uOttawa

Development of ^{19}F NMR methods for the study of GlpG rhomboid protease in detergents and lipid nanoparticle systems

Anwar I. Hassan

A thesis submitted

in partial fulfillment of the requirements

for the Masters of Science degree in Chemistry

Department of Chemistry and Biomolecular Sciences

Faculty of Science

Abstract

Rhomboids are a family of intramembrane serine proteases that cleave transmembrane protein substrates within the lipid membrane. They are involved in a wide range of biological processes, including signal transduction, parasite invasion, bacterial quorum sensing and apoptosis. While previous X-ray crystal structures and functional studies have provided some detailed insights into the mechanism of intramembrane hydrolysis, it is still not clear how the transmembrane substrate can gain access into the active site from the lipid environment. While several modes of action have been suggested, one hypothesis proposes a lateral movement of the fifth transmembrane helix, causing a displacement that would allow transmembrane substrates to enter the rhomboid active site. A powerful method that has the potential to yield insights into rhomboid dynamics is solution NMR; however, the large size of rhomboid protease samples has complicated conventional methods typically used to assess protein structure and dynamics. ^{19}F NMR could allow the study of rhomboid conformational dynamics by providing a simplified spectrum with high sensitivity to changes in local chemical environments. In this thesis various methods of ^{19}F incorporation were evaluated for utility in studying rhomboid conformational dynamics, focusing on the GlpG rhomboid from *E. coli*. First, GlpG samples were prepared with ^{19}F incorporated into tryptophan sidechains, and 1D ^{19}F NMR spectra were acquired. While spectra with decent spectral dispersion were obtained, the assignment process was complicated by low signal-to-noise, and multiple changes in the spectrum introduced by the mutation. Chemoselective labelling of cysteine residues with probes containing a trifluoromethyl group was also investigated and found to give rise to well resolved ^{19}F NMR spectra with promising characteristics. In addition, protocols for incorporation of trifluoromethyl-phenylalanine using unnatural amino acid incorporation at introduced amber codon sites were also explored, since one of the long-term goals

of this work is to study ^{19}F-labelled GlpG in its native lipid environment. For this purpose, some protocol development was also performed to introduce GlpG into lipid nanoparticles using styrene maleic acid co-block polymers. However, low expression yields of trifluoromethyl-phenylalanine-labelled GlpG and the large size of the lipid nanoparticles are not yet compatible with solution NMR. Nonetheless, this thesis lays the groundwork for further development of these samples to allow the future study of conformational exchange of GlpG in native lipid membranes.

Acknowledgements

First and foremost, I give many thanks to my supervisor, Dr. Natalie Goto, for sparking my interest in scientific research, and providing me with guidance during the conduct of my graduate research project. I would like to thank her for her constant and unwavering support, encouragement, and patience during the smooth and rough moments in both my professional and personal life. I would also like to thank her for the countless hours she has spent helping me to prepare presentations and written documents. She has provided me with many opportunities including a rich environment for my scientific development, and it has truly been a pleasure.

I would like to acknowledge the contributions from Daniel Read, Amina Fartas, and Mathias Bell; undergraduate students who assisted in various experiments throughout my research. I thank them for their hard work and support, while contributing to a fun and engaging work environment. I also would like to acknowledge professors Jeffrey Keillor, Roberto Chica, Adam Schuhendler, Eva Hemmer and John Pezacki, for allowing me to use their instruments.

I am also thankful to previous members of the Goto lab, Alexander Foo and Saud Ayed for their helpful discussions and support in the early stages of my research project. I owe a debt of gratitude to Adam Cloutier, he helped with various aspects of the research process, including troubleshooting the many challenges the project presented.

Last but not least, I give my deepest thanks to my family. My mother Saada, who passed away in the early stages of my graduate studies, has provided me with endless amounts of love and support my entire life. Finally, my sisters, Serwa and Sahwan, along with my father, Ismail who have always encouraged me and kept me going through life’s challenges.

Table of Contents

List of abbreviations

5FW	5-fluorotryptophan
7FI	7-fluoroindole
Aβ	Amyloid Beta Peptide
AarA	*Providencia stuartii* rhomboid protease
AcrB	Inner Membrane Component Acriflavine Resistance B
AEBSF	4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride
APP	Amyloid Precursor Protein
BmrA	*Bacilus subtilis* Multidrug Resistance ABC Transporter
bR	Bacteriorhodopsin
BTFA	3-bromo-1,1,1-trifluoroacetone
BTFMA	2-bromo-4-(trifluoromethyl)acetanilide
CBB	Coomassie Brilliant Blue
CD	Circular Dichroism
cNDs	Covalently Circularized Nanodiscs
CTD	Catalytic Transmembrane Domain
cytb5	Cytochrome-b5
CytD	Cytoplasmic Domain
DDM	n-dodecyl β-D-maltoside
DLS	Dynamic Light Scattering
DMSO	Dimethyl Sulfoxide
DPC	n-dodecylphosphocholine
***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*

FPLC	Fast Protein Liquid Chromatography
GlpR	Glycerol-3-phosphate Regulon Repressor
GPCR	G-protein Coupled Receptor
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HSQC	Heteronuclear Single Quantum Coherence
IMM	Inner Mitochondrial Membrane
IMP	Intramembrane Protease
IPTG	Isopropyl B-D-Thiogalactopyranoside
KCNE1	Potassium Voltage-Gated Channel Subfamily E Member 1
kDa	Kilodalton
L1	Loop 1 Region
L5	Loop 5 Region
MjTyrRS	Methanococcus jannaschii suppressor tRNA
MPP	Mitochondrial Processing Peptidase
MSP	Membrane Scaffold Protein
MW	Molecular Weight
Ni-NTA	Nickel-Nitrilotriacetic Acid
NMR	Nuclear Magnetic Resonance
NOE	Nuclear Overhauser Effect
NTR	Nitroreductase
NTR1	Rat Neurotensin Receptor 1
OD	Optical Density
OMM	Outer Mitochondrial Membrane
OmpX	Bacterial Outer Membrane Protein X
PagP	Lipid A Palmitoyltransferase
PARL	Presenilin-associated Rhomboid-like Protease
Pcp1	*Schizosaccharomyces pombe* mitochondrial rhomboid
PDB	Protein Database (RCSB)
PDC	Protein-Detergent Complex
pfRom1	*P. falciparum* Rhomboid 1

pfRom4	*P. falciparum* Rhomboid 4
PINK1	PTEN-Induced Putative Kinase 1
PRE	Paramagnetic Relaxation Enhancement
RHBDL2	Human Rhomboid Related Protein 2
RHBDL4	Human Rhomboid Related Protein 4
Rhom-1	*Drosophila* Rhomboid-1 protease
Rho-7	*Drosophila melanogaster* Mitochondrial Rhomboid 7
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SEC	Size Exclusion Chromatography
SMA	Styrene-Maleic Acid
SMALPs	Styrene-Maleic Acid Lipid Particles
SMAn	Styrene-Maleic Anhydride
Tat	Twin Arginine Translocase
TatA	Twin Arginine Translocase, Component A
TatB	Twin Arginine Translocase, Component B
TatC	Twin Arginine Translocase, Component C
TBS-T	Tris-buffered saline with 0.01% Tween20
TCEP	Tris(2-carboxyethyl)phoshine
TET	Trifluoroethanethiol
tfmF	Trifluoromethyl-L-phenylalanine
TGFα	Transforming Growth Factor Alpha
TIM	Translocase of the Inner Membrane
TOM	Translocase of the Outer Membrane
TM	Transmembrane
TMD	Transmembrane Domain
TMD81	ecGlpG Transmembrane Domain Residues 81-276
TRIS	2-Amino-2-(hydroxymethyl)propane-1,3-diol
tRNA	Transfer ribonucleic acid
UIM	Ubiquitin Interacting Motif
WT	Wild Type

List of Figures

List of Tables

1. Introduction

1.1 Intramembrane proteases

It has been 25 years since the discovery of several families of unusual enzymes called intramembrane proteases (IMP).[1–4] These proteins have the unique ability to cleave other transmembrane proteins within the plane of the phospholipid bilayer. The idea of intramembrane proteolysis was initially met with skepticism, as the need for water in proteolysis seemed incompatible with the hydrophobic environment of the lipid bilayer. However, despite the hydrophobicity of the lipid bilayer, it can be crossed by water molecules.[5] Also, crystal structures of various families of IMPs have revealed bound water molecules near the catalytic machinery.[6,7] Today it is clear that the catalytic residues of IMP are indeed located within the plane of the membrane since the X-ray crystal structure of all the 4 main types of IMPs[8] have been solved. 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. This unique ability of IMPs enables them to participate in a diverse range of biological roles such as the propagation of cell signalling pathways, and the processing and degradation of other integral membrane proteins. Since their initial discovery, four different classes of intramembrane proteases have been identified, namely metallo-protease IMPs, aspartyl IMPs, glutamyl IMPs and serine IMPs. The existence of different IMP families suggests that intramembrane proteolysis is not an odd event but rather one that evolved to perform critical tasks in cells.

1.2 Rhomboid protease

Rhomboids are serine proteases present in all kingdoms of life, except in some species of bacteria and organisms with small genomes,[9] which suggests wide biological significance. The rhomboid serine protease family is the most ubiquitous of IMP family, being widely distributed in almost every biological membrane system in eukaryotic cells (i.e., the endoplasmic reticulum, the Golgi complex, endosomes, lysosomes, mitochondria, and the plasma membrane).[10]

1.2.1 Role in cell signalling

The first rhomboid protease was discovered for its role in epidermal growth factor (EGF) receptor signaling in the fruit fly, *Drosophila melanogaster* (Fig. 1.1A). There is only a single EGF receptor present in *Drosophila*, and its main activating ligand is the protein Spitz, which is bound to the membrane and requires proteolytic cleavage to become a functional signalling molecule.[4,11] The *Drosophila* Rhomboid-1 protease (Rhom-1) was found to be responsible for Spitz cleavage[12] leading to the secretion of the soluble fragment of Spitz from the cell, where it can activate the EGF receptor to regulate cell growth and development.[4,11] Mammalian rhomboids are also involved in EFGR signalling, as was observed by the cleavage of a mammalian homologue, EGF and transforming growth factor alpha (TGFα), by human rhomboids RHBDL2 and RHBDL4, respectively.[13,14]

While the prokaryotic rhomboid AarA from *Providencia stuartii* is also involved in cell signaling, its mechanism is completely different from eukaryotic organisms. The AarA rhomboid removes a short N-terminal inhibitory sequence form the twin-arginine translocase component A (TatA) enabling its complex formation with TatB and TatC.[15–18] Although it is not yet fully

understood, the complex is known to be involved in the release of virulence factors for quorum sensing and biofilm formation.

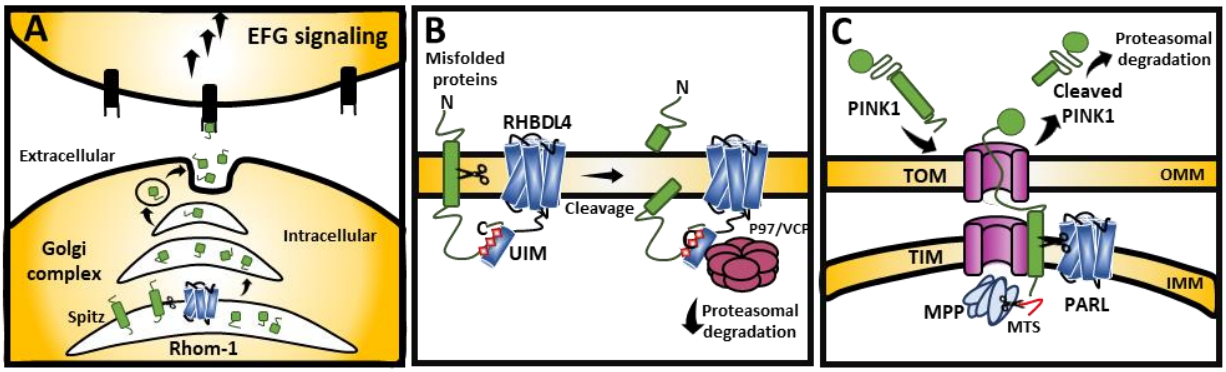

Figure 1.1: The biological roles of various rhomboid proteases. A) The Rhomboid-1 (blue) cleaves the Spitz transmembrane domain (green) for release of the signalling molecule to be exported out of the cell for subsequent epidermal growth factor (EGF) signalling. B) Ubiquitinated, misfolded membrane proteins are recognized by the Ubiquitin Interacting Motifs (UIM) of the human endoplasmic reticulum rhomboid RHBDL4 (blue). Cleavage of these substrates and recruitment of P97/VCP facilitates their excision and degradation via the ERAD pathway. C) Processing of PINK1 (green) by the inner-membrane mitochondrial rhomboid PARL (blue). PINK1 is imported to the IMM through TOM and TIM translocators (purple), the mitochondrial processing protease (MPP) cleaves PINK1 to remove the MTS. Cleavage of PINK1 results in its release from the mitochondrial membrane, facilitating its degradation.

1.2.2 Role in protein degradation

In recent years, the rhomboid protease has been gaining increasing attention for its role in protein homeostasis.[19] For example, the main role of mammalian rhomboid RHBDL4 is in protein homeostasis (Fig 1.1B). RHBDL4 resides in the ER and is involved in the degradation of misfolded proteins associated with the ER-Associated Degradation (ERAD).[20,21] RHBDL4 binds ubiquitinated TM substrates through its ubiquitin interacting motif (UIM) and facilitates their cleavage. In addition, RHBDL4 contains a valosin-binding motif that recruits the p97 AAA+ ATPase, which is required to facilitate the cleavage and subsequent proteasomal degradation of TM substrates.[21,22]

A recent study from the Munter group showed that RHBDL4 can cleave the ectodomain of APP.[23] The cleavage of APP by other protease generates Aβ peptides, which are implicated in Alzheimer's disease. It was shown that the secretion of Aβ peptides was significantly lower in cells expressing RHBDL4.[23] However, it is not yet clear whether APP cleavage by RHBDL4 leads to degradation through ERAD.

1.2.3 Mitochondrial rhomboids

The mammalian mitochondrial rhomboid, presenilin-associated rhomboid-like (PARL), is localized to the inner mitochondrial membrane (IMM) of mammalian cells with the N-terminus in the mitochondrial matrix and the C-terminus facing the inter-membrane space (Fig 1.1C).[24] PARL was initially thought to interact with presenilin, due to results from a yeast-two hybrid screen,[25] leading to its name. The most studied substrate of PARL is the phosphatase and tensin homolog (PTEN)-induced putative kinase 1 (PINK1). The cleavage of PINK1 serves as an indicator or sensor of mitochondrial health.[26] Once PINK1 is imported to the IMM through TOM and TIM translocators, the mitochondrial processing protease cleaves PINK1 to remove the mitochondrial targeting sequence (MTS) before its subsequent cleavage by PARL.[27] PINK1 cleavage mediated by PARL is required to rapidly turn over PINK1 in the cell which conveys the signal that the mitochondrial pool is healthy.[28] Damaged mitochondria, have PINK1 accumulation in the OMM, preventing PARL cleavage, and ultimately leading to apoptosis, with implications for neurodegenerative diseases such as Parkinson's.[29,30] Orthologs of PARL that have been studied include the yeast Pcp1 and the Drosophila Rho-7 protease.[31,32] Both of these proteases are involved in the maintenance of mitochondrial morphology and function.[31,33]

1.2.4 Role in pathogenicity

Rhomboid proteases are also known for their role in the pathogenicity of apicomplexan parasites such as *Taxoplasma gondii*. The cell-surface adhesins are the substrates of the rhomboid protease and are cleaved during the invasion process.[34] The adhesins bind receptors on the host cell to initiate entry into the cell. It was also shown that the inhibition of this cleavage prevents the internalization process.[35] The adhesins in *Plasmodium falciparum*, the causative agent of malaria, are also cleaved by rhomboids. The cleavage-mediated by pfRom1 and pfRom4 rhomboids are required for the release of the mature parasite and to complete the invasion process.[36]

To date, the role of the *E. coli* rhomboid homologue (*ec*GlpG) is not well defined. It was first suggested that *ec*GlpG could potentially play a role in the regulation of glycerol metabolism due to its gene location within the *glp*EGR operon. However, its activity was not required for growth.[37,38] In recent years, *ec*GlpG knockouts in extraintestinal pathogenic strains of *E. coli* (ExPEC) resulted in a decrease of gut colonization and pathogenicity *in vivo*.[39] *Ex vivo* experiments on mucous substrates showed a decrease in viability. Additionally, knockouts of GlpR, a transcriptional repressor for the glycerol-3-phosphate assimilation operon, showed similar results under the same conditions, while viability could be recovered with the addition of glycerol-3-phosphate.[15] Here, results suggest that *ec*GlpG contributes to ExPEC pathogenicity via the activation of glycerol metabolic pathways, although the mechanism through which this occurs, and the substrates involved in this process, have yet to be discovered.

Taken together, the studies presented in this section serve to highlight the ability of rhomboid proteases to participate in a wide range of biological processes, all of which are dependent upon their crucial ability to perform hydrolytic cleavage within the plane of the

phospholipid bilayer. For this reason, there is significant interest in elucidating the structural mechanism by which this unique proteolytic event can occur.

1.3 Rhomboid mechanism of proteolysis

Proteases are enzymes that catalyze the hydrolysis of peptide bonds. To do this, rhomboids use a serine nucleophile to perform an attack on the peptide carbonyl carbon, resulting in the formation of a negatively charged tetrahedral intermediate that is stabilized by an oxyanion hole (Fig 1.2A). A covalent acyl-enzyme intermediate is formed following the collapse of the tetrahedral intermediate which releases the N-terminal cleavage product. The acyl-enzyme intermediate is subsequently hydrolyzed via a second tetrahedral intermediate to release the C-terminal cleavage product and to restore the active serine side chain. Interestingly, most serine proteases utilize a catalytic triad, in which an aspartic residue aligns the imidazole side chain of the histidine, for the abstraction of the hydroxyl proton to activate the serine. In contrast, rhomboid proteases lack this Asp residue and rely on a catalytic Ser-His dyad instead.[40–42]

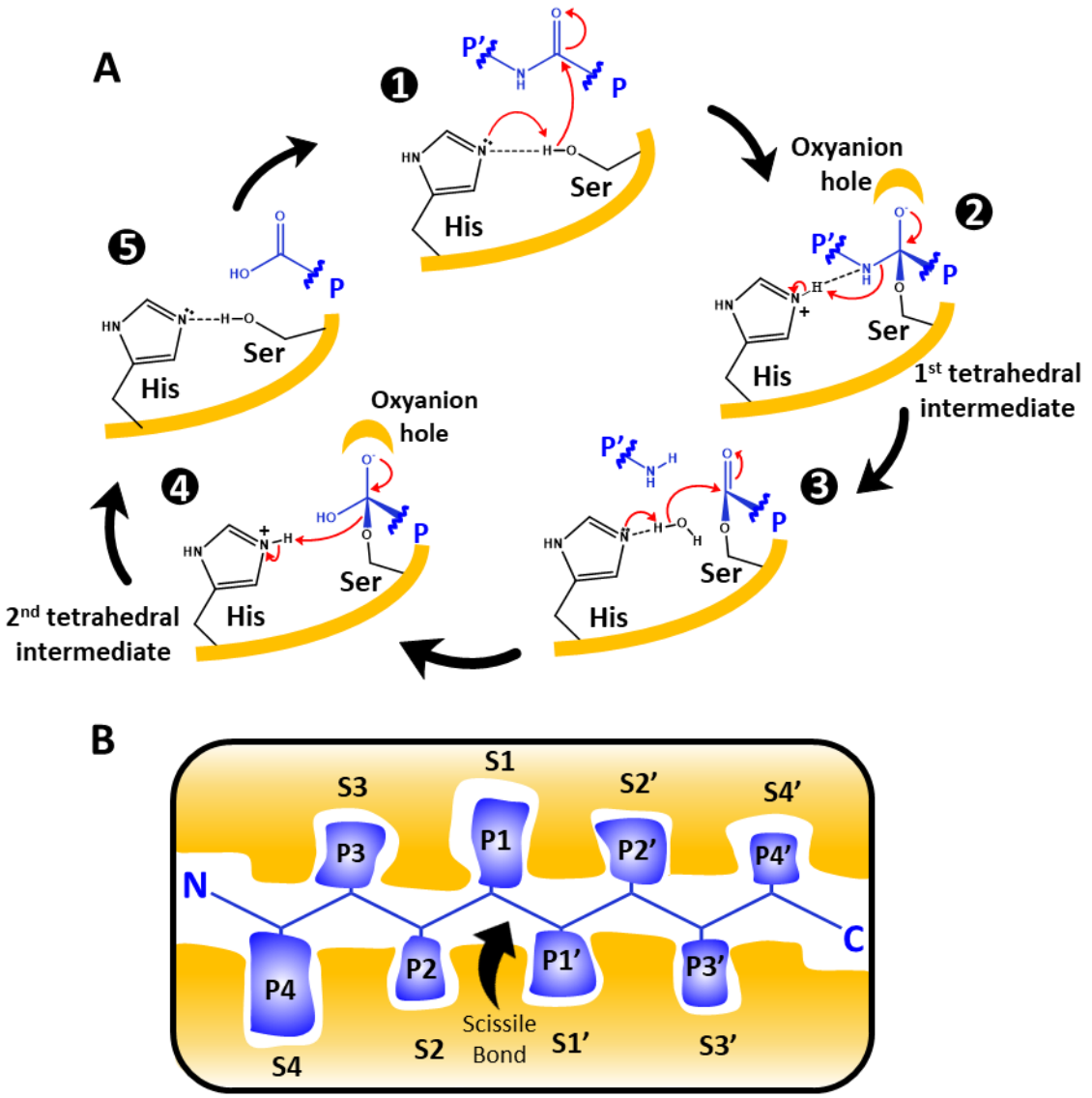

Figure 1.2: Rhomboid mechanism of peptide hydrolysis.[43] A) illustrating the initial nucleophilic attack (1), formation of the 1st tetrahedral intermediate (2) the acyl-enzyme intermediate (3), a 2nd tetrahedral intermediate and subsequent hydrolysis (4) to release the cleavage products and regenerate the rhomboid active site (5). B) Schematic representation of the nomenclature of protease active site and substrate, showing the interactions between substrate side chains (P4-P4') and the substrate binding pockets (S4-S4').

Since proteases typically cleave peptide substrates with specific sequences and that bind to specific sites in the protease structure, a classic nomenclature is often used to describe the site of cleavage and respective protein binding sites. According to this system, illustrated in Figure 1.2B,

cleavage occurs between P1 and P1' positions, with Px or Px' used to describe the residues from the cut site towards the N-terminal or C-terminal sides, respectively. To describe the interactions between the substrate Px/Px' side chains and protease substrate-binding pockets a corresponding Sx/Sx' nomenclature is used.

1.4 Rhomboid structure and function

1.4.1 Rhomboid structure

While *ec*GlpG contains 6 TM helices, other members of the rhomboid family such as the *Providencia stuartii* AarA and *Drosophila* Rhomboid-1 have one additional TM helix.[44] *ec*GlpG is often used as a model system to study rhomboid function due to the wealth of structural information available for this enzyme and the highly conserved nature of the catalytic domain[6,42,45–59] an example of which is shown in Figure 1.3. The active site is positioned in a water-filled cavity located approximately 10 Å below the surface of the cell membrane, bordered by α2, α4, α5 and α6 with the catalytic serine provided by transmembrane helix 4 (α4), and a histidine from α6.[6]

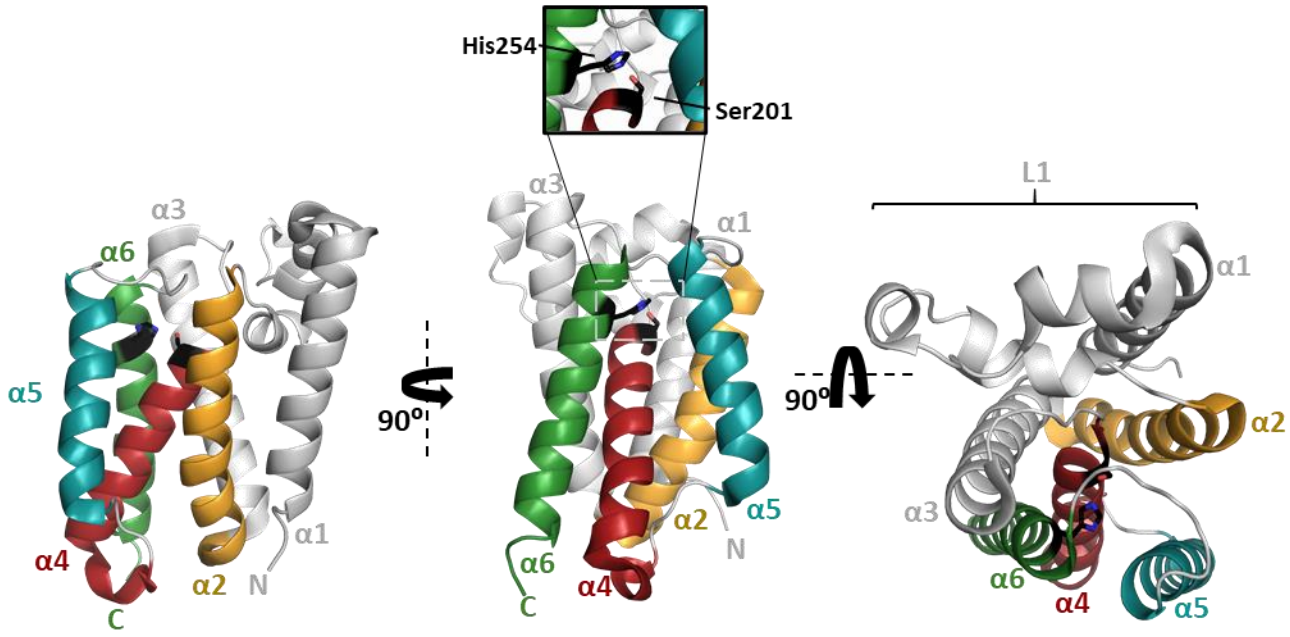

Figure 1.3: X-ray crystal structure of *ec*GlpG (2IC8).[6] The core domain includes six transmembrane helices (labeled α1- α6) forming a hydrophilic cavity. The active site is composed of a catalytic dyad, S201 and H254 is shown in black, on α6 and α4, respectively. A small globular domain also protrudes from the main body of the structure between α1 and α2, called L1.

The active site also contains conserved GXSX and AHXXGXXXG motifs that play a role in maintaining the close packing interactions of α4 and α6, ensuring that catalytic serine and histidine side chains remain within hydrogen bonding distance.[60] An earlier study of *ec*GlpG structure in the presence of isocoumarin inhibitors suggested that the oxyanion hole was formed by the backbone amide protons of L200 and S201, including the side chain hydrogen bonds of H150 and N154.[51] However, more recent studies with a peptide-aldehyde inhibitor which provides a better approximation of the tetrahedral intermediate suggests that this oxyanion hole is formed with only the backbone amide of S201 and shorter hydrogen bonds with H150 and N154 side chains as shown in Figure 1.4.[57] In contrast to other serine protease systems, the nucleophilic attack by the serine occurs at the *si* face of the scissile peptide bond.[44,51]

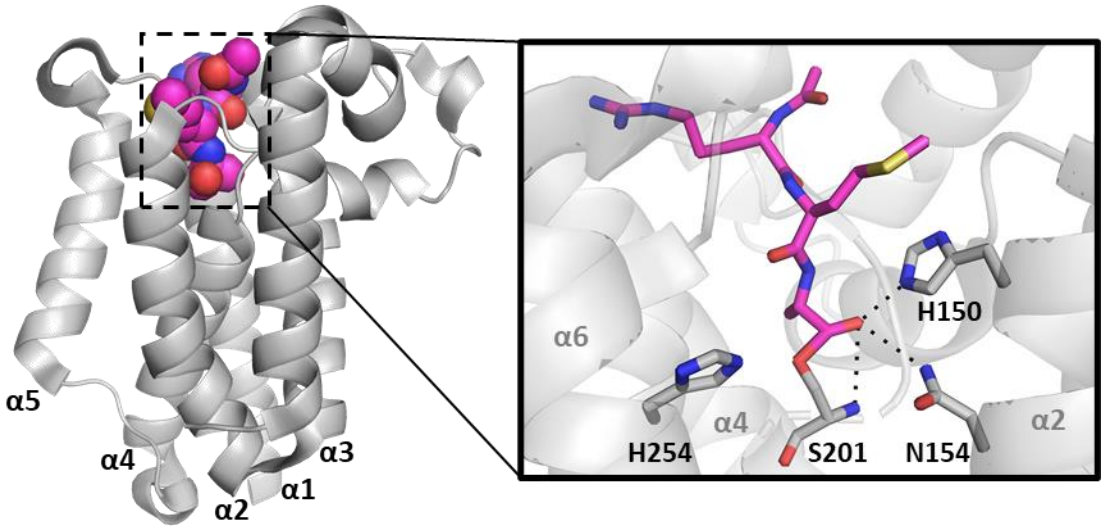

Figure 1.4: X-ray structure of *ec*GlpG in complex with a peptide-aldehyde inhibitor (RMA) illustrating the position of the oxyanion hole formed by hydrogen bonding interactions between the inhibitor oxyanion and the side chain protons from H150 and N154 along with the backbone amide proton of Ser 201 (PDB 5F5G).[57] Note that the *si* face of the peptide bond is facing the viewer.

Most rhomboids including *ec*GlpG possess one or more soluble cytoplasmic domains (CytD). It has been proposed that these cytoplasmic domains regulate rhomboid activity and specificity.[61,62] For example, the mammalian RHBDL2 rhomboid requires its N-terminal CytD to cleave thrombomodulin, a functional anticoagulant endothelial cell receptor.[62] The functional role of the *ec*GlpG CytD has yet to be determined since its removal does not appear to disrupt the active site integrity or be deleterious to its catalytic activity.[63] Interestingly, the *ec*GlpG CytD has been shown to form domain-swapped dimers,[64] and the tertiary structure of *ec*GlpG can be disrupted by the interactions of the CytD with detergent micelles.[65] To avoid complications in structural studies of the catalytic transmembrane domain (CTD), a truncated variant of *ec*GlpG consisting of residues 81-276 (TMD81), which corresponds to the isolated transmembrane domain (TMD), was used in this thesis to conserve the stability of *ec*GlpG catalytic core.

1.4.2 Rhomboid substrate selectivity

The sequence selectivity of *ec*GlpG is broad, consisting of a large hydrophobic residue at the P4 position, a small residue at the P1 position next to the cleavage site, and a hydrophobic residue at the P2' position.[57,66–68] X-ray structures of *ec*GlpG in complex with peptide-based inhibitors have provided some insight into the structural basis underlying this sequence specificity. A recent structure depicts *ec*GlpG in complex with a covalently bound tetrapeptide aldehyde with a sequence derived from the N-terminal cleavage product of the Gurken TM substrate (VRMA), shown in Figure 1.5. The P1-Ala side chain residue of the VRMA peptide points into the large cavity delineated by residues Q189, F197, G198, H141 and S185. Only a single residue, G198 was observed to form a hydrogen bond with the P1-A, via the backbone amide forming a contact with the P1-A carbonyl (Fig. 1.5A).[57] The S1 site shows a large cavity capable of accommodating a large side chain at the P1 position; however, both the peptide bound and apo structures showed the presence of a water molecule within this cavity suggesting that it functions as a water retention site, which would explain the requirement for small side chains at the P1 site.[57,69]

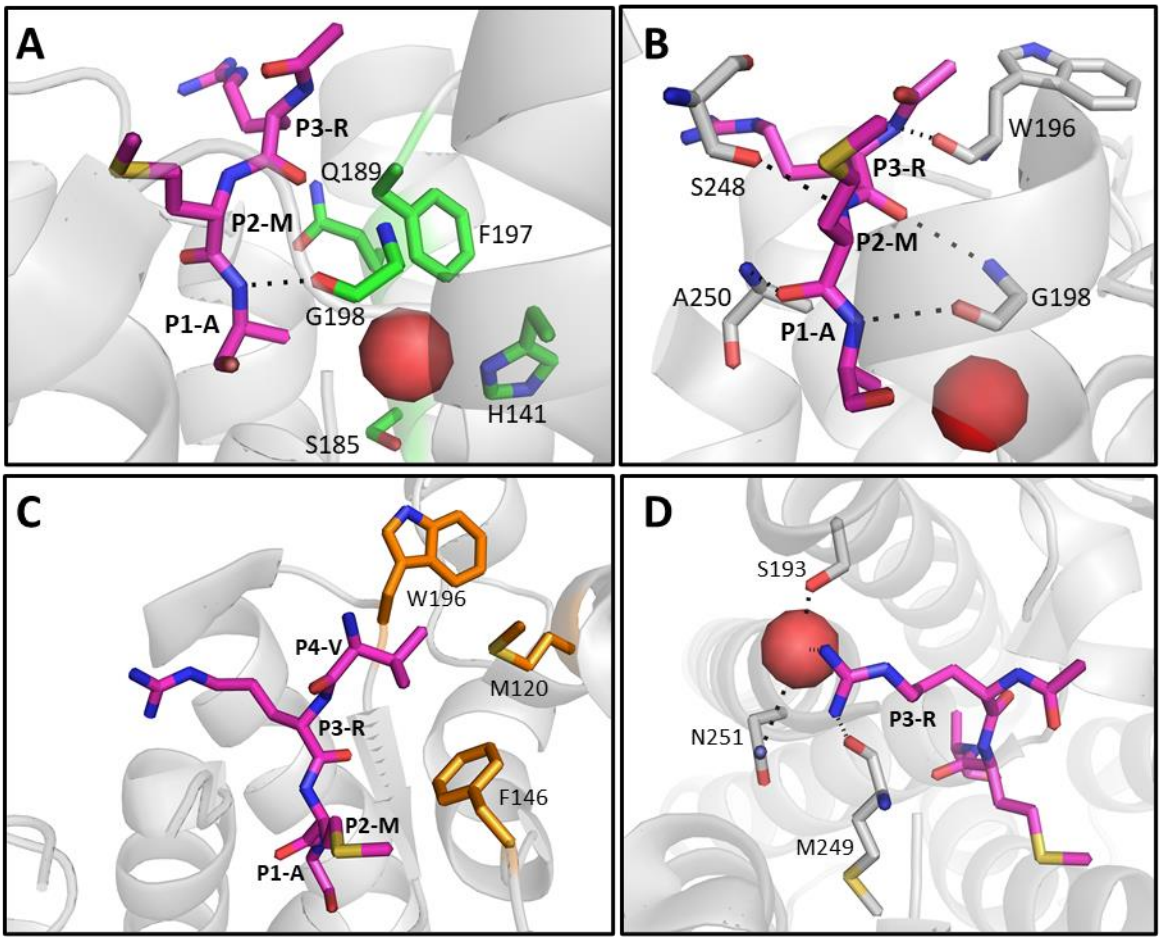

Figure 1.5: X-ray crystal structure of *ec*GlpG in complex with peptide-based aldehyde inhibitor Ac-VRMA-CHO showing the water-retention/P1 pocket (A). Residues forming direct hydrogen bonds with the peptide backbone (B), or the substrate P3 arginine side chain (D). Panel C shows residues (orange) forming a shallow hydrophobic pocket required for the large hydrophobic side chain at the P4 site. Water molecules bound into the P1/water retention site (A) or forming hydrogen bonds with the P3-R side chain (D) are shown as red spheres. (PDB 5F5J).[57]

The *ec*GlpG structure was also able to establish interactions further upstream, with the P2, P3 and P4 residues of the peptide inhibitor. Specifically, this included additional backbone hydrogen bonds with residues Ala250 and Ser248 with P2-Met, and Trp196 and Gly198 with P3-Arg (shown in Figure 1.5B),[57,70] all of which are consistent with earlier structures showing *ec*GlpG in complex with a TatA-derived peptidyl-chloromethylketone inhibitor.[56] Interestingly, the

arginine side chain of the P3 residue also forms additional interactions through a bridging water molecule with the side chains of Ser193 , Ser248 and Asn251 and forms a direct hydrogen bond with the backbone of Met249 (Fig. 1.5D). The mutation of P3-Arg to P3-Ala showed a 10-fold decrease in binding affinity, and since P3-Arg is not found in other TM substrates, its interactions maybe unique to *ec*GlpG.[58] While the P4-Val side chain does not make direct hydrogen bonds with rhomboid, its large side chain fits into a deep hydrophobic pocket delineated by Phe146, Met120 and Trp196 (Fig. 1.5C), which could explain the preference for large hydrophobic side chain at the P4 site.[57] Although kinetic studies with a longer peptide aldehyde inhibitor that includes the Gurken P5 and P6 residues (RKVRMA) have suggested the presence of additional contacts, electron density could not be resolved for residues at these positions.[57] When in complex with a longer peptide substrate that extends beyond the P1 position (RKVRMAAIVFSFP), no electron density could be resolved for residues beyond the P1' position, consistent with the structural and functional studies that indicate that residues beyond the P2' position do not form specific interactions with the active site.[57,66,67,70]

The substrate consensus sequence for *ec*GlpG defined by these interactions is quite broad, allowing it to cleave various physiological substrates.[66] Since the rhomboid is constitutively active, it is likely that there are additional factors that guide rhomboid selectivity.[66]

1.4.3 Substrate gating in rhomboid protease

The wide range of biological functions attributed to the rhomboid superfamily is dependent upon the ability to carry out hydrolytic cleavage of TM segments within the hydrophobic membrane. The rhomboid TMD structure limits exposure of the hydrophilic active site to the

hydrophobic membrane. However, the mechanism through which TM substrates can transition from this hydrophobic environment into the hydrophilic rhomboid active site remains unknown. It was initially thought, that the L1 region might, act as a potential substrate gate. This was supported by its location between α1 and α3, blocking the opening between these helices, with mutations in that region having detrimental effects on activity.[58] However, an X-ray crystal structure for a loss-of-function L1 mutant suggests that this region is highly rigid, and possess characteristics that are not compatible with the role of substrate gating.[45,48] Alternatively, it has been suggested that dynamics in α5 may facilitate substrate entry into the active site. In support of this idea, protease activity measurements on mutants presumed to disrupt the interaction between α5 and α2 have shown an increase in protease activity.[54,71–73] Furthermore, an X-ray crystal structure showed a different *ec*GlpG conformation in which α5 was rotated by ~35° and was essentially tilted away from α2. This was, accompanied by an upward movement of the L5 loop and a slight distortion to α6 that exposed a gap large enough to provide access to TM substrates (Figure 1.6A).[72] Despite, providing a plausible structural model of substrate gating, this 'open' structure has not been seen in any of the other structures captured for *ec*GlpG. This raises questions as to whether the observed tilt in α5 was caused by crystal-packing interactions process rather than reflecting a physiologically relevant open state for substrate entry. Subsequent, X-ray crystal structures with covalently-bound inhibitors depicted a lateral shift in α5 by ~1 Å, giving rise to an alternate model for substrate gating and entry.[51] In 2016, the Urban group provided more evidence for this structural model with a series of X-ray crystal structures that showed α5 displaced laterally from α2 by ~5 Å (Figure 1.6B).[57] While this structure was obtained in complex with a Gurken-derived peptide aldehyde, it could only be captured with the functionally impaired mutant Y205F, and only when solubilized in phospholipid bicelles. Interestingly, the Y205F mutant in detergent micelles, failed to produce

this α5-shift, despite previous studies suggesting that solubilization in detergent micelles would enhance rhomboid gating dynamics.[68] A subsequent study from the Urban group in 2019, with the WT ecGlpG in complex with an extended Gurken derived peptide substrate, showed the entire α5 in a completely disordered state. While the detailed substrate and enzyme interactions involved in α5 disruption are still not clear, it does support the idea that gating by large lateral displacement of α5 can occur.[70]

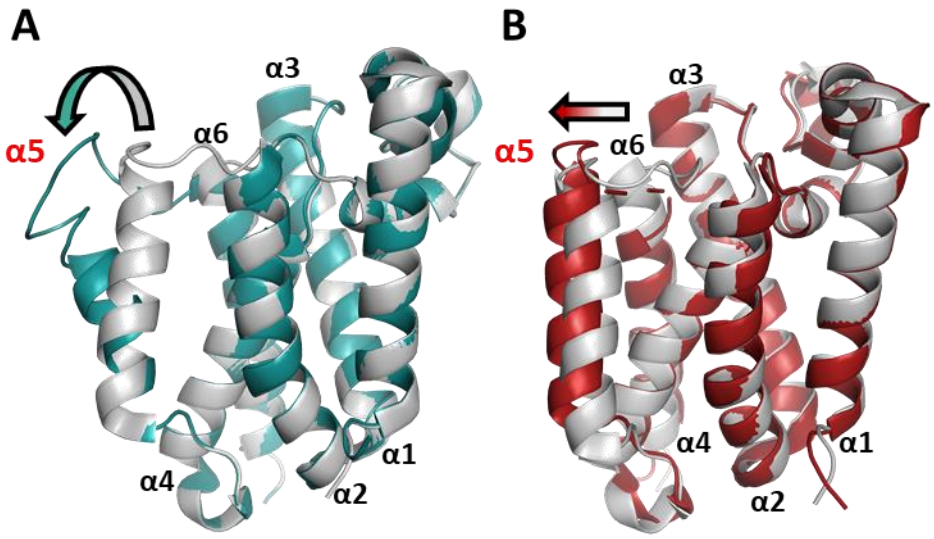

Figure 1.6: *ec*GlpG gate-open conformation showing α5 in a bent conformation (A) or laterally displaced (B) shown in teal and red respectively, superimposed on the closed conformation (gray). (PDB 2IC8[6], 2NRF[45], 5F5D[57]).

Together, the composite studies conducted represent a growing body of evidence that supports the key role of α5 in substrate gating. However, the physiological relevance of this gate-open conformation and how its gating mechanism allows for substrate to enter the active site remains a question. This thesis will explore these questions through the use of solution NMR, the theorical basis of which will be introduced in the next section.

1.5 NMR spectroscopy

This thesis employs solution-state NMR to provide structural and dynamic information on rhomboid proteases. Solution ^{19}F NMR is a technique that can be used to obtain information on the conformational dynamics of proteins, allowing it to complement the structural information obtained from X-ray crystallography studies. In this thesis, specific ^{19}F signals were used to monitor the structure and dynamics of the α5 substrate gate of *ec*GlpG structure.

1.5.1 Basic principles of NMR spectroscopy

Solution nuclear magnetic resonance (NMR) spectroscopy can be used to characterize protein structure and dynamics. This technique involves subjecting NMR-active nuclei to an external magnetic field and radiofrequency pulses of energy. For nuclei that possess spin angular momentum, such as ^{1}H, ^{15}N, ^{13}C, and ^{19}F exposure to the external magnetic field (B_0) induces a dipole in these spins orienting them in a parallel or antiparallel configuration relative to B_0. Absorption of radiofrequency energy can then lead to transitions between these states. The local chemical environment influences the frequency of radiation absorbed and is reported as a chemical shift from a known standard, in units of parts per million (ppm). Although ^{1}H NMR is one of the most popular types of NMR spectroscopy to be performed, the ^{1}H NMR spectrum of larger molecules such as proteins is less informative because of the extensive overlap in peaks arising from hundreds of protons. These densely populated spectra can be resolved using multidimensional NMR experiments, which provide correlations between nuclei that are either directly bonded, or near each other in space. The most common multidimensional NMR experiment used for proteins is the 2D ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) spectrum. Each peak in the ^{1}H-^{15}N HSQC represents a through bond correlation between an amide ^{1}H and ^{15}N atom, with the peak position being given by the chemical shift of the proton in one

dimension and the chemical shift of the directly bonded ^{15}N atom in the second dimension. Each non-proline residue possesses an amide NH group in its backbone, giving rise to a single peak in this type of spectrum.

A useful NMR-active isotope is ^{19}F, which exists in 100% natural abundance and has a large gyromagnetic ratio (83% of that of ^{1}H), allowing it to be highly sensitive in NMR spectroscopy. 1D ^{19}F NMR has been extensively used in studies of protein structure and dynamics, often as a complementary tool to other methods.[74] Since ^{19}F is absent in biological systems, the incorporation of ^{19}F labels at targeted sites in a protein can provide information on the local chemical environment of these labels.[75] Its high sensitivity allows ^{19}F NMR to be used to report on changes related to van der Waals interactions, and other local electrostatic or hydrogen bonding interactions.[76] A particularly useful application of ^{19}F NMR is to probe proteins of known structure, since changes in the ^{19}F local environment can reveal important structural and kinetic information about protein conformational changes. ^{19}F NMR has been used with great success to characterize dynamic conformational processes in the 7-TM helix GPCR protein family.[76–79] Together, these studies have illustrated the potential for solution ^{19}F NMR to be used to probe α5 dynamics for *ec*GlpG.

1.5.2 ^{19}F labelling of proteins

Since ^{19}F does not occur naturally in proteins, its use in structural studies relies on incorporation of extrinsic ^{19}F-labels at structurally interesting locations of the targeted protein. ^{19}F labels can be incorporated into proteins in the following ways: 1) biosynthetically, where the expression media is supplemented with fluorinated amino acid analogs,[80] 2) through chemical

modification of specific side chains, (e.g., thiol groups in cysteine residues)[81] and 3) the site-specific incorporation of fluorinated amino acids at amber suppressor codons through the introduction of engineered tRNAs and aminoacyl tRNA synthetases.[82–84] This thesis employs these three methods for fluorine labelling, and the following sections aim to provide some introduction to their practical use.

1.5.2.1 Biosynthetic incorporation of fluorinated amino acid analogs

Several methods exist for incorporating fluorine probes into proteins. Most commonly, fluorinated amino acids are uniformly incorporated biosynthetically by bacterial protein expression in the presence of the desired fluorinated amino acid analogue. Fluorinated aromatic amino acids can be easily incorporated into a bacterial expression system by including it in the growth medium. Monofluorinated aromatic amino acids represent the most common group of fluorinated analogues used in ^{19}F NMR studies of proteins.[75] In *E. coli*, monofluorinated tryptophan, tyrosine and phenylalanine are commercially available, and can be routinely incorporated using well-established protocols.[75,85,86] Some protocols also recommend eliminating the ability of the cell to endogenously synthesize the amino acid of interest by using a bacterial strain that is auxotrophic for the amino acid of interest.[87] If the target is an aromatic amino acid then it is also possible to add glyphosate to the cell expression media, since this will inhibit the biosynthesis of aromatic amino acids.[85] These types of strategies can allow labelling efficiency to exceed 90%.[80] It should be noted that variation in codon usage can lead to the differences in levels of incorporation at different sites in the protein.[88] High levels of ^{19}F amino acids incorporation reduces the heterogeneity of the population,[80] although heterogeneity will still tend to increase with protein size.[83]

1.5.2.2 Chemical modification of cysteine side chains

The biosynthetic incorporation of fluorinated amino acids can be impractical when different probe locations are desired or when overlapping ^{19}F signals cause challenges with obtaining peak assignments. In these cases, chemoselective ^{19}F labelling of purified proteins may be an attractive alternative. A fluorinated reagent can be used to covalently modify specific protein sidechains. The most common amino acid target for covalent modification is cysteine, due to the high nucleophilicity of the sidechain sulfhydryl group.[75] The conveniently low frequency of cysteine occurrence in natural protein sequences offers specificity and straightforward assignment, while site-directed mutagenesis can allow labels to be introduced at desired sites.[89]

^{19}F labelling agents tend to contain a trifluoromethyl group, as this offers increased signal and narrower line widths due to the presence of three degenerate fluorine atoms that undergo rapid rotation about the methyl symmetry axis, making them particularly attractive in the investigation of large protein complexes and membrane proteins.[90] Commonly employed tags (shown in Figure 1.7) include trifluoroethanethiol,[91] 3-bromo-1,1,1-trifluoroacetone[89,92] and 2-bromo-4-(trifluoromethyl)acetanilide.[78] The high sensitivity of these fluorine tags enables ^{19}F spectra to be obtained at micromolar concentrations with acquisition times facilitating real-time observation of biologically relevant processes.[78] The combination of these characteristics might make ^{19}F tags particularly useful in the analysis of high molecular weight systems, including membrane proteins. Interestingly, these tags are different in terms of chemical shift sensitivity (*i.e.* overall dispersion). Improved chemical shift dispersion has been observed by making use of the trifluoromethylphenyl moiety. The effects of solvent environments on ^{19}F NMR chemical shift have been investigated over a range of polarities and for a wide variety trifluoromethyl reporters.[93] It was proposed that

the electron withdrawing environment of the phenyl group rendered the adjacent trifluoromethyl reporter more sensitive to changes in polarity.

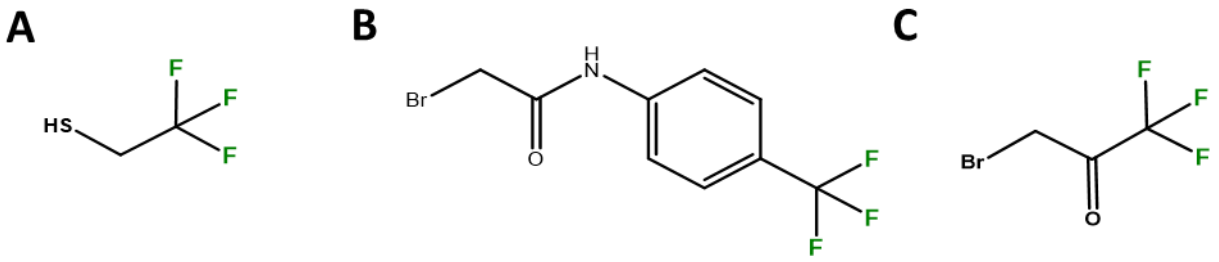

Figure 1.7: Chemical structures for A) trifluoroethanethiol, B) 2-bromo-4-trifluoromethyl acetanilide and C) 3-bromo-1,1,1-trifluoroacetone

1.5.2.3 Site-specific incorporation of fluorinated amino acids

The *in vivo* incorporation of fluorinated unnatural amino acids has also been well-established for over a decade now and is part of the biosynthetic arsenal that has the advantage of introducing the fluorine probe site-specifically. This approach requires an orthogonal tRNA/aminoacyl-tRNA synthetase pair that specifically incorporates the unnatural amino acid into the target protein in response to a TAG amber codon. The unnatural amino acid is typically added to the culture media just prior to induction of protein expression. The ability to incorporate fluorinated amino acids site-specifically was first demonstrated by Furter *et al.* In this study, they employed the first orthogonal tRNA/synthetase pair to incorporate a 4-fluorophenylalanine into single sites of dihydrofolate reductase using the amber suppressor codon in *E. coli*.[94] While this study was a great step for the site-directed incorporation of unnatural amino acids, they were only able to achieve 64-75% incorporation at a single site, which was too low for what is needed for

protein structure characterization. Also 3-7% of all-natural phenylalanine residues were replaced by 4-fluorophenylalanine.[94]

In 2001, Wang *et al.* generated a unique tRNA/aminoacyl-tRNA pair that lead to 99% incorporation at a single site and a minimum 95% incorporation purity.[95–97]A modified *Methanococcus jannaschii* suppressor tRNA/aminoacyl-tRNA synthetase pair (MjTyrRS/$tRNA^{Tyr}$) led to the site-specific incorporation of UAAs at sites with the TAG amber codon. This technology was then used by the Mehl group to engineer a MjTyrRS/$tRNA^{Tyr}$ that could incorporate trifluoromethyl-L-phenylalanine (tfmF) into the TAG amber codon in *E. coli*.[83] This synthetase/tRNA pair was shown to function with high translational efficiency and fidelity, mirroring the results in Wang *et al*.[95]

In this study, tfmF was incorporated at multiple interface sites of the homodimer nitroreductase (NTR). Since NTR has its active site at the dimeric interface, structural changes within the interface due to substrate and inhibitor binding could be assessed by solution ^{19}F NMR. The NTR-tfmF-124 variant was sensitive to conformational changes and the complete reduction of the active site FMN, which caused a upfield shift of the ^{19}F signal. Additionally, they were able to monitor separate peaks representing the relative population of the oxidized and reduced states and observed the change in these populations as protein was oxidized by the stepwise addition of substrates. The ability to characterize conformational changes at the functional interface shows the power of tfmF to monitor structural changes in proteins by ^{19}F NMR.

1.5.3 ^{19}F NMR in the study of conformational exchange

^{19}F NMR spectroscopy can be also used to provide information on dynamic processes such as conformational exchange. ^{19}F probes in locations 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. For example, if the ^{19}F probe undergoes exchange between two different states (Fig. 1.8), with an exchange rate (k_{ex}) that is smaller than the frequency difference ($|\Delta v|$) between these states, two distinct peaks will be observed, with the relative intensity reflecting the population of the two states. However, if the exchange rate is rapid relative to the frequency difference, then the resulting NMR spectrum will yield a single resonance at the population-weighted average chemical shift between both species. Intermediate exchange processes where the exchange rate approximately equals the NMR frequency difference, will give rise to a broad peak of low intensity centred on the weighted average chemical shift of the two states.[98] Slow exchange processes typically have exchange rates on the millisecond to second timescale, whereas fast exchange processes occur on the microsecond range and faster. However, because the NMR frequency difference is directly related to B_o the effect of exchange on the quality of an NMR spectrum can sometimes be affected by using a spectrometer of different field strength.

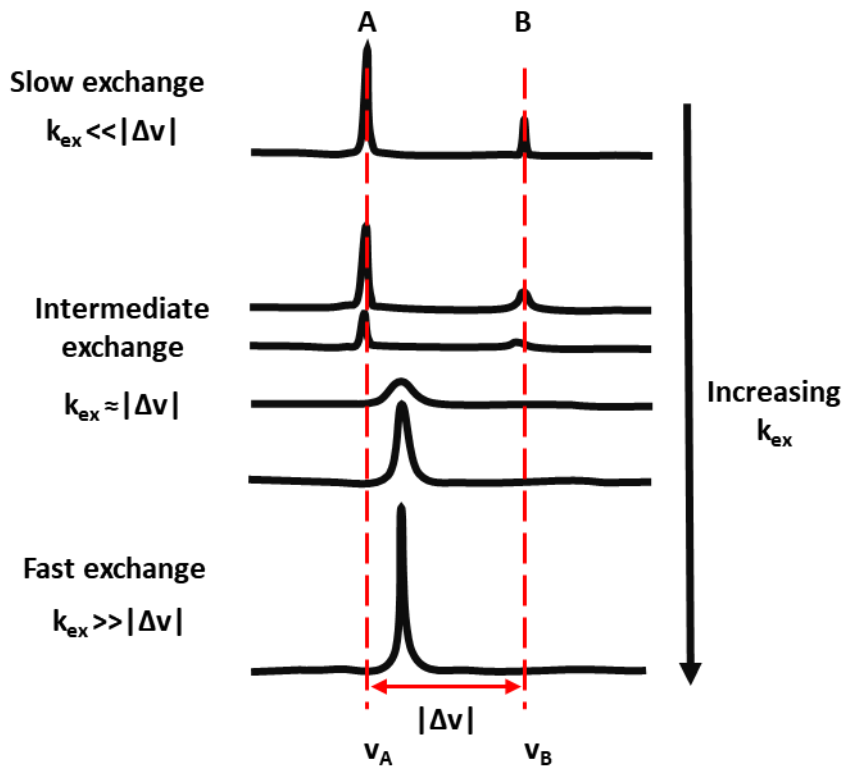

Figure 1.8: The effect of chemical exchange on the NMR spectrum. v_A and v_B represent the frequencies for the two exchanging species (A and B) in units of Hz, while $|\Delta v|$ represents the difference between v_A and v_B. k_{ex} is the rate constant for exchange between the two states.

1.5.4 Solution NMR of membrane proteins

The use of solution NMR to study membrane proteins such as rhomboid proteases, requires solubilization in a membrane mimetic system that maintains the functional fold in liquid state. This thesis employs detergent micelle and lipid nanoparticle systems to perform solution-state ^{19}F NMR studies of *ec*GlpG. Both techniques are described in the following sections.

1.5.4.1 Detergent micelles

The detergent micelle is the most frequently used membrane mimetic solvent for solution NMR studies of membrane proteins due to its relatively small size.[99] The micelles are formed from the spontaneous assembly of amphipathic detergent molecules that organize to form spherically shaped complexes with the polar detergent headgroups facing the aqueous solvent to minimize the exposure of the hydrophobic alkyl chains. Membrane proteins inserted into micelles have the detergent molecules aggregate around the hydrophobic regions of their TM segments to reduce their exposure to the aqueous phase (Fig. 1.9). While the protein detergent complex (PDC) can mimic the general features of the phospholipid bilayer, there are some significant differences. The most common detergents used for solution NMR have a single alkyl chain, causing a high degree of surface curvature in the micelle that is not representative of the native membrane. The PDC structures have higher levels of entropy than that observed in the phospholipid bilayer, which ultimately leads to reduced packing efficiencies that enables water molecules to penetrate the hydrophobic core.[72] Additionally, the packing defects often reduces the lateral pressure applied on the membrane proteins.[100,101] Despite these challenges, detergent micelles have successfully been used to characterize a range of membrane proteins by solution NMR and remains the most common membrane mimetic system for many applications.

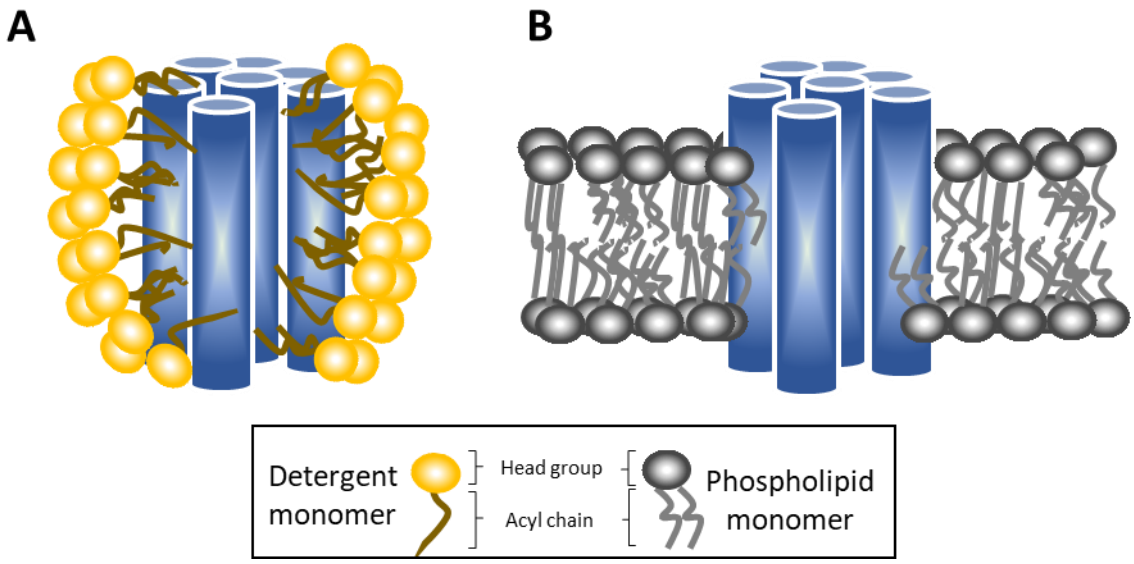

Figure 1.9: Schematic illustration of the solubilization of integral membrane proteins in detergent micelles (A), and phospholipid bilayers (B). Detergent molecules are shown in yellow, and phospholipid molecules are shown in grey.

1.5.4.2 Introduction to the native lipid nanodisc

Over the past 40 years, detergent solubilization has been one of the most common methods used to purify membrane proteins. Detergents have been convenient for solution NMR of membrane proteins due to the relatively small size of the protein-detergent complex that minimizes peak broadening due to transverse relaxation. However, its significant differences from the native lipid membrane can have influence on membrane protein structure and function.

An alternative to detergent micelles that provides a better mimic of the native bilayer environment and may also be compatible with solution NMR is the nanodisc, where lipid bilayer fragments are made soluble by proteins or polymers that cap the hydrophobic edges of the fragment. One type of nanodisc can even extract membrane proteins, along with native lipids

directly from its source membrane (Fig. 1.10), making it possible to obtain more near-native structure, function, and stability.

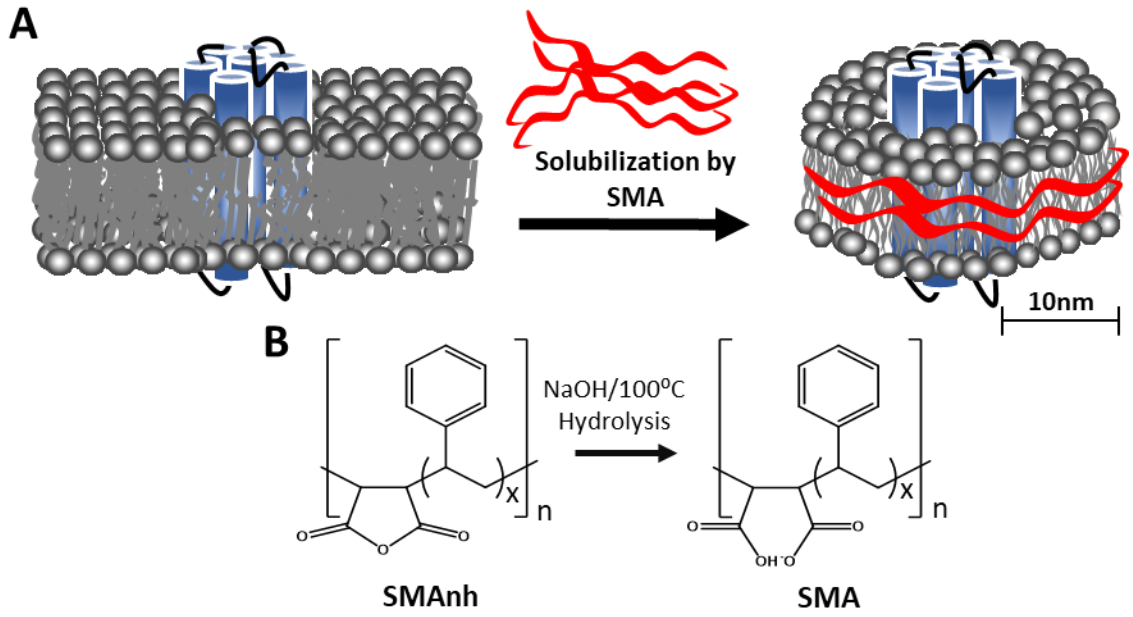

Figure 1.10: Schematic illustration of SMALPs. A) The synthetic styrene–maleic acid (SMA) co-polymer forms discs by encapsulating lipid fragments and protecting the hydrophobic phase from solvent water. SMA can directly extract integral membrane proteins from their native lipid environments. B) The chemical conversion of styrene maleic anhydride co-polymer to the SMA co-polymer.

This can be done with the polymer styrene-maleic acid (SMA), which is the product of the hydrolysis of the precursor styrene–maleic anhydride (SMAn). As shown in Figure 1.10, SMA is an amphipathic copolymer containing distinct ratios of styrene:maleic acid, depending on the polymerization reaction utilized to produce the SMAn precursor. When incubated with lipid membranes, SMA can make discoidal lipid-polymer particles called SMALPs, where the outer ring of the lipid bilayer is protected from the aqueous phase by SMA with styrene:maleic acid

ratios typically in the range of 2:1-3:1.[102] The SMALPs assemble spontaneously in solution, forming a disc with approximately 140 lipid molecules and a diameter of ~10 nm.[102–106]

Interestingly, the SMA co-polymer can directly solubilize intramembrane proteins from their native membrane environments without the use of detergents by isolating patches of the membrane bilayer with native lipids surrounding the membrane proteins. The advantage of this method is that it can provide useful information about associated proteins and a means for identifying the endogenous lipids surrounding membrane proteins.[107–110]

To date, several membrane proteins have been extracted using SMA co-polymer. Among these are bacteriorhodopsin[111], the multidrug efflux protein AcrB[104,112], the KCNE potassium channel 1[113], the P-glycoprotein transporter PgP[114], the KcsA potassium channel[115], penicillin-binding protein PBP2A[116] and even the adenosine A2AR receptor.[117,118] Despite these successes, solution NMR studies of membrane proteins in SMALPs have not yet been demonstrated. This might be due to some of the limitations of SMALPs, for example, SMA-based SMALPs can only form above pH 6.5, since SMA is insoluble below this value. For any type of NMR based on amide proton detection (e.g., ^{1}H-^{15}N HSQC), lower pH conditions are more favourable as this reduces loss of signal from amide solvent exchange. Therefore, NMR experiments with samples solubilized in SMALPs would need to be performed within a narrow pH range of 6.8-7.5.[119] Another limitation of SMA-based SMALPs is that this polymer is a strong chelator of divalent cations (e.g., Mg^{2+} and Ca^{2+}), which if present in high concentrations (e.g., above 5 mM) would lead to SMA destabilization and dissociation from SMALPs. However, the most important limitation of SMALPs is the size heterogeneity commonly observed, making it difficult to control the size of the SMALP, and giving rise to a subset of complexes that might be too large to be

compatible with solution NMR. Despite these limitations there is growing interest in attempting to use this technology, for solution NMR, as it has a unique ability to allow the study of membrane proteins structure and dynamics in a native lipid environment.

1.6 Thesis objectives

While structural and kinetic studies over the last 15 years have provided a detailed understanding of the rhomboid catalytic cycle, many questions remain concerning how rhomboids can gain access to TM substrates from the lipid phase. It has been suggested that dynamics in α5 may facilitate substrate entry into the active site; however, the detailed dynamic changes involved in α5 are still not clear. In this thesis, we developed solution ^{19}F NMR strategies to probe α5 dynamics. Since we were interested in characterizing changes in NMR spectra that arise from α5 dynamics, peptide binding studies were performed, and a TMD81 open-gate mutant was characterized. We also describe the progress made towards using lipid nanoparticles in solution ^{19}F NMR studies of *ec*GlpG. Taken together, results from this thesis lay the groundwork for elucidating *ec*GlpG conformational ensemble and delineating states involved with α5 using ^{19}F solution NMR studies and lipid nanodisc technology.

2. Material and Methods

2.1 Mutagenesis

The C-terminal transmembrane domain consisting of residues 81-276 (TMD81) of *E. coli* GlpG rhomboid in a pET25b vector[87,120,121] was used as the template to make all cysteine and amber codon TAG mutants using the Quick-Change site-directed mutagenesis protocol (Stratagene). All primers (Table 2.1 and 2.2) were synthesized by Sigma Aldrich or Eurofins Genomics. Mutagenesis reactions were performed using the KOD Xtreme Hot Start polymerase kit (Novagen). The reaction mixture contained 1X of Xtreme buffer, 0.4 mM dNTPs, 0.3 µM of the forward and reverse primers, 10 ng of template DNA and 0.02 U/µl of KOD Xtreme Hot Start DNA polymerase in a total reaction volume of 50 µL. PCR tubes were placed in a S1000 Thermal Cycler (Bio-Rad) or an Eppendorf MasterCycler Personal running the program shown in Table 2.3.

PCR products were run on a 0.7% agarose gel containing 1X SYBR Safe (Invitrogen). Electrophoresis was carried out in buffer containing 36 mM boric acid and 10 mM NaOH for 75 minutes at 120 V on an OWL EasyCast (Thermo Scientific) gel electrophoresis system. PCR products were identified by UV light exposure using a ChemiDoc XRS+ imaging system (Bio-Rad). Reactions containing PCR products were then incubated with DPNI (NEB) for 60 minutes at 37 ºC and cleaned using E.Z.N.A. Cycle Pure Kit (Omega Bio-Tek) according to the manufacturer's protocols. Plasmids were sequenced at the StemCore DNA Sequencing facility at the Ottawa Hospital Research Institute.

Table 2.1: DNA primer sequences used to generate cysteine mutants and the W236A gate-open mutant.

Mutants[a]	Sequence[b]
F153C	Forward: 5'-GCTGATGCATATCCTCTGTAACCTGCTCTGGTGGT-3' Reverse: 5'-ACCACCAGAGCAGGTTACAGAGGATATGCATCAGC-3'
Y160C	Forward: 5'-GCTCTGGTGGTGGTGTCTCGGCGGTGC-3' Reverse: 5'-GCACCGCCGAGACACCACCACCAGAG C-3'
W236A	Forward 5'-ATTATCTTTGCGCTGATCGCGATTGTCGCCGGATGGTTTGATTTG-3' Reverse 5'CAAATCAAACCATCCGGCGACAATCGCGATCAGCGCAAAGATAAT-3'

[a]All mutant GlpG sequences are in the pET25b plasmid
[b]Mutated nucleotides and residue codons are underlined

Table 2.2: DNA primer sequences used to generate the amber TAG mutants

Mutants	Sequence[a]
W241TAG	Forward: 5'-CTGGATTGTCGCCGGATAGTTTGATTTGTTTGGGAT-3 Reverse: 5'-ATCCCAAACAAATCAAACTATCCGGCGACAATCCAG-3'
W212TAG	Forward: 5'-GCGCTGATGGGCTACGTCTAGCTACGTGGC-3' Reverse: 5'-GCCACGTAGCTAGACGTAGCCCATCAGCGC-3'

[a]Mutated nucleotides and residue codons are underlined

Table 2.3: PCR thermocycler program to generate cysteine and amber TAG mutants

Step	Temperature (ºC)	Time (s)
1	94	120
2	98	10
3	73	30
4	68	364
5	Go to step 2 (25 cycles)	
6	68	360
7	4	Hold

2.2 Transformation

Plasmids were transformed into chemically competent DH5α and C43 (DE3) cells (Lucigen) for mutagenesis and protein expression, respectively. 10 ng of plasmid DNA was introduced into a 40 µL cell aliquot, mixed and allowed to incubate on ice for 30 minutes. Following incubation, cells were heat-shocked for 45 seconds at 42 ºC and immediately returned to ice for 2 minutes. Cells were then transferred to 900 µL of warm Luria-Bertani broth (Lennox, Bioshop) and placed in a MaxQ 4000 Benchtop Orbital Shaker (Thermo Scientific) for 60 minutes at 37 ºC and 220 rpm. Cells were then spread on an LB agar plate containing 100 µg/mL ampicillin (Bioshop) and incubated for 12-16 hours at 37 ºC.

2.3 Plasmid DNA preparation

A single colony of freshly transformed DH5α was used to inoculate 30 mL of LB media in a 250 mL flask with 100 µg/mL of ampicillin. Cultures were shaken at 220 rpm and 37 ºC for 16-18 hours. Cells were then harvested by centrifugation using a Spinchron 15 Centrifuge (Beckman Coulter) with a C0650 fixed angle rotor at 5000 x *g* and 25 ºC for 10 min. The plasmids were purified from pellets using the E.Z.N.A. Plasmid DNA Mini Kit II (Omega Bio-Tek). Plasmid DNA stocks were stored at -20 ºC.

2.4 Expression of *ec*GlpG

E. coli C43 (DE3) cells were transformed with an expression vector encoding TMD81 (residues 81-276 of *ec*GlpG). This construct was overexpressed using previously published protocols.[63,122] In brief, a transformed colony was inoculated into LB liquid medium with 100

µg/mL of ampicillin and incubated for 16 hours at 37 ºC with shaking at 220 rpm. The culture was then centrifuged and resuspended in 1 L of M9 minimal media (47.9 mM Na_2HPO_4, 22.0 mM KH_2PO_4, 8.5 mM NaCl, 1 mM $MgSO_4$, 0.1% (w/v) D-glucose, 0.1% (w/v) NH_4Cl, 100 uM $CaCl_2$, 0.01X LB, and 100 µg/mL ampicillin). To monitor culture growth the optical density of the culture was measured at 600 nm using an Ultrospec 2100 Pro spectrophotometer (GE healthcare). When the optical density of the culture reached 0.5-0.7, expression was induced with 1 mM of isopropyl-β-D-thiogalactopyranoside (IPTG). When performing the biosynthetic ^{19}F labelling of tryptophan[86] in *ec*GlpG, 60 mg of 7-fluoroindole (Sigma-Aldrich) or 5-flurorotryptophan (Sigma-Aldrich) pre-dissolved in 1 mL dimethyl sulfoxide was added with the IPTG. The temperature was reduced to 16 ºC and the cultures incubated for 16 hours. Cells were harvested by centrifugation using an Avanti J-E centrifuge (Beckham Coulter), for 10 min at 6000 x *g* and 4 °C. Pellets were stored at -20 °C for purification.

2.5 Expression of fluorinated GlpG by amber codon suppression

A plasmid encoding the L-4-trifluoromethylphenylalanine (tfmF) aminoacyl-tRNA synthetase/suppressor tRNA pair[87,120,121] was purchased from Addgene. The plasmid was transformed into C43 (DE3) chemically competent *Escherichia coli* cells (Lucigen) that were selected with tetracycline (25 µg/mL). Transformants were made competent in the presence of tetracycline and then transformed with the pET25b-TMD81 expression vector containing the TAG mutation at designated sites. Cells were grown on plates containing tetracycline (25 µg/mL) and ampicillin (100 µg/mL). Cells transformed with the TMD81-TAG mutants were grown at 37 ºC in LB broth in the presence of ampicillin (100 µg/mL) and tetracycline (25 µg/mL) until an optical density of 0.5-0.6 was reached. Protein expression was induced with the addition of 1 mM IPTG in the presence of 1 mM tfmF (SynQuest) for 4 h at 37 ºC. An aliquot (0.2 mL) was taken out

every hour post induction for western blot analysis. For this purpose, aliquots were pelleted by centrifugation for 1 min at 16,000 x g, then resuspended in 100 µL of water and mixed with 100 µL of 2X SDS loading buffer.

Cell lysates were resolved by 12% SDS-PAGE and transferred onto a polyvinylidene fluoride (PVDF) membrane using the XCell II Blot Module (Invitrogen) with a transfer buffer (12 mM tris base, 96 mM glycine and 0.8% (v/v) methanol). The transfer was performed at 30 V for 70 min. Membranes were blocked for 1 hour in 3% non-fat milk in Tris-buffered saline with 0.01% Tween20 (TBS-T) and then incubated for an hour with a 6X-His-tag horseradish peroxidase-conjugated antibody (Thermofisher). Membranes were washed for 3 X 10 min in TBS-T and then incubated for 5 min with Clarity ECL Western substrate (Bio-rad) and visualized on a ChemiDoc XRS+ imaging system (Bio-Rad).

2.6 Purification of GlpG in detergent micelles

Frozen cell pellets were resuspended in cold 20 mM HEPES buffer (pH 7.4) containing 100 mM NaCl, 5 mM $MgCl_2$, 10% glycerol, and 1X protease inhibitor cocktail (1 mg/mL 4-(2-aminoethyl) benzene sulfonyl fluoride hydrochloride (AEBSF), 2 µg/mL pepstatin, 1.2 µg/mL E-64, 1 µg/mL bestatin, 1.8 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. The lysate was then centrifuged using an Avanti J-E centrifuge (Beckham Coulter) at 50,000 x *g* for 1 hour to isolate the insoluble membrane fraction. The pellet was mechanically resuspended 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 0.1% w/v n-dodecyl β-D-maltoside (DDM). The supernatant containing the detergent-solubilised protein was applied to 2.5 mL nickel NTA agarose affinity resin (Qiagen), equilibrated in 50 mM HEPES buffer (pH 7.4) containing

200 mM NaCl, 5 mM imidazole and 10% w/v glycerol. Impurities were removed through 2 sequential wash steps using a wash buffer (pH 7.4) containing 50 mM HEPES, 50 mM NaCl and 0.1% w/v of *n*-dodecylphosphocholine (DPC). The protein was eluted using the wash buffer supplemented with 250 mM imidazole. The eluted protein was applied to a Superdex-200 10/300 GL size exclusion chromatography column run on an AKTA FPLC (GE Healthcare) equilibrated in GlpG buffer containing 50 mM Tris pH 7.4 50 mM NaCl, 100 µM EDTA, and 0.1% w/v DPC monitored by absorbance at 280 nm. Fractions of the main peak were collected, and the resulting protein concentrations and purity were assessed by SDS-PAGE analysis.

2.7 SDS-PAGE analysis

The SDS-PAGE was done using the Laemmli protocol[123] to assess the purity of protein samples. In brief, 1 part of protein sample was mixed with 1 part of 2X SDS loading dye (125 mM Tris pH 6.8, 20% glycerol (v/v), 4% SDS, 0.01% (v/v) bromophenol blue, 0.02% (v/v) 2-mercaptoethanol). 10 µL of each sample and 2 µL EZ-RUN prestained Rec protein marker (Fisher) was loaded onto a 12% SDS-polyacrylamide gel with 5% stacking gel both containing 12 or 5% (v/v) acrylamide mix (acrylamide: bis-acrylamide, 29:1) (Fisher), 0.1% (w/v) SDS, 0.1% (w/v) ammonium persulfate (Fisher), 0.06% (v/v) TEMED (Fisher) in 1 M Tris pH 8.8, or 1.5 M Tris pH 6.8. Gels were run in a Mini-PROTEAN electrophoresis cell (Bio-Rad) for 70 minutes at 150 V in Tris-glycine running buffer containing 0.3 % (w/v) Tris pH 8.5, 11.44% (w/v) glycine and 0.1% (w/v) SDS.

Gels were stained by Coomassie blue stain solution containing 0.1% (w/v) Coomassie brilliant blue R-250 (CBB), 50% (v/v) methanol and 10% (v/v) acetic acid and heated for 30

seconds in a microwave and left on a rocking platform for 60 minutes. For destaining, gels were first rinsed with double-distilled water, then soaked in destaining solution containing 50% (v/v) methanol and 10% (v/v) acetic acid for 60 minutes followed by storage in ddH_2O before imaging.

2.8 Estimation of protein concentrations

GlpG protein concentrations were determined by SDS-PAGE gels using Coomassie brilliant blue R-250 staining. Standards and dilutions of the GlpG protein sample were prepared using the Laemmli protocol and resolved on a 12% gel as described above. The staining intensity of the CBB was captured with a ChemiDoc XRS+ (Bio-Rad) and analyzed by using ImageJ analysis software. Estimates of protein concentration from stained SDS-PAGE gels were based on duplicate gels. The standard curve was generated from the mean intensity. The total amount of protein in a band was estimated by measuring the intensity across the whole area of the band.

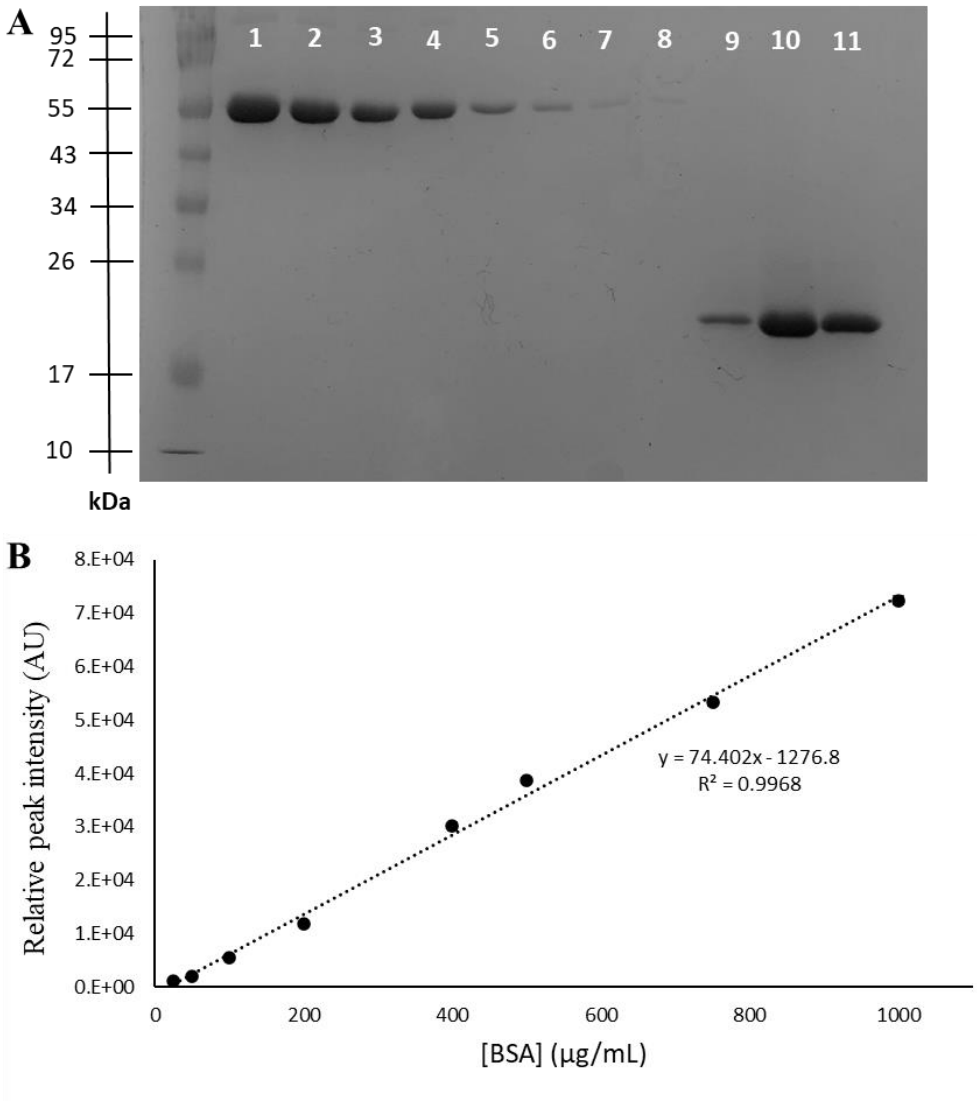

Figure 2.1: Example of protein concentration determination by CBB–stained SDS-PAGE gel (12%). Left lane: Molecular-weight markers. Lanes 1-8: 25, 50, 100, 200, 400, 500, 750 and 1000ug/ml BSA, respectively. Lane 9-11: 6.7µL of WT TMD81, F153C and Y160C mutants, respectively (A). The intensity across the whole area of stained bands of BSA was determined with ImageJ analysis software. A standard curve was generated to calculate the concentration of TMD81 samples, using mean band intensity obtained from duplicate gels (B).

2.9 Fluorine labelling of cysteine residues in GlpG

To label wild type TMD81 or the cysteine mutants with a trifluoromethyl group, 50-100 µM of purified protein in GlpG buffer was first thoroughly reduced with 1 mM tris(2-

carboxyethyl) phosphine (TCEP) for 1 hour at room temperature. A 5-fold molar excess of 3-bromo-1,1,1-trifluoroacetone (BTFA) or 2-bromo-N-(4(trifluoromethyl)phenyl) acetamide (BTFMA) was then added and incubated overnight at 4 ºC. Unreacted labelling reagents were removed by buffer exchange using a Superdex-200 10/300 GL size exclusion column on an AKTA FPLC (GE Healthcare) equilibrated with GlpG buffer.

2.10 ^{19}F NMR spectroscopy

The ^{19}F labelled protein samples were concentrated using a 30,000 kDa MWCO centrifugal filter unit (Amicon) to 50-100 µM in a 375 µL volume containing 10% (v/v) D_2O. The samples were then inserted into a Bruker shaped NMR sample tube of 0.25 mm wall thickness or in a 5 mm Shigemi tube.

NMR experiments were performed at various temperatures using a Bruker 500 MHz Avance spectrometer. 1D ^{19}F NMR spectra were collected using 90 pulse width of around 15 µs. Each 1D ^{19}F NMR experiment was acquired over 3 h using 6800 scans, an acquisition time of 576 ms and a 1s interscan delay time. The ^{19}F chemical shifts were indirectly referenced to the lock signal of the solvent at 4.7 ppm relative to the ^{1}H signal of TMS. The experimental spectra were deconvoluted using Topspin 3.6.2. NMR titrations were carried out by acquiring 1D ^{19}F NMR spectra on a sample of 50-100 µM of TMD81 and cysteine mutants with the addition of increasing amounts of peptides. The peptides were synthesized by CanPeptide and are both N-acetylated and C-amidated. The peptide sequences were derived from the *Drosophila* Gurken rhomboid substrate (RKVRMA) and were diluted in GlpG buffer to produce a ~40 mM stock prior to use.

2.11 SMA copolymer preparation

Styrene maleic acid co-polymer 2000 (SMA) was prepared from styrene maleic anhydride co-polymer 2000 (poly (styrene/maleic anhydride) (2:1), MW 7500, Cray Valley) as described by Lee *et al*.[124] In brief, 25 g of styrene maleic anhydride co-polymer was solubilized in 1.0 M sodium hydroxide and the hydrolysis reaction carried out by refluxing the solution for 2 hours. The solution was cooled to room temperature and the pH reduced to < 5 by the addition of concentrated HCl to precipitate the SMA co-polymer. The precipitate was washed three times with water followed by separation using centrifugation at 11,000 x *g* at RT for 15 minutes. The SMA co-polymer was then resuspended in 0.6 M NaOH. The solution was precipitated and washed again and finally resuspended in 0.6 M NaOH. The pH was then adjusted to pH 8. The polymer was then lyophilized using a freeze-dryer (Labconco). Completion of the reaction was confirmed by Fourier transform infrared (FTIR) spectroscopy of the lyophilized product (Appendix, Figure 7.1A).

2.12 Preparation of protein-free lipid nanoparticles

To prepare lipid vesicles, 20 mg of *E. coli* total lipid extract (Avanti) was dissolved in 1 mL of chloroform. The lipid solution was then dried to a thin film under a gentle argon gas flow. The residual traces of chloroform were removed by drying the pellet under continuous vacuum in a desiccator overnight. The lipid film was rehydrated in lipid nanodisc buffer to give a concentration of 20 mg/mL, the mixture was warmed to 50 ºC and vortexed to obtain complete dissolution.

Small unilamellar vesicles were prepared with a previously established protocol.[125] The lipid suspension was drawn into a 250 µL gas-tight syringe and incubated at 50 ºC for 5 minutes.

The sample was extruded through a 1 µm membrane 30 times using a Mini-Extruder (Avanti). The extruded sample was then extruded again through a 0.1 µm filter. A 5% (w/v) stock of the polymer of interest in lipid nanodisc buffer was gradually added to the extruded lipid to a final concentration of 2.5% (w/v).

2.13 Isolation of GlpG in lipid nanoparticle

This purification was performed as described in Reading *et al.*[126] with a few modifications. Stored cell pellets were thawed and resuspended in 20 mM Tris (pH 7.4) buffer supplemented with 300 mM NaCl and a 1X protease inhibitor cocktail (1 mg/mL AEBSF, 2 µg/mL pepstatin, 1.2 µg/mL E-64, 1 µg/mL bestatin, 1.8 µg/mL phosphoramidon). The cell suspension was passed twice through a French Press (Avestin Emulsifex-B15), at 80 psi. The large insoluble material was removed by centrifugation twice using an Avanti J-E centrifuge with 25.5 rotor (Beckham Coulter) at 20,000 x *g* for 25 minutes at 4 ºC. Membranes were then pelleted by centrifugation at 150,000 x *g* for 1.5 hour at 4 ºC using the Optima™ XL-100K Ultracentrifuge with the type 70 Ti rotor (Beckman Coulter). Membranes were resuspended to 40 mg/mL in cold lipid nanodisc buffer (50 mM Tris, 500 mM NaCl, 10% (v/v) glycerol, pH 7.4) supplemented with the protease inhibitor cocktail. TMD81 was solubilized and purified from SMA 2000 (Cray Valley).

The polymer of interest was added to the suspension at a final concentration of 2.5% (w/v) to solubilize the membranes. The membrane suspension was incubated for 2 hours with gentle agitation at 37 ºC, followed by 1 hour of centrifugation at 100,000 x *g* at 4 ºC. The solubilized membrane suspension was added to a gravity column containing 2.5 mL nickel NTA agarose affinity resin (Qiagen), equilibrated with lipid nanodisc buffer overnight at 4 ºC. Impurities were

removed through 2 sequential wash steps using the lipid nanodisc buffer supplemented with 20mM imidazole. The protein was eluted using 500 mM imidazole and applied to a Superdex-200 10/300 GL size exclusion column using an AKTA FPLC (GE Healthcare) equilibrated in modified lipid nanodisc buffer containing 50 mM Tris pH 7.4 150 mM NaCl and 10% (v/v) glycerol. Fractions of the main peak were collected and stored at 4 ºC. The *ec*GlpG lipid nanodisc purification was assessed by SDS-PAGE analysis.

2.14 Circular dichroism spectroscopy

Circular dichroism (CD) spectra were collected with 5 µM of purified rhomboid in 0.1% (w/v) DPC detergent using GlpG buffer, or in native lipid nanodiscs with lipid nanodisc buffer using a Jasco J-815 spectrometer. Spectra were recorded at 37 ºC, from 200 to 250 nm. Eight accumulations were collected and averaged for each sample using a data pitch of 0.2 nm and a scan rate of 20 nm/min. At least two spectra were acquired for each construct.

2.15 Dynamic light scattering

Dynamic light scattering was performed at 20 ºC using 50 µM of GlpG integrated in lipid nanodisc in 50 mM Tris pH 7.4 150 mM NaCl on a Zetazizer APS (Malvern) with a flat-bottomed 96-well microplate (Greiner, Frikenhausen, Germany). Three accumulations were collected for each sample. Data was recorded in 10 s reads and 11 readings were averaged.

3. Results

3.1.1 Overexpression and purification of uniformly ^{19}F-labelled TMD81

The 11 well distributed tryptophan residues within the *ec*GlpG TMD81 sequence (Fig. 3.1) provide a broad sampling of structural environments, ranging from highly exposed dynamic loops to the more restricted protein core. Incorporating fluorinated probes at the tryptophan sites could therefore be a powerful tool to study *ec*GlpG dynamics. Fluorinated tryptophan amino acids or precursors have been successfully incorporated biosynthetically using well established protocols,[87,120,121] wherein the expression medium is supplemented with fluorinated tryptophan analogs. The monofluorinated aromatics are the most common group of commercially available fluorinated analogues. These include, 5-fluorotryptophan (5FW) and 7-fluoroindole (7FI), both used in this thesis to generate ^{19}F labelled TMD81.

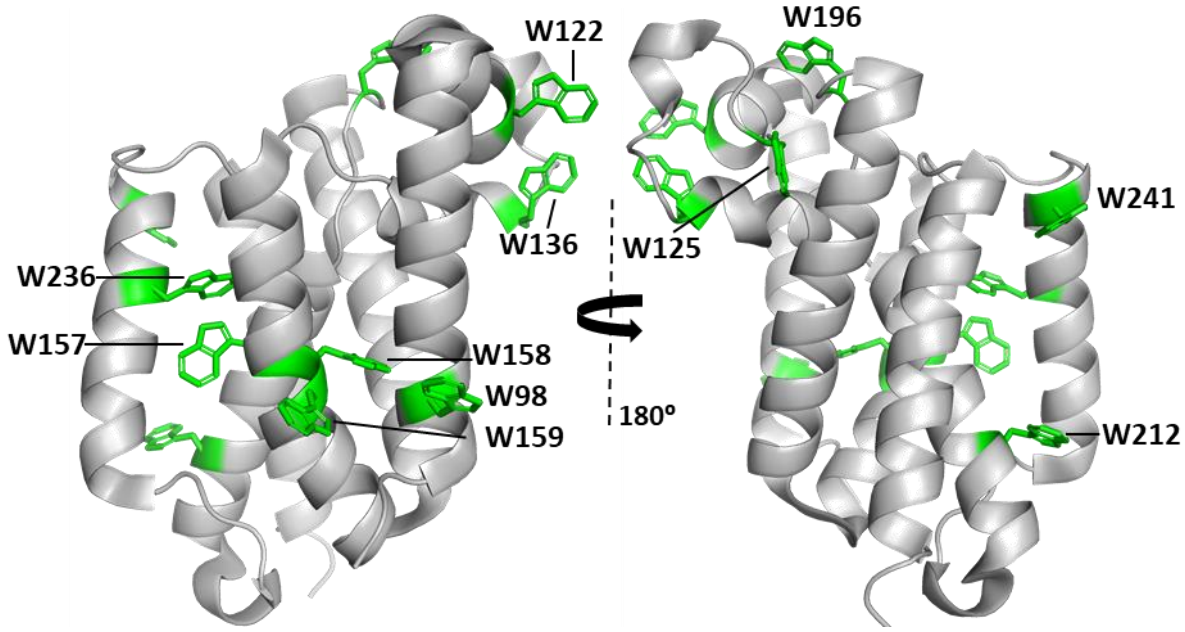

Figure 3.1: Two views of the *ec*GlpG TMD (PDB 2IC8)[6] showing the position of all 11 native tryptophan residues.

^{19}F labelled and unlabelled TMD81 were overexpressed with a C-terminal hexahistidine (6XHis) tag in *E. coli* C43 (DE3) cells. To monitor the expression, samples of the growth medium were taken before and after inducing TMD81 expression. Samples were analyzed using western blotting (Fig. 3.2), where a band corresponding to a 6xHis tag species with a ~24 kDa molecular weight, which corresponds to the expected mass of TMD81, appears 4 hours after expression is induced. Expression levels were significantly lower when the medium was supplemented with 5FW or 7FI.

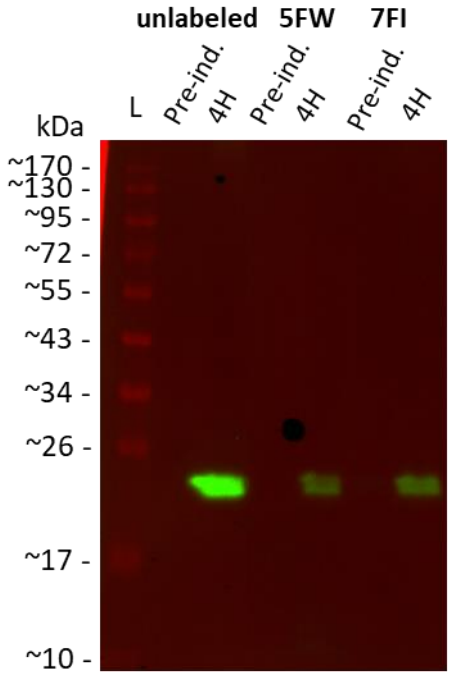

Figure 3.2: Western blot monitoring expression of TMD81 grown in minimal media, with no supplementation, or supplemented with, 5FW or 7FI, with samples taken before (Pre-ind) or 4 hours after induction of expression with IPTG (4H). The ladder is denoted as L.

TMD81 protein samples were purified from cell pellets by nickel affinity and size exclusion chromatography, using previously described protocols.[63,122] Coomassie-stained gels of various samples taken during the purification process are shown in Figure 3.3A, B and C for

unlabelled, 7FI and 5FW labelled TMD81, respectively. An intense band is present in the detergent soluble fraction (DSF) that is absent in the flow-through (FT) from the nickel affinity column. Two sequential wash steps with increasing concentrations of imidazole, were also lacking TMD81 (W1 and W2) while 250 mM of imidazole eluted TMD81 (E).

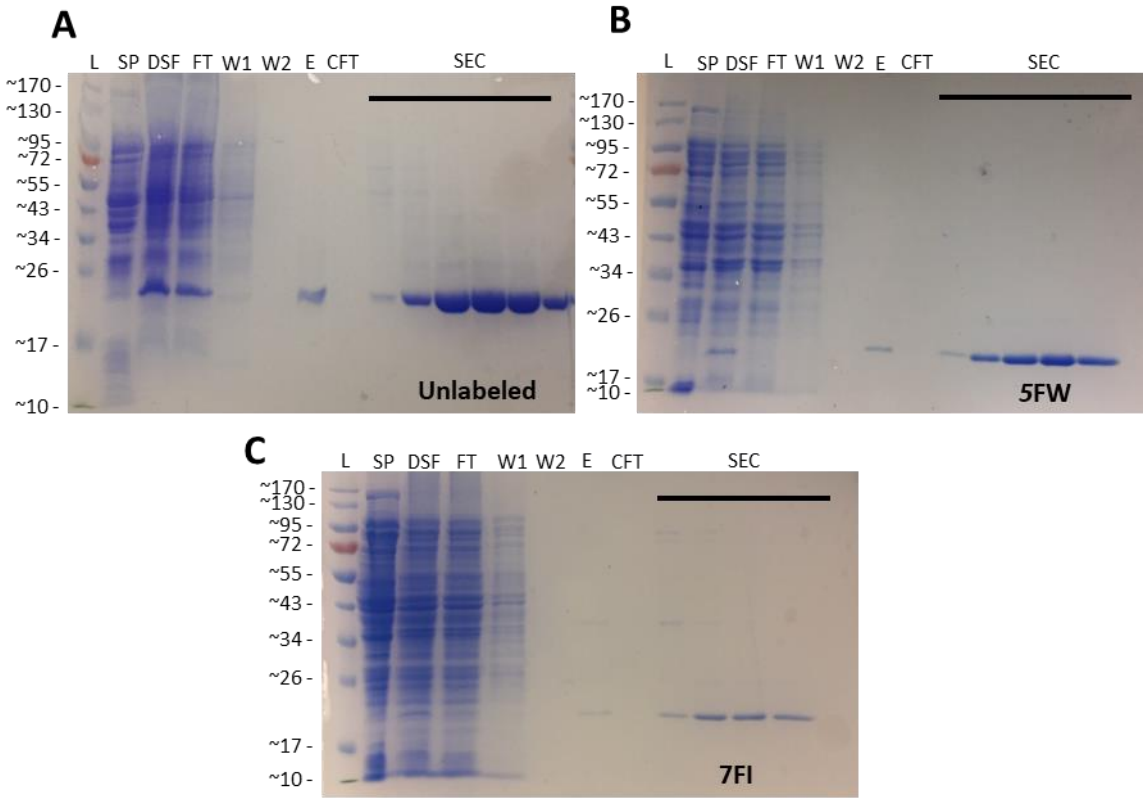

Figure 3.3: Coomassie-stained SDS-PAGE gel of fractions taken during the various stages of the purification of A. unlabeled, B. 5FW- and C.7FI-labelled TMD81. L: molecular weight ladder in kDa, SP: soluble protein fraction, DSF: detergent soluble fraction, FT: flow-through, W1 and W2: contaminants removed during the Ni-NTA wash step, E: elution fraction from Ni-NTA column, CFT: concentrator flow-through, SEC: fractions obtained after size exclusion chromatography containing purified TMD81.

TMD81 samples were then applied to a Superdex-200 10/300 GL size exclusion chromatography column for a final purification step. Elution was monitored by absorbance at 280 nm and found to give rise to a profile with TMD81 eluting in a peak at approximately 15 mL (Fig.

3.4), as has been previously observed.[122] In most cases, TMD81 eluted as asymmetrical peak with contaminants eluting from 11-13 mL. To maximize sample purity only fractions corresponding to the major peak (14-16 mL) were pooled.

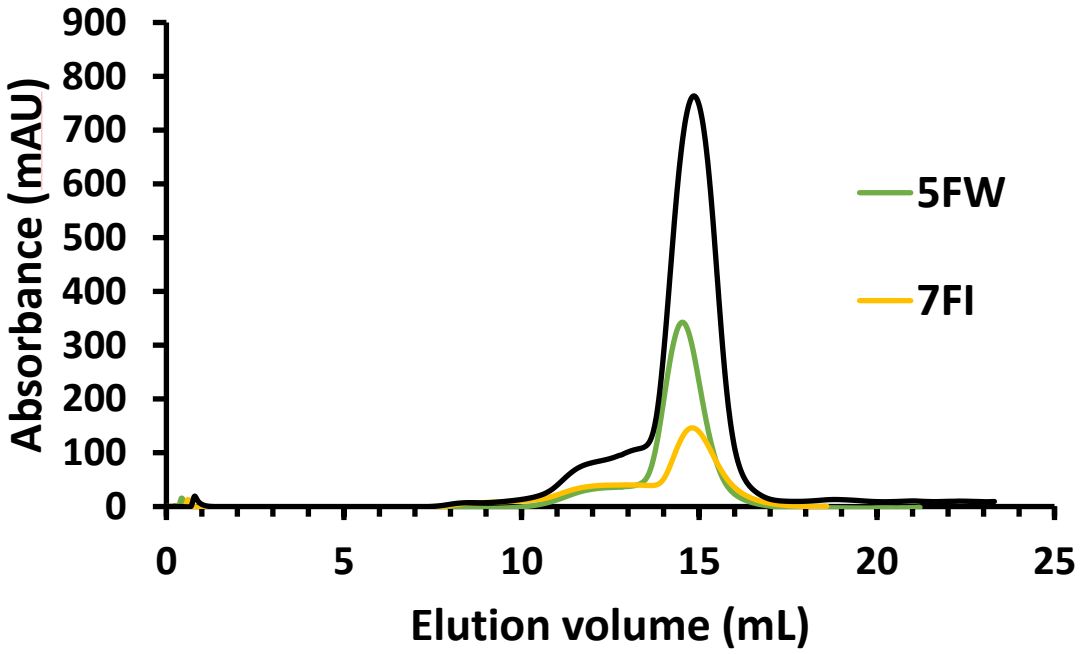

Figure 3.4: Representative size exclusion chromatography profiles of nickel-affinity purified TMD81 that is unlabelled (black), labelled with 5FW (green), or 7FI (yellow), monitored by absorbance at 280 nm.

Both 5FW- and 7FI-labelled samples had significantly lower yields relative to the unlabeled species. This is likely due to the well-known toxicity issues associated with fluorinated precursors and their inhibitory effects on growth.[86,120,121] This is supported by the lower cell mass obtained from 1 L growth cultures containing the fluorinated analogues with ~1.5 g (wet weight) compared to ~2.7 g from growths with untreated media.

3.1.2 NMR spectroscopy of uniformly ^{19}F labelled TMD81

To assess the suitability of the fluorinated tryptophan sidechains for NMR studies of *ec*GlpG, 1D ^{19}F NMR spectra were acquired for TMD81 labelled with 5FW or 7FI. As shown in Figure 3.5, 5FW-labelled TMD81 yielded a spectrum with low signal-to-noise, and poorly dispersed peaks. It was not possible to discern 11 distinct peaks as would be expected from the number of tryptophan residues in the sequence, although there was one higher intensity peak ~125 ppm (indicated with the asterisk in Fig. 3.5). Overall, the extensive spectral overlap and low peak intensities indicate that 5FW-labelled TMD81 is not amenable to ^{19}F resonance assignment and protein dynamics studies.

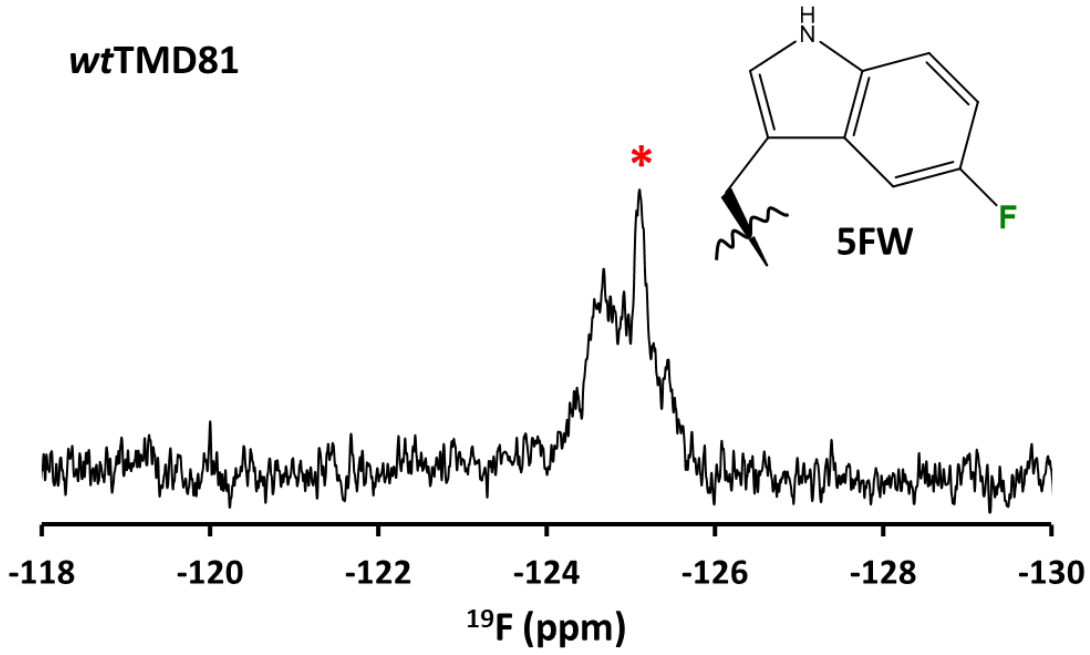

Figure 3.5: ^{19}F NMR spectrum of 5FW-labelled TMD81, recorded at 45 ºC, pH 7.4 with a TMD81 concentration of 60 µM. The chemical structure of 5-fluorotryptophan is shown.

Despite the high toxicity levels and low expression yields associated with fluorinated precursors, there was a sufficient amount of 7FI-labelled TMD81 to allow ^{19}F NMR spectra to be

acquired. As shown in Figure 3.6, peak number, and chemical shift dispersion (6 ppm) was larger for this sample relative to the 5FW-TMD81 spectrum. Although some peak broadening and overlap made it difficult to identify exactly the number of peaks displayed in the spectrum, some peaks were well-resolved, raising the possibility that some peak assignments might be made. For this purpose, ^{19}F spectra were collected for tryptophan knockout mutants W236G and W212F, both of which were chosen for their proximity to the gating region. Although there are significant differences between the wild-type and mutant spectra, it is not possible to straightforwardly identify the peaks associated with either W236 or W212. For example, the spectrum for W236G (Fig. 3.6A) showed loss of a peak at -135 ppm (labelled peak b) while a new peak (a) appeared around -133 ppm and the central envelope of peaks also changed, becoming more intense with small changes in chemical shift. It is possible that the new peak arises from W157 since it is in close proximity to W236 in the WT protein and its local chemical environment would be changed by the substitution of Trp for Gly. However, due to these changes, it was not possible to determine if peak b in the WT spectrum corresponds to W236. Further complicating assignment is the spectrum of W212F showing a loss of intensity in the vicinity of peak b for this mutant. In addition, in the WT spectrum there may be a very broad peak around -136ppm (peak c), although the low signal to noise makes it difficult to be certain. There is a peak at this position in the W236G spectrum, while in the W212F spectrum this peak is either of very low intensity or absent. Ultimately the low signal-to-noise and multiple changes in the spectrum caused by tryptophan knockout mutants indicate that it is not possible to straightforwardly assign peaks in the 7FI-labelled spectrum.

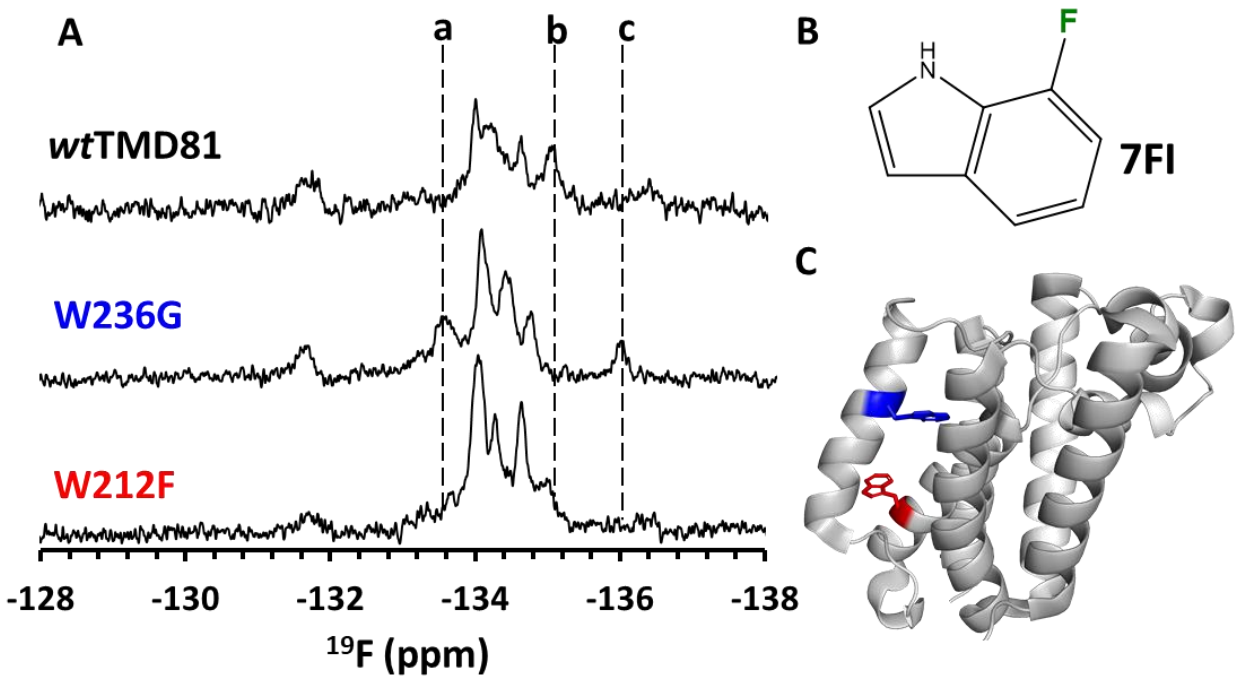

Figure 3.6: ^{19}F NMR spectra of TMD81 and mutants uniformly labelled with 7-fluroindole. Spectra were recorded at 45 ºC, pH 7.4 with TMD81 concentrations of 60 µM (W236G), 70 µM (W212F) and 50 µM (*wt*TMD81). b) Chemical structure of 7-fluoroindole. C) Structure of *ec*GlpG (PDB 2IC8) showing the position of W236 (blue) and W212 (red).

3.2 Studies of the chemoselective ^{19}F-labelling of TMD81

The site specific ^{19}F labelling of proteins, post-expression, is an attractive strategy, when biosynthetic labelling yields complications with spectral overlap and challenges associated with obtaining assignments through mutagenesis. The limited number of cysteine residues within *ec*GlpG sequence makes it a great target for chemical conjugation with small ^{19}F-containing molecules. *ec*GlpG has a single native cysteine residue (C104) buried within the hydrophobic core, which may protect it from derivatization by sulfhydryl-reactive agents, while site directed mutagenesis can introduce other cysteine residues at desirable locations around the gating region. These can then by derivatized with the trifluoromethyl-containing probes (Fig. 3.7) 3-bromo-1,1,1-trifluoroacetone (BTFA) or 2-bromo-4'-(trifluoromethyl) acetanilide (BTFMA). These have

higher sensitivity than the fluoro-indole probes used for tryptophan labelling as they contain 3 fluorine atoms versus the single fluorine atom in 5FW or 5FI. As these fluorine atoms are part of a methyl group, the chemical shift anisotropy will also be smaller compared to the indole-based probes.[127]

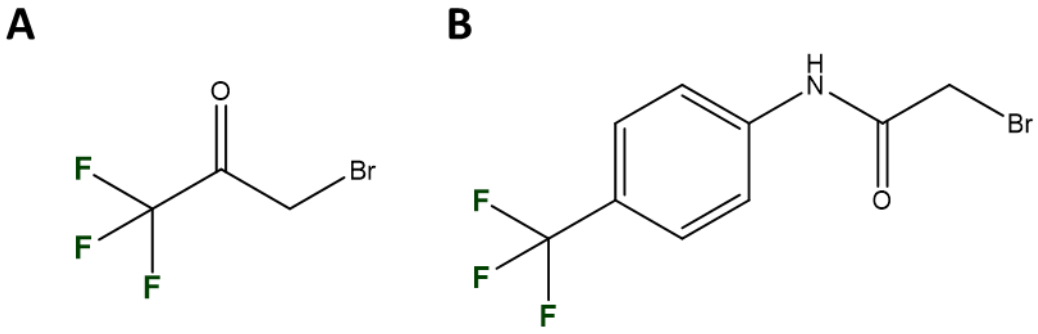

Figure 3.7: Chemical structure for A) 3-bromo-1,1,1-trifluoroacetone (BTFA) and B) 2-bromo-4'-(trifluoromethyl) acetanilide (BTFMA).

3.2.1 Solution state ^{19}F NMR of BTFA-labelled TMD81

^{19}F spectra were collected for *wt*TMD81 labelled with BTFA to determine whether the native core cysteine residue was accessible to reaction with this probe (Fig. 3.8). This spectrum was then compared to that of the spectrum of our GlpG buffer with and without the BTFA label free in solution. A single distinct peak at -84 ppm was observed in the vicinity of the expected chemical shift for free BTFA[77,128] (-83.3 ppm) but was much broader, confirming that this site is accessible enough to react with the label. The broadness of this peak indicates that the ^{19}F probe is now associated with a large molecular weight complex, as expected for labeled TMD81.

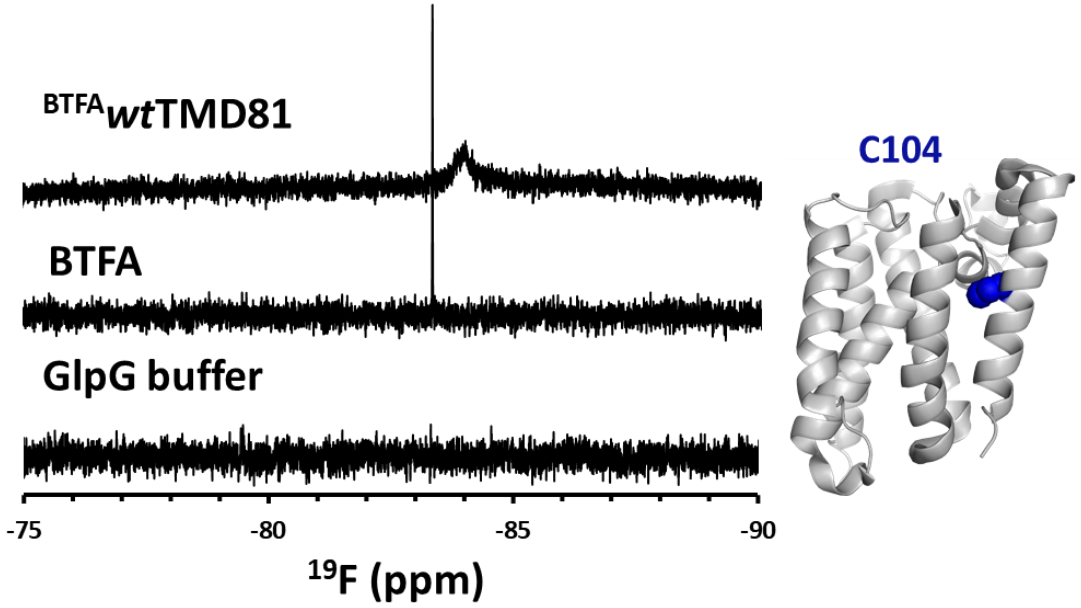

Figure 3.8: ^{19}F NMR spectrum of 60 µM wtTMD81 labelled with BTFA at 25 ºC. Spectra of GlpG buffer with and without BTFA (100 µM) were acquired at 25 ºC. B) Structure of *ec*GlpG (2IC8)[6] showing the position of its native core cysteine residue (blue).

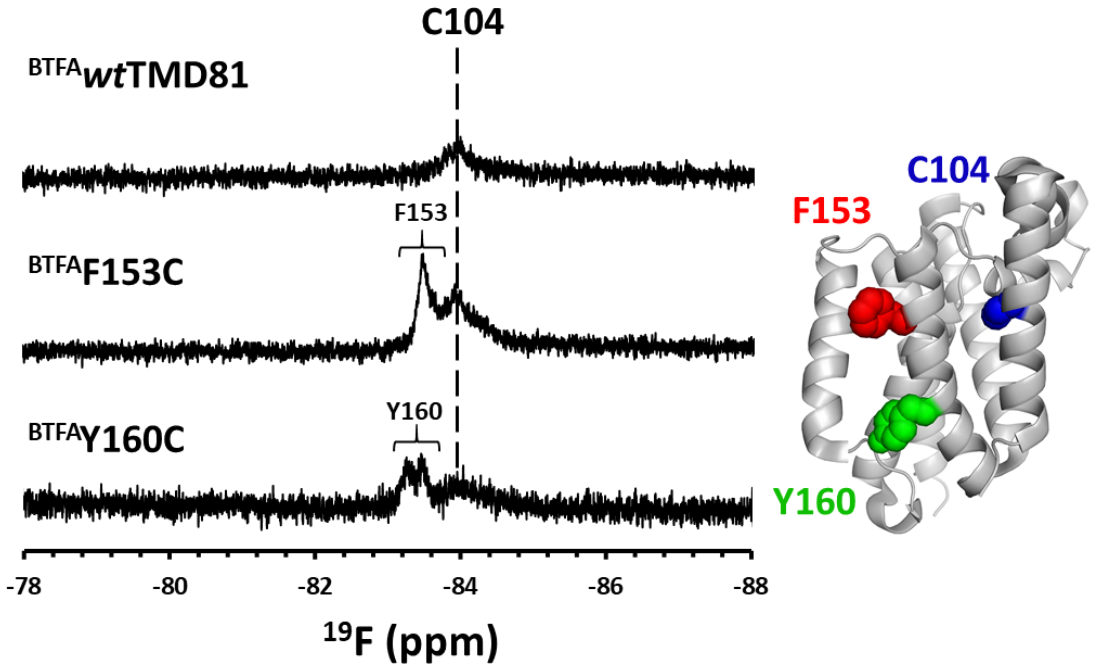

Figure 3.9: ^{19}F peak assignments for cysteine mutants of TMD81 labelled with BTFA. A dashed line is shown for the peak arising from Cys104. ^{19}F spectra were acquired at 25 ºC, with concentrations of 45 µM (Y160C) and 55 µM (F153C) and 60 µM (*wt*TMD81)

To see if BTFA-labelled GlpG could report on α5 dynamics, cysteine mutants F153C and Y160C were generated, both of which were chosen for their proximity to the gating region. F153 from α2 is known to interact with W236 on α5. Y160 is located on α2 and lies near L229 on α5. ^{19}F NMR spectra were collected for BTFA-labelled F153C and Y160C (Fig. 3.9). To assign ^{19}F peaks to those specific positions within *ec*GlpG sequence, the spectra were compared to that of the BTFA-labelled *wt*TMD81. The F153C-BTFA spectrum showed two distinct peaks, one arising from Cys104, and the other at -83.47 ppm being assigned to F153. This single narrow peak suggests that the probe is more dynamic at F153 compared to Cys104. For the Y160 mutant, the peak arising from the mutation site appears to be split into two. This may arise from slow exchange between two different local chemical environments around the Y160 side chain.

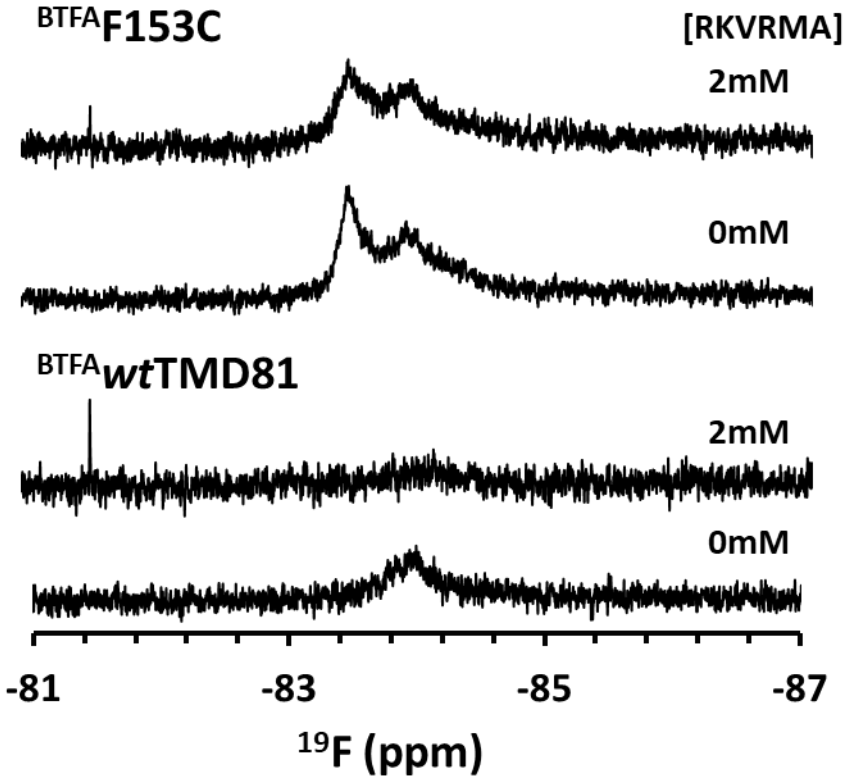

Figure 3.10: Effect of peptide binding on TMD81. ^{19}F NMR spectrum of BTFA labelled F153C and *wt*TMD81 in the presence of 0 mM and 2 mM of the Gurken-derived peptide fragment.

In addition to assignment information, ^{19}F NMR spectra of the BFTA labelled mutants could potentially provide some information on substrate gating dynamics. To explore whether substrate binding influences *ec*GlpG gating, ^{19}F NMR was used to monitor TMD81 interactions with a peptide derived from the Gurken transmembrane substrate. This peptide shares the same sequence as a peptide aldehyde inhibitor known to be a potent inhibitor of *ec*GlpG.[57] Curiously, 1D ^{19}F spectra of BFTA-labelled F153C (Fig. 3.10) obtained in the presence of this peptide showed no significant change in chemical shift. These results suggest that substrate binding may not facilitate the displacement of α5 and stimulate gate opening. However, there was some broadening of the F153C-BTFA peak in the presence of 2 mM peptide, potentially showing a change in the dynamic properties of the environment around this residue upon peptide binding.

The effects of peptide binding on *ec*GlpG were also assessed for BFTA-labelled Y160C (Fig. 3.11). Although there was a loss of signal intensity due to sample dilution, there it appears that there was no change to the chemical shift, or peak broadness. This may suggest that substrate binding does not stimulate any conformational changes at this site.

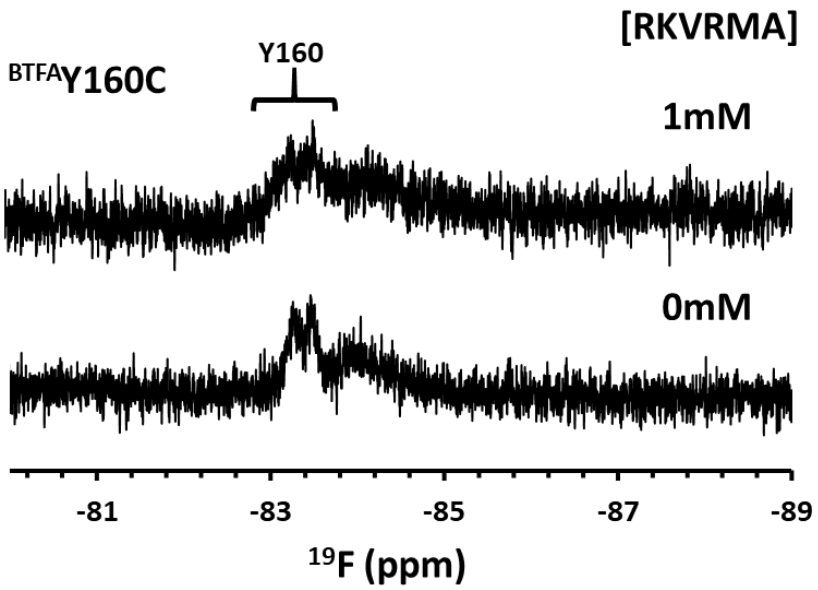

Figure 3.11: Effect of peptide binding on TMD81 spectra. ^{19}F NMR spectrum of BTFA labelled Y160C in the presence of 0 mM and 1 mM of the Gurken-derived peptide fragment.

We also tested BTFMA since its large size could provide additional van der Waals contacts that could replace those lost by the mutation of Trp to Cys. The ability of the larger probe to label *wt*TMD81 was tested and monitored by ^{19}F NMR. The spectrum of this sample was compared to spectra obtained with free BTFMA in the presence and absence DPC detergent micelles (Figure 3.12). BTFMA-labeled TMD81 showed a single broad peak at -61.9 ppm, which is in the proximity of the expected chemical shift window characteristic of BTFMA (δ -61 to -63 ppm).[78] This indicates that Cys104 remains accessible to react with the BTFMA label. Curiously, the spectrum of free BTFMA showed three unique resonances, suggesting the presence of possible contaminants in our commercial supply. It is likely these contaminants are removed during our purification process after the labelling reaction since only one species is observed in the BTFMA-labelled TMD81 spectrum.

3.2.2 Solution state ^{19}F NMR of BTFMA-labelled TMD81

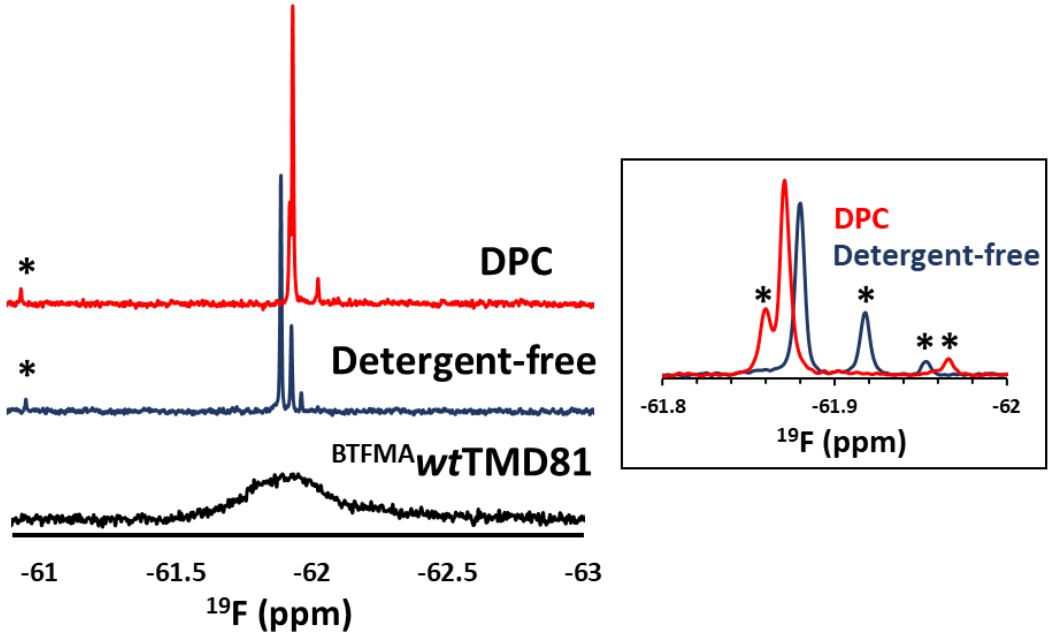

Figure 3.12: ^{19}F NMR spectrum of wtTMD81 labelled with BTFMA at 25 ^{o}C, with a concentration of 70 µM. Spectra of 100 µM BTFMA labelling reagent with and without 0.1% w/v DPC detergent were acquired at 25 ^{o}C. Note that peaks corresponding to contaminants are indicated by an asterisk. An expanded view of the free BFMA spectra is shown on the right.

To assess the suitability of this probe to study gating dynamics within *ec*GlpG, the F153C mutant was treated with BTFMA and the 1D ^{19}F NMR spectrum was recorded (Fig. 3.13). At 25 ^{o}C, the spectrum showed three distinctive resonances, with one peak being attributable to the Cys104 label (peak #3). The spectrum showed that the ^{19}F peak corresponding to Phe153 is split into two peaks, suggesting that the probe samples two states that have slightly different local chemical environments (peaks #1 and #2 at -61.3 ppm and -61.6 ppm, respectively). The chemical shift of peak #1 may reflect a state that has lower mobility since its value is more typical for hydrophobic environments. This might reflect a more closed state in which W236 in α5 is interacting with the side chain at position 153 in α2. The chemical shift of peak #2 is more consistent with a solvent exposed environment, which may reflect an open state.

To explore the influence of temperature on these states, spectra were acquired on the same sample at both lower and higher temperatures. As the temperature was increased, peak #2 became more narrow, which is consistent with a temperature-dependent increase in mobility. Interestingly, this effect was not observed for resonance #1 which instead decreased in intensity, both when the temperature was increased to 45 °C, and lowered to 8 °C. However, the signal-to-noise of spectra in both cases was lower overall, which made it more difficult to observe the lower intensity peaks.

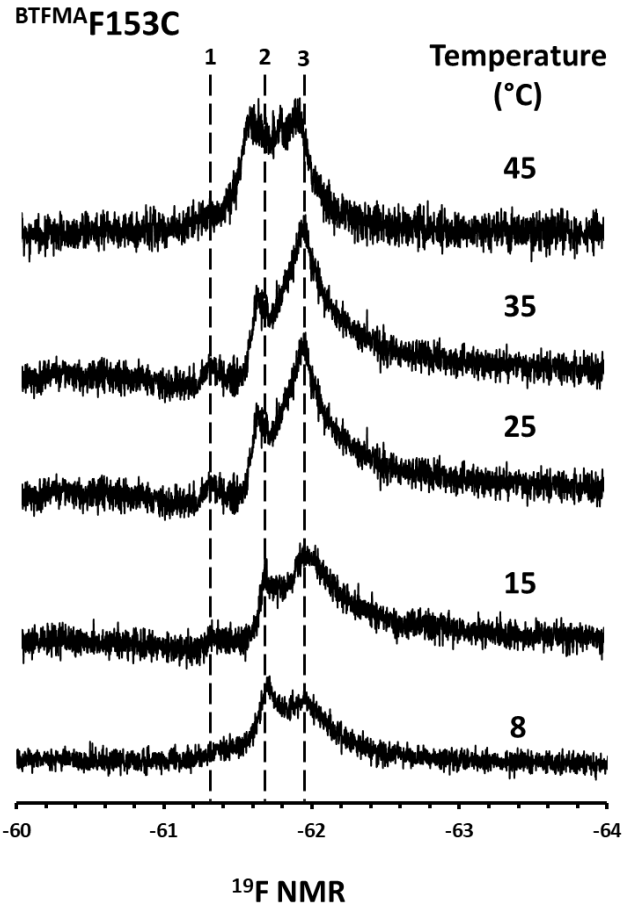

Figure 3.13: ^{19}F NMR spectra of BTFMA-labelled F153C collected at various temperatures.

To determine whether binding of the N-terminal fragment of the substrate peptide would lead to a change in *ec*GlpG dynamics, 1D ^{19}F NMR spectra were acquired for BTFMA-labelled F153C in the presence of the N-Gurken peptide. As shown in Figure 3.14, as peptide was titrated into the sample, a new resonance appeared at -61.8 ppm (indicated with the asterisk). However, there was no change in the relative intensity of peak #2 that may correspond to a more open conformation of α5. Similarly, peak #1 that may reflect a more closed conformer also showed no change in relative intensity. Spectra acquired at 45 ºC with 2 mM peptide, showed an even stronger peak for this species, while the closed gate resonance was reduced in intensity.

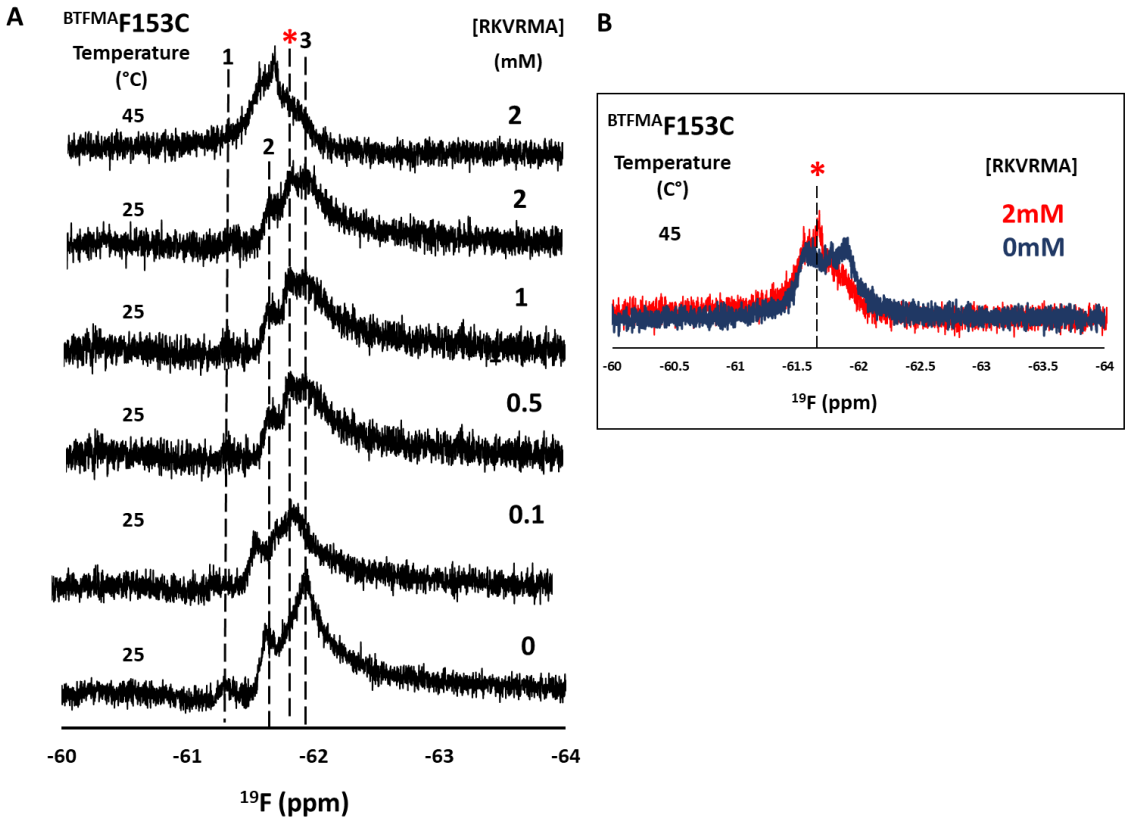

Figure 3.14: Effect of peptide binding on TMD81. ^{19}F NMR spectrum of BTFMA labelled F153C in the presence of the Gurken-derived peptide fragment at various concentrations. The asterisk indicates the position of the new peak that appears when peptide is added at 25 ºC.

1D ^{19}F spectra of BTFMA-labelled Y160C were also recorded at various temperatures (Fig. 3.15). The spectrum recorded at 25 ºC shows multiple peaks which may be a sign of conformational exchange at this site. There is also a peak at the Cys104 position as expected, however, it is not as broad as observed in the WT spectrum. This may reflect a change in the dynamics at this site in this mutant, or it may also arise from overlap with a species from the label at position 160. As the temperature was increased, this peak became more intense. In addition, peaks at smaller chemical shifts that presumably arise from the label at position 160 appear to coalesce and increase in intensity as the temperature is increased. Interestingly, a similar phenomenon was also observed when the temperature was decreased, although with a broader peak width overall. These results suggest some exchange process is being detected by the label at this site, although it is too complex to understand its origins at this time.

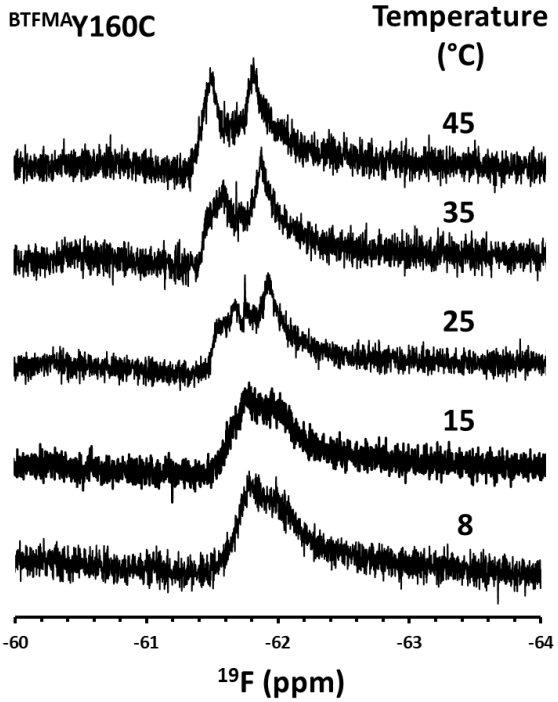

Figure 3.15: ^{19}F NMR spectra of BTFMA labelled Y160C collected at various temperatures.

The effects of peptide binding on this mutant were also assessed. ^{19}F NMR spectra of BTFMA-labelled Y160 collected in the presence of 0.5 mM and 1 mM peptide (Fig. 3.16) showed little to no changes, suggesting that this structural probe is not involved in any conformational changes near the gating region.

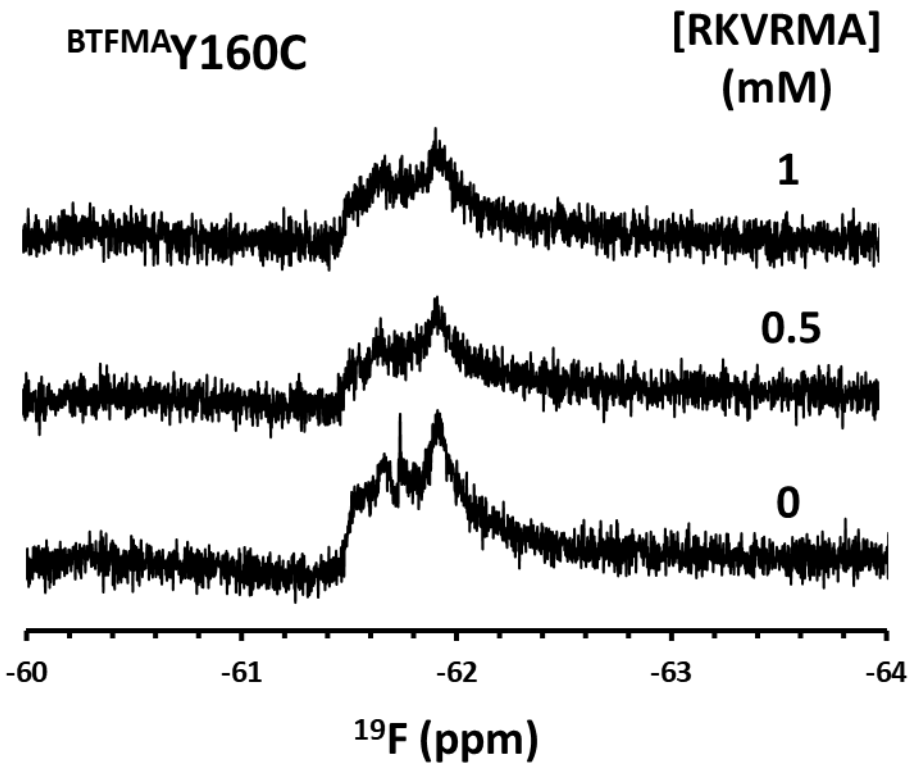

Figure 3.16: ^{19}F NMR spectrum of BTFMA-labelled Y160C in the presence of 0 mM, 0.5 mM, and 1 mM of the Gurken-derived peptide fragment.

3.2.3 ^{19}F NMR of an activated TMD81 mutant

To investigate the gate-open state of TMD81, a W236A substitution mutant was generated into the F153C mutant, as its substitution to alanine was previously found to increase proteolytic activity by potentially promoting opening of the lateral gate.[63,72] As shown in Figure 3.17 the spectrum of BFTMA-labelled W236A/F153C has five peaks, with peaks at -60.5 ppm (A) and -

60.56 ppm (B) in the 25 ºC spectrum being unique to this open gate mutant. These peaks may arise from the newly accessible rotameric states of the F153C side chain caused by W236A mutation in which the small Ala side chain is less sterically encumbering. The spectra were also recorded at 45 ºC and 8 ºC (not shown), which showed similar effects to previous spectra, where broadening or narrowing of all resonances is observed at lower and higher temperatures, respectively.

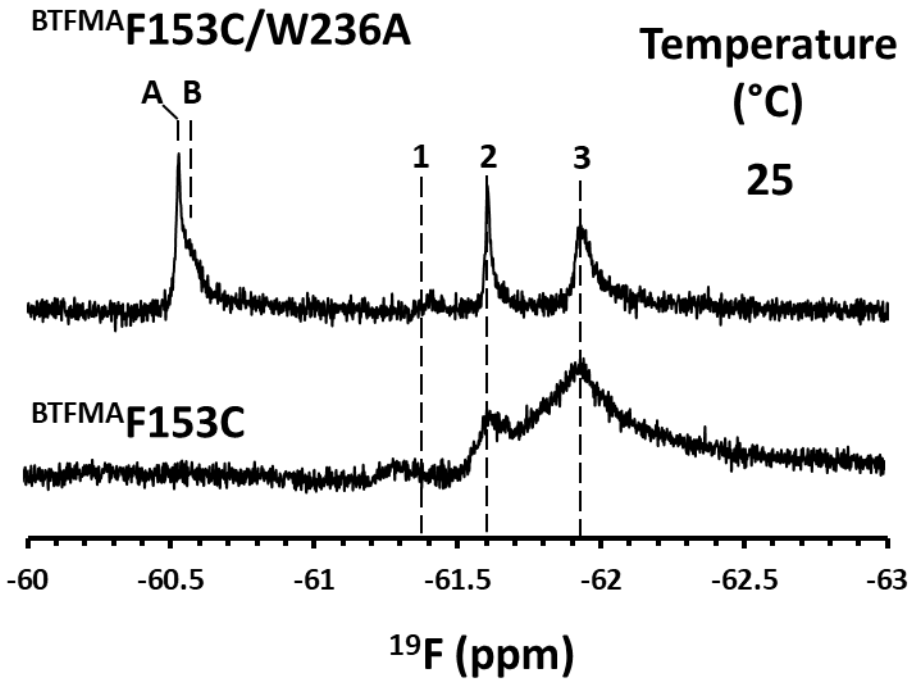

Figure 3.17: ^{19}F NMR spectra of BTFMA-labelled F153C/W236A and F153C collected at 25 ºC.

The effects of peptide binding were also assessed on this mutant, and showed the peaks corresponding to F153 becoming more broad upon peptide binding, mirroring results obtained with F153C. The native cysteine side chain resonance remained unchanged over the course of the titration (Fig. 3.18). These results also suggest that the probe at this location senses the peptide

interaction. Meanwhile, peak #1 remains unchanged over the course of the titration suggesting no change in gating if this peak corresponds to the closed state.

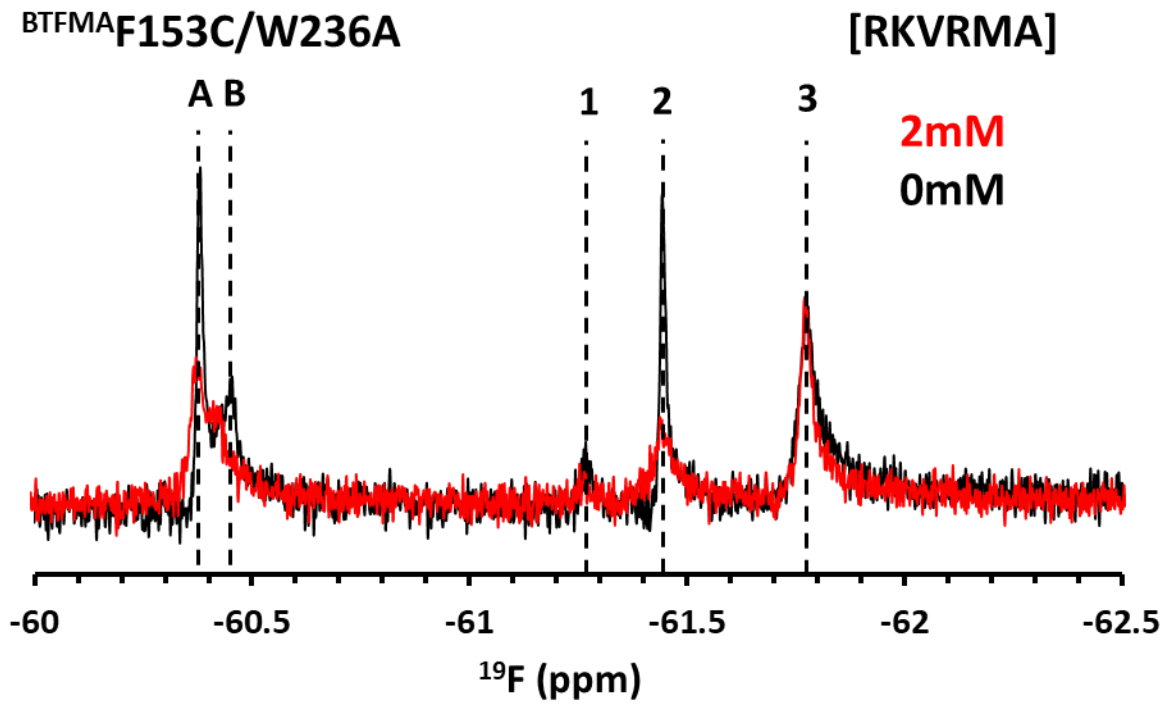

Figure 3.18: Effect of peptide binding on ^{19}F NMR spectrum of BTFMA-labelled F153C/W236A in the presence 0 mM (black) and 2 mM (red) of the Gurken-derived peptide fragment.

3.3 Towards ^{19}F-labelled TMD81 using unnatural amino acid incorporation

The incorporation of unnatural amino acids by amber codon suppression has the potential alternative to allow incorporation of ^{19}F labels during the expression of *ec*GlpG. The method relies on an orthogonal aminoacyl-tRNA synthetase/tRNA$_{CUA}$ that incorporates L-4-trifluoro-methylphenylalanine (tfmF) (Fig. 3.20B) whenever the amber codon UAG occurs in a translating sequence.[83,129] The orthogonal synthease/tRNA pair that is encoded on the plasmid pDule-tfmF was co-transformed into C43(DE3) *E. coli* with the TMD81 expression plasmid containing an amber codon at the desired site of incorporation, and tfmF was added to the culture prior to

induction of expression. The mutation sites were chosen specifically for their proximity to the gating region. Specifically, W212 is located on α4 and is closely packed against I230 located on α5 in the crystal structure of the closed state. W241 is located on α5 packed against N251 on α6 (Fig. 3.19). The incorporation of a fluorinated probe at this position could provide some dynamic information concerning substrate gating.

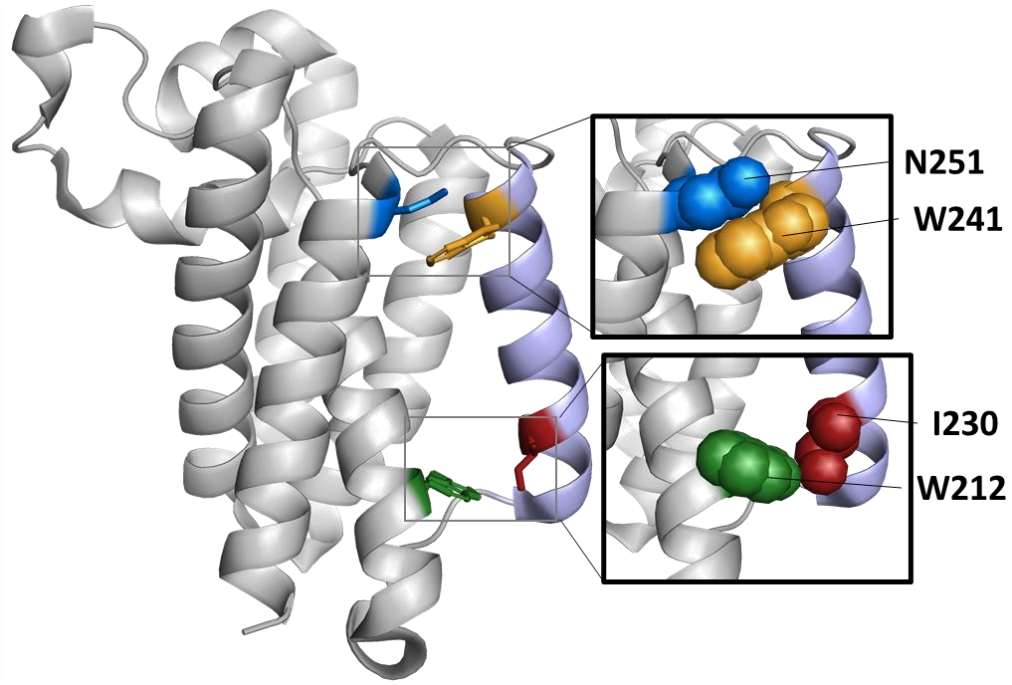

Figure 3.19: X-ray structure of *ec*GlpG (PDB 3B45) showing the location of W212 relative to I230 and W241 relative to N251. Residues are shown as spheres or as sticks.

To determine if the tfmF probe was successfully incorporated into TMD81, samples from growth medium with or without 1 mM tfmF were analysed by western blots. As shown in Figure 3.20A, bands corresponding to species with the expected molecular weight of TMD81 appeared only in cultures that were supplemented with tfmF. To determine the optimal expression time, samples from a 1 L culture were taken at various time points after induction of expression and analysed by western blot. This showed that W212*tfmF* appears 2 hours post-induction and

increased to a maximum after about four hours, with the culture reaching an optical density of ~1.0 (4 h post induction).

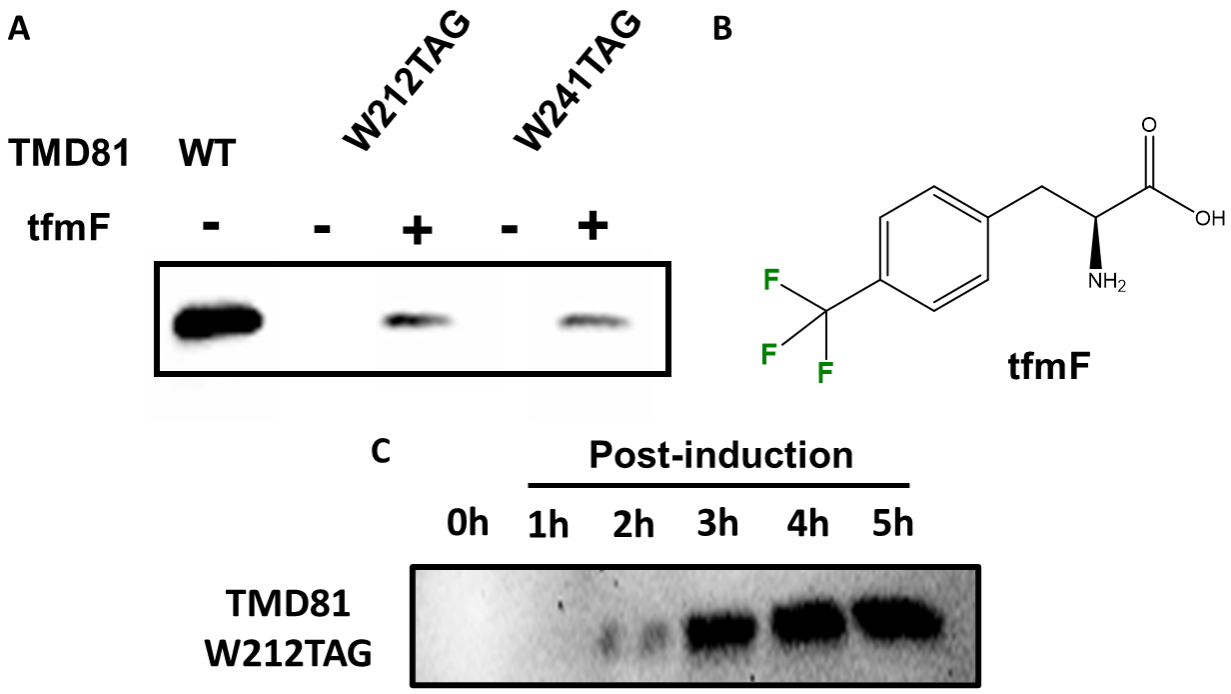

Figure 3.20: (A) Western blot analysis of W212TAG and W241TAG expressed in the presence (+) and absence (-) of 1 mM tfmF (B) Chemical structure of L-4-trifluormethylphenylalanine (B) Time course of W212TAG expression in large scale culture with tfmF, monitored by western blot.

TfmF-labelled W212TAG was subsequently purified and assessed for its structural integrity. As shown in the size exclusion chromatography elution profiles in Figure 3.21, a peak eluted at the expected volume for WT TMD81 (~15 mL). As has been observed in most size exclusion chromatography profiles for TMD81 purifications, a larger molecular weight species was also seen at 11-13mL; however, only fractions corresponding to the major peak (14-17 mL) were pooled for further analysis.

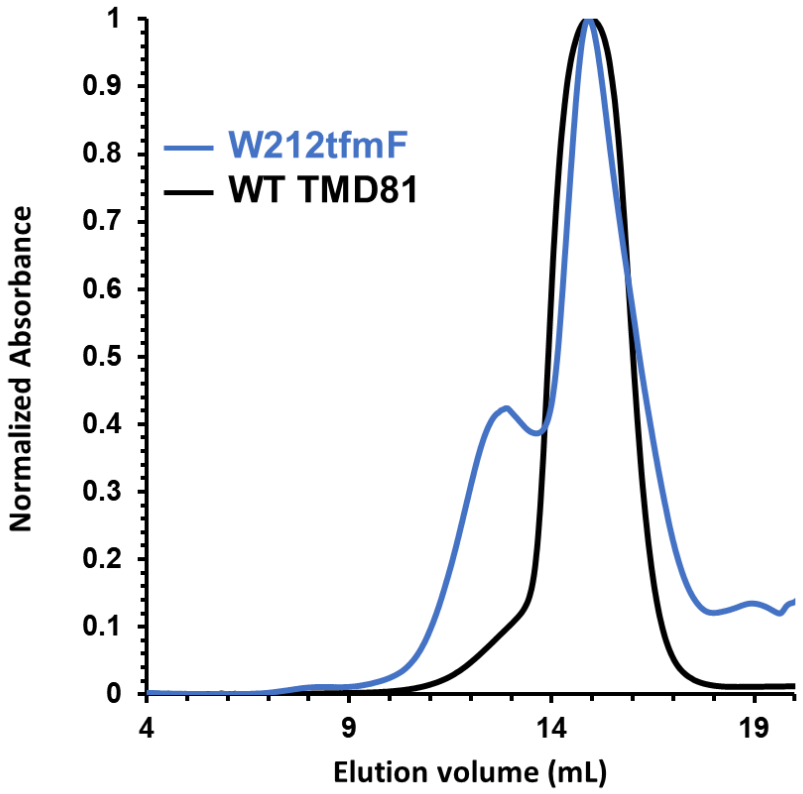

Figure 3.21: Size exclusion chromatography profiles of WT TMD81 and tfmF-labelled W212TAG solubilized in dodecylphosphocholine (DPC) detergent micelles in black and blue, respectively. Samples were purified by FPLC using a Superdex SP200 column in GlpG buffer. Absorbance at 280 nm was monitored.

CD spectroscopy was employed to determine if any structural perturbation was induced by the side chain trifluoromethyl group. The CD spectra for tfmF-labelled W212TAG (Fig. 3.22) showed characteristic minima at 208 and 222 nm, consistent with the largely folded α-helical structure for a properly folded *ec*GlpG structure, suggesting that overall rhomboid structural integrity was not significantly altered for this ^{19}F-labelled mutant. While a structurally active tfmF-labeled TMD81 was expressed, yields obtained were too low with only 6 µM of 375 µL NMR sample being produced from a 1 L culture. This is far from the $\geq$ 50 µM concentrations typically needed for ^{19}F NMR, making it impossible to go forward with NMR experiments at this time.

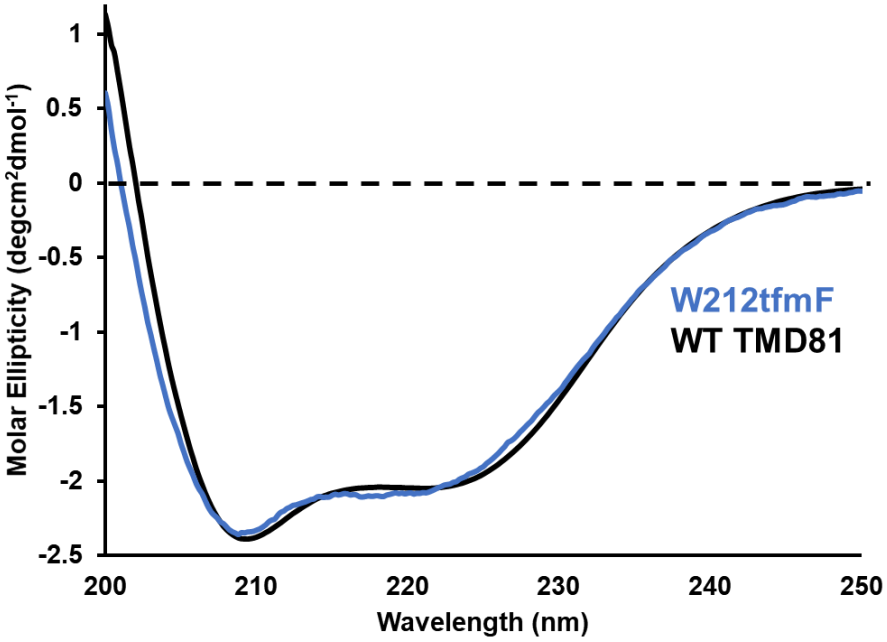

Figure 3.22: Circular dichroism spectrum of 5 µM tfmF-labelled W212TAG (blue) superimposed on the spectrum of 5 µM WT TMD81.

3.4 The isolation of WT TMD81 into synthetic lipid particles

To study protein structure and dynamics, membrane proteins are typically solubilized and purified from the native lipid bilayer using detergent micelles. This approach has been successful for *ec*GlpG; however, detergent-solubilized systems lack information about interactions with their native lipid environments. The composition of the membrane is thought to have a stabilizing effect on the structure of rhomboid.[126] In order to preserve interactions with the native lipid environment, it is possible to use extract membrane proteins directly from the membrane using amphipathic polymers containing maleic acid.[124] These polymers can disrupt membranes in a way that creates nanoparticles containing membrane proteins surrounded by its lipid bilayer. Rhomboid proteases have been successfully isolated with their native membrane environment using styrene maleic acid

(SMA) and diisobutylene maleic acid (DIBMA).[126,130] This raised the possibility that these polymers could be used for solution NMR studies of *ec*GlpG in its native lipid environment.

For this purpose, styrene maleic acid co-polymer was prepared from styrene maleic anhydride co-polymer as previously described.[124] Fourier transform infrared (IR) spectroscopy was used to assess the hydrolysis of SMAn to SMA (Fig. 3.23A). The IR spectrum for SMAnh showed the presence of the maleic anhydride carbonyl groups (1774 cm^{-1}) which are subsequently converted to carboxylate carbonyl groups (1554 cm^{-1}) upon hydrolysis to form SMA. SMA lipid particles were prepared using *E. coli* lipid extract. A lipid suspension of large unilamellar vesicles was prepared by extrusion. Dynamic light scattering performed on the lipid vesicle suspension revealed an average hydrodynamic diameter of 175.4 ± 41.7 nm (Fig. 3.24B). When 2.5% (w/v) SMA co-polymer was added to this lipid suspension, the solution changed from a cloudy solution to a clear one (Fig. 3.23C) with an average hydrodynamic diameter of 10.1 ± 2.5 nm.

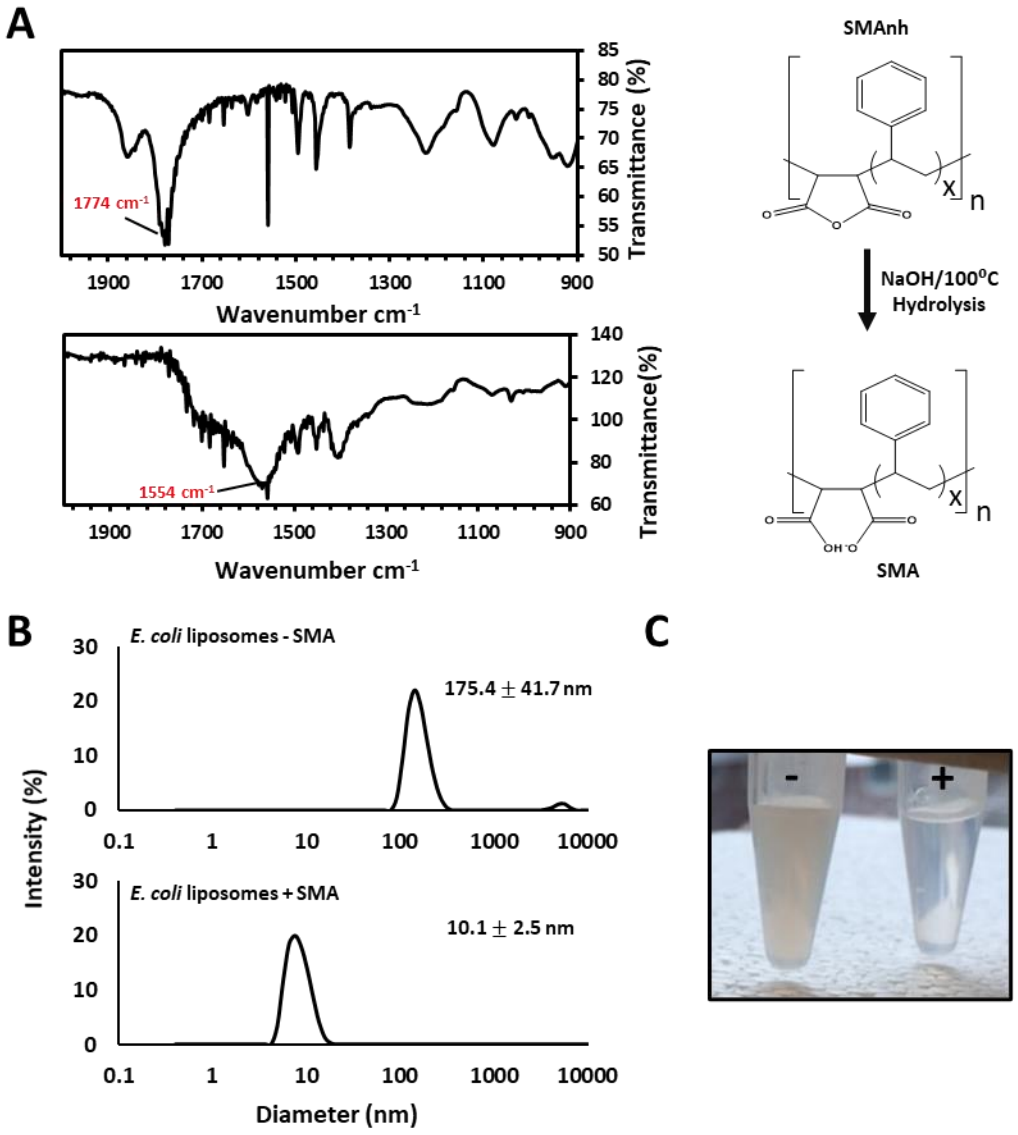

Figure 3.23: Preparation of stryrene maleic acid (SMA) from styrene maleic anhydride co-polymer 2000 (SMAnh). A) SMA was converted to SMAnh by hydrolysis. The infrared spectrum of SMAnh (top) shows the presence of maleic anhydride carbonyl groups (1774 cm^{-1}) which are subsequently converted to carboxylate carbonyl groups (1554 cm^{-1}) upon hydrolysis (Bottom) to form SMA. B) Dynamic light scattering (DLS) measurements of liposomes derived from *E. coli* lipid extracts with (+) and without (-) SMA addition at 2.5% (w/v). C) *E. coli* liposomes, with and without SMA.

TMD81 protein samples were purified and isolated into SMA lipid particles from cell pellets by successive nickel affinity and size exclusion chromatography steps, using previously

established protocols[126]. Coomassie-stained gels of samples taken at various stages during the purification process are shown in Figure 3.24. A faint band was observed in the SMA soluble fraction (SSF) that is absent in the flow-through (FT) from the nickel affinity column. Two sequential wash steps were done with increasing concentrations of imidazole (W1 and W2), and then 500 mM of imidazole was used to elute TMD81 (E) with yields of ~100 µM per litre of culture. As seen in the Coomassie-stained gel, in addition to the TMD81 band at ~24 kDa, the SMA polymer can be identified as a diffuse band at ~8 kDa. This band was present in all fractions and became less intense over the course of the purification, indicating that excess polymer was separated from the TMD81. Samples from size exclusion chromatography (SEC) were also shown to assess sample purity.

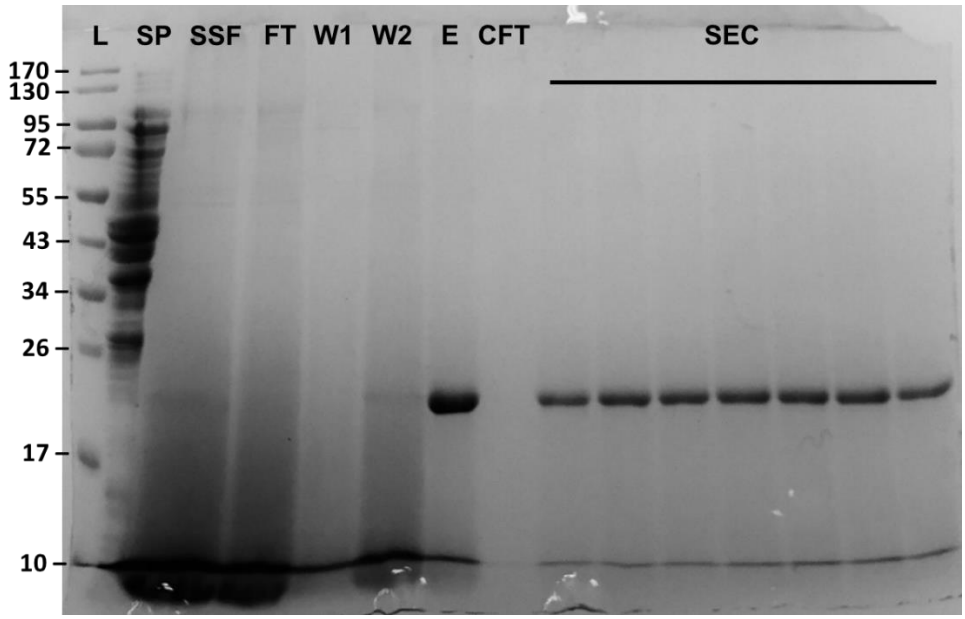

Figure 3.24: Coomassie-stained SDS-PAGE gel of fractions taken during various stages of the purification of TMD81 extracted from the membrane using SMA. L: Molecular weight ladder in kDa. SP: Soluble protein fraction, SSF: SMA solubilized membrane protein fraction using 2.5% (w/v) SMA. FT: Ni-NTA column flow through, W1 and W2: contaminants removed during Ni-NTA wash steps. E: purified *wt*TMD81-SMALPs eluted using 500 mM imidazole. SEC:

Representative fractions obtained after size-exclusion chromatography containing purified TMD81 in SMALPs.

The size exclusion chromatography profile shown in Figure 3.25 was compared to the elution profile for TMD81 solubilized in DPC detergent micelles. This detergent produces a relatively small protein-detergent complex that makes it appropriate for solution NMR studies. The profile obtained for TMD81 in SMALPS was found to be a broad peak compared to the DPC-solubilized samples, with the elution volume spanning ~5 mL (~10-15 mL). This suggests that the rhomboid-containing SMALPs are heterogenous in size. Since a large part of the elution occurred at smaller elution volumes than those for the detergent-solubilized samples, TMD81 in SMALPs are much larger in molecular weight. However, there is some overlap each between the two profiles at ~13-15 mL, suggesting a nanoparticle size that could potentially be amenable for solution NMR. Therefore, the fractions that eluted between 14 and 15 mL were retained to assess for structural integrity and SMALPs size.

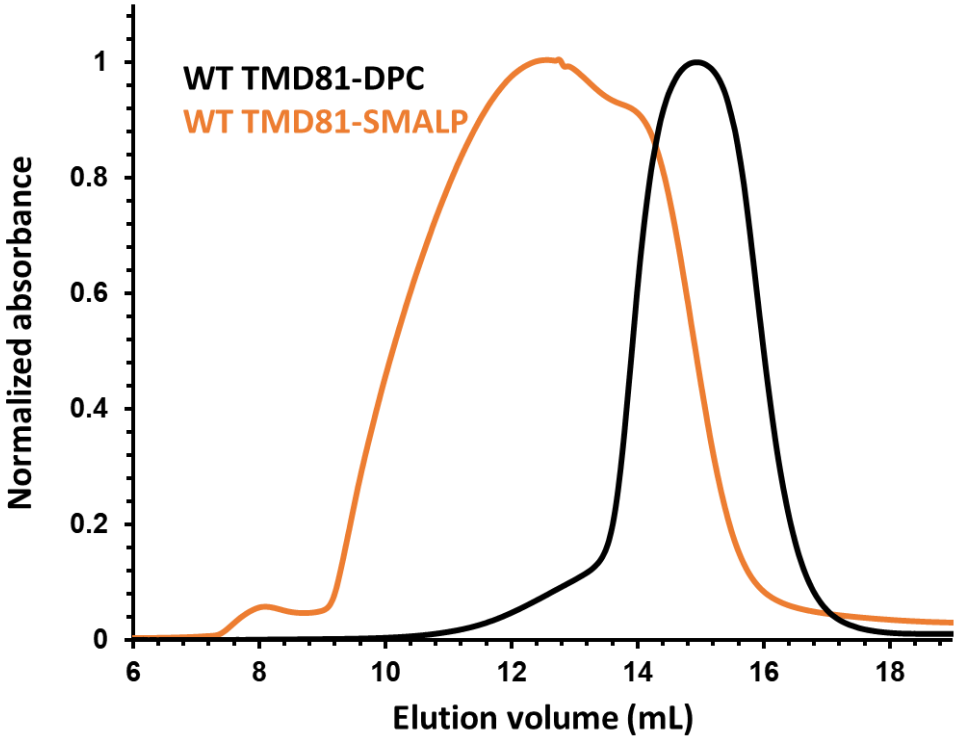

Figure 3.25: Size exclusion chromatography profile of TMD81 solubilized in 0.1% (w/v) DPC detergent (black) and 2.5% (w/v) SMA (orange). Samples were analysed by FPLC using a Superdex SP200 column. Absorbance at 280 nm was monitored.

As seen in Figure 3.26A, The CD spectrum obtained for TMD81 purified into SMALPs showed a high degree of similarity to the spectrum obtained in DPC detergent micelles, suggesting that overall structural integrity was not altered by incorporation into SMALPs. Dynamic light scattering performed on this sample revealed an average hydrodynamic radius of 16.6 ± 4.1 nm for the TMD81-SMALPs (Fig. 3.26B), although there was still significant variability in size (PDI = 0.18). This is much larger than the more homogeneous 10 nm size of SMALP that we could obtain from this preparation of SMA using vesicles made from *E. coli* phospholipids (Fig. 3.23B). Since large disc sizes (>10 nm) and heterogeneous samples are unfavorable for NMR, it was not possible to move forward with ^{19}F labelling and solution NMR experiments.

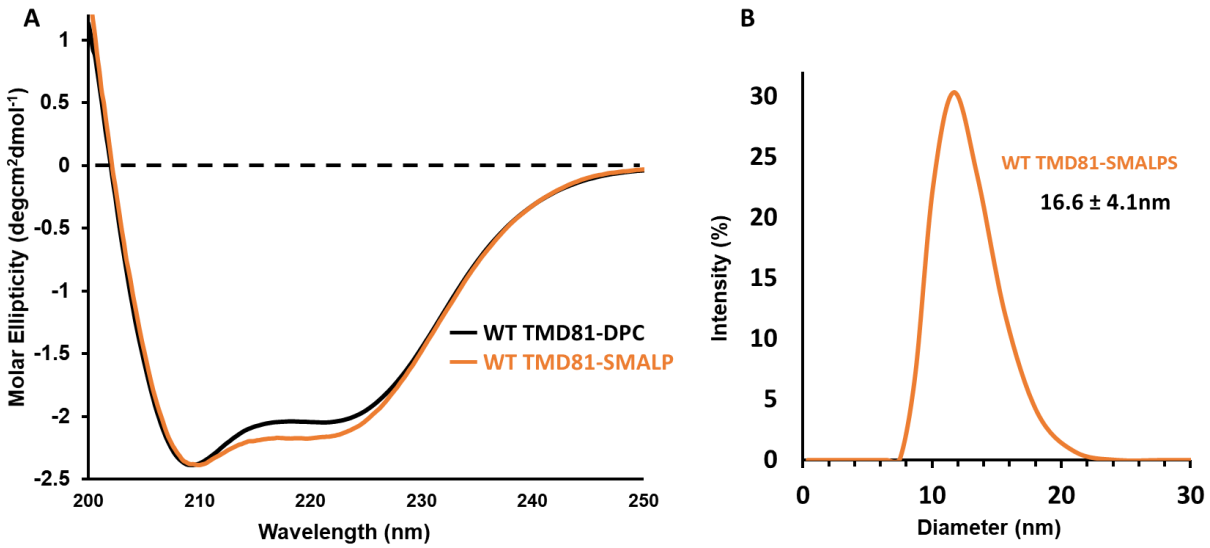

Figure 3.26: Characterization of TMD81 incorporated into SMALPs. A) Representative CD spectrum obtained for 5 µM TMD81 in 0.1% (w/v) DPC (black) or 2.5% (w/v) SMA (orange) at 37 °C. B) Dynamic light scattering of *wt*TMD81 in native SMA nanodiscs. Reported are the average and the standard error of the mean from repeated measurements (n = 3).

4. Discussion

4.1 Challenges with NMR of uniformly ^{19}F-labelled *ec*GlpG

The relatively small chemical shift dispersion for all ^{19}F spectra obtained is surprising considering the 11 well distributed tryptophan residues within *ec*GlpG (Fig. 3.1) provide a broad sampling of structural environments. For example, a residue of *ec*GlpG like W196 is aqueous solvent exposed whereas residues like W236 and W158 are in the more restricted protein core. The inherent sensitivity of the ^{19}F chemical shift to small changes in local chemical environment is expected to yield relatively large chemical shift dispersion in resonances associated with residues from these different chemical environments.[131] The low spectral dispersion ^{19}F spectra observed in 5FW-TMD81 could suggest that the protein is not properly folded. The uniform labelling of all tryptophan sites might affect the global protein stability in TMD81. However, this has only been reported for monofluorinated derivatives of tyrosine and histidine, in which the ring fluorine alters the side chain pKa.[132,133] It is unlikely that the sample is completely unfolded, however, since this should give rise to a spectrum with a single peak as has been seen for other unfolded proteins.[134,135] However, without a functional assay performed on these samples, it is not possible to determine if the structure of *ec*GlpG has been perturbed by the introduction of these fluorine atoms.

The wider spectral dispersion in the 7FI-labelled TMD81 raised the possibility of chemical shift assignment through tryptophan knockout mutations. The spectrum of the single tryptophan to glycine substitution exhibited substantial chemical shift perturbations. Unfortunately, this did not work out for the tested samples as there was low signal-to-noise for some of the peaks, and multiple changes in the spectrum introduced by the mutation. In the case of W236G, the drastic

nature of the substitution could have given rise to additional changes in structure that complicated the spectrum. This mutant was chosen as its functional properties have already been characterized[71,73]; however, it would have been better to use a more conservative mutation (e.g., W236F).

It might have also been useful to mutate residues that are proximal to targeted Trp residues, to look for a change in chemical shift of the corresponding peak.[136,137] It is also possible to differentiate solvent exposed versus buried residues by use of paramagnetic additives,[136,138,139] a strategy that could be used in the future for these mutants.

The site-specific ^{19}F labelling of proteins, post-expression can be a powerful strategy when the biosynthetic incorporation of fluorinated amino acids is impractical, due to challenges associated with obtaining assignments. This is an attractive approach for *ec*GlpG studies, since this would reduce the potential for cumulative perturbations resulting from uniform labelling. Additionally, this method would benefit from monitoring a single resonance from a strategically positioned probe within *ec*GlpG sequence.

4.2 Investigating the utility of solution ^{19}F NMR in *ec*GlpG

Site-specific labelling of TMD81 with probes containing a trifluoromethyl group gave rise to well resolved ^{19}F NMR spectra that allowed for peak assignment. The single peak corresponding to BTFA-labelled F153 was surprising given the location of the probe at the interface of α2 and α5. This may indicate that a single conformation was being sensed by the probe, or that the rate of exchange was faster than the frequency difference between the environments being sampled. However, it is also possible that the probe altered gating dynamics. This hypothesis is supported

by the F153A mutation that was previously shown to increase the proteolytic activity of *ec*GlpG by potentially promoting the α5 open gate conformer.[42,54,72,73] The narrow nature of this peak would suggest that the probe is more dynamic, which would be consistent with an open gate conformer in which the lateral shift of the entire α5 helix away from α2, would free the probe from the tightly packed interactions in the closed state. However, further structural, and functional studies of the derivatized mutant would be required to confirm this hypothesis, which was beyond the scope of this thesis.

^{19}F NMR spectra of a BTFMA-labelled F153C showed splitting of the signal into 2 peaks, potentially corresponding to different conformations undergoing slow exchange. The narrow peak was upfield, which may indicate a more solvent-exposed open state. The broad peak that appeared at the downfield shift may represent a closed state, with the aromatic ring of the probe interacting potentially interacting with the aromatic ring W236. The temperature dependence of the ^{19}F NMR spectra showed that the downfield state disappeared to near background levels at 4 ºC and 45 ºC. However, it was not clear whether this loss was accompanied by a corresponding increase in the upfield peak as would be expected if the two states were undergoing exchange.

To limit ambiguities in the spectral analysis it would be necessary to remove the native cysteine. This peak only limits our ability to assess gating dynamics, therefore its mutation to an Ala or Ser could allow us to focus on peaks arising from a unique position. As mentioned previously, structural, and functional studies are required to assess the effects of the labels on TMD81. While the chemoselective Cys-conjugation method can be advantageous over uniform labelling, the potential to introduce structural perturbation is greater, because of the relatively large size of the ^{19}F labels. To minimize the potential structural perturbation, one could use a smaller label, such the classic TET label (Fig. 1.7).

The peptide-binding studies suggest that the Gurken-derived tetrapeptide does not stimulate any changes to the gating region of TMD81, even at saturating conditions. While ^{19}F NMR spectra of BTFMA-labelled F153C and W236A/F153C are consistent with the presence of a largely open-gate state, the peptide titration did not seem to affect the populations of the open state. However, future studies with label incorporated at other locations around the interface of α2 and α5 may allow changes in gating to be observed.

Since the short Gurken-derived peptide did not seem to stimulate a conformational change gate opening is not required for entry of some substrates. This is consistent with the available X-ray structures which also depict the closed conformation of *ec*GlpG in complex with peptide aldehyde inhibitors with a similar sequence to the peptide employed in this thesis.[57,70] However, a recent study showed that a longer peptide containing both the N and C-terminal portion of the substrate sequence gave rise to an X-ray structure where the entire α5 was in a disordered state.[70] This suggests that TM substrates and inhibitors or short peptides may take different paths to enter the active site.[57,73] In the future, ^{19}F NMR studies should be done with longer peptide sequences to evaluate this hypothesis. This would require that an inactive mutant be created to prevent hydrolysis of the peptide, for example, by substituting the active site serine residue with threonine. In addition, different sites for probe conjugation might be more appropriate since F153 may be involved in interactions that are important for gating. One possibility is I230 which is on α5 facing away from the α5-α2 interface. While it does not participate in significant packing interactions, it is proximal to α4 residue W212, and so changes in the position of α5 could potentially alter the local chemical environment around this residue.

In future studies, ^{19}F-^{19}F NOEs could be used to establish proximity between probes and relate it to the three-dimensional structure. This has been done for the membrane protein

rhodopsin, where close contacts in the three-dimensional structure of rhodopsin were quantified by collecting ^{19}F–^{19}F NOEs between pairs of TET-labels.[81] Similarly, NOE measurements between the ^{19}F-labeled cysteine residues between α5 and α2 could give some more structural details concerning gate opening. For example, these measurements could be performed if labels were introduced at positions 153 and 236 to measure α5 displacement upon gate-opening, and whether this displacement is uniform across the entire helix, as was done using disulfide-bond mapping in previous studies on *ec*GlpG.[54]

4.3 Alternate strategies for ^{19}F-labelling of *ec*GlpG

The incorporation of an ^{19}F label at a single site in a protein using an orthogonal aminoacyl-tRNA synthetase/$tRNA_{CUA}$ that incorporates L-4-trifluoromethylphenylalanine (tfmF) has the potential to be a powerful tool[83,129] to produce large quantities of ^{19}F-labelled proteins without post-translational modifications. The results presented here demonstrate the expression of site-specifically tfmF-labeled TMD81. W212tfmF was purified with a CD spectrum that suggests that is folds properly, and that the probe does not significantly alter structure. However, yields obtained were too low, as 1 litre of culture yielded ~6 μM in the volume required for NMR, while concentrations of ≥ 50 μM are required to obtain sufficient resolution in the NMR spectra.

Although, the use of pET vector and C43(DE3) for overexpression has previously been used to successfully obtain protein concentrations >100 μM[140], it can also be the source for low yields, as it has been reported that the pET vectors with IPTG induction typically produces low expression yields in this labelling system,[129] which is consistent with our results. Instead, when the pDule-tfmF plasmid is paired with a pBad overexpression vector, high yields of ^{19}F labelled

proteins is produced in autoinduction media.[141] The expression in pBad is turned on by L-arabinose, and when the expression is performed in *E. coli* DH10B strains expression can be done for >20 h, as this strain is not capable of metabolizing L-arabinose. In future studies, it may be useful to put TMD81 expression under the control of a standard arabinose-inducible operator by subcloning this sequence into the pBad vector and perform the overexpression in an autoinduction medium.

4.4 Towards solution NMR of membrane proteins in native lipid environments

The results demonstrates that TMD81 can be purified into a SMALP system with high concentrations >100 µM while maintaining similar α-helical content as DPC-detergent solubilized TMD81. The Ni^{2+}-NTA affinity purification of TMD81-SMALPs produced yields that were comparable to those obtained with DPC detergent micelles. However, the purity achieved using SMA was higher than that achieved using detergents, as has been reported previously for other proteins.[114,130,142]

The SMALPs gave rise to a heterogenous population observed by size exclusion chromatography with DLS experiments suggesting particle sizes greater or equal to ~16 nm. For NMR applications, a larger size (>10 nm) is unfavorable as the slower molecular tumbling rate increases the transverse relaxation rate of the NMR signal, leading to peak broadening and lower resolution. However, previous studies have shown that it is possible to isolate full-length *ec*GlpG in SMALPs[126] with a more homogenous size distribution than we observed here. One possible explanation for this difference is the different lysis method used in this study. A high-pressure microfluidizer (Microfluidics) that can significantly reduce particle size was used in the previous

study, whereas a French press was used in this thesis. Also, we observed that if cells were lysed by sonication, only large aggregates were formed, and protein extraction was poor.

Several studies have found that SMALP size can be tuned according to the ratio of polymer to lipid[103,143,144] or the ratio of styrene:maleic acid with the polymer.[145] In this thesis, an excess of SMA2000 (2:1) polymer at 2.5% (w/v) with membranes at 40 mg/mL (wet pellet weight) were used. It would be interesting in the future to investigate whether conditions could be tailored to produce smaller size SMALPs while still extracting sufficient protein. In a separate study, *ec*GlpG was solubilized in alternative polymers, DIBMA,[145] and XIRAN SL30010 S30 (2.3:1) and revealed a narrow size distribution of the lipid nanodiscs with an average diameter of 12.7 nm and 10.8 nm, respectively.[130] While the size and polydispersity of these nanodiscs can offer a considerable advantage for NMR applications, their suitability for solution NMR of *ec*GlplG remains to be tested.

To date, the only successful NMR application of membrane protein NMR with SMALPs was reported by the Ramamoorthy group in 2020, which performed the direct extraction of cytochrome-b5 (cytb5) from *E. coli* using a styrene-co-maleic acid-ethanolamine (SMA-EA) polymer.[146] SMA-EA is a derivatized version of SMA in which the carboxylic acid groups have been covalently conjugated to ethanolamine groups via amide bonds.[147] While this derivatized version offers larger stability by remaining soluble over a broad pH range (pH 2.5-10), they tend to form medium to larger particles (10-30 nm). Interestingly, the Ramamoorthy group were able to solubilize cytb5 in SMALPs particles sizes of ~5nm with SMA-EA. In this work, a 2D ^{1}H-^{15}N spectrum of ^{15}N-labelled cytb5 was acquired; however, the observable peaks only arose from the large soluble cytoplasmic domain of cytb5, with no peaks being detected from its single transmembrane helix. This suggests that the particle size was still too large to see signals from the

membrane-embedded portions of the SMALP, and that further work is needed to further reduce particle size. It was possible to observe residues in the transmembrane domain by ^{19}F NMR obtained from ^{19}F-labelled cytb5 in peptide nanodiscs.[148]

4.5 Alternate membrane mimetic systems

Nanodiscs are detergent-free discoidal shaped systems similar to SMALPs, except the phospholipid bilayer surrounding the membrane protein is capped by two anti-parallel amphipathic helical membrane scaffold proteins (MSPs) (Fig. 4.1). The use of a "protein belt" to constrain the dimensions of the bilayer ensures the system adopts monodisperse disc size distribution. In contrast to SMALPs, nanodiscs require the removal of native bilayer, and must be reconstituted in detergent phospholipids. The MSPs are formed of 10 amphipathic helical repeats, where variation in the number of repeats allows the formation of nanodiscs with different diameters. The first thermodynamically stable disc with a size under 10 nm was made with an MSP called Δ4H5, and was found to have a diameter of 7.3 nm with an overall molecular mass of ~70 kDa.[149] A smaller nanodisc has since been developed, called D7, which had three terminal helices removed (ΔH8-H10) and is 7nm in diameter with a molecular mass of ~62 kDa.[150] While nanodics have been successfully used in NMR studies of various integral membrane proteins,[151–154] the best NMR spectrum was obtained with the Δ5 construct, with a slightly larger diameter of 8.4 nm, used for deducing the structure of OmpX.[149] The number of MSP variants has continued to expand introducing a wider range of options with various advantages. Among these, are the covalently circularized nanodiscs (cNDs), which have the N and C-termini of MSPs covalently linked, further increasing homogeneity in nanodisc size. The Wagner group in 2016 used a small cNDs with a diameter of 8.5 nm to study the rat neurotensin receptor 1 (NTR1) by NMR and demonstrated that

higher stability and homogeneity were obtained compared to detergent micelles.[155] While nanodiscs are not made with native lipid bilayers, they can provide a competitive edge over SMALPs systems for solution NMR studies due to their solubility, higher stability and monodispersity, illustrating their potential for the structural characterization of *ec*GlpG and other integral membrane proteins.

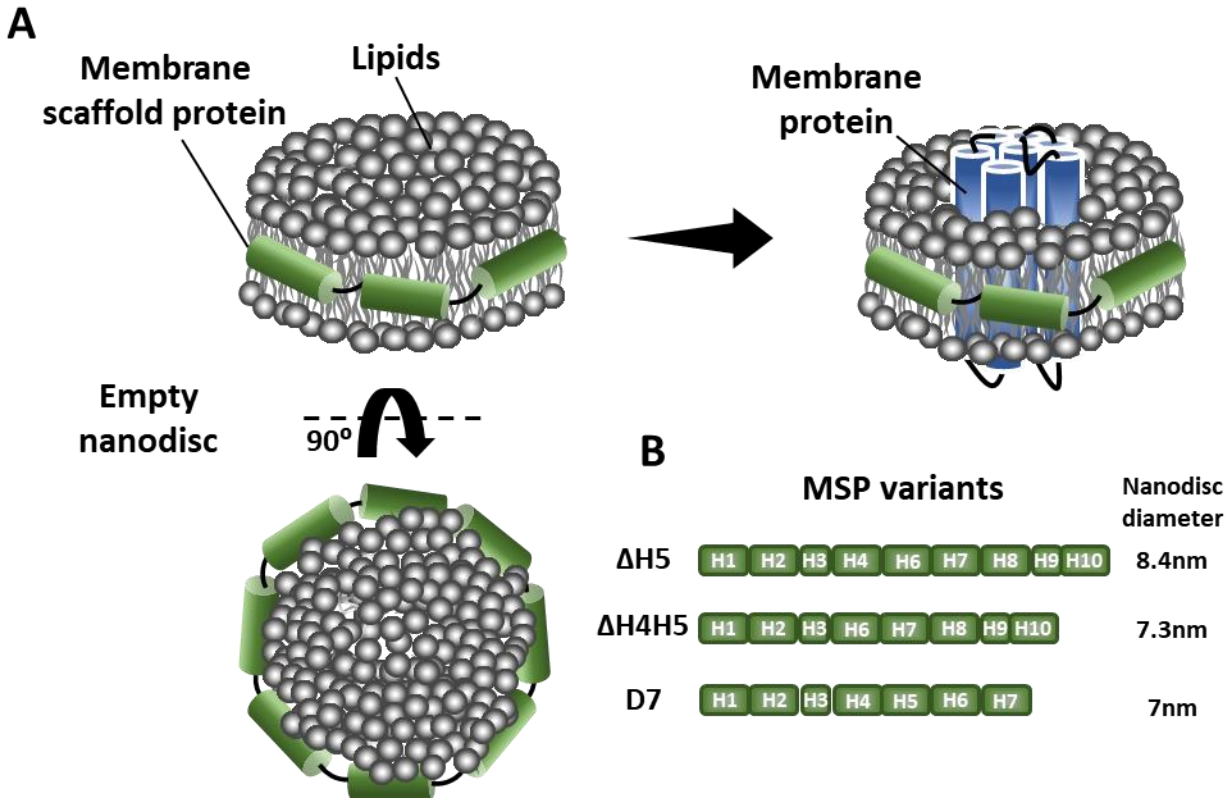

Figure 4.1: Schematic representation of a nanodisc. A) The membrane scaffold protein (green) is surrounding a patch of phospholipids (gray). that is either empty or contains an integral membrane protein (bleu)**.** B) Schematic diagram of the length and disc diameters for several MSP variants.

5. Conclusion and Outlook

Investigations into the mechanism of substrate gating have been ongoing since the first *ec*GlpG crystal structure was reported 15 years ago.[6] Despite the wealth of information, the mechanism remains a topic of significant interest and discussion. In this thesis, we sought to develop techniques to monitor the structure and dynamics of the *ec*GlpG catalytic TMD using solution ^{19}F NMR, which were employed to explore the structural nature of substrate gating and its role in rhomboid function. While uniform ^{19}F labelling strategies might offer the advantage of providing a broad sampling of structural environments, the results in section 4.1 showed that uniformly ^{19}F labelled *ec*GlpG has severe challenges with low spectral dispersion, and challenges associated with obtaining assignments. Site-specific labelling of TMD81 with probes containing a trifluoromethyl group gave rise to well resolved ^{19}F NMR spectra that allowed for peak assignment. While the solution ^{19}F NMR spectra of BTFMA-labelled F153C and W236A/F153C are consistent with the presence of a largely open-gate state, the peptide titration did not seem to affect the populations of the open state. Since the short Gurken-derived peptide did not seem to stimulate a conformational change, gate opening is not required for entry of some substrates. Future ^{19}F NMR studies should be done with longer peptide sequences more likely to stimulate gate opening. We also demonstrated the expression of site-specifically tfmF-labelled TMD81. While the yields obtained were too low for solution NMR to be performed, placement of expression under the control of a standard arabinose-inducible operator and an autoinduction medium may help increase to yields. Preliminary investigations of the utility of SMALPs for TMD81 show that sufficiently high protein concentrations can be obtained for solution NMR (>100 μM) while maintaining similar α-helical content as DPC-detergent solubilized TMD81. However, the heterogenous SMALPs disc size posed a challenge for solution NMR. The MSP-

based nanodisc systems may offer a more homogenous disc size population more amenable for solution NMR studies of *ec*GlpG and should be explored in the future.

6. References

1. Rawson, R. B. *et al.* Complementation cloning of S2P, a gene encoding a putative metalloprotease required for intramembrane cleavage of SREBPs. *Molecular Cell* **1**, 47–57 (1997).
2. Wolfe, M. S. *et al.* Two transmembrane aspartates in presenilin-1 required for presenilin endoproteolysis and γ-secretase activity. *Nature* **398**, 513–517 (1999).
3. Weihofen, A., Binns, K., Lemberg, M. K., Ashman, K. & Martoglio, B. Identification of signal peptide peptidase, a presenilin-type aspartic protease. *Science* **296**, 2215–2218 (2002).
4. Urban, S., Lee, J. R. & Freeman, M. Drosophila Rhomboid-1 Defines a Family of Putative Intramembrane Serine Proteases Sinisa. *Cell* **107**, 173–182 (2001).
5. Deamer, D. W. & Bramhall, J. Permeability of lipid bilayers to water and ionic solutes. *Chemistry and Physics of Lipids* **40**, 167–188 (1986).
6. Wang, Y., Zhang, Y. & Ha, Y. Crystal structure of a rhomboid family intramembrane protease. *Nature* **444**, 179–183 (2006).
7. Manolaridis, I. *et al.* Mechanism of farnesylated CAAX protein processing by the intramembrane protease Rce1. *Nature* **504**, 301–305 (2013).
8. Strisovsky, K. Structural and mechanistic principles of intramembrane proteolysis - Lessons from rhomboids. *FEBS Journal* **280**, 1579–1603 (2013).
9. 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 biology* **4**, (2003).
10. Urban, S. Mechanisms and cellular functions of intramembrane proteases. *Biochimica et Biophysica Acta - Biomembranes* **1828**, 2797–2800 (2013).
11. Bang, A. G. & Kintner, C. Rhomboid and star facilitate presentation and processing of the Drosophila TGF-α homolog Spitz. *Genes and Development* **14**, 177–186 (2000).
12. Lee, J. R., Urban, S., Garvey, C. F. & Freeman, M. Regulated intracellular ligand transport and proteolysis control EGF signal activation in Drosophila. *Cell* **107**, 161–171 (2001).
13. Adrain, C. *et al.* Mammalian EGF receptor activation by the rhomboid protease RHBDL2. *EMBO Reports* **12**, 421–427 (2011).
14. Song, W. *et al.* Rhomboid domain containing 1 promotes colorectal cancer growth through activation of the EGFR signalling pathway. *Nature Communications* **6**, (2015).
15. 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.

Proceedings of the National Academy of Sciences of the United States of America **99**, 8312–8317 (2002).

16. Palmer, T. & Berks, B. C. The twin-arginine translocation (Tat) protein export pathway. *Nature Reviews Microbiology* **10**, 483–496 (2012).

17. Gohlke, U. *et al.* The TatA component of the twin-arginine protein transport system forms channel complexes of variable diameter. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 10482–10486 (2005).

18. Stevenson, L. G. *et al.* Rhomboid protease AarA mediates quorum-sensing in Providencia stuartii by activating TatA of the twin-arginine translocase. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 1003–1008 (2007).

19. Kandel, R. R. & Neal, S. E. The role of rhomboid superfamily members in protein homeostasis: Mechanistic insight and physiological implications. *Biochimica et Biophysica Acta - Molecular Cell Research* **1867**, 118793 (2020).

20. Greenblatt, E. J., Olzmann, J. A. & Kopito, R. R. Making the cut: Intramembrane cleavage by a rhomboid protease promotes ERAD. *Nature Structural and Molecular Biology* **19**, 979–981 (2012).

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

22. Lim, J. J. *et al.* Structural insights into the interaction of p97 N-Terminus domain and VBM in rhomboid protease, RHBDL4. *Biochemical Journal* **473**, 2863–2880 (2016).

23. Paschkowsky, S., Hamzé, M., Oestereich, F. & Munter, L. M. Alternative processing of the amyloid precursor protein family by rhomboid protease RHBDL4. *Journal of Biological Chemistry* **291**, 21903–21912 (2016).

24. Jeyaraju, D. v. *et al.* Phosphorylation and cleavage of presenilin-associated rhomboid-like protein (PARL) promotes changes in mitochondrial morphology. *Proceedings of the National Academy of Sciences of the United States of America* **103**, 18562–18567 (2006).

25. Pellegrini, L. *et al.* PAMP and PARL, two novel putative metalloproteases interacting with the COOH-terminus of Presenilin-1 and -2. *Journal of Alzheimer's Disease* **3**, 181–190 (2001).

26. Meissner, C., Lorenz, H., Weihofen, A., Selkoe, D. J. & Lemberg, M. K. The mitochondrial intramembrane protease PARL cleaves human Pink1 to regulate Pink1 trafficking. *Journal of Neurochemistry* **117**, 856–867 (2011).

27. Deas, E. *et al.* PINK1 cleavage at position A103 by the mitochondrial protease PARL. *Human Molecular Genetics* **20**, 867–879 (2011).

28. Meissner, C., Lorenz, H., Hehn, B. & Lemberg, M. K. Intramembrane protease PARL defines a negative regulator of PINK1- and PARK2/Parkin-dependent mitophagy. *Autophagy* **11**, 1484–1498 (2015).

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

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

31. McQuibban, G. A., Lee, J. R., Zheng, L., Juusola, M. & Freeman, M. Normal Mitochondrial Dynamics Requires Rhomboid-7 and Affects Drosophila Lifespan and Neuronal Function. *Current Biology* **16**, 982–989 (2006).

32. Esser, K., Tursun, B., Ingenhoven, M., Michaelis, G. & Pratje, E. A novel two-step mechanism for removal of a mitochondrial signal sequence involves the mAAA complex and the putative rhomboid protease Pcp1. *Journal of Molecular Biology* **323**, 835–843 (2002).

33. 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. *Journal of Biological Chemistry* **278**, 27781–27788 (2003).

34. Brossier, F., Jewett, T. J., Sibley, L. D. & Urban, S. A spatially localized rhomboid protease cleaves cell surface adhesins essential for invasion by Toxoplasma. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 4146–4151 (2005).

35. Conseil, V., Soête, M. & Dubremetz, J. F. Serine protease inhibitors block invasion of host cells by Toxoplasma gondii. *Antimicrobial Agents and Chemotherapy* **43**, 1358–1361 (1999).

36. Baker, R. P., Wijetilaka, R. & Urban, S. Two Plasmodium rhomboid proteases preferentially cleave different adhesins implicated in all invasive stages of malaria. *PLoS Pathogens* **2**, 0922–0932 (2006).

37. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: The Keio collection. *Molecular Systems Biology* **2**, (2006).

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

39. 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. *Infection and Immunity* **85**, 1–15 (2017).

40. Lemberg, M. K. *et al.* Mechanism of intramembrane proteolysis investigated with purified rhomboid proteases. *EMBO Journal* **24**, 464–472 (2005).

41. Arutyunova, E. *et al.* Allosteric regulation of rhomboid intramembrane proteolysis. *The EMBO Journal* **33**, 1869–1881 (2014).

42. 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).

43. Hedstrom, L. Serine protease mechanism and specificity. *Chemical Reviews* **102**, 4501–4523 (2002).

44. Ha, Y., Akiyama, Y. & Xue, Y. Structure and mechanism of rhomboid protease. *Journal of Biological Chemistry* **288**, 15430–15436 (2013).

45. Wu, Z. *et al.* Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry. *Nature Structural and Molecular Biology* **13**, 1084–1091 (2006).

46. Ben-Shem, A., Fass, D. & Bibi, E. Structural basis for intramembrane proteolysis by rhomboid serine proteases. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 462–466 (2007).

47. Wang, Y. & Ha, Y. Open-cap conformation of intramembrane protease GlpG. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 2098–2102 (2007).

48. Wang, Y., Maegawa, S., Akiyama, Y. & Ha, Y. The Role of L1 Loop in the Mechanism of Rhomboid Intramembrane Protease GlpG. *Journal of Molecular Biology* **374**, 1104–1113 (2007).

49. Vinothkumar, K. R. *et al.* The structural basis for catalysis and substrate specificity of a rhomboid protease. *EMBO Journal* **29**, 3797–3809 (2010).

50. Vinothkumar, K. R. Structure of rhomboid protease in a lipid environment. *Journal of Molecular Biology* **407**, 232–247 (2011).

51. Xue, Y. & Ha, Y. Catalytic mechanism of rhomboid protease GlpG probed by 3,4-dichloroisocoumarin and diisopropyl fluorophosphonate. *Journal of Biological Chemistry* **287**, 3099–3107 (2012).

52. Xue, Y. *et al.* Conformational change in rhomboid protease GlpG induced by inhibitor binding to its S′ subsites. *Biochemistry* **51**, 3723–3731 (2012).

53. Vosyka, O. *et al.* Activity-based probes for rhomboid proteases discovered in a mass spectrometry-based assay. *Proceedings of the National Academy of Sciences of the United States of America* **110**, 2472–2477 (2013).

54. Xue, Y. & Ha, Y. Large lateral movement of transmembrane helix S5 is not required for substrate access to the active site of rhomboid intramembrane protease. *Journal of Biological Chemistry* **288**, 16645–16654 (2013).

55. 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**, 1051–1058 (2013).

56. Zoll, S. *et al.* Substrate binding and specificity of rhomboid intramembrane protease revealed by substrate–peptide complex structures. *The EMBO Journal* **33**, 2408–2421 (2014).

57. 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. *Molecular Cell* **61**, 329–340 (2016).

58. 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 porteolysis. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 750–754 (2007).

59. Brooks, C. L., Lazareno-Saez, C., Lamoureux, J. S., Mak, M. W. & Lemieux, M. J. Insights into substrate gating in H. influenzae rhomboid. *Journal of Molecular Biology* **407**, 687–697 (2011).

60. Urban, S., Schlieper, D. & Freeman, M. Conservation of intramembrane proteolytic activity and substrate specificity in prokaryotic and eukaryotic rhomboids. *Current Biology* **12**, 1507–1512 (2002).

61. del Rio, A., Dutta, K., Chavez, J., Ubarretxena-Belandia, I. & Ghose, R. Solution Structure and Dynamics of the N-terminal Cytosolic Domain of Rhomboid Intramembrane Protease from Pseudomonas aeruginosa: Insights into a Functional Role in Intramembrane Proteolysis. *Journal of Molecular Biology* **365**, 109–122 (2007).

62. Lohi, O., Urban, S. & Freeman, M. Diverse Substrate Recognition Mechanisms for Rhomboids: Thrombomodulin Is Cleaved by Mammalian Rhomboids. *Current Biology* **14**, 236–241 (2004).

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

64. Lazareno-Saez, C., Arutyunova, E., Coquelle, N. & Lemieux, M. J. Domain swapping in the cytoplasmic domain of the escherichia coli rhomboid protease. *Journal of Molecular Biology* **425**, 1127–1142 (2013).

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

66. Akiyama, Y. & Maegawa, S. Sequence features of substrates required for cleavage by GlpG, an Escherichia coli rhomboid protease. *Molecular Microbiology* **64**, 1028–1037 (2007).

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

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

69. Vinothkumar, K. R. & Freeman, M. Intramembrane proteolysis by rhomboids: Catalytic mechanisms and regulatory principles. *Current Opinion in Structural Biology* **23**, 851–858 (2013).

70. Cho, S., Baker, R. P., Ji, M. & Urban, S. Ten catalytic snapshots of rhomboid intramembrane proteolysis from gate opening to peptide release. *Nature Structural and Molecular Biology* **26**, 910–918 (2019).

71. Baker, R. P. & Urban, S. Architectural and thermodynamic principles underlying intramembrane protease function. *Nature Chemical Biology* **8**, 759–768 (2012).

72. 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. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 8257–8262 (2007).

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

74. Danielson, M. A. & Falke, J. J. Use of 19F NMR to probe protein structure and conformational changes. *Annual Review of Biophysics and Biomolecular Structure* **25**, 163–195 (1996).

75. Kitevski-LeBlanc, J. L. & Prosser, R. S. Current applications of 19F NMR to studies of protein structure and dynamics. *Progress in Nuclear Magnetic Resonance Spectroscopy* **62**, 1–33 (2012).

76. Didenko, T., Liu, J. J., Horst, R., Stevens, R. C. & Wüthrich, K. Fluorine-19 NMR of integral membrane proteins illustrated with studies of GPCRs. *Current Opinion in Structural Biology* **23**, 740–747 (2013).

77. Kim, T. H. *et al.* The role of ligands on the equilibria between functional states of a G protein-coupled receptor. *Journal of the American Chemical Society* **135**, 9465–9474 (2013).

78. Manglik, A. *et al.* Structural insights into the dynamic process of β2-adrenergic receptor signaling. *Cell* **161**, 1101–1111 (2015).

79. Horst, R., Liu, J. J., Stevens, R. C. & Wüthrich, K. β2-adrenergic receptor activation by agonists studied with 19F NMR spectroscopy. *Angewandte Chemie - International Edition* **52**, 10762–10765 (2013).

80. Frieden, C., Hoeltzli, S. D. & Bann, J. G. The Preparation of 19F-Labeled Proteins for NMR Studies. *Methods in Enzymology* **380**, 400–415 (2004).

81. Loewen, M. C. *et al.* Solution 19F nuclear Overhauser effects in structural studies of the cytoplasmic domain of mammalian rhodopsin. *Proceedings of the National Academy of Sciences of the United States of America* **98**, 4888–4892 (2001).

82. Jones, D. H. *et al.* Site-specific labeling of proteins with NMR-active unnatural amino acids. *Journal of Biomolecular NMR* **46**, 89–100 (2010).

83. Jackson, J. C., Hammill, J. T. & Mehl, R. A. Site-specific incorporation of a 19F-amino acid into proteins as an NMR probe for characterizing protein structure and reactivity. *Journal of the American Chemical Society* **129**, 1160–1166 (2007).

84. Shi, P. *et al.* Site-specific 19F NMR chemical shift and side chain relaxation analysis of a membrane protein labeled with an unnatural amino acid. *Protein Science* **20**, 224–228 (2011).

85. Kim, H. W., Perez, J. A., Ferguson, S. J. & Campbell, I. D. The specific incorporation of labelled aromatic amino acids into proteins through growth of bacteria in the presence of glyphosate. Application to fluorotryptophan labelling to the H+-ATPase of Escherichia coli and NMR studies. *FEBS Letters* **272**, 34–36 (1990).

86. Crowley, P. B., Kyne, C. & Monteith, W. B. Simple and inexpensive incorporation of 19F-Tryptophan for protein NMR spectroscopy. *Chemical Communications* **48**, 10681–10683 (2012).

87. Sykes, B. D. & Hull, W. E. Fluorine nuclear magnetic resonance studies of proteins. *Methods in Enzymology* **49**, 270–295 (1978).

88. Luck, L. A. & Falke, J. J. Open Conformation of a Substrate-Binding Cleft: 19F NMR Studies of Cleft Angle in the D-Galactose Chemosensory Receptor. *Biochemistry* **30**, 6484–6490 (1991).

89. Luchette, P. A., Prosser, R. S. & Sanders, C. R. Oxygen as a paramagnetic probe of membrane protein structure by cysteine mutagenesis and 19F NMR spectroscopy. *Journal of the American Chemical Society* **124**, 1778–1781 (2002).

90. Kim, H. J., Howell, S. C., van Horn, W. D., Jeon, Y. H. & Sanders, C. R. Recent advances in the application of solution NMR spectroscopy to multi-span integral membrane proteins. *Progress in Nuclear Magnetic Resonance Spectroscopy* **55**, 335–360 (2009).

91. Klein-Seetharaman, J., Getmanova, E. v., Loewen, M. C., Reeves, P. J. & Khorana, H. G. NMR spectroscopy in studies of light-induced structural changes in mammalian rhodopsin: Applicability of solution 19F NMR. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 13744–13749 (1999).

92. Chung, K. Y. *et al.* Role of Detergents in Conformational Exchange of a G. *Journal of Biological Chemistry* **287**, 36305–36311 (2012).

93. Ye, L., Larda, S. T., Frank Li, Y. F., Manglik, A. & Prosser, R. S. A comparison of chemical shift sensitivity of trifluoromethyl tags: optimizing resolution in 19F NMR studies of proteins. *Journal of Biomolecular NMR* **62**, 97–103 (2015).

94. Furter, R. Expansion of the genetic code: Site-directed p-fluoro-phenylalanine incorporation in Escherichia coli. *Protein Science* **7**, 419–426 (1998).

95. Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of Escherichia coli. *Science* **292**, 498–500 (2001).

96. Xie, J. *et al.* The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. *Nature Biotechnology* **22**, 1297–1301 (2004).

97. Xie, J. & Schultz, P. G. An expanding genetic code. *Methods* **36**, 227–238 (2005).

98. Kleckner, I. R. & Foster, M. P. An introduction to NMR-based approaches for measuring protein dynamics. *Biochimica et Biophysica Acta - Proteins and Proteomics* **1814**, 942–968 (2011).

99. Poget, S. F. & Girvin, M. E. Solution NMR of membrane proteins in bilayer mimics: Small is beautiful, but sometimes bigger is better. *Biochimica et Biophysica Acta - Biomembranes* **1768**, 3098–3106 (2007).

100. Garavito, R. M. & Ferguson-Miller, S. Detergents as Tools in Membrane Biochemistry. *Journal of Biological Chemistry* **276**, 32403–32406 (2001).

101. van den Brink-Van Der Laan, E., Antoinette Killian, J. & de Kruijff, B. Nonbilayer lipids affect peripheral and integral membrane proteins via changes in the lateral pressure profile. *Biochimica et Biophysica Acta - Biomembranes* **1666**, 275–288 (2004).

102. Knowles, T. J. *et al.* Membrane proteins solubilized intact in lipid containing nanoparticles bounded by styrene maleic acid copolymer. *Journal of the American Chemical Society* **131**, 7484–7485 (2009).

103. Vargas, C., Arenas, R. C., Frotscher, E. & Keller, S. Nanoparticle self-assembly in mixtures of phospholipids with styrene/maleic acid copolymers or fluorinated surfactants. *Nanoscale* **7**, 20685–20696 (2015).

104. Postis, V. *et al.* The use of SMALPs as a novel membrane protein scaffold for structure study by negative stain electron microscopy. *Biochimica et Biophysica Acta - Biomembranes* **1848**, 496–501 (2015).

105. Jamshad, M. *et al.* Structural analysis of a nanoparticle containing a lipid bilayer used for detergent-free extraction of membrane proteins. *Nano Research* **8**, 774–789 (2015).

106. Scheidelaar, S. *et al.* Effect of Polymer Composition and pH on Membrane Solubilization by Styrene-Maleic Acid Copolymers. *Biophysical Journal* **111**, 1974–1986 (2016).

107. Prabudiansyah, I., Kusters, I., Caforio, A. & Driessen, A. J. M. Characterization of the annular lipid shell of the Sec translocon. *Biochimica et Biophysica Acta - Biomembranes* **1848**, 2050–2056 (2015).

108. Lee, A. G. How lipids and proteins interact in a membrane: A molecular approach. *Molecular BioSystems* **1**, 203–212 (2005).

109. Landreh, M., Marty, M. T., Gault, J. & Robinson, C. v. A sliding selectivity scale for lipid binding to membrane proteins. *Current Opinion in Structural Biology* **39**, 54–60 (2016).

110. Lee, A. G. How lipids affect the activities of integral membrane proteins. *Biochimica et Biophysica Acta - Biomembranes* **1666**, 62–87 (2004).

111. Orwick-Rydmark, M. *et al.* Detergent-free incorporation of a seven-transmembrane receptor protein into nanosized bilayer lipodisq particles for functional and biophysical studies. *Nano Letters* **12**, 4687–4692 (2012).

112. Parmar, M. *et al.* Using a SMALP platform to determine a sub-nm single particle cryo-EM membrane protein structure. *Biochimica et Biophysica Acta - Biomembranes* **1860**, 378–383 (2018).

113. Sahu, I. D. *et al.* DEER EPR measurements for membrane protein structures via bifunctional spin labels and lipodisq nanoparticles. *Biochemistry* **52**, 6627–6632 (2013).

114. Gulati, S. *et al.* Detergent-free purification of ABC (ATP-binding-cassette) transporters. *Biochemical Journal* **461**, 269–278 (2014).

115. Dörr, J. M. *et al.* Detergent-free isolation, characterization, and functional reconstitution of a tetrameric K+ channel: The power of native nanodiscs. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 18607–18612 (2014).

116. Paulin, S. *et al.* Surfactant-free purification of membrane protein complexes from bacteria: Application to the staphylococcal penicillin-binding protein complex PBP2/PBP2a. *Nanotechnology* **25**, (2014).

117. Routledge, S. J. *et al.* Ligand-induced conformational changes in a SMALP-encapsulated GPCR. *Biochimica et Biophysica Acta - Biomembranes* **1862**, (2020).

118. Jamshad, M. *et al.* G-protein coupled receptor solubilization and purification for biophysical analysis and functional studies, in the total absence of detergent. *Bioscience Reports* **35**, 1–10 (2015).

119. Puthenveetil, R. & Vinogradova, O. Solution NMR: A powerful tool for structural and functional studies of membrane proteins in reconstituted environments. *Journal of Biological Chemistry* **294**, 15914–15931 (2019).

120. Gerig, J. T. Fluorine NMR of proteins. *Progress in Nuclear Magnetic Resonance Spectroscopy* **26**, 293–370 (1994).

121. Gerig, J. T. Fluorine nuclear magnetic resonance of fluorinated ligands. *Methods in Enzymology* **177**, 3–23 (1989).

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

123. Laemmli, U. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680–685 (1970).

124. Lee, S. C. *et al.* A method for detergent-free isolation of membrane proteins in their local lipid environment. *Nature Protocols* **11**, 1149–1162 (2016).

125. MacDonald, R. C. *et al.* Small-volume extrusion apparatus for preparation of large, unilamellar vesicles. *BBA - Biomembranes* **1061**, 297–303 (1991).

126. Reading, E. *et al.* Interrogating Membrane Protein Conformational Dynamics within Native Lipid Compositions. *Angewandte Chemie - International Edition* **56**, 15654–15657 (2017).

127. Hull, W. E. & Sykes, B. D. Fluorotyrosine alkaline phosphatase: Internal mobility of individual tyrosines and the role of chemical shift anisotropy as a 19F nuclear spin relaxation mechanism in proteins. *Journal of Molecular Biology* **98**, 121–153 (1975).

128. Chrisman, I. M. *et al.* Defining a conformational ensemble that directs activation of PPARγ. *Nature Communications* **9**, 1–16 (2018).

129. Hammill, J. T., Miyake-Stoner, S., Hazen, J. L., Jackson, J. C. & Mehl, R. A. Preparation of site-specifically labeled fluorinated proteins for 19f-nmr structural characterization. *Nature Protocols* **2**, 2601–2607 (2007).

130. Barniol-Xicota, M. & Verhelst, S. H. L. Stable and Functional Rhomboid Proteases in Lipid Nanodiscs by Using Diisobutylene/Maleic Acid Copolymers. *Journal of the American Chemical Society* **140**, 14557–14561 (2018).

131. Pearson, J. G., Oldfield, E., Lee, F. S. & Warshel, A. Chemical Shifts in Proteins: A Shielding Trajectory Analysis of the Fluorine Nuclear Magnetic Resonance Spectrum of the Escherichia coli Galactose Binding Protein Using a Multipole Shielding Polarizability-Local Reaction Field-Molecular Dynamics Approach. *Journal of the American Chemical Society* **115**, 6851–6862 (1993).

132. Schiller, P. W. *et al.* Direct electrophilic fluorination of tyrosine in dermorphin analogues and its effect on biological activity, receptor affinity and selectivity. **37**, 430–439 (1991).

133. Taylor, H. C. *et al.* "Active" conformation of an inactive semi-synthetic ribonuclease-S. *Journal of Molecular Biology* **149**, 313–317 (1981).

134. Feeney, J. *et al.* 19F nuclear magnetic resonance chemical shifts of fluorine containing aliphatic amino acids in proteins: Studies on Lactobacillus casei dihydrofolate reductase containing (2S,4S)-5-fluoroleucine. *Journal of the American Chemical Society* **118**, 8700–8706 (1996).

135. Hoeltzli, S. D. & Frieden, C. 19F NMR Spectroscopy of [6-19F]Tryptophan-Labeled Escherichia coli Dihydrofolate Reductase: Equilibrium Folding and Ligand Binding Studies. *Biochemistry* **33**, 5502–5509 (1994).

136. Danielson, M. A., Falke, J. J., Biemann, H. P. & Koshland, D. E. Attractant- and Disulfide-Induced Conformational Changes in the Ligand Binding Domain of the Chemotaxis Aspartate Receptor: A 19F NMR Study. *Biochemistry* **33**, 6100–6109 (1994).

137. Bourret, R. B., Drake, S. K., Chervitz, S. A., Simon, M. I. & Falke, J. J. Activation of the phosphosignaling protein CheY. II. Analysis of activated mutants by 19F NMR and protein engineering. *Journal of Biological Chemistry* **268**, 13089–13096 (1993).

138. 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. *Journal of the American Chemical Society* **128**, 4389–4397 (2006).

139. Hoai-Thu Truong, N., Pratt, E. A. & Ho, C. Interaction of the Membrane-Bound D-Lactate Dehydrogenase of Escherichia coli with Phospholipid Vesicles and Reconstitution of Activity Using a Spin-Labeled Fatty Acid as an Electron Acceptor: A Magnetic Resonance and Biochemical Study. *Biochemistry* **30**, 3893–3898 (1991).

140. Ahmed, N., Foss, D. v, Powdrill, M. H. & Pezacki, J. P. Site-Specific Cross-Linking of a p19 Viral Suppressor of RNA Silencing Protein and Its RNA Targets Using an Expanded Genetic Code. **58**, 3520–3526 (2019).

141. Studier, F. W. Protein production by auto-induction in high density shaking cultures. *Protein expression and purification* **41**, 207–234 (2005).

142. Morrison, K. A. *et al.* Membrane protein extraction and purification using styrene-maleic acid (SMA) copolymer: Effect of variations in polymer structure. *Biochemical Journal* **473**, 4349–4360 (2016).

143. Cuevas Arenas, R., Klingler, J., Vargas, C. & Keller, S. Influence of lipid bilayer properties on nanodisc formation mediated by styrene/maleic acid copolymers. *Nanoscale* **8**, 15016–15026 (2016).

144. Zhang, R., Sahu, I. D., Bali, A. P., Dabney-Smith, C. & Lorigan, G. A. Characterization of the structure of lipodisq nanoparticles in the presence of KCNE1 by dynamic light scattering and transmission electron microscopy. *Chemistry and Physics of Lipids* **203**, 19–23 (2017).

145. Oluwole, A. O. *et al.* Solubilization of Membrane Proteins into Functional Lipid-Bilayer Nanodiscs Using a Diisobutylene/Maleic Acid Copolymer. *Angewandte Chemie - International Edition* **56**, 1919–1924 (2017).

146. Krishnarjuna, B., Ravula, T. & Ramamoorthy, A. Detergent-free extraction, reconstitution and characterization of membrane-anchored cytochrome-b5 in native lipids. *Chemical Communications* **56**, 6511–6514 (2020).

147. Ravula, T., Ramadugu, S. K., di Mauro, G. & Ramamoorthy, A. Bioinspired, Size-Tunable Self-Assembly of Polymer–Lipid Bilayer Nanodiscs. *Angewandte Chemie - International Edition* **56**, 11466–11470 (2017).

148. Bai, J. *et al.* Expression, purification, and functional reconstitution of 19F-labeled cytochrome b5 in peptide nanodiscs for NMR studies. *Biochimica et Biophysica Acta - Biomembranes* **1862**, 183194 (2020).

149. Hagn, F., Etzkorn, M., Raschle, T. & Wagner, G. Optimized phospholipid bilayer nanodiscs facilitate high-resolution structure determination of membrane proteins. *Journal of the American Chemical Society* **135**, 1919–1925 (2013).

150. Puthenveetil, R. & Vinogradova, O. Optimization of the design and preparation of nanoscale phospholipid bilayers for its application to solution NMR. *Proteins: Structure, Function and Bioinformatics* **81**, 1222–1231 (2013).

151. 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).

152. 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).

153. Raschle, T. *et al.* Structural and functional characterization of the integral membrane protein VDAC-1 in lipid bilayer nanodiscs. *Journal of the American Chemical Society* **131**, 17777–17779 (2009).

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

155. Nasr, M. L. *et al.* Covalently circularized nanodiscs for studying membrane proteins and viral entry. *Nature Methods* **14**, 49–52 (2016).