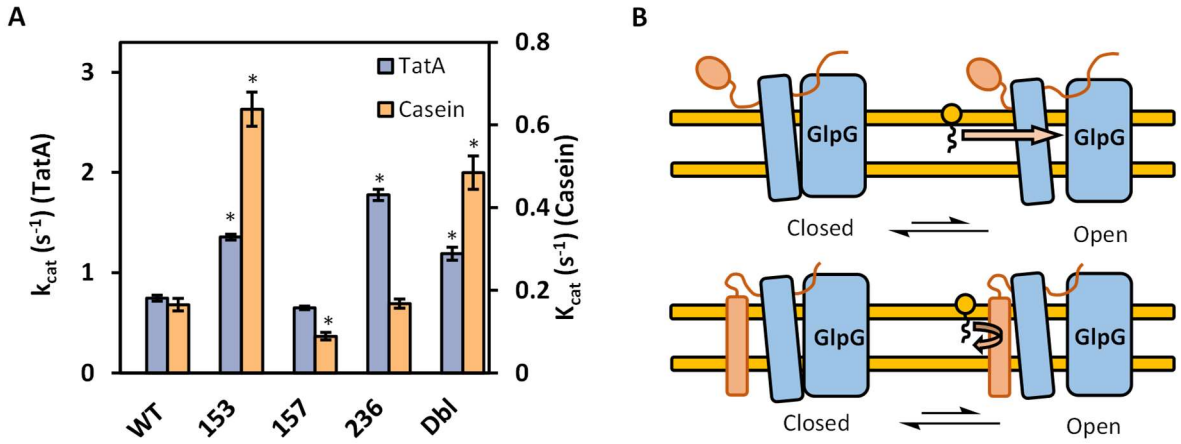


Figure 5.8: Effect of gate-opening on *ec*GlpG catalytic activity A) Catalytic turnover (k_{cat}) values for the wild type TMD81, F153A, W157F, W236G and F153A+W236G double mutant (Dbl) assessed using the PKS187 TM substrate (TatA) or a soluble BODIPY-casein substrate. Kinetic parameters that were significantly different from WT ($p < 0.05$) are indicated with *. B) Model summarizing the impact of gate-opening on catalytic activity against TM and soluble substrates. Corresponding data for all figures can be found in Table 5.1 and 5.2.

Table 5.2: Kinetic parameters obtained for TMD81 mutants against BODIPY-casein in 0.1% Fos-12.

Mutant	Kinetics with casein: k_{cat} (min^{-1})	Kinetics with casein: K_m (µM)	Kinetics with TatA TM substrate: k_{cat} (min^{-1})	K_m (µM)	n_H	T_m (ºC)	% α-helix
WT	0.16±0.02	2.0±0.4†	0.5±0.1	40±10	2.4±0.7	73.5±0.4†	92±2†
F153A	0.64±0.09*	3.0±0.6	1.0±0.1*	50±10	2.2±0.8	73.9±2.8	98±2
W157F	0.09±0.02†*	1.3±0.8†	0.45±0.08	39±8	2.4±0.9	76.9±1.8†	90±2†
W236G	0.17±0.02†	1.2±0.3†*	1.2±0.2*	44±6	2.2±0.7	74.2±0.4†	92±3†
F153A+W236G	0.5±0.1*	4.7±0.7*	1.2±0.2*	48±5	2.5±0.8	71.4±0.8	94±3
W212F	0.145±0.006†	2.1±0.8†	0.44±0.09	45±18	1.7±0.3	77±3†	88±5†
I230A	0.075±0.008*	2.7±0.7	0.16±0.04*	66±5*	1.2±0.5*	78±1	96±3

† Data reported in Table 5.1, and included for comparison. Kinetic parameters that were significantly different from WT ($p < 0.05$) are indicated with *

Table 5.3: Kinetic parameters obtained for TMD81 mutants against BODIPY-casein in 0.1% Mal-12.

Mutant	k_{cat} (min^{-1})	K_m (μM)
WT	0.20±0.01†	1.27±0.04†
F153A	0.40±0.02*	0.9±0.2
W157F	0.01±0.05*	1.6±0.2
W236G	0.14±0.03	1.0±0.3
F153A+W236G	0.37±0.05*	0.8±0.3*

† Data acquired as part of previous study described in Chapter 3, and included for comparison.
* Values which were significantly different from WT ($p < 0.05$).

It is possible that the presence of helical TM regions within the model substrate may help insulate the active site from the hydrophobic membrane environment, potentially overcoming the deleterious effects of gate-opening on proteolysis (Fig. 5.8B). To explore this possibility, catalytic activity of the TMD81 mutants was assessed using a TM substrate generated by extending the native AarA TatA substrate by an $(SG)_4$ repeat to increase the MW difference between the un-cleaved substrate and the C-terminal fragment used to determine proteolysis rates.[69] A plot of the resulting v_o values against substrate concentration yielded a sigmoidal curve (Fig. 5.9), suggesting a cooperative mode of substrate cleavage which could be accurately modeled using the Hill Equation to obtain values of k_{cat}, $K_{0.5}$ and n_H (Table 5.2).

A previous study carried out using the full-length *ec*GlpG construct in Mal-12 detergent micelles reported similar values of k_{cat} against TatA, suggesting that catalytic activity is unaffected by the use of truncated TMD81 mutants.[52] Curiously, the same study showed little cooperativity ($n_H = 1.2$) for *ec*GlpG against TatA, in contrast to the value of 2.4±0.7 reported here for TMD81 against a similar TatA-based model substrate.[52] To determine whether this discrepancy could be attributed to differences in the local micelle environment, the activity of TMD81 was measured in the same Mal-12 detergent system

employed in this study. The resulting kinetic parameters compared favourably to those obtained in Fos-12 (k_{cat} = 0.5 min^{-1}, K_M =40 μM, n_H = 2) suggesting that the absence of cooperative binding observed for the full-length *ec*GlpG may reflect the ability of the soluble cytoplasmic domain to interfere with communication between the two binding sites.

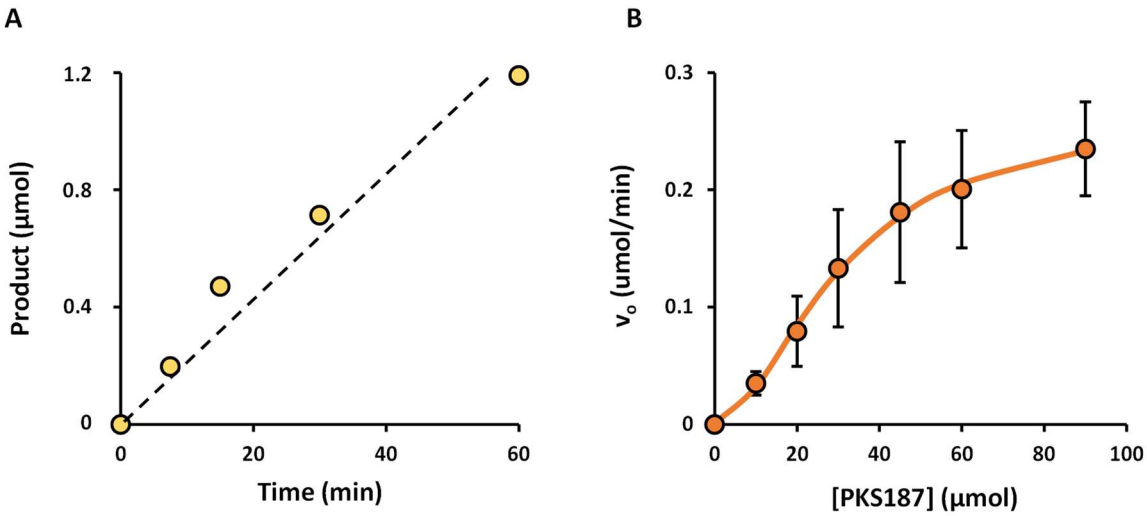


Figure 5.9: Representative data used to obtain rhomboid kinetic parameters against TatA model substrate. A) Representative data showing the formation of the BODIPY-casein cleavage product over time, from which initial rates of proteolysis can be obtained B) Plot of the initial proteolysis rates against substrate concentration. The resulting data was fit to the Hill Equation (solid line) to obtain values of K_M, V_{max}, and n_H. Data shown was obtained for TMD81 in Fos-12

While I230A displayed a significant loss of cooperativity, analytical size-exclusion chromatography profiles for this mutant are comparable to those obtained for the WT suggesting that its dimeric state is maintained. All other gate-open mutants showed a similar degree of cooperativity to the wild-type TMD81. K_M also remained relatively constant across all other mutants tested, (Table 5.2) with changes in catalytic activity being reflected primarily by changes in k_{cat} alone. The F153A, W236G and W236G + F153A gate-open mutants

displayed a twofold increase in activity over the wild type, mirroring results reported previously using the full-length *ec*GlpG (Table 1.2),[70,53,80,92] and providing evidence for a functional correlation between the open state observed via solution-state NMR and proteolytic activity. Curiously, no recovery in catalytic activity was observed in I230A suggesting a different mechanism of inactivation, potentially related to the unique loss of cooperativity observed in this mutant.

The ability of TM substrates to insulate the rhomboid active site from the hydrophobic membrane environment could potentially overcome the detrimental effects of hydrophobic mismatch discovered in Chapter 3. To test this hypothesis, catalytic activity of TMD81 against TatA was measured in mixed micelles composed of varying concentrations of Fos-12 and Fos-14 micelles. No significant loss in catalytic activity was observed across the range of conditions tested (Figure 5.10) despite the hydrophobic mismatch present at higher Fos-14 concentrations. This contrasts results obtained in Chapter 3 using the water-soluble BODIPY-casein, which showed a 3-fold decrease in activity over a similar range of mixed-micelle conditions, providing further evidence for the hypothesis that substrate TM sequences can facilitate catalysis by insulating the active site from the membrane environment in response to conditions which might otherwise impair catalysis such as hydrophobic mismatch or α5-gate opening.

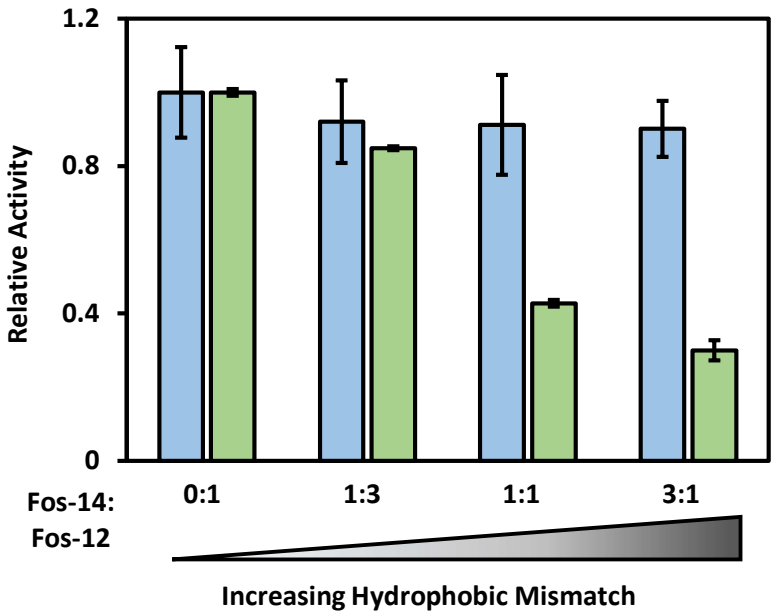


Figure 5.10: Relative rates of proteolysis of PKS187 by TMD81 in mixed micelles comprised of Fos-12 and Fos-14 assessed using a soluble casein substrate (green) and a transmembrane model substrate (blue). Data for the casein substrate obtained from prior publication, and reported in Chapter 3 with permission.[81]

5.3: Discussion

5.3.1: Insight into the Role of α5 Dynamics in the Rhomboid Catalytic Cycle

Over the past 10 years, there has been a growing body of evidence suggesting a role for α5 in the rhomboid catalytic cycle, giving rise to various structural models in which this region functions as a lateral gate for substrate entry. The solution-NMR data presented in this chapter indicates that the *ec*GlpG α5 can undergo a conformation change consistent with a "gate-open" structure under conditions which support proteolytic function. While further studies are required to obtain the exact structure of the gate-open conformation, the observed perturbation around W212 is consistent with a lateral shift of the entire intact α5 away from α2, rather than the 35^{o} rotation proposed in previous models.

Outside of *the E. coli* GlpG, the only other rhomboid whose structure has been successfully solved is the *hi*GlpG from *H. influenzae*.[82] Here, the entire α5 is in a more disordered state over its full length, suggesting that the entire helix is capable of displacing to form the open conformation (Figure 5.11A) in line with the results presented in this chapter.[84] This gating process is also similar to the one proposed for the Site-2 Protease (S2P) family of intramembrane metalloproteases (Figure 5.11B) where the intact gating helices α1, α6 and, to a lesser degree, α5 are displaced laterally while remaining parallel to one another,[197] suggesting that similar mechanisms might be employed by other IMP families to allow substrate access into the active site while minimizing exposure to the membrane environment.

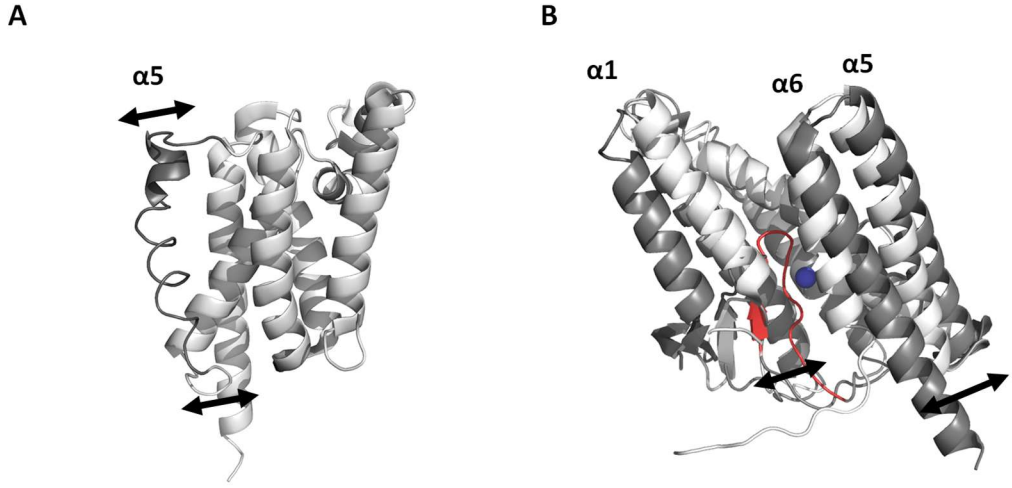


Figure 5.11: Substrate Gating in hiGlpG and S2P A) X-ray crystal structure of the rhomboid protease from *H. influenzae* (PDB 2NR9) in which α5 (grey) is present in a disordered state across the majority of its length. B) X-ray crystal structure of the Site-2-protease from *M. jannaschii* (PDB 3B4R) showing the putative gating helices TM1 and TM6 in their open (grey) and closed (white) conformation. The active site is denoted by the bound Zn^{2+} cofactor (blue), while the substrate-binding exosite identified in previous studies is highlighted in red.[198]

The results from the peptide-binding studies presented in this chapter show incomplete (~40%) conversion of TMD81 into the α5-shifted conformation at saturating conditions suggesting that gate opening facilitates, but is not required for binding. This is consistent with available X-ray structures which depict both the open and closed conformations of *ec*GlpG in complex with peptide-aldehyde inhibitors with a similar sequence to the peptide employed in this thesis,[10] potentially reflecting its small size and lack of TM segments. A similar phenomenon was observed with the Spitz transmembrane substrate, whose solvent-exposed cut site is able to access the rhomboid active site even in the presence of cysteine cross-links preventing opening of the lateral gate.[92] While further NMR-binding studies employing an extended peptide containing both the N and C-terminal portions of the substrate sequence are required to confirm this hypothesis, the correlation observed between peptide binding and gate-opening provide valuable evidence for the functional role of α5 dynamics in facilitating substrate gating.

The loss of activity observed in the gate-open mutants indicates that there are functional consequences associated with adopting this conformation, potentially resulting from an increased exposure of the hydrophilic active site to the hydrophobic membrane environment. This hypothesis is supported by an X-ray structure which depicts a phospholipid molecule bound to the active site of *ec*GlpG (Fig. 3.8A). Here, α5 is displaced slightly from the its position in the apo-structure, suggesting a potential correlation between the α5-open conformation and the entry of lipid or detergents into the active site.[85,89] Alternatively, X-ray structures of the gate-open conformation show a subtle rearrangement of other TM helices around the active site even in the absence of lipid or detergent binding (Fig. 1.15). These rearrangements include a loss of secondary structure in the extracellular-facing region of α6

accompanied by a rearrangement of the catalytic H254 side chain, providing another mechanism through catalytic activity may be compromised upon gate-opening. In this case, interactions with TM substrates could help maintain the rhomboid active site in its functional configuration, giving rise to the recovery in proteolytic activity observed against the TatA TM substrate.

5.3.2: New Insights into *ec*GlpG Selectivity

As shown in Chapter 3, TMD81 displays negligible activity against casein when reconstituted into PC-14 phospholipid bilayers whose hydrophobic dimensions approximate those of the native *E. coli* membrane, raising questions as to how rhomboids are able to carry out their biological functions under such conditions. Our observation that cleavage of the TM model substrate was unaffected by mismatch conditions is consistent with previous studies in which proteolysis rates of AarA and *ec*GlpG in *E. coli* lipids against a TM substrate were found to be comparable to those obtained in Mal-12 micelles (k_{cat} = 1.06 min^{-1} vs 0.8 min^{-1} for AarA. k_{cat} = 0.36 min^{-1} vs 0.1 min^{-1} for *ec*GlpG).[53,52] Since all known rhomboid substrates contain TM segments, these results raise the possibility that hydrophobic mismatch could prevent proteolysis of non-target sequences, as exemplified by the water-soluble casein, *in vivo* while maintaining *ec*GlpG activity against its physiological TM substrates.

The observation that gate-opening mutations enhanced *ec*GlpG activity against substrates with TM segments provides an additional avenue through which selectivity can be regulated *in vivo*. This bears some similarities to results reported in a previous study in which opening of the *Drosophila* Rhomboid-4 lateral gate selectively enhanced cleavage of its physiological Spitz TM substrate while cleavage of both water-soluble and TM substrates

whose cut site lies outside of the membrane environment was unaltered.[80] At the time, the authors attributed these findings to the proposed ability of soluble or extramembrane substrate sequences to enter the rhomboid active site via the L5 loop region, thereby bypassing the α5 gate.[80,92] However, the functional assays and solution-NMR studies reported in this chapter indicate that the binding and proteolysis of peptide sequences which do not contain TM segments is still correlated with the lateral displacement of α5. Instead, the ability of TM substrate sequences to limit the exposure of the gate-open rhomboid active site to the hydrophobic membrane environment is likely responsible for this enhancement.

It has been proposed that exosite binding and cooperativity could provide additional levels of regulation to modulate rhomboid selectivity *in vivo*. This is supported by a study reported by the Lemieux group demonstrating the presence of an exosite in the *P. Stuartii* AarA rhomboid which selectively enhances k_{cat} and K_M against its physiological TatA TM substrate, potentially by facilitating cooperative interactions.[52] The same study reported similar levels of cooperativity for *hi*GlpG against TatA indicating that cooperativity is not restricted to a given rhomboid's cognate substrates and instead represents a general property of TM, substrate sequences. However, the same study failed to detect evidence of exosite binding and cooperativity for *ec*GlpG against TatA, which appears to contradict the results presented in this chapter in which cooperative cleavage of the model TatA substrate was observed for TMD81 in both Mal-12 and Fos-12. This may be explained by the potential influence of the cytoplasmic domain on the ability of the two sites to communicate, particularly given its tendency to interact with Mal and Fos detergent micelles.

Previous observations of the absence of cooperativity in AarA, *ec*GlpG and *hi*GlpG cleavage of casein led to speculation that changes in rhomboid oligomeric state can alter

proteolysis in a substrate-dependant fashion, contributing to *in vivo* selectivity. While this thesis did not study the effect of TMD81 oligomeric state on proteolysis of the model TM TatA substrate, it is curious to note that the only mutant which was catalytically impaired against this substrate (I230A) was also the sole construct to display a significant loss of cooperativity. This is in line with previous studies demonstrating that the monomeric *ec*GlpG was catalytically impaired against TatA while proteolysis of casein was independent of oligomeric state.[52] However, it should also be noted that the full-length *ec*GlpG construct employed in these experiments did not display cooperativity under the conditions tested, and that other studies report significant levels of protease activity against the Gurken TM substrate under conditions which give rise to the monomeric form of *ec*GlpG suggesting that further studies, potentially using the TMD81 construct described in this thesis, are required to fully explore the role of cooperativity in rhomboid selectivity.[88,91]

Taken together, the results presented in this chapter provide evidence that *ec*GlpG is able to adopt an open conformation that is consistent with a lateral displacement of the intact α5 helix, and that gate opening could be correlated to the binding of substrate-derived peptides to the active site. However, mutants that promoted formation of the α5-open state even in the absence of peptide binding were found to have compromised catalytic activity. This loss in activity was found to be dependent on the presence of TM substrate segments that may potentially insulate the rhomboid active site from the hydrophobic membrane environment. Such substrates were able to overcome the deleterious effects of hydrophobic mismatch on proteolytic activity, and could be cleaved by *ec*GlpG in a cooperative manner potentially providing additional mechanisms by which rhomboid selectivity can be enhanced *in vivo*

5.4: Materials and Methods

5.4.1: DNA and Constructs:

The C-terminally hexahistidine tagged TMD81 construct in the pET25b vector was prepared by Dr. Sherratt as described in the preceding chapter. Constructs for the W98F, W159F, W196F, and I230A mutants were created and sequenced through the Civic Bioscience mutagenesis service, and all remaining tryptophan substitution mutants were created using the Quikchange site-directed mutagenesis kit (Quiagen) with the primer sequences shown in Table 5.4, using either the wild-type or W236G TMD81 sequence as a template. Constructs were transformed into chemically-competent *E. coli* DH5α cells for amplification, and verified by sequencing performed by the Stemcore DNA sequencing facility at the Ottawa Hospital Research Institute.

Table 5.4: Primers used to generate the tryptophan substitution mutants along with the F153A gate-open construct employed in this chapter. The location of the altered codon is underlined and in bold.

Mutation	Primer
W122F	Forward: 5'-GAT CAG GAA GTG ATG TTA **TTT** CTG GCC TGG CCA TTC GAT-3' Reverse: 5'-ATC GAA TGG CCA GGC CAG **AAA** TAA CAT CAC TTC CTG ATC-3'
W125F	Forward: 5'- GTG ATG TTA TGG CTG GCC **TTC** CCA TTC GAT CCA ACA CTG-3' Reverse: 5'- CAG TGT TGG ATC GAA TGG **GAA** GGC CAG CCA TAA CAT CAC-3'
W136Y	Forward: 5'- CCA ACA CTG AAA TTT GAG TTC **TAC** CGT TAC TTC ACC CAC GC-3' Reverse: 5'- GC GTG GGT GAA GTA ACG **GTA** GAA CTC AAA TTT CAG TGT TGG-3'
F153A	Forward: 5'- C TTC TCG CTG ATG CAT ATC CTC **GCT** AAC CTG CTC TGG TGG TGG TAT C-3' Reverse: 5'- G ATA CCA CCA CCA GAG CAG GTT **AGC** GAG GAT ATG CAT CAG CGA GAA G-3'
W157F	Forward: 5'- CAT ATC CTC TTT AAC CTG CTC **TTC** TGG TGG TAT CTC GGC GGT GCG-3' Reverse: 5'- CGC ACC GCC GAG ATA CCA CCA **GAA** GAG CAG GTT AAA GAG GAT ATG-3'
W158F	Forward: 5'- CAT ATC CTC TTT AAC CTG CTC TGG **TTC** TGG TAT CTC GGC GGT GCG -3' Reverse: 5'- CGC ACC GCC GAG ATA CCA **GAA** CCA GAG CAG GTT AAA GAG GAT ATG-3'
W212F	Forward: 5'- CTG ATG GGC TAC GTC **TTT** CTA CGT GGC GAA CGC-3' Reverse: 5'- GCG TTC GCC ACG TAG **AAA** GAC GTA GCC CAT CAG-3'

Table 5.4 (Continued): Primers used to generate the tryptophan substitution mutants along with the F153A gate-open construct employed in this chapter. The location of the altered codon is underlined and in bold.

W236G	Forward: 5'- ATT ATC TTT GCG CTG ATC **GGG** ATT GTC GCC GGA TGG TTT GAT TTG-3' Reverse: 5'- CAA ATC AAA CCA TCC GGC GAC AAT **CCC** GAT CAG CGC AAA GAT AAT-5'
W241A	Forward: 5'- CTG ATC TGG ATT GTC GCC GGA **GCG** TTT GAT TTG TTT GGG-3' Reverse: 5'- CCC AAA CAA ATC AAA **CGC** TCC GGC GAC AAT CCA GAT CAG-3'

5.4.2: Protein Expression and Purification

Unlabeled and ^{15}N-labeled TMD81 constructs were expressed and purified into GlpG (pH 7.4) or modified GlpG (pH 7.0) buffer as described in Chapters 3 and 4. The PKS187 TatA model substrate was expressed in LB media, and purified using the same method as the TMD81 constructs. SDS-PAGE analysis of the resulting purified protein fractions reveals a single band corresponding to the expected MW of the TatA construct (Fig. 5.12)

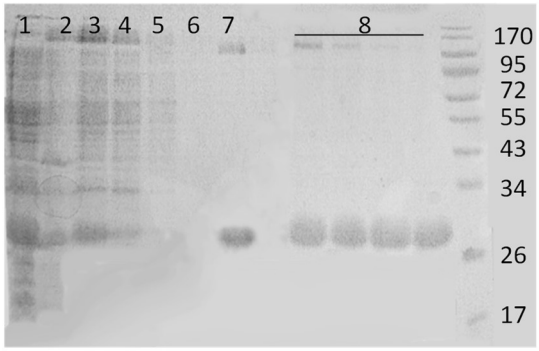


Figure 5.12: SDS-PAGE gel of fractions taken during various stages of the purification of PKS187. L: Molecular weight ladder in kDa. 1: Soluble protein fraction. 2: Insoluble protein fraction. 3: Mal-12-solubilized membrane protein fraction. 4: Ni-NTA column flow through. 5 and 6: Non-specifically bound contaminants removed during Ni-NTA wash step. 7: Purified PKS187 eluted using 250 mM imidazole. 8: Representative fractions obtained after size-exclusion chromatography containing the final, purified TMD81 in GlpG buffer.

5.4.3: Enzyme Kinetics Measurements:

The EnzChek Protease Assay Kit (Invitrogen) was used to evaluate k_{cat} against a soluble casein substrate as described in Chapter 3. For the TM TatA assay, TMD81 in GlpG buffer was incubated with varying concentrations of purified TatA PKS187, with the concentration of rhomboid varied between 0.25 μM to 2 μM to ensure sufficient amounts of cleavage product could be detected while ensuring that the values measured corresponded to the initial rate of proteolysis (<10% substrate consumption). Samples were removed at regular time intervals (0-60 min) and the amount of cleavage product produced was quantified as described in Chapter 3. The TMD81 concentration-normalized initial rates of hydrolysis were plotted against substrate concentration and fit to the Hill equation to determine values of K_m, k_{cat} and n_H.

To assess the rate of proteolysis against a TM substrate, TMD81 in GlpG buffer was incubated with varying concentrations of purified TatA PKS187, with the concentration of rhomboid varied between 0.25 μM to 2 μM to ensure sufficient amounts of cleavage product could be detected while ensuring that the values measured corresponded to the initial rate of proteolysis (<10% substrate consumption). Samples were removed at regular time intervals (0-60 min), quenched by addition of in SDS-PAGE sample loading buffer followed by boiling for 1 minute, and analysed using Coomassie blue stained SDS-PAGE (Fig. 3.2). The amount of the C-terminal PKS187 cleavage product formed (P) was quantified using imageJ as shown in Equation 5.1, where S and P is the intensity of the intact substrate and C-terminal product band reported by imageJ respectively, and S^o is the amount of substrate present at the start of the reaction. Values were calculated from samples corresponding to at least three different time intervals from the same reaction and plotted over time to obtain the initial rate of proteolysis. Due to its small size, the N-terminal cleavage product could not be resolved via

the SDS-PAGE techniques employed in this thesis and was thus not used in the calculation of proteolysis rates.[69] Activity in mixed-Fos-12/Fos-14 micelles was acquired using 0.5 μM of TMD81 and a fixed concentration (50 μM) of PKS187

$$\textbf{5.1} \quad P = \left(\frac{P}{S+P}\right) * S^{o}$$

5.4.4: NMR Spectrometry

All NMR spectra presented in this chapter were acquired using the 600 MHz Bruker Avance III spectrometer with cryoprobe at 45°C as discussed in Chapter 4. For the 5-DOXYL-steric acid free radical titration, a 33 mM stock of the PRE probe was prepared in DMSO and added to the NMR samples to produce the desired PRE: Protein mole ratio, with a maximum DMSO concentration of 3%. As reported in Chapter 4, NMR spectra collected in up to 8% DMSO showed no significant perturbations. An N-acetylated, C-amidated peptide with sequence RKVRMA derived from the N-terminal cleavage product of the *Drosophila* Gurken rhomboid substrate was synthesized by Biobasic Inc. and diluted in modified GlpG buffer to produce a 12 mM stock prior to use.

Chapter 6: Discussion and Closing Remarks

Preamble

In this thesis, novel interactions were identified that could modulate the function of the *E. coli* GlpG rhomboid protease, with potential implications for its regulation and selectivity *in vivo.* While rhomboids possess a limited ability to alter the local membrane environment to accommodate its TM hydrophobic dimensions, the results detailed in Chapter 3 suggest that the energetic cost of membrane thinning is detrimental to *ec*GlpG activity against a water-soluble casein model substrate. This may reflect structural rearrangements within the active site, or impaired substrate entry resulting from its increased exposure to the hydrophobic membrane environment. Curiously, this phenomenon was found to severely hinder proteolytic activity in phospholipid bicelles whose hydrophobic dimensions better emulate that of the native *E. coli* membrane environment, raising questions as to how rhomboids are able to carry out their biological functions under such conditions. Chapter 4 sought to develop techniques to monitor the structure and dynamics of the *ec*GlpG catalytic TMD using solution NMR, which were employed in the subsequent chapter (Chapter 5) to explore the structural nature of substrate gating and its role in rhomboid function. The gate-open mutants W236G and W157F were found to produce a change in the position of α5 consistent with the lateral shift model of substrate gating first demonstrated by the Urban group in 2016. Solution-NMR was used to correlate the lateral displacement of α5 to the binding of substrate-derived peptide fragments into the rhomboid active site, providing structural evidence for its role in substrate gating. Kinetic parameters obtained for the W157F and W236G mutants against a casein model substrate suggest that opening of the lateral gate may have a detrimental effect on proteolysis, though use of a TM model substrate was able to overcome these deleterious effects, potentially through insulating the now exposed active site

from the hydrophobic membrane environment. In line with this hypothesis, the use of TM substrates was also found to restore catalytic activity under mismatch conditions. While the physiological substrate of *ec*GlpG has yet to be identified, all known rhomboid substrates contain TM segments, providing a mechanism for the prevention of non-physiological substrates cleavage *in vivo*.

6.1: Future Prospects for Solution-NMR Studies of Rhomboid Gate-dynamics and Allostery

Investigations into the mechanism of substrate gating have been ongoing since the first *ec*GlpG crystal structure was reported in 2006.[78] Despite the wealth of information, the process through which this occurs remains a topic of significant interest and discussion within the rhomboid community. While the results produced in this thesis provide some insight into the contribution of α5 dynamics to substrate gating, the low quality of the available NMR spectra prevents further characterization of the gate-open conformation. In particular, questions remain concerning the extent to which α5 is displaced, and whether this is accompanied by rearrangements in other regions of *ec*GlpG both within the active site and beyond. One possibility for consideration in futures studies is the selective labeling of TMD81 samples using ^{13}C, ^{15}N, or ^{19}F – enriched amino acids, which can be used to introduce additional structural probes to study specific regions of rhomboid protease.[114,199] For example, established protocols for the selective incorporation of ^{13}C-methyl groups into methionine residues could be used to introduce a series of structural probes across the structure of *ec*GlpG (Fig. 6.1).[200–203] Additionally, M208 and M247 lie in close proximity to several residues on α5 and L5 including W236, allowing them to report on changes across the entire length gate region. Similar methods could be used to introduce ^{13}C-labeled methyl

groups into isoleucine, leucine and valine side chains, potentially allowing α5 displacement to be detected and quantified via Nuclear-Overhauser Effect (NOE) measurements between the W212 indole NH and the selectively-labeled I230 methyl group, and for other structural perturbations to be monitored. Both strategies could be combined with deuteration techniques in which all non-exchangeable hydrogens in the protein sample, except for those introduced site-selectively into the targeted methyl groups, are replaced with deuterium. The lower gyromagnetic ratio of deuterium greatly reduces its dipolar interactions with nearby nuclei, thereby reducing peak broadening resulting from T_2 relaxation processes while simultaneously enhancing the transfer of magnetization in multidimensional NMR experiments. This gives rise to enhanced sensitivity and resolution in triple resonance experiments on uniformly ^{13}C-^{15}N-labeled samples, and has been used in previous studies to obtain peak assignments for large proteins (i.e. > 50 kDa).[204] Several selective labeling strategies have also been developed to facilitate the study of large proteins and biomolecular complexes. For example, the Wagner group reported a technique to facilitate the selective incorporation of protonated, ^{14}N-amino acids that are ^{13}C labeled at the backbone carbonyl position within an otherwise uniformly ^{15}N labeled, deuterated background.[205] This would simplify the HNCO triple-resonance NMR spectra such that only amide protons from amino acids containing the ^{13}C-label would give rise to a correlation with an ^{15}N-labeled amide proton on the subsequent residue in the protein sequence.[205,206] This technique was used to facilitate backbone chemical shift assignments for the >60 kDa VDAC1 integral membrane protein, highlighting the potential of selective backbone labelling for other IMPs such as the rhomboid. In addition to providing a more accurate understanding of the gate-open structure, these tools could be used to identify and characterize structural perturbations

occurring at other regions of *ec*GlpG upon gate-opening, potentially revealing novel pathways through which α5 dynamics could be regulated through allosteric interactions occurring beyond the active site.

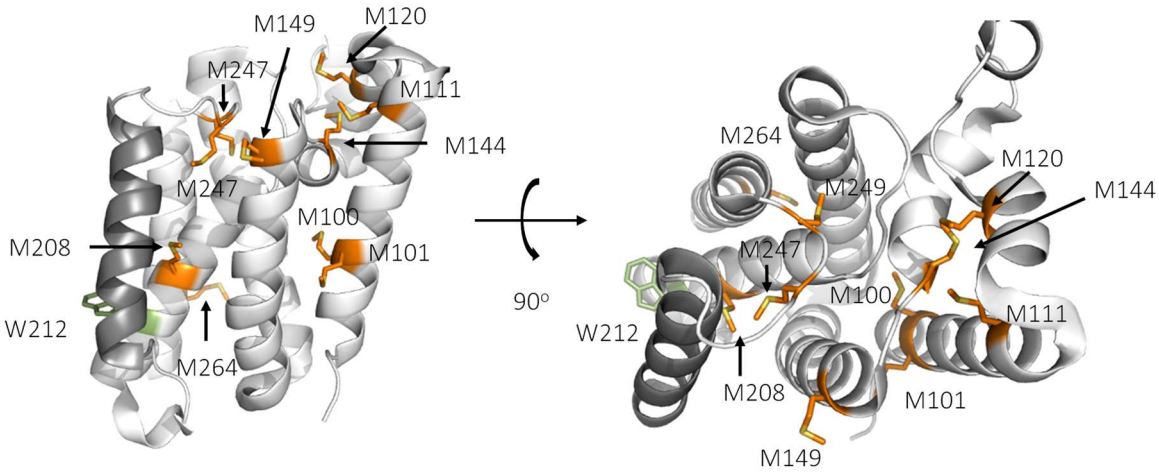


Figure 6.1: Structure of *ec*GlpG (PDB 3B45) showing the location of all Met residues in orange. α5 is highlighted in grey, while W212 is shown in green

6.2: Potential Role for CytD in the Regulation of α5 Dynamics and Gating

Currently, a mechanism for the regulation of α5-opening has yet to be discovered in *ec*GlpG. However, this process can be stimulated in the *Drosophila* Rhomboid-4 through the binding of Ca^{2+} to its cytosolic N-terminal domain, resulting in substrate-specific changes in proteolytic activity both *in vitro* and *in vivo*.[80] This raises the possibility that the *ec*GlpG CytD could play a similar role. In a previous study, removal of the *ec*GlpG cytoplasmic domain by chymotrypsin digestion did not significantly alter its catalytic parameters against both water-soluble and TM substrates when solubilised in Mal-12 detergent micelles.[79] However, the loss of activity in the full-length *ec*GlpG relative to the isolated TMD when solubilised in Fos-12, coupled with the recovery of cooperative proteolysis observed in the TMD81 construct

suggests that *ec*GlpG cytoplasmic domain can indeed exert an influence on rhomboid function depending on the local membrane environment. The ability of the CytD from some rhomboids such as the human RHBDL4 to promote proteolysis through interactions with its native substrate raises the possibility that the *ec*CytD could also play a similar role in the *ec*GlpG proteolytic cycle through facilitating substrate recognition while simultaneously opening the lateral gate upon binding. Efforts to further examine this hypothesis are currently hindered by the lack of information concerning the physiological substrate(s) of the *E. coli* rhomboid. However, recent evidence for its role in glycerol metabolism could expedite the search for potential candidates.[50]

6.3: Regulation of Rhomboid Proteolysis by Cooperativity and Allostery

The ability of AarA to cleave its cognate TatA substrate in a cooperative manner as reported by the Lemieux group in 2014 was initially attributed to its ability to induce a conformation change in the neighboring subunit upon binding to an exosite on the TM surface of AarA, though the same study found no evidence of an exosite in *hi*GlpG and *ec*GlpG.[52] However, it is possible that the cytoplasmic domain masked exosite binding in the full-length construct, potentially explaining the lack of cooperativity previously observed. MD simulations reported in the literature also suggest that *ec*GlpG interactions with TM substrates involve an exosite, since substrate capture was found to involve several sites on the surface of *ec*GlpG.[109] Moreover, kinetics studies on *ec*GlpG examining interactions with a reversible covalent peptide-aldehyde inhibitor showed that inhibition of TatA cleavage was non-competitive.[10] Therefore the TM substrate forms an enzyme-substrate complex that does not

involve interactions with the active site, providing strong evidence that *ec*GlpG also has an exosite.

In addition to facilitating substrate recognition, exosite binding in rhomboids such as AarA and *ec*GlpG could potentially limit the intrusion of lipid molecules into the active site. Exosite binding is employed by other IMP families including the S2P and γ-secretases to facilitate cleavage of TM substrates.[198,207] While the mechanism of this enhancement has yet to be determined, the exosite for both IMP families is located directly in front of the substrate gate (Figure 5.11B) suggesting that binding of TM segments to this region could also help minimize unfavourable interactions between the active site and the lipid phase.[198,207,208] Furthermore, X-ray structures obtained for Signal Peptide Peptidase (SPP), a homologue of the γ-secretase catalytic subunit, show the two catalytic aspartic acid residues positioned in a manner which is incompatible with the known proteolytic mechanism.[209] However, substrate binding to the γ-secretase exosite is accompanied by a conformational change within the catalytic domain which may serve to restore the catalytic integrity of the γ-secretase/SPP aspartic acid dyad, raising the possibility that similar substrate-exosite interactions in *ec*GlpG could help counteract the structural perturbations potentially resulting from gate-opening or hydrophobic mismatch.[207,208,210–212] Future studies examining the ability of BODIPY-casein or the Gurken-derived peptide fragment presented in Chapter 5 to non-competitively inhibit proteolysis of the TM PKS187 TatA substrate by either the WT or gate-open mutant of TMD81 could be used to verify the existence of a substrate-binding exosite in *ec*GlpG. This would open the door to an investigation of the role of the exosite in regulating rhomboid α5 opening, activity, and selectivity.

While previous studies have identified a consensus required for rhomboid proteolysis, the results presented in this thesis open up new avenues of investigation into the regulation of rhomboid activity and selectivity involving interactions beyond the active site including α5 dynamics and hydrophobic mismatch, potentially coupled with allosteric interactions involving the rhomboid CytD, exosite binding, and cooperativity. The conserved nature of rhomboid TMD suggest that the results and analyses provided in this thesis can be applied to better understand the regulation of other members of the rhomboid family, providing valuable insight into the diverse range of biological pathways in which they participate.

Chapter 7: References

1. Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. Predicting Transmembrane Protein Topology with a Hidden Markov model: Application to Complete Genomes. *J. Mol. Biol.* **305,** 567–580 (2001).

2. Wallin, E. & von Heijne, G. Genome-wide analysis of integral membrane proteins from eubacterial, archaean, and eukaryotic organisms. *Protein Sci.* **7,** 1029–38 (1998).

3. Lemberg, M. K. Sampling the membrane: function of rhomboid-family proteins. *Trends Cell Biol.* 1–8 (2013). doi:10.1016/j.tcb.2013.01.002

4. Bracey, M. H., Cravatt, B. F. & Stevens, R. C. Structural commonalities among integral membrane enzymes. *FEBS Lett.* **567,** 159–165 (2004).

5. Snijder, H. J. *et al.* Structural evidence for dimerization-regulated activation of an integral membrane phospholipase. *Nature* **401,** 717–721 (1999).

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

7. Nüsslein-Volhard, C., Wieschaus, E. & Kluding, H. Mutations affecting the pattern of the larval cuticle in Drosophila melanogaster I. Zygotic loci on the second chromosome. *Roux's Arch Dev Biol* **193,** 267–282 (1984).

8. Bang, A. G. & Kintner, C. Rhomboid and Star facilitate presentation and processing of the Drosophila TGF-alpha homolog Spitz. *Genes Dev.* **14,** 177–86 (2000).

9. Urban, S., Lee, J. . R. & Freeman, M. Drosophila rhomboid-1 defines a family of putative intramembrane serine proteases. *Cell* **107,** 173–82 (2001).

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

11. Lemberg, M. K. & Freeman, M. Functional and evolutionary implications of enhanced genomic analysis of rhomboid intramembrane proteases. *Genome Res.* **17,** 1634–46 (2007).

12. Arutyunova, E. *et al.* Probing catalytic rate enhancement during intramembrane proteolysis. *Biol. Chem.* **397,** 907–919 (2016).

13. Wasserman, J. D., Urban, S. & Freeman, M. A family of rhomboid-like genes: Drosophila rhomboid-1 and roughoid/rhomboid-3 cooperate to activate EGF receptor signaling. *Genes Dev.* **14,** 1651–1663 (2000).

14. Sturtevant, M. A., Roark, M. & Bier, E. The Drosophila rhomboid gene mediates the localized formation of wing veins and interacts genetically with components of the EGF-R signaling pathway. *Genes Dev.* **7,** 961–973 (1993).

15. Adrain, C. *et al.* Mammalian EGF receptor activation by the rhomboid protease RHBDL2. *EMBO Rep.* **12,** 421–427 (2011).

16. Song, W. *et al.* Rhomboid domain containing 1 promotes colorectal cancer growth through activation of the EGFR signalling pathway. *Nat. Commun.* **6,** 8022 (2015).

17. Wang, Y. *et al.* A novel member of the Rhomboid family, RHBDD1, regulates BIK-mediated apoptosis. *Cell. Mol. Life Sci.* **65,** 3822–9 (2008).

18. Wan, C. *et al.* Exosome-related multi-pass transmembrane protein TSAP6 is a target of rhomboid protease RHBDD1-induced proteolysis. *PLoS One* **7,** (2012).

19. Palmer, T. & Berks, B. C. The twin-arginine translocation (Tat) protein export pathway. *Nat. Rev. Microbiol.* **10,** 483–496 (2012).

20. Ochsner, U. a, Snyder, A., Vasil, A. I. & Vasil, M. L. Effects of the twin-arginine translocase on secretion of virulence factors, stress response, and pathogenesis. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 8312–8317 (2002).

21. Stevenson, L. G. *et al.* Rhomboid protease AarA mediates quorum-sensing in Providencia stuartii by activating TatA of the twin-arginine translocase. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 1003–1008 (2007).

22. Gohlke, U. *et al.* The TatA component of the twin-arginine protein transport system forms channel complexes of variable diameter. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 10482–6 (2005).

23. Fleig, L. *et al.* Ubiquitin-Dependent Intramembrane Rhomboid Protease Promotes ERAD of Membrane Proteins. *Mol. Cell* **47,** 558–569 (2012).

24. Greenblatt, E. J., Olzmann, J. a & Kopito, R. R. Making the cut: intramembrane cleavage by a rhomboid protease promotes ERAD. *Nat. Struct. Mol. Biol.* **19,** 979–81 (2012).

25. Paschkowsky, S., Hamze, M., Oestereich, F. & Munter, L. M. Alternative processing of the amyloid precursor protein family by rhomboid protease RHBDL4. *J. Biol. Chem.* **291,** 21903–21912 (2016).

26. O'Brien, R. J. & Wong, P. C. Amyloid precursor protein processing and Alzheimer's disease. *Annu. Rev. Neurosci.* **34,** 185–204 (2011).

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

28. Michaelis, G. *et al.* Mitochondrial signal peptidases of yeast: The rhomboid peptidase Pcp1 and its substrate cytochrome c peroxidase. *Gene* **354,** 58–63 (2005).

29. Orrenius, S. & Zhivotovsky, B. Cardiolipin oxidation sets cytochrome c free. *Nat. Chem. Biol.* **1,** 188–189 (2005).

30. Kluck, R. M., Bossy-Wetzel, E., Green, D. R. & Newmeyer, D. D. The release of cytochrome c from mitochondria: a primary site for Bcl-2 regulation of apoptosis. *Science (80-.).* **275,** 1132–1136 (1997).

31. Herlan, M., Vogel, F., Bornhövd, C., Neupert, W. & Reichert, A. S. Processing of Mgm1 by the rhomboid-type protease Pcp1 is required for maintenance of

mitochondrial morphology and of mitochondrial DNA. *J. Biol. Chem.* **278,** 27781–27788 (2003).

32. Schafer, A. *et al.* Intramembrane proteolysis of Mgm1 by the mitochondrial rhomboid protease is highly promiscuous regarding the sequence of the cleaved hydrophobic segment. *J. Mol. Biol.* **401,** 182–193 (2010).

33. Escobar-Henriques, M. & Langer, T. Mitochondrial shaping cuts. *Biochim. Biophys. Acta* **1763,** 422–429 (2006).

34. Zick, M. *et al.* Distinct roles of the two isoforms of the dynamin-like GTPase Mgm1 in mitochondrial fusion. *FEBS Lett.* **583,** 2237–2243 (2009).

35. Shi, G. *et al.* Functional alteration of PARL contributes to mitochondrial dysregulation in Parkinson's disease. *Hum. Mol. Genet.* **20,** 1966–1974 (2011).

36. Pellegrini, L. & Scorrano, L. A cut short to death: Parl and Opa1 in the regulation of mitochondrial morphology and apoptosis. *Cell Death Differ.* **14,** 1275–1284 (2007).

37. Chan, E. Y. L. & McQuibban, G. A. The mitochondrial rhomboid protease: Its rise from obscurity to the pinnacle of disease-relevant genes. *Biochim. Biophys. Acta - Biomembr.* **1828,** 2916–2925 (2013).

38. Phasukkijwatana, N. *et al.* Genome-wide linkage scan and association study of PARL to the expression of LHON families in Thailand. *Hum. Genet.* **128,** 39–49 (2010).

39. Walder, K. *et al.* The mitochondrial rhomboid protease PSARL is a new candidate gene for type 2 diabetes. *Diabetologia* **48,** 459–68 (2005).

40. Jewett, T. J. & Sibley, L. D. Aldolase forms a bridge between cell surface adhesins and the actin cytoskeleton in apicomplexan parasites. *Mol. Cell* **11,** 885–894 (2003).

41. Olshina, M. A., Wong, W. & Baum, J. Holding back the microfilament - Structural insights into actin and the actin-monomer-binding proteins of apicomplexan parasites. *IUBMB Life* **64,** 370–377 (2012).

42. Baker, R. P., Wijetilaka, R. & Urban, S. Two Plasmodium rhomboid proteases preferentially cleave different adhesins implicated in all invasive stages of malaria. *PLoS Pathog.* **2,** e113 (2006).

43. Buguliskis, J. S., Brossier, F., Shuman, J. & Sibley, L. D. Rhomboid 4 (ROM4) affects the processing of surface adhesins and facilitates host cell invasion by Toxoplasma gondii. *PLoS Pathog.* **6,** e1000858 (2010).

44. Dowse, T. J., Pascall, J. C., Brown, K. D. & Soldati, D. Apicomplexan rhomboids have a potential role in microneme protein cleavage during host cell invasion. *Int. J. Parasitol.* **35,** 747–56 (2005).

45. Riestra, A. M. *et al.* A Trichomonas vaginalis Rhomboid Protease and Its Substrate Modulate Parasite Attachment and Cytolysis of Host Cells. *PLoS Pathog.* **11,** 1–25 (2015).

46. Rugarabamu, G., Marq, J. B., Guérin, A., Lebrun, M. & Soldati-Favre, D. Distinct

contribution of Toxoplasma gondii rhomboid proteases 4 and 5 to micronemal protein protease 1 activity during invasion. *Mol. Microbiol.* **97,** 244–262 (2015).

47. Zeng G. et al. Repressor for the sn-glycerol 2-phosphate regulon of Escherichia coli K-12: primary structure and identification of the DNA-binding domain. *Am. Soc. MIirobiology* **178,** 7080–7089 (1996).

48. Maegawa, S., Ito, K. & Akiyama, Y. Proteolytic action of GlpG , a rhomboid protease in the Escherichia coli cytoplasmic membrane. *Biochemistry* **44,** 13543–13552 (2005).

49. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2,** 2006.0008 (2006).

50. Russell, C. W., Richards, A. C., Chang, A. S. & Mulvey, M. A. The Rhomboid Protease GlpG Promotes the Persistence of Extraintestinal Pathogenic Escherichia coli within the Gut. *Infect. Immun.* **85,** IAI.00866-16 (2017).

51. Lemberg, M. K. *et al.* Mechanism of intramembrane proteolysis investigated with purified rhomboid proteases. *EMBO J.* **24,** 464–72 (2005).

52. Arutyunova, E. *et al.* Allosteric regulation of rhomboid intramembrane proteolysis. *EMBO J.* 1–13 (2014). doi:10.15252/embj.201488149

53. Dickey, S. W., Baker, R. P., Cho, S. & Urban, S. Proteolysis inside the membrane is a rate-governed reaction not driven by substrate affinity. *Cell* **155,** 1270–1281 (2013).

54. Koonin, E. V *et al.* The rhomboids: a nearly ubiquitous family of intramembrane serine proteases that probably evolved by multiple ancient horizontal gene transfers. *Genome Biol.* **4,** R19 (2003).

55. Southan, C. A genomic perspective on human proteases. *FEBS Lett.* **498,** 214–218 (2001).

56. Freeman, M. Proteolysis within the membrane: rhomboids revealed. *Nat. Rev. Mol. Cell Biol.* **5,** 188–97 (2004).

57. Zettl, M., Adrain, C., Strisovsky, K., Lastun, V. & Freeman, M. Rhomboid family pseudoproteases use the ER quality control machinery to regulate intercellular signaling. *Cell* **145,** 79–91 (2011).

58. Populo, H., Lopes, J. M. & Soares, P. The mTOR signalling pathway in human cancer. *Int. J. Mol. Sci.* **13,** 1886–1918 (2012).

59. Morin, S. *et al.* Tumor Necrosis Factor Signaling Requires iRhom2 to Promote Trafficking and Activation of TACE. *Science (80-.).* **335,** 225–228 (2013).

60. Rami Tzafriri, A. & Edelman, E. R. Quasi-steady-state kinetics at enzyme and substrate concentrations in excess of the Michaelis-Menten constant. *J. Theor. Biol.* **245,** 737–748 (2007).

61. Briggs, G. E. & Haldane, J. B. S. A Further Note on the Kinetics of Enzyme Action. *Biochem. J.* **19,** 1037–1038 (1925).

62. Urban, S., Schlieper, D. & Freeman, M. Conservation of intramembrane proteolytic activity and substrate specificity in prokaryotic and eukaryotic rhomboids. *Curr. Biol.* **12,** 1507–12 (2002).

63. Sherratt, A. R., Blais, D. R., Ghasriani, H., Pezacki, J. P. & Goto, N. K. Activity-based protein profiling of the E. coli GlpG rhomboid protein delineates the catalytic core. *Biochemistry* **51,** 7794–803 (2012).

64. Vosyka, O. *et al.* Activity-based probes for rhomboid proteases discovered in a mass spectrometry-based assay. *Proc. Natl. Acad. Sci. U. S. A.* 1–6 (2013). doi:10.1073/pnas.1215076110

65. Wolf, E. V., Seybold, M., Hadravová, R., Strisovsky, K. & Verhelst, S. H. L. Activity-Based Protein Profiling of Rhomboid Proteases in Liposomes. *ChemBioChem* **16,** 1616–1621 (2015).

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

67. Zoll, S. *et al.* Substrate binding and specificity of rhomboid intramembrane protease revealed by substrate-peptide complex structures. *EMBO J.* **33,** 1–14 (2014).

68. Urban, S. & Baker, R. P. In vivo analysis reveals substrate-gating mutants of a rhomboid intramembrane protease display increased activity in living cells. *Biol. Chem.* **389,** 1107–1115 (2008).

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

70. Baker, R. P., Young, K., Feng, L., Shi, Y. & Urban, S. Enzymatic analysis of a rhomboid intramembrane protease implicates transmembrane helix 5 as the lateral substrate gate. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 8257–8262 (2007).

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

72. Horne, D. S. Casein structure, self-assembly and gelation. *Curr. Opin. Colloid Interface Sci.* **7,** 456–461 (2002).

73. Chakraborty, A. & Basak, S. Effect of surfactants on casein structure: a spectroscopic study. *Colloids Surf. B. Biointerfaces* **63,** 83–90 (2008).

74. Portnaya, I. *et al.* Micellization of bovine beta-casein studied by isothermal titration microcalorimetry and cryogenic transmission electron microscopy. *J. Agric. Food Chem.* **54,** 5555–61 (2006).

75. Caro, A. L., Rodríguez Niño, M. R. & Rodríguez Patino, J. M. Dynamics of penetration of dipalmitoyl-phosphatidyl-choline (DPPC) monolayers by β-casein. *Colloids Surfaces A Physicochem. Eng. Asp.* **341,** 134–141 (2009).

76. Gallier, S., Gragson, D., Jiménez-Flores, R. & Everett, D. W. β-Casein-phospholipid monolayers as model systems to understand lipid-protein interactions in the milk fat globule membrane. *Int. Dairy J.* **22,** 58–65 (2012).

77. Sokolovski, M., Sheynis, T., Kolusheva, S. & Jelinek, R. Membrane interactions and lipid binding of casein oligomers and early aggregates. *Biochim. Biophys. Acta - Biomembr.* **1778,** 2341–2349 (2008).

78. Wang, Y., Zhang, Y. & Ha, Y. Crystal structure of a rhomboid family intramembrane protease. *Nature* **444,** 179–80 (2006).

79. Lazareno-Saez, C., Arutyunova, E., Coquelle, N. & Lemieux, M. J. Domain Swapping in the Cytoplasmic Domain of the Escherichia coli Rhomboid Protease. *J. Mol. Biol.* **425,** 1127–42 (2013).

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

81. Foo, A. C. Y., Harvey, B. G. R., Metz, J. J. & Goto, N. K. Influence of hydrophobic mismatch on the catalytic activity of E. coli GlpG rhomboid protease. *Protein Sci.* **24,** 464–473 (2015).

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

83. Brooks, C. L., Lazareno-Saez, C., Lamoureux, J. S., Mak, M. W. & Lemieux, M. J. Insights into substrate gating in H. influenzae rhomboid. *J. Mol. Biol.* **407,** 687–97 (2011).

84. Wu, Z. *et al.* Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry. *Nat. Struct. Mol. Biol.* **13,** 1084–1091 (2006).

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

86. Wang, Y. & Ha, Y. Open-cap conformation of intramembrane protease GlpG. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 2098–102 (2007).

87. Wang, Y., Maegawa, S., Akiyama, Y. & Ha, Y. The role of L1 loop in the mechanism of rhomboid intramembrane protease GlpG. *J. Mol. Biol.* **374,** 1104–13 (2007).

88. Vinothkumar, K. R. *et al.* The structural basis for catalysis and substrate specificity of a rhomboid protease. *EMBO J.* **29,** 3797–809 (2010).

89. Vinothkumar, K. R. Structure of rhomboid protease in a lipid environment. *J. Mol. Biol.* **407,** 232–47 (2011).

90. Xue, Y. & Ha, Y. Catalytic mechanism of rhomboid protease GlpG probed by 3,4-dichloroisocoumarin and diisopropyl fluorophosphonate. *J. Biol. Chem.* **287,** 3099–

107 (2012).

91. Xue, Y. *et al.* Conformational change in rhomboid protease GlpG induced by inhibitor binding to its S' subsites. *Biochemistry* **51,** 3723–31 (2012).

92. Xue, Y. & Ha, Y. Large lateral movement of transmembrane helix S5 is not required for substrate access to the active site of rhomboid intramembrane protease. *J. Biol. Chem.* **288,** 16645–54 (2013).

93. Vinothkumar, K. R., Pierrat, O. A., Large, J. M. & Freeman, M. Structure of rhomboid protease in complex with β-lactam inhibitors defines the S2' cavity. *Structure* **21,** 1051–8 (2013).

94. Ha, Y., Akiyama, Y. & Xue, Y. Structure and mechanism of rhomboid protease. *J. Biol. Chem.* **288,** 15430–15436 (2013).

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

96. Ghasriani, H. *et al.* Micelle-Catalyzed Domain Swapping in the GlpG Rhomboid Protease Cytoplasmic Domain. *Biochemistry* **53,** 5907–5915 (2014).

97. Sherratt, A. R., Braganza, M. V, Nguyen, E., Ducat, T. & Goto, N. K. Insights into the effect of detergents on the full-length rhomboid protease from Pseudomonas aeruginosa and its cytosolic domain. *Biochim. Biophys. Acta* **1788,** 2444–53 (2009).

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

99. Kim, T. *et al.* Influence of hydrophobic mismatch on structures and dynamics of gramicidin a and lipid bilayers. *Biophys. J.* **102,** 1551–60 (2012).

100. Mondal, S., Khelashvili, G., Shi, L. & Weinstein, H. The cost of living in the membrane: a case study of hydrophobic mismatch for the multi-segment protein LeuT. *Chem. Phys. Lipids* **169,** 27–38 (2013).

101. van den Brink-van der Laan, E., Killian, J. A. & de Kruijff, B. Nonbilayer lipids affect peripheral and integral membrane proteins via changes in the lateral pressure profile. *Biochim. Biophys. Acta* **1666,** 275–88 (2004).

102. daCosta, C. J. B., Dey, L., Therien, J. P. D. & Baenziger, J. E. A distinct mechanism for activating uncoupled nicotinic acetylcholine receptors. *Nat. Chem. Biol.* **9,** 701–707 (2013).

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

104. Moin, S. M. & Urban, S. Membrane immersion allows rhomboid proteases to achieve specificity by reading transmembrane segment dynamics. *Elife* **1,** 1–16 (2012).

105. Vinothkumar, K. R. & Freeman, M. Intramembrane proteolysis by rhomboids: catalytic mechanisms and regulatory principles. *Curr. Opin. Struct. Biol.* 1–8 (2013). doi:10.1016/j.sbi.2013.07.014

106. Zhou, Y., Moin, S. M., Urban, S. & Zhang, Y. An internal water-retention site in the rhomboid intramembrane protease GlpG ensures catalytic efficiency. *Structure* **20,** 1255–63 (2012).

107. Langosch, D., Scharnagl, C., Steiner, H. & Lemberg, M. K. Understanding intramembrane proteolysis: From protein dynamics to reaction kinetics. *Trends Biochem. Sci.* **40,** 318–327 (2015).

108. Sampathkumar, P. *et al.* Oligomeric state study of prokaryotic rhomboid proteases. *Biochim. Biophys. Acta - Biomembr.* **1818,** 3090–3097 (2012).

109. Reddy, T. & Rainey, J. K. Multifaceted substrate capture scheme of a rhomboid protease. *J. Phys. Chem. B* **116,** 8942–54 (2012).

110. Pervushin, K., Riek, R., Wider, G. & Wüthrich, K. Attenuated T2 relaxation by mutual cancellation of dipole–dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc. Natl. Acad. Sci.* **94,** 12366–12371 (1997).

111. Liang, B., Bushweller, J. H. & Tamm, L. K. Site-directed parallel spin-labeling and paramagnetic relaxation enhancement in structure determination of membrane proteins by solution NMR spectroscopy. *J Am Chem Soc* **128,** 4389–4397 (2006).

112. Iwahara, J. & Clore, G. M. Detecting transient intermediates in macromolecular binding by paramagnetic NMR. *Nature* **440,** 1227–1230 (2006).

113. MacRaild, C. A., Howlett, G. J. & Gooley, P. R. The structure and interactions of human apolipoprotein C-II in dodecyl phosphocholine. *Biochemistry* **43,** 8084–8093 (2004).

114. Tugarinov, V., Kanelis, V. & Kay, L. E. Isotope labeling strategies for the study of high-molecular-weight proteins by solution NMR spectroscopy. *Nat. Protoc.* **1,** 749–754 (2006).

115. Fernández, C. & Wider, G. TROSY in NMR studies of the structure and function of large biological macromolecules. *Curr. Opin. Struct. Biol.* **13,** 570–580 (2003).

116. Kleckner, I. R. & Foster, M. P. An introduction to NMR-based approaches for measuring protein dynamics. *Biochim. Biophys. Acta* **1814,** 942–68 (2011).

117. Kloiber, K., Spitzer, R., Grutsch, S., Kreutz, C. & Tollinger, M. Longitudinal exchange: An alternative strategy towards quantification of dynamics parameters in ZZ exchange spectroscopy. *J. Biomol. NMR* **51,** 123–129 (2011).

118. Igumenova, T. I., Brath, U., Akke, M. & Palmer, A. G. Characterization of chemical exchange using residual dipolar coupling. *J. Am. Chem. Soc.* **129,** 13396–13397 (2007).

119. Skrynnikov, N. R., Dahlquist, F. W. & Kay, L. E. Reconstructing NMR spectra of 'invisible' excited protein states using HSQC and HMQC experiments. *J. Am. Chem. Soc.* **124,** 12352–12360 (2002).

120. Hansen, A. L., Lundström, P., Velyvis, A. & Kay, L. E. Quantifying millisecond exchange dynamics in proteins by CPMG relaxation dispersion NMR using side-chain 1H probes. *J. Am. Chem. Soc.* **134,** 3178–89 (2012).

121. Poget, S. F. & Girvin, M. E. Solution NMR of membrane proteins in bilayer mimics: small is beautiful, but sometimes bigger is better. *Biochim. Biophys. Acta* **1768,** 3098–106 (2007).

122. Garavito, R. M. & Ferguson-Miller, S. Detergents as Tools in Membrane Biochemistry. *J. Biol. Chem.* **276,** 32403–32406 (2001).

123. Morrison, E. A. & Henzler-Wildman, K. A. Reconstitution of integral membrane proteins into isotropic bicelles with improved sample stability and expanded lipid composition profile. *Biochim. Biophys. Acta* **1818,** 814–820 (2012).

124. Chung, K. Y. *et al.* Role of Detergents in Conformational Exchange of a G Protein-coupled Receptor. *J. Biol. Chem.* **287,** 36305–11 (2012).

125. Laurent J. Catoire, X. L. W. and D. E. W. in *Membrane proteins production for structural analysis* 315–345 (2014). doi:10.1007/978-1-4939-0662-8

126. Lipfert, J., Columbus, L., Chu, V. B., Lesley, S. A. & Doniach, S. Size and shape of detergent micelles determined by small-angle X-ray scattering. *J. Phys. Chem. B* **111,** 12427–38 (2007).

127. Oliver, R. C. *et al.* Dependence of micelle size and shape on detergent alkyl chain length and head group. *PLoS One* **8,** e62488 (2013).

128. Columbus, L. *et al.* Mixing and matching detergents for membrane protein NMR structure determination. *J. Am. Chem. Soc.* **131,** 7320–6 (2009).

129. Stangl, M., Veerappan, A., Kroeger, A., Vogel, P. & Schneider, D. Detergent properties influence the stability of the glycophorin a transmembrane helix dimer in lysophosphatidylcholine micelles. *Biophys. J.* **103,** 2455–2464 (2012).

130. Peric, M., Alves, M. & Bales, B. L. Combining precision spin-probe partitioning with time-resolved fluorescence quenching to study micelles. Application to micelles of pure lysomyristoylphosphatidylcholine (LMPC) and LMPC mixed with sodium dodecyl sulfate. *Chem. Phys. Lipids* **142,** 1–13 (2006).

131. Ericsson, C. A., Söderman, O., Garamus, V. M., Bergström, M. & Ulvenlund, S. Effects of temperature, salt, and deuterium oxide on the self-aggregation of alkylglycosides in dilute solution. 2. n-Tetradecyl-beta-D-maltoside. *Langmuir* **21,** 1507–1515 (2005).

132. Seddon, A. M., Curnow, P. & Booth, P. J. Membrane proteins, lipids and detergents: not just a soap opera. *Biochim. Biophys. Acta* **1666,** 105–17 (2004).

133. Vinogradova, O., Sönnichsen, F. & Sanders, C. R. On choosing a detergent for solution NMR studies of membrane proteins. *J. Biomol. NMR* **11,** 381–386 (1998).

134. Tian, C. *et al.* Solution NMR spectroscopy of the human vasopressin V2 receptor, a

G protein-coupled receptor. *J. Am. Chem. Soc.* **127,** 8010–1 (2005).

135. Koehler, J. *et al.* Lysophospholipid micelles sustain the stability and catalytic activity of diacylglycerol kinase in the absence of lipids. *Biochemistry* **49,** 7089–99 (2010).

136. Chill, J. H., Louis, J. M., Miller, C. & Bax, A. NMR study of the tetrameric KcsA potassium channel in detergent micelles. *Protein Sci.* **15,** 684–698 (2006).

137. Heginbotham, L., Odessey, E. & Miller, C. Tetrameric stoichiometry of a prokaryotic K+ channel. *Biochemistry* **36,** 10335–10342 (1997).

138. Ahn, V. E. *et al.* A hydrocarbon ruler measures palmitate in the enzymatic acylation of endotoxin. *EMBO J.* **23,** 2931–2941 (2004).

139. Hwang, P. M., Bishop, R. E. & Kay, L. E. The integral membrane enzyme PagP alternates between two dynamically distinct states. *Proc. Natl. Acad. Sci. U. S. A.* **101,** 9618–23 (2004).

140. Klingler, J., Vargas, C., Fiedler, S. & Keller, S. Preparation of ready-to-use small unilamellar phospholipid vesicles by ultrasonication with a beaker resonator. *Anal. Biochem.* **477,** 10–12 (2015).

141. Warschawski, D. E. *et al.* Choosing membrane mimetics for NMR structural studies of transmembrane proteins. *Biochim. Biophys. Acta - Biomembr.* **1808,** 1957–1974 (2011).

142. Morrison, E. a *et al.* Antiparallel EmrE exports drugs by exchanging between asymmetric structures. *Nature* **481,** 45–50 (2012).

143. Poget, S. F., Cahill, S. M. & Girvin, M. E. Isotropic bicelles stabilize the functional form of a small multidrug-resistance pump for NMR structural studies. *J. Am. Chem. Soc.* **129,** 2432–2433 (2007).

144. Luchette, P. A. *et al.* Morphology of fast-tumbling bicelles: a small angle neutron scattering and NMR study. *Biochim. Biophys. Acta* **1513,** 83–94 (2001).

145. Mineev, K. S., Nadezhdin, K. D., Goncharuk, S. A. & Arseniev, A. S. Characterization of Small Isotropic Bicelles with Various Compositions. *Langmuir* **32,** 6624–6637 (2016).

146. Goodman, M., Verdini, A. S., Toniolo, C., Phillips, W. D. & Bovey, F. A. Sensitive Criteria for the Critical Size for Helix Formation in Oligopeptides. *PNAS* **64,** 444–450 (1969).

147. Whitmore, L. & Wallace, B. A. Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases. *Biopolymers* **89,** 392–400 (2008).

148. Woody, R. W. & Sreerama, N. On the analysis of membrane protein circular dichroism spectra. *Protein Sci.* **13,** 100–112 (2004).

149. Greenfield, N. & Fasman, G. D. Computed Circular Dichroism Spectra for the Evaluation of Protein Conformation. *Biochemistry* **8,** 4108–4116 (1969).

150. Provencher, S. W. & Glöckner, J. Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* **20,** 33–37 (1981).

151. Wallace, B. A., Lees, J. G., Orry, A. J. W., Lobley, A. & Janes, R. W. Analyses of circular dichroism spectra of membrane proteins. 875–884 (2003). doi:10.1110/ps.0229603.structural

152. Sreerama, N. & Woody, R. W. Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set. *Anal. Biochem.* **287,** 252–260 (2000).

153. Williamson, I. M., Alvis, S. J., East, J. M. & Lee, A. G. Interactions of phospholipids with the potassium channel KcsA. *Biophys. J.* **83,** 2026–38 (2002).

154. Lee, A. G. How lipids interact with an intrinsic membrane protein: the case of the calcium pump. *Biochim. Biophys. Acta* **1376,** 381–390 (1998).

155. Jensen, M. & Mouritsen, O. G. Lipids do influence protein function - The hydrophobic matching hypothesis revisited. *Biochim. Biophys. Acta - Biomembr.* **1666,** 205–226 (2004).

156. Harroun, T. a, Heller, W. T., Weiss, T. M., Yang, L. & Huang, H. W. Experimental evidence for hydrophobic matching and membrane-mediated interactions in lipid bilayers containing gramicidin. *Biophys. J.* **76,** 937–45 (1999).

157. Hong, H. & Tamm, L. K. Elastic coupling of integral membrane protein stability to lipid bilayer forces. *Proc. Natl. Acad. Sci.* **101,** 4065–4070 (2004).

158. Ryba, N. J. & Marsh, D. Protein rotational diffusion and lipid/protein interactions in recombinants of bovine rhodopsin with saturated diacylphosphatidylcholines of different chain lengths studied by conventional and saturation-transfer electron spin resonance. *Biochemistry* **31,** 7511–7518 (1992).

159. Wang, H., Andersen, K. K., Vad, B. S. & Otzen, D. E. Biochimica et Biophysica Acta OmpA can form folded and unfolded oligomers. *Biochem. Biophys. Acta* **1834,** 127–136 (2013).

160. Engelman, D. M. Membranes are more mosaic than fluid. *Nature* **438,** 578–580 (2005).

161. Brossier, F., Jewett, T. J., Sibley, L. D. & Urban, S. A spatially localized rhomboid protease cleaves cell surface adhesins essential for invasion by Toxoplasma. **102,** (2005).

162. Shen, B., Buguliskis, J. S., Lee, T. D. & David Sibley, L. Functional analysis of rhomboid proteases during Toxoplasma invasion. *MBio* **5,** 1–14 (2014).

163. Cellular, A. J. O. F. An Entamoeba histolytica rhomboid protease with atypical specificity cleaves a surface lectin involved in phagocytosis and immune evasion. **22,** 1636–1646 (2008).

164. Cordy, J. M., Hooper, N. M. & Turner, A. J. The involvement of lipid rafts in

Alzheimer's disease. *Mol. Membr. Biol.* **23,** 111–122 (2006).

165. Wahrle, S. *et al.* Cholesterol-dependent gamma-secretase activity in buoyant cholesterol-rich membrane microdomains. *Neurobiol. Dis.* **9,** 11–23 (2002).

166. Urano, Y. *et al.* Association of active gamma-secretase complex with lipid rafts. *J. Lipid Res.* **46,** 904–12 (2005).

167. Winkler, E. *et al.* Generation of Alzheimer disease-associated amyloid B42/43 peptide by Y-secretase can be inhibited directly by modulation of membrane thickness. *J. Biol. Chem.* **287,** 21326–21334 (2012).

168. Kestler, C. *et al.* Chronophin dimerization is required for proper positioning of its substrate specificity loop. *J. Biol. Chem.* **289,** 3094–3103 (2014).

169. De Angelis, A. A. & Opella, S. J. Bicelle samples for solid-state NMR of membrane proteins. *Nat. Protoc.* **2,** 2332–2338 (2007).

170. Chae, P. S. *et al.* Glucose-Neopentyl Glycol (GNG) amphiphiles for membrane protein study. *Chem. Commun. (Camb).* **23,** 2287–2289 (2013).

171. Cho, K. H., Byrne, B. & Chae, P. S. Hemifluorinated Maltose-Neopentyl Glycol (HF-MNG) Amphiphiles for Membrane Protein Stabilisation. *ChemBioChem* **14,** 452–455 (2013).

172. Chae, P. S. *et al.* Maltose–neopentyl glycol (MNG) amphiphiles for solubilization, stabilization and crystallization of membrane proteins. *Nat. Methods* **7,** 1003–1008 (2010).

173. Jiang, X. *et al.* Evidence for an intermediate conformational state of LacY. *Proc. Natl. Acad. Sci. U. S. A.* **109,** E698-704 (2012).

174. Fernández, C., Hilty, C., Wider, G. & Wüthrich, K. Lipid-protein interactions in DHPC micelles containing the integral membrane protein OmpX investigated by NMR spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **99,** 13533–7 (2002).

175. Gautier, A., Mott, H. R., Bostock, M. J., Kirkpatrick, J. P. & Nietlispach, D. Structure determination of the seven-helix transmembrane receptor sensory rhodopsin II by solution NMR spectroscopy. *Nat. Struct. Mol. Biol.* **17,** 768–774 (2010).

176. Krueger-Koplin, R. D. *et al.* An evaluation of detergents for NMR structural studies of membrane proteins. *J. Biomol. NMR* **28,** 43–57 (2004).

177. Kallick, D., Tessmer, M., Watts, C. & Li, C. The Use of Dodecylphosphocholine Micelles in Solution NMR. *J. Magn. Reson.* **109,** 60–65 (1995).

178. Wolf, E. V., Zeissler, A. & Verhelst, S. H. L. Inhibitor Fingerprinting of Rhomboid Proteases by Activity-Based Protein Profiling Reveals Inhibitor Selectivity and Rhomboid Autoprocessing. *ACS Chem. Biol.* **10,** 2325–2333 (2015).

179. Qureshi, T. & Goto, N. K. Contemporary methods in structure determination of membrane proteins by solution NMR. *Top. Curr. Chem.* **326,** 123–185 (2012).

180. Kielec, J. M., Valentine, K. G., Babu, C. R. & Wand, A. J. Reverse Micelles in Integral Membrane Protein Structural Biology by Solution NMR Spectroscopy. *Structure* **17,** 345–351 (2009).

181. Leitz, A. J., Bayburt, T. H., Barnakov, A. N., Springer, B. A. & Sligar, S. G. Functional reconstitution of β2-adrenergic receptors utilizing self-assembling Nanodisc technology. *Biotechniques* **40,** 601–612 (2006).

182. Park, S. H. *et al.* Nanodiscs versus macrodiscs for NMR of membrane proteins. *Biochem.* **50,** 8983–8985 (2011).

183. Denisov, I. G., Grinkova, Y. V, Lazarides, A. A. & Sligar, S. G. Directed Self-Assembly of Monodisperse Phospholipid Bilayer Nanodiscs with Controlled Size Directed Self-Assembly of Monodisperse Phospholipid Bilayer Nanodiscs with Controlled Size. *J. Am. Chem. Soc* **126,** 3477–3487 (2004).

184. Raschle, T. *et al.* Structural and functional characterization of the integral membrane protein VDAC-1 in lipid bilayer nanodiscs. *J. Am. Chem. Soc.* **131,** 17777–9 (2009).

185. Yu, T. Y., Raschle, T., Hiller, S. & Wagner, G. Solution NMR spectroscopic characterization of human VDAC-2 in detergent micelles and lipid bilayer nanodiscs. *Biochim. Biophys. Acta - Biomembr.* **1818,** 1562–1569 (2012).

186. Sušac, L., Horst, R. & Wüthrich, K. Solution-NMR characterization of outer-membrane protein A from E. coli in lipid bilayer nanodiscs and detergent micelles. *ChemBioChem* **15,** 995–1000 (2014).

187. Tzitzilonis, C., Eichmann, C., Maslennikov, I., Choe, S. & Riek, R. Detergent/Nanodisc Screening for High-Resolution NMR Studies of an Integral Membrane Protein Containing a Cytoplasmic Domain. *PLoS One* **8,** 2–9 (2013).

188. Hagn, F., Etzkorn, M., Raschle, T. & Wagner, G. Optimized Phospholipid Bilayer Nanodiscs Facilitate High- Resolution Structure Determination of Membrane Proteins. *J. Am. Chem. Soc.* **135,** 1919–1925 (2013).

189. Tribet, C., Audebert, R. & Popot, J.-L. Amphipols: Polymers that keep membrane proteins soluble in aqueous solutions. *Proc. Natl. Acad. Sci.* **93,** 15047–15050 (1996).

190. Gohon, Y. *et al.* Bacteriorhodopsin/amphipol complexes: structural and functional properties. *Biophys. J.* **94,** 3523–3537 (2008).

191. Polovinkin, V. *et al.* High-Resolution Structure of a Membrane Protein Transferred from Amphipol to a Lipidic Mesophase. *J. Membr. Biol.* **247,** 997–1004 (2014).

192. Elter, S. *et al.* The Use of Amphipols for NMR Structural Characterization of 7-TM Proteins. *J. Membr. Biol.* **247,** 957–964 (2014).

193. Bazzacco, P. *et al.* Non-ionic homopolymeric amphipols: Application to membrane protein folding, cell-free synthesis, and solution NMR. *Biochemistry* **51,** 120123152316008 (2012).

194. Dahmane, T., Giusti, F., Catoire, L. J. & Popot, J. L. Sulfonated amphipols:

Synthesis, properties, and applications. *Biopolymers* **95,** 811–823 (2011).

195. Delaglio, F. *et al.* Nmrpipe - a Multidimensional Spectral Processing System Based On Unix Pipes. *J. Biomol. NMR* **6,** 277–293 (1995).

196. Johnson, B. A. & Blevins, R. A. NMR View: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4,** 603–614 (1994).

197. Feng, L. *et al.* Structure of a site-2 protease family intramembrane metalloprotease. *Science (80-.).* **318,** 1608–12 (2007).

198. Akiyama, K. *et al.* Roles of the membrane-reentrant β-hairpin-like loop of RseP protease in selective substrate cleavage. *Elife* **4,** 1–18 (2015).

199. Sarker, M. *et al.* Tracking Transitions in Spider Wrapping Silk Conformation and Dynamics by 19F Nuclear Magnetic Resonance Spectroscopy. *Biochemistry* **55,** 3048–3059 (2016).

200. Godoy-Ruiz, R., Guo, C. & Tugarinov, V. Alanine methyl groups as NMR probes of molecular structure and dynamics in high-molecular-weight proteins. *J. Am. Chem. Soc.* **132,** 18340–18350 (2010).

201. Zheng, X., Mueller, G. A., DeRose, E. F. & London, R. E. Solution characterization of [methyl-13C]methionine HIV-1 reverse transcriptase by NMR spectroscopy. *Antiviral Res.* **84,** 205–214 (2009).

202. Fischer, M. *et al.* Synthesis of a 13C-methyl-group-labeled methionine precursor as a useful tool for simplifying protein structural analysis by NMR spectroscopy. *ChemBioChem* **8,** 610–612 (2007).

203. Wiesner, S. & Sprangers, R. Methyl groups as NMR probes for biomolecular interactions. *Curr. Opin. Struct. Biol.* **35,** 60–67 (2015).

204. Shan, X. *et al.* Assignment of 15N, 13C(α), 13C(β), and HN resonances in an 15N,13C,2H labeled 64 kDa trp repressor-operator complex using triple-resonance NMR spectroscopy and 2H-decoupling. *J. Am. Chem. Soc.* **118,** 6570–6579 (1996).

205. Takeuchi, K., Ng, E., Malia, T. J. & Wagner, G. 1-13C amino acid selective labeling in a 2 H15N background for NMR studies of large proteins. *J. Biomol. NMR* **38,** 89–98 (2007).

206. Sibille, N. *et al.* Selective backbone labelling of ILV methyl labelled proteins. *J. Biomol. NMR* **43,** 219–227 (2009).

207. Fukumori, A. & Steiner, H. Substrate recruitment of γ-secretase and mechanism of clinical presenilin mutations revealed by photoaffinity mapping. *EMBO J.* **35,** 1628–1643 (2016).

208. Kornilova, A. Y., Bihel, F., Das, C. & Wolfe, M. S. The initial substrate-binding site of -secretase is located on presenilin near the active site. *Proc. Natl. Acad. Sci.* **102,** 3230–3235 (2005).

209. Li, X. *et al.* Structure of a presenilin family intramembrane aspartate protease. *Nature*

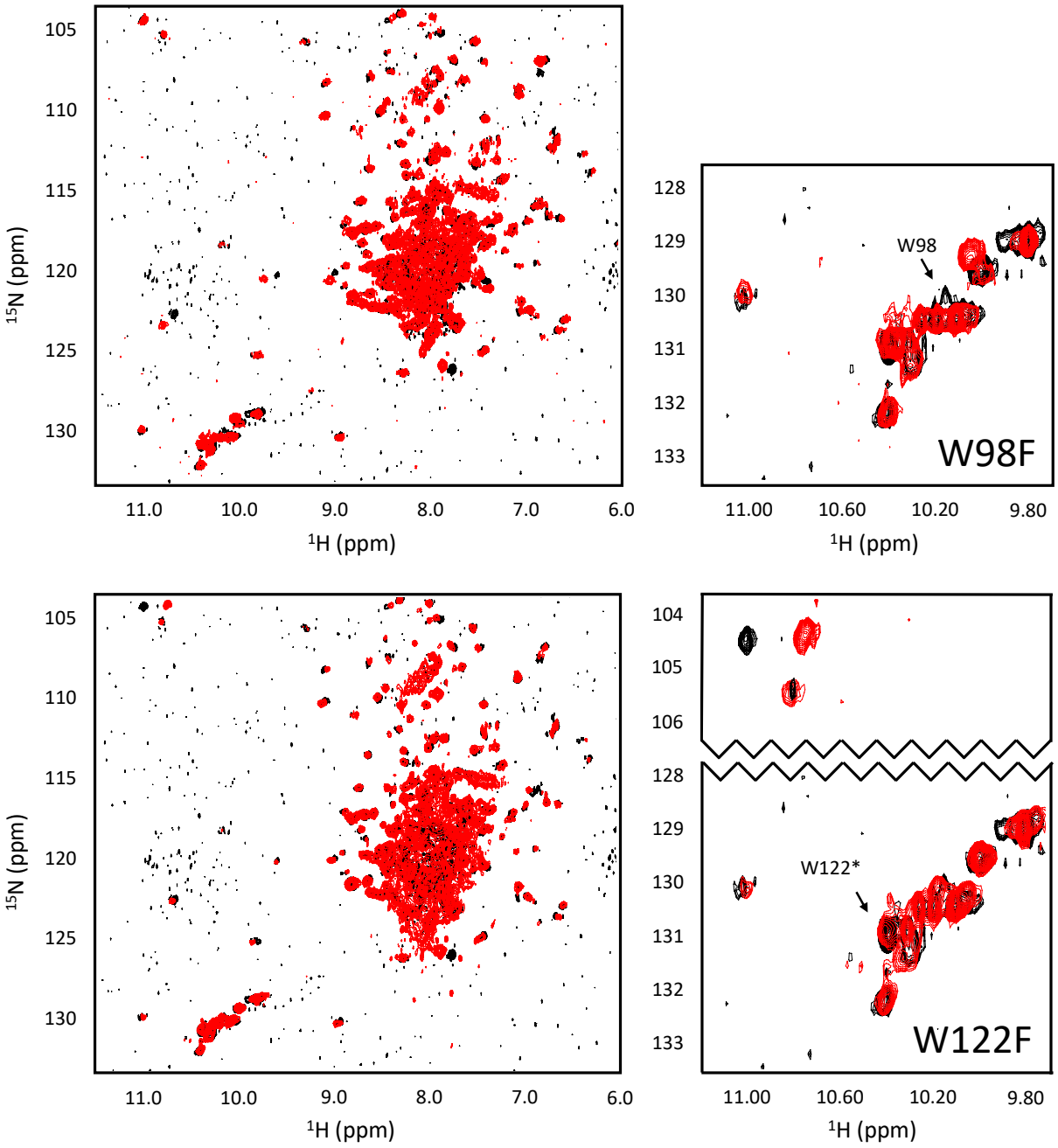
493, 56–61 (2013).

210. Elad, N. *et al.* The dynamic conformational landscape of gamma-secretase. *J. Cell Sci.* **128,** 589–98 (2015).

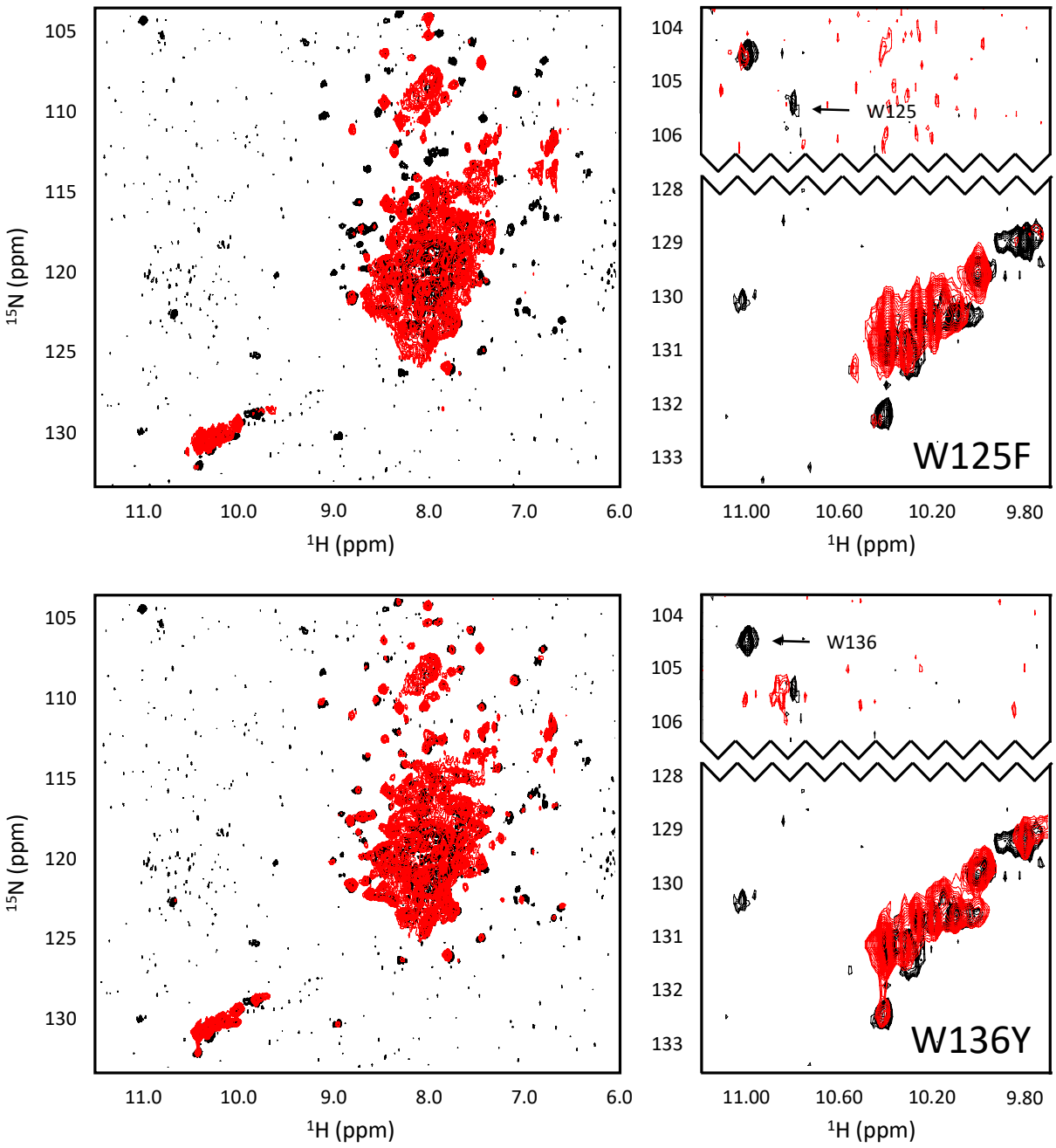
211. Li, Y. *et al.* Structural interactions between inhibitor and substrate docking sites give insight into mechanisms of human PS1 complexes. *Structure* **22,** 125–135 (2014).

212. Uemura, K. *et al.* Substrate docking to γ-secretase allows access of γ-secretase modulators to an allosteric site. *Nat. Commun.* **1,** 130 (2010).

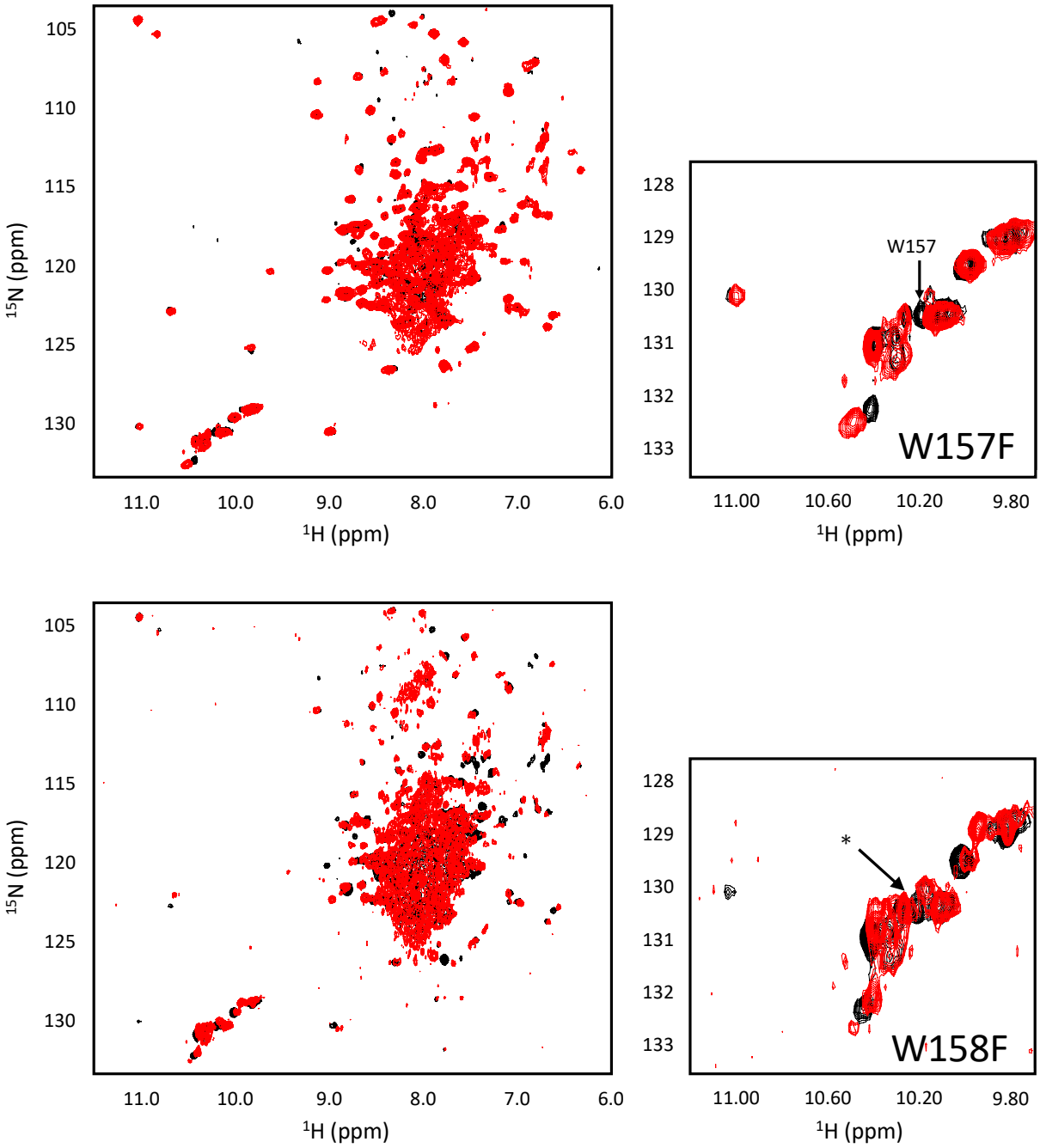
Appendix



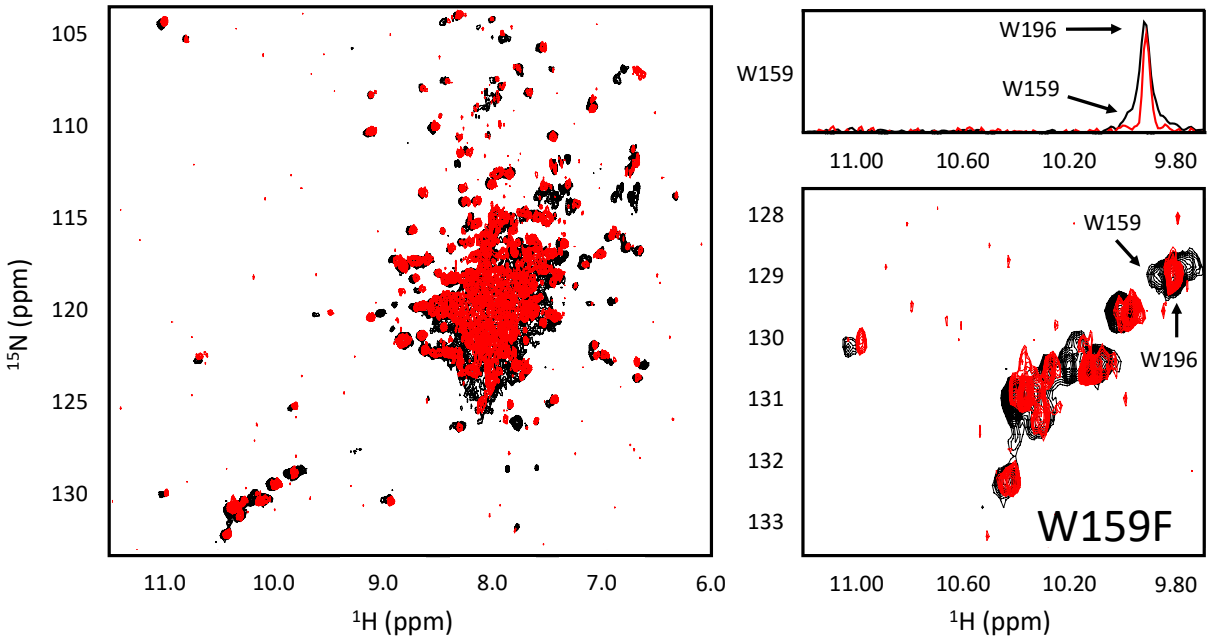
* Peak experiences ~2x reduction in intensity relative to the WT spectrum



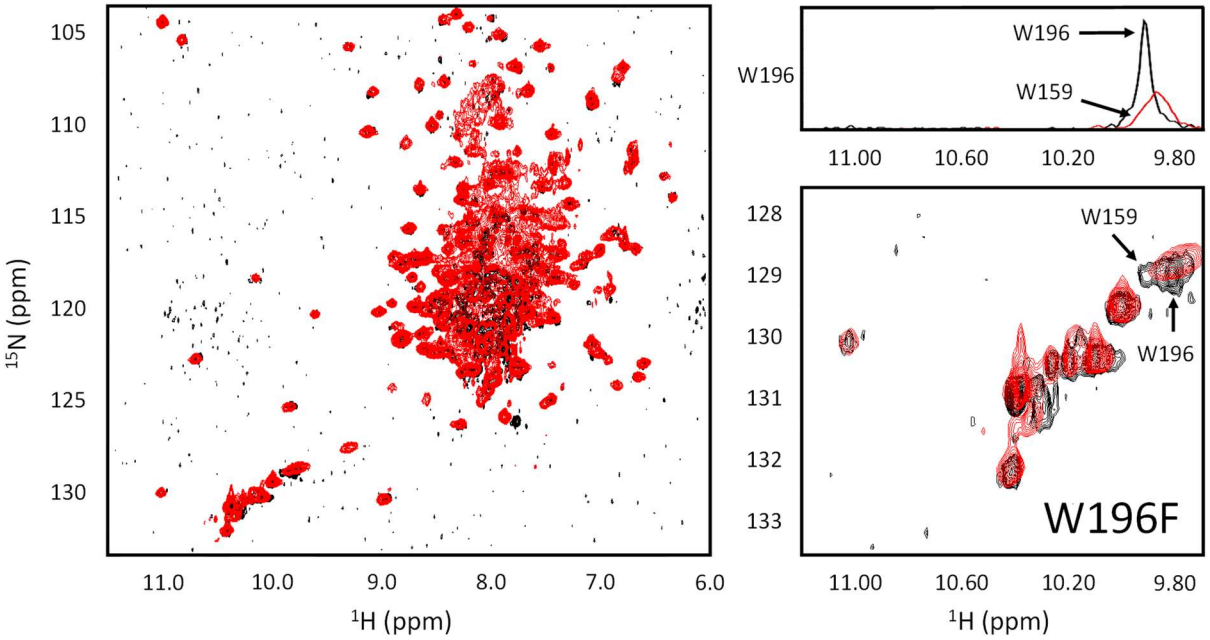
Note that the top right inset has been magnified 3x for the sake of clarity.



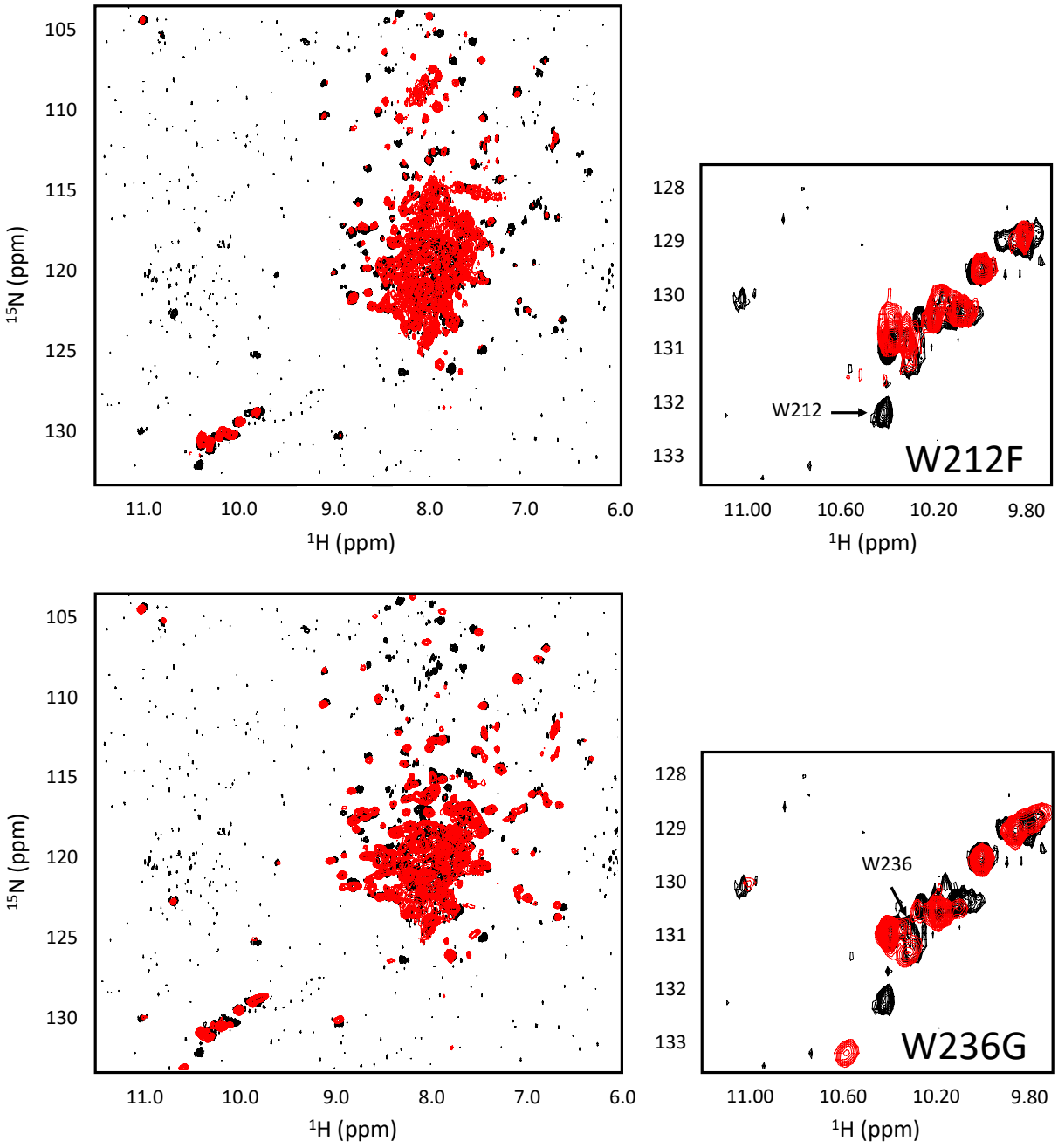
Note that no peak could be assigned in this spectrum. The peak denoted by * loses intensity in W158F spectrum. However, this is accompanied by rearrangements in other Trp side chain peaks. This same peak is eliminated completely in the W157F mutant without the accompanying perturbation of the other Trp side chain peaks. As such, it is more confidently assigned to the later.

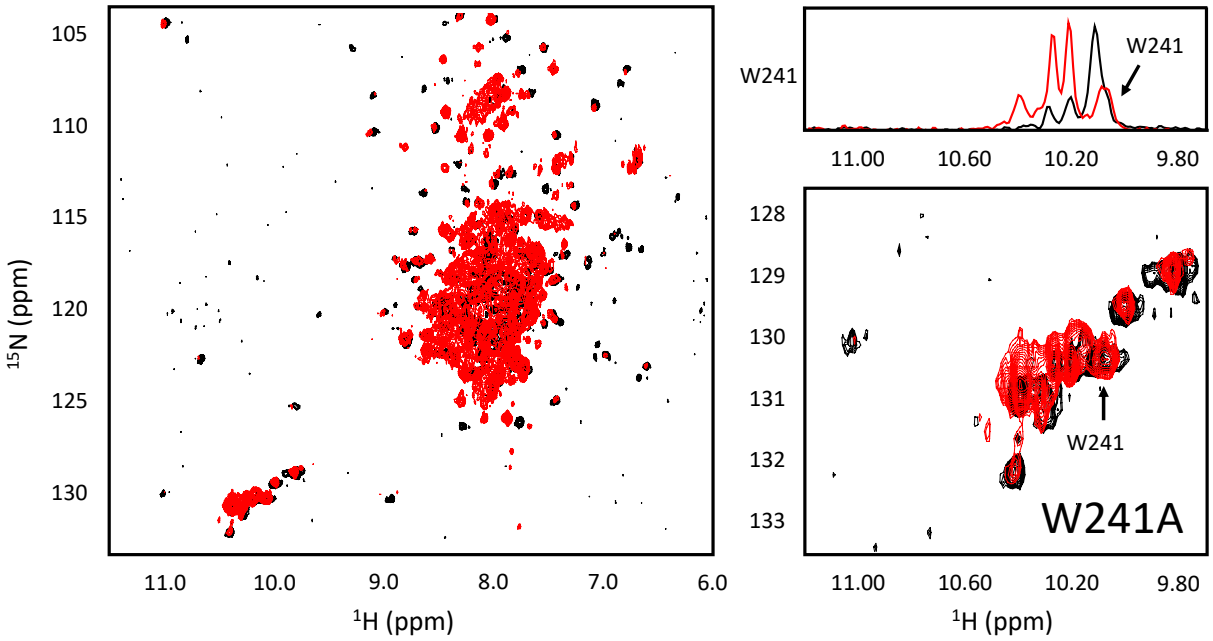


W159 is assigned to the broad peak overlapping the sharper W196 peak. The signal intensity of the W159F spectrum is reduced by 2x in this plot relative to that of the WT for the sake of clarity. A 1D view of the W159/W196 peak is shown to the top right



W196 is assigned to the sharp peak overlapping the broader W159 peak. The signal intensity of the W196F spectrum is reduced by 2x in this plot relative to that of the WT for the sake of clarity. A 1D view of the W159/W196 peak is shown to the top right


105
110
115
120
125
130
^{15}N (ppm)
11.0
10.0
9.0
8.0
7.0
6.0
^{1}H (ppm)
128
129
130
131
132
133
W212
W212F
11.00
10.60
10.20
9.80
^{1}H (ppm)
105
110
115
120
125
130
^{15}N (ppm)
11.0
10.0
9.0
8.0
7.0
6.0
^{1}H (ppm)
128
129
130
131
132
133
W236
W236G
11.00
10.60
10.20
9.80
^{1}H (ppm)



The signal intensity of the W241A spectrum is reduced by 2x in this plot relative to that of the WT for the sake of clarity. A 1D view of the W241 peak region is shown to the top right