uOttawa

Investigation of GlpG rhomboid protease using site-specific ^{19}F NMR

Medhani Mohottalage

A thesis submitted in partial fulfillment of the requirements

for the Master of Science degree in Chemistry

Department of Chemistry and Biomolecular Sciences

Faculty of Science

Abstract

The rhomboid family of intramembrane serine proteases are capable of cleaving transmembrane protein substrates embedded within the phospholipid membrane. This allows them to play a critical role in a wide range of biological processes enabling their use as potential targets for the development of therapeutic compounds. Although X-ray crystallography and functional studies have revealed insights into the rhomboid mechanism of proteolysis, questions remain regarding how transmembrane substrates gain access to the water-filled active site from the hydrophobic lipid environment. Currently, there exists a growing body of evidence suggesting that helix 5 (α5) undergoes conformational dynamics that facilitate substrate entry into the rhomboid active site. In this thesis, 1D ^{19}F solution-state NMR spectroscopy was used to monitor dynamics around the α5 gate of the rhomboid protease from *E. coli*, *ec*GlpG. This was achieved using site-specific labelling of cysteine residues with thiol-reactive probes bearing trifluoromethyl groups. A site in the interface of the α5 gate was labeled and well resolved ^{19}F NMR spectra were obtained, showing evidence of two-state exchange. Although promising, complications from background labeling of a single endogenous cysteine residue required the generation of a cysteine knockout mutant. The cysteine knockout retained function and stability, but at lower levels than seen for the wild-type protein. Overall, these results demonstrate the potential for ^{19}F NMR to be used to monitor α5 gating dynamics, although further work is needed to optimize labels and mutation sites.

Acknowledgements

First and foremost, I would like to express my deepest appreciation to my supervisor, Dr. Natalie Goto, for accepting me into her research lab as an undergraduate student, and for providing me with her guidance and mentorship throughout my masters research project. She has provided me with copious opportunities to advance my scientific development by providing a welcoming, enriched, and unbiased lab environment. I am grateful for her encouragement, her patience and her direction throughout the trials and tribulations of my research work. I would also like to thank her for her generosity and the time she has spent helping me prepare for presentations and written documents including this thesis. Dr. Goto's assistance throughout the years has really shaped and developed my professionalism both in written and oral scientific language to which I am endlessly thankful for.

To James Davey, and Anwar Hassan, previous members of the Goto lab, thank you for your helpful discussions and support. To Jamie, I am grateful for your philanthropy and to have learned a multitude of experimental practises surrounding mutagenesis and cell culture work. To Anwar, thank you for his guidance and direction in the early stages of my research project. To Brittany, I've enjoyed troubleshooting machine mishaps and our helpful scientific conversations – thank you for being such an inviting, and easygoing lab mate to work alongside. I would also like to acknowledge and thank the contributions from Amanda Featherstone, an undergraduate student for her assistance with various experiments throughout my research.

I would also like to thank Dr. Roberto Chica and Dr. Jeffery Keillor who provided our lab generous use of their instruments being essential for the work conducted in this thesis. Additionally, my deepest thanks go out to Patrick Szell and Peter Pallister from the NMR facility – thank you both for your unwavering help with experimental troubleshooting in regards to the NMR spectrometers.

Last, but not least, I give my deepest thanks to my close friends and family, specially, to my father Susantha, my mother Dayani, and sister Dhanuddara. Thank you for your patience, support, and endless amounts of love during my very stressful mornings and tiresome nights – you have never failed to be understanding, considerate and encouraging.

Table of contents

List of abbreviations

5FW	5-fluorotryptophan
7FI	7-fluoroindole
Aβ	Amyloid beta peptide
AarA	*Providencia stuartii* rhomboid protease
AEBSF	4-(2-aminoethyl)benzenesulfonyl fluoride hydrochloride
APP	Amyloid precursor protein
BSA	Bovine serum albumin
BODIPY	Boron dipyrromethene
CBB	Coomassie brilliant blue
CD	Circular dichroism
CMC	Critical micelle concentration
CPEC	Circular polymerase extension cloning
CSA	Chemical shift anisotropy
CytD	Cytoplasmic domain
DFP	Diisopropylfluorophosphate
DDM	n-dodecyl β-D-maltoside
DMSO	Dimethyl sulfoxide
DPC	n-dodecylphosphocholine
DPS	2,2'-dipyridyl disulfide
***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
EPR	Electron paramagnetic resonance
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
***hi*GlpG**	Haemophilus influenzae rhomboid
HSQC	Heteronuclear single quantum coherence
IMP	Intramembrane protease
IPTG	Isopropyl-β-D-thiogalactopyranoside
kDa	Kilodalton
L1	Loop 1 region
L5	Loop 5 region
LB	Lennox broth
M2M	1,2-bis(methylsulfonylsulfanyl)ethane
MSP	Membrane scaffold protein
MW	Molecular weight
Ni-NTA	Nickel-nitrilotriacetic acid
NMR	Nuclear magnetic resonance
OD	Optical density
PARL	Presenilin-associated rhomboid-like protease
PDB	Protein database (RCSB)
PDC	Protein detergent complex
pfRom1	*P. falciparum rhomboid 1*
pfRom4	*P. falciparum rhomboid 4*
PRE	Paramagnetic relaxation enhancement
RF	Radio frequency
RHBDL2	Human rhomboid related protein 2
RHBDL4	Human rhomboid related protein 4
Rhom-1	*Drosophila* rhomboid-1 protease
Rhom-4	*Drosophila* rhomboid-4 protease

SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SEC	Size exclusion chromatography
SOE	Splicing by overlap extension
ssNMR	Solid-state nuclear magnetic resonance
Tat	Twin arginine translocase
TatA	Twin arginine translocase, component A
TatB	Twin arginine translocase, component B
TatC	Twin arginine translocase, component C
TCEP	Tris(2-carboxyethyl)phosphine
TEMED	Tetramethylethylenediamine
TET	2,2,2-trifluoroethanethiol
TFMB	3,5-bis(trifluoromethyl)benzyl bromide
T. gondii	*Toxoplasma gondii*
***tg*Rom4**	Rhomboid 4 from *toxoplasma gondii*
***tg*Rom5**	Rhomboid 5 from *toxoplasma gondii*
TMD	Transmembrane domain
TM	Transmembrane
TMD81	*ec*GlpG Transmembrane domain residues 81-276
TRIS	2-Amino-2-(hydroxymethyl)propane-1,3-diol
tRNA	Transfer ribonucleic acid
WT	Wild type

List of figures

List of tables

List of equations

1 Introduction

1.1 Intramembrane proteases

Intramembrane proteases (IMP) are unique enzymes that transverse the membrane bilayer and cleave peptide bonds of transmembrane substrates (TM). The hydrophobic nature of the bilayer presents a unique challenge for the entry of water molecules essential for peptide bond cleavage. However, crystal structures have resolved that water molecules can gain access to the active site for hydrolysis to take place[1-3,4]. The catalytic residues of the protease active site are localized within the membrane in the multi-spanning transmembrane domain (TMD)[4]. Depending on the protease, the active site is capable of cleaving both solvent exposed and/or buried regions of transmembrane substrates as they become unstructured when exposed to the aqueous environment of the catalytic cavity[4-6]. This unique capability is utilized in a range of biological processes such as membrane protein homeostasis (proteostasis) as well as the propagation of cell signalling events[4,5,7,8]. There are 4 families of IMPs classified by their catalytic mechanism: metallo-, aspartyl-, glutamyl- and serine IMPs[8]. Metallo- and aspartyl-IMPs use an activated water molecule, bound to either a metal ion or aspartate residue(s) as their nucleophile to perform proteolysis on the carbonyl carbon of their peptide substrate[9,10]. Glutamyl- and serine IMPs contain their namesake residue as the nucleophile within their active sites[11,12]. In this thesis the focus has been a member of the serine IMP family, namely the rhomboid protease.

1.2 The rhomboid protease

Of all the IMPs, rhomboid proteases are the most widespread[13], being found in almost all genomes of archaea, bacteria, and eukaryotes[14] playing a role in a wide range of biological processes[15].

The first rhomboid to be discovered is the *Drosophila* Rhomboid-1 (Rhom-1) that plays a role in cell signalling through the cleavage of a membrane-anchored signalling molecule – Spitz. Cleavage of Spitz by Rhom-1 releases its soluble epidermal growth factor (EGF) domain which is transported out of the cell to activate the EGF receptor (EGFR) in surrounding cells to regulate growth and development[16] (Fig. 1.0A). The prokaryotic rhomboid AarA from *Providencia stuartii* also plays a role in cell signalling as it is responsible for the proteolytic removal of an N-terminal extension in twin arginine translocase component A (TatA)[17]. This is required for TatA to participate in the formation of the complete twin-arginine translocation complex with components TatB and TatC, allowing the export of virulence factors for quorum sensing and biofilm formation[18-20].

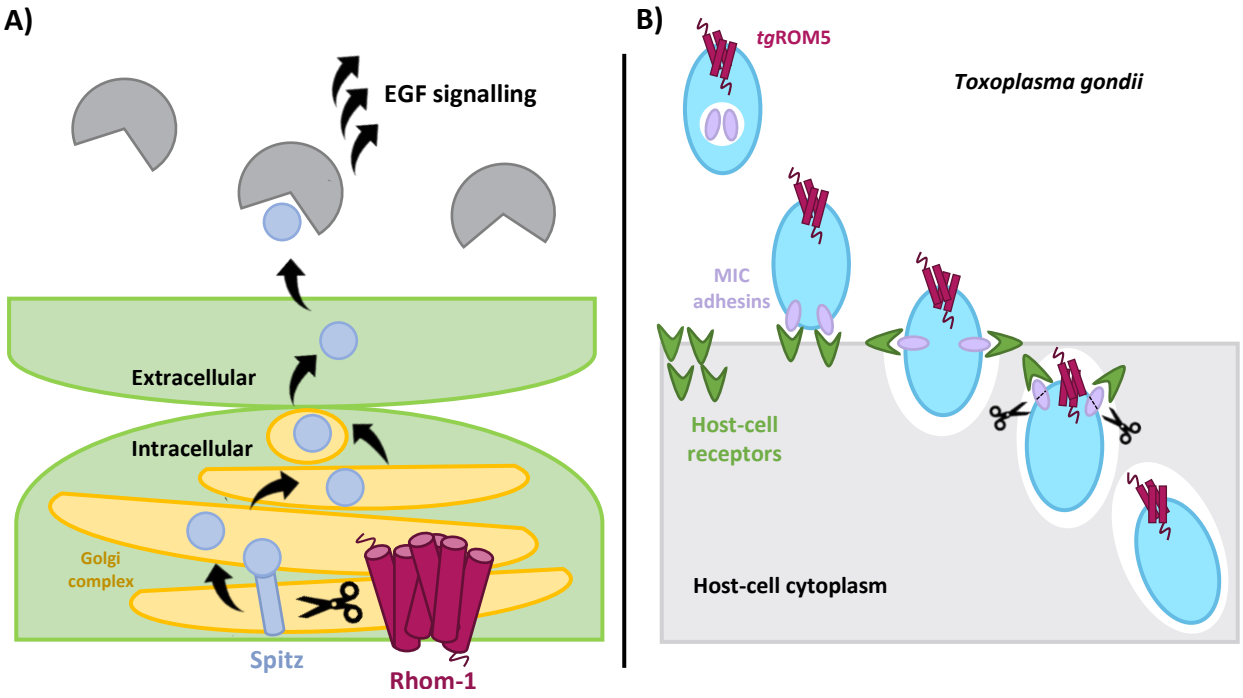


Figure 1.0: Schematic representations illustrating the biological roles of rhomboids (A) The *Drosophila* Rhomboid-1 (magenta) cleaves the Spitz transmembrane domain (purple) to release the signaling molecule out of the cell for downstream epidermal growth factor (EGF) signaling. **(B)** Invasion of the *T. gondii* parasite into host cells is mediated by the *tg*Rom5 (magenta). *T. gondii* cell surface adhesins (purple) bind to host-cell receptors (green) and are cleaved during host-cell invasion to release the parasite.

Rhomboids have also been shown to be key players in protein degradation. For example, the human rhomboid RHBDL4 is localized to the endoplasmic reticulum (ER) where it plays a role in the degradation of incorrectly folded proteins as part of the ER-associated degradation (ERAD) pathway.[21,22] RHBDL4 is upregulated during ER stress where its conserved C-terminal domain is capable of binding ubiquitinated TM substrates to target them for cleavage. Additionally, RHBDL4 directly interacts with the p97 AAA+ ATPase which acts as a mediator downstream of substrate cleavage to facilitate the excision of substrate fragments through the proteasome.[21] Recently, it was also reported that the RHBDL4 rhomboid could cleave a key molecule in the etiology of Alzheimer's disease - the amyloid precursor protein (APP) which could circumvent the cleavage by secretases in the plasma membrane that generates the toxic amyloid-forming cleavage product Aβ.[23]

Rhomboids also play an important role in the pathogenicity of the apicomplexan parasite – *Toxoplasma gondii* (*T. gondii*) which encodes five *T. gondii* rhomboids. The cell surface *T. gondii* rhomboids; *tg*Rom5 and *tg*Rom4 were found to be responsible for the cleavage of *T. gondii* surface adhesins – identified as rhomboid substrates[24]. These adhesins are cleaved as they bind to the host-cell receptors which is required for their entry into the host cell during internalization (Fig. 1.0B)[25]. Likewise, the *Plasmodium falciparum* rhomboids – *pf*Rom4 and *pf*Rom1 homologues of *tg*Rom5 have a similar role whereby cleavage of diverse cell surface adhesins by these rhomboids were found to be required for the release of the mature parasite into the host cell, completing the invasion process[26]. These cleavage events are vital as these adhesins have been identified as causative agents of malaria.

To date, the role of the *E. coli* rhomboid homologue – *ec*GlpG is not as well defined. Initially, it was suggested that *ec*GlpG may play a role in glycerol metabolism since the gene encoding *ec*GlpG was localized in the glpEGR operon which also encodes a GlpR transcriptional repressor responsible for glycerol-3-phophate regulation[27]. However, *ec*GlpG activity was not found to be required for *E. coli* growth and replication[28,29]. In more recent years, it was found that disruption in *ec*GlpG production in extraintestinal pathogenic strains of *E. coli* (ExPEC)

impaired gut colonization and pathogenicity *in vivo* and had implications on the activity of glpR *in vitro*[30]. These results suggest a novel biological role for *ec*GlpG in ExPEC pathogenicity through the activation of glycerol metabolic pathways, though the precise mechanism of this regulation and the endogenous substrates involved remain unknown.

Taken together, the studies highlighted in this section show rhomboids as members of a diverse range of systems participating in various biological processes, all of which are dependent upon its ability to perform hydrolytic cleavage within the membrane environment. Understanding these systems can allow rhomboids to be used as potential targets for the development of therapeutic compounds in processes such as the modulation of signal transduction pathways or the inhibition of parasite invasion. Correspondingly, there is significant interest in understanding the structural mechanism by which this unique proteolytic event can occur.

1.3 Rhomboid structure and function

1.3.1 Catalytic mechanism

Rhomboids use a serine protease mechanism to catalyze the hydrolysis of peptide bonds. To achieve this, an activated serine side chain acts as a nucleophile, attacking the peptide carbonyl carbon which produces a negatively charged tetrahedral intermediate stabilized by an oxyanion hole (Fig. 1.1). Collapse of the tetrahedral intermediate generates an acyl-enzyme intermediate while releasing the N-terminal cleavage product. This intermediate is subsequently hydrolyzed via formation of a second tetrahedral intermediate that releases the C-terminal cleavage product and regenerates the nucleophilic serine. Unlike typical serine proteases, the rhomboid active site is composed of a catalytic dyad instead of a triad[3] lacking the aspartic residue required to align and activate the indole side chain of the histidine. Early studies on *Drosophila* Rhomboid-1 with the Spitz substrate suggested a conserved asparagine

residue (N154) from $\alpha 2$ as a prospective catalytic triad member[16] however, crystal structures have shown that N154 is not proximal to the activating histidine in order to stabilize it in the correct tautomeric configuration.

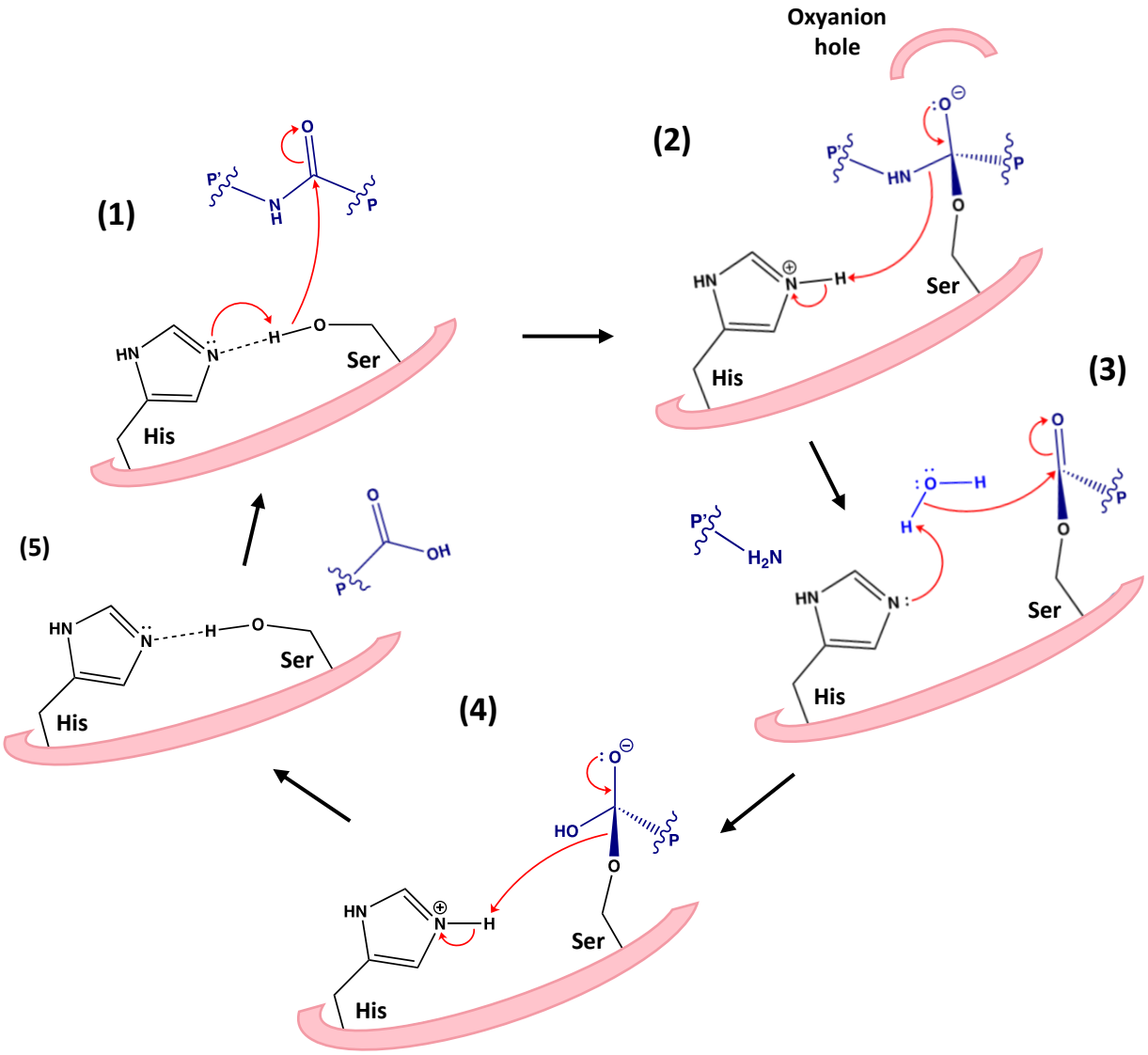


Figure 1.1: The mechanism of rhomboid peptide hydrolysis. (1) Initial nucleophilic attack on peptide by histidine-activated serine nucleophile **(2)** Formation of the first tetrahedral intermediate stabilized by an oxyanion hole **(3)** Collapse of the tetrahedral intermediate leads to the formation of the acyl-enzyme intermediate, the release of the N-terminal cleavage product and subsequent hydrolysis **(4)** The formation of a second tetrahedral intermediate **(5)** Release of the cleavage products and regeneration of the rhomboid active site.

Classic nomenclature is typically used to describe the substrate cleavage site and protease binding sites. According to this classification (Fig. 1.2), cleavage occurs between the P1 and P1' positions of the substrate, where Px or Px' represents the residues of increasing distance N-terminal or C-terminal to the cut site respectively. This same terminology can be used to describe interaction pockets, called Sx/Sx', that interact with the substrate Px/Px' side chains.

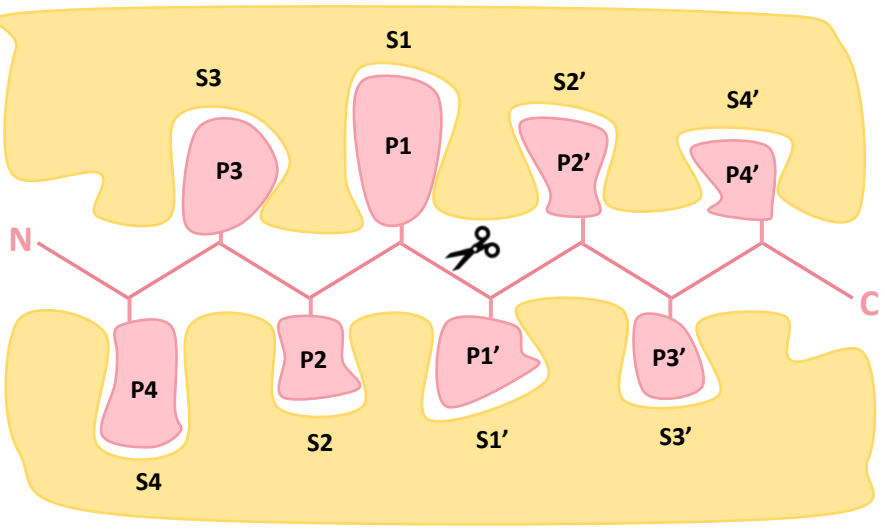


Figure 1.2: Schematic representation of the protease cleavage site showing the interactions between substrate side chains (P4-P4') and the substrate-binding pockets (S4-S4'). The location of the sissle bond is represented by the scissors.

1.3.2 Rhomboid structure

While there exist a variety of rhomboids in various kingdoms of life, the *Escherichia coli* rhomboid protease (*ec*GlpG) has been widely used as a model system for the study of rhomboid function and structure. In 2006, the first high-resolution X-ray crystal structures of *ec*GlpG became available[3]. To date, over 30 structures have been acquired, all resolving the catalytic TMD showing a bundle of six TM helices connected by small loops with a larger L1 loop that

protrudes out of the main body connecting helices 1 and 2 (Fig. 1.3). All structures were acquired using detergent micelles[31-34,35,36,37] or bicelles[33] and some in complex with inhibitors[31-34,36,38]. The highly conserved active site Ser-His catalytic dyad suggests that many of the findings made with this system will reveal general insights that apply for all members of this family. Meanwhile, there is an accumulated wealth of structural and biochemical information on this protease due to the relative ease of its purification and characterization. While the majority of rhomboids, including *ec*GlpG, contain 6 TM helices, there are some members, such as *Providencia* AarA, *Drosophila* Rhomboid-1 and the mitochondrial rhomboid PARL that are suggested to have an additional TM helix.[42]

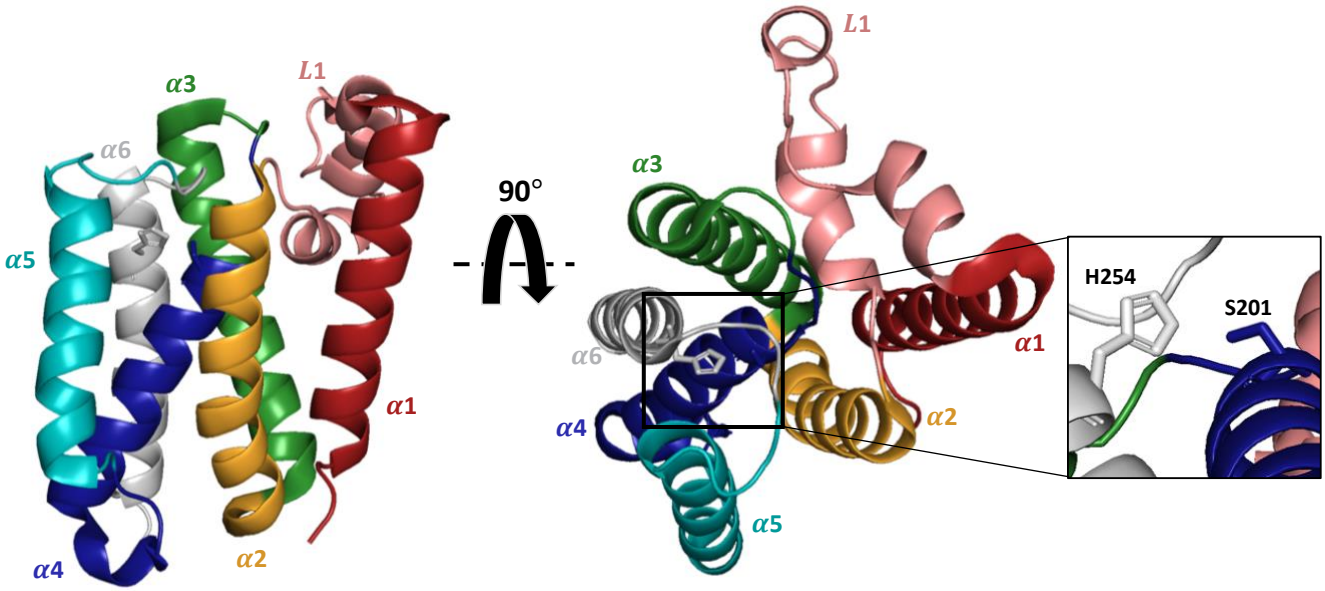


Figure 1.3: Crystal structure of *ec*GlpG (PBD 2ic8) showing the core domain that contains six transmembrane helices (α1-α6) encompassing the hydrophilic active site. The catalytic dyad is shown in the inset to include Ser201 and His254 on α6 and α4 respectively.

Many rhomboids including *ec*GlpG have one or more soluble domain(s), usually on the cytoplasmic side of the membrane, that may regulate rhomboid function. For example, *Drosophila* Rhom-4 contains a calcium-binding motif in a cytoplasmic domain that aids in the transport of calcium ions reported to increase the cleavage rate of Spitz by promoting substrate gating.[43] Likewise, removal of an N-terminal cytoplasmic domain in the mammalian rhomboid

RHBDL2 significantly reduced its activity against thrombomodulin which is a potential substrate for this rhomboid.[44,22] The bacterial rhomboid *ec*GlpG has a unique cytoplasmic domain at its N-terminus that does not seem to be required to maintain active site integrity and catalytic activity[39]. Interestingly, it was found that the structural integrity of *ec*GlpG is compromised by interactions between this cytoplasmic domain and detergent micelles which can affect its activity [40]. For this reason, a truncated variant of *ec*GlpG, namely TMD81 (residues 81-276) consisting of just the transmembrane domain was used for structural and functional analysis in this thesis. This truncation mutant has similar activity to *ec*GlpG in detergent micelles against a water-soluble fluorogenic casein substrate[39,41] and in crude membrane extracts using a fluorophosphonate-rhodamine probe[39].

The catalytic machinery of *ec*GlpG is located at a depth that is approximately 10Å below the membrane surface in a water-filled cavity. A conserved GxSx motif near the active serine on α4 interacts with the AHxxGxxxG motif on α6, which helps to keep these two catalytic residues within hydrogen bonding distance to one another.[45] Crystallographic studies of GlpG in complex with isocoumarin inhibitors suggested that the oxyanion hole is formed by side chains of N154 and H150, along with backbone atoms from L200 and S201.[34,38] Subsequent studies with a β-lactam inhibitor[32] and a peptide-aldehyde inhibitor[33] that better mimics the tetrahedral intermediate provided additional support for the role of these residues, with the backbone amide of S201 being within hydrogen bonding distance (2.7Å and 3Å) to the side chains of H150 and N154 (Fig. 1.4).

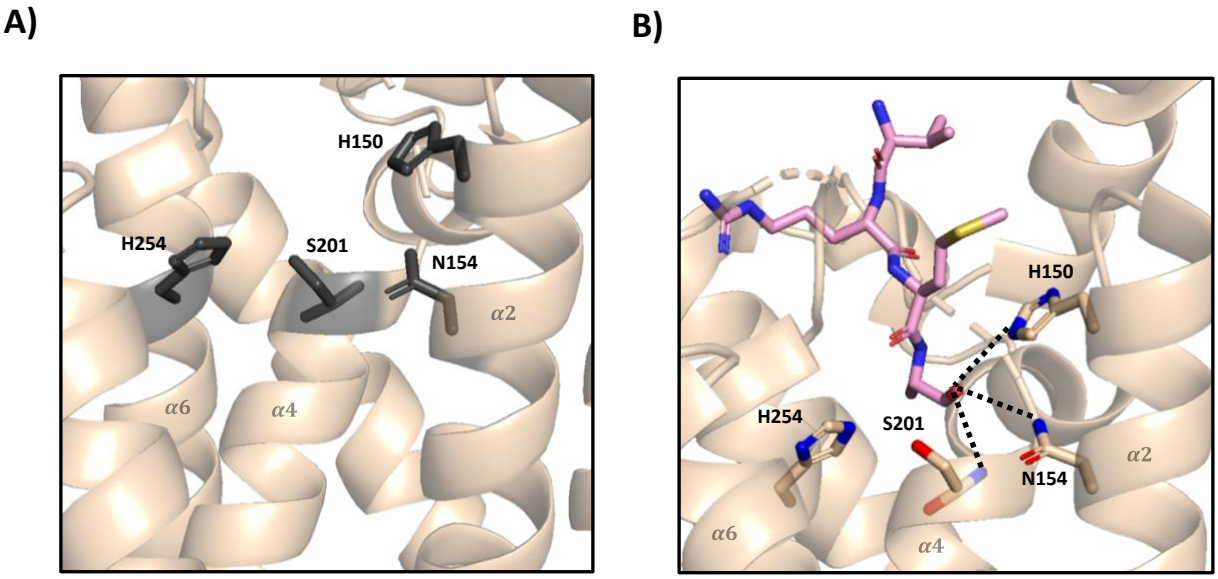


Figure 1.4: (A) Detailed view of the *ec*GlpG active site and residues postulated to encompass the *ec*GlpG oxyanion hole. **(B)** Crystal structure of *ec*GlpG in complex with a peptide-aldehyde inhibitor illustrating the position of the oxyanion hole formed by hydrogen bonding interactions between the inhibitor oxyanion and the side chain protons from H250 and N154 with the backbone amide proton of Ser201 (PDB 5F5K).

The sequence selectivity of *ec*GlpG is relatively broad, consisting of a large hydrophobic residue at the P4 position, a small residue at the P1 position adjacent to the cleavage site and a hydrophobic residue at the P2' position as illustrated in Fig. 1.5.

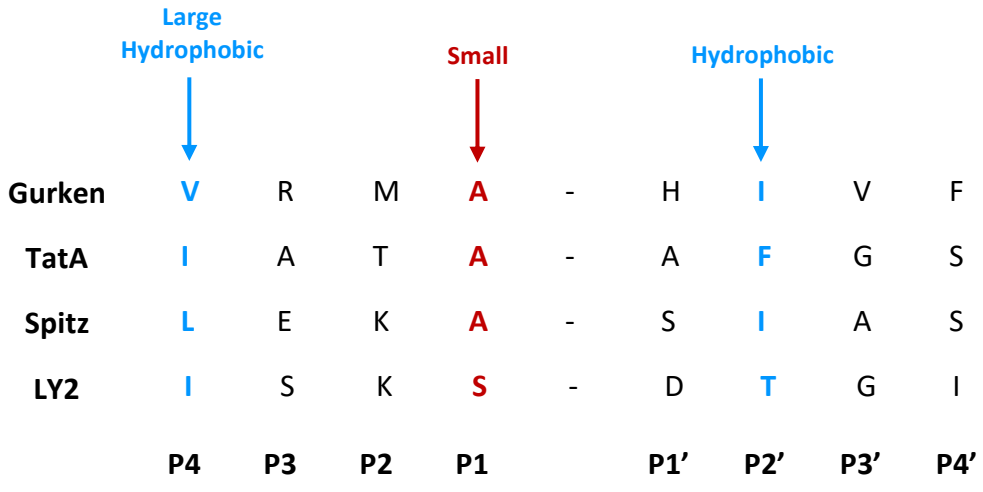


Figure 1.5: ***ec*GlpG cleavage sites for various commonly employed TM substrates**, illustrating the sequence features associated with *ec*GlpG rhomboid substrates[47]. Note that the P2' residue in LY2 is not considered to be a hydrophobic residue.

X-ray crystal structures of *ec*GlpG in complex with peptide-based inhibitors have provided insight into the structural basis for this sequence specificity. A recent structure shows *ec*GlpG in complex with a covalently bound Gurken-derived tetrapeptide (sequence VRMA) with an aldehyde replacing the C-terminal carboxyl group where its P1-Ala side chain carbonyl was found to hydrogen bond to the backbone amide of the G198 residue[33]. Although peptide-bound and apo structures of the S1 site show a large cavity, it has been suggested that this site functions as both a water retention site and the S1 binding pocket, giving rise to the requirement for smaller side chain residues in the P1 position[33,46]. Additional hydrogen bonding interactions were found between the Gurken substrate at P2-Met with residues A250 and S248 as well as P3-Arg with residues G198 and W196. Interestingly, a mutation at the P3-Arg to a small alanine residue as occurs naturally in the TatA substrate displayed a 10-fold decrease in binding affinity, potentially indicating the importance of its interactions. The large side chain of the P4-Val side chain was found to interact with a deep hydrophobic pocket in the active site, which explained the observed preference for larger hydrophobic side chains at this site. Kinetic studies with longer peptide substrates derived from Gurken including P5 and P6 positions (sequence RKVRMA) showed stronger binding, which suggested additional contacts may be made with rhomboid, although electron density was not resolved for residues at these positions. When in complex with a longer Gurken peptide substrate, electron density could not be resolved for residues beyond the P1' position. This is consistent with structural and functional studies indicating that residues beyond P2' do not directly interact with the active site[33,47,48]. Overall, the consensus sequence for *ec*GlpG is quite broad, allowing it to cleave various physiological substrates[49].

1.4 Substrate gating in rhomboid protease

The diverse range of biological functions mediated by rhomboids in all kingdoms of life relies on their ability to cleave peptide bonds in membrane protein substrates. Although the

mechanism of rhomboid proteolysis is well-conserved across various organisms[3,50], how TM substrates enter the water-filled active site from the hydrophobic phase of the lipid membrane is still not well understood. Various models have been suggested, with conformational dynamics in α5 identified as the most likely mechanism for substrate entry into the active site. This gating hypothesis is supported by activity assays with mutations that disrupt interactions between α5 and its neighbouring α2 helix showing a drastic increase in protease activity. In contrast, mutations that allow covalent cross-links to be formed between two helices abrogated activity[39,51-55] (Table 1.0). Further evidence for the role of α5 dynamics in substrate gating was obtained by electron paramagnetic resonance (EPR) studies with a nitroxide spin-labelled residue in the α5 gate. This study showed increased α5 dynamics for more active mutants over mutants having lower activity, and also increased dynamics in micelles relative to lipid bilayers which also correlates with similar differences in activity.[6] These studies suggest that conformational dynamics contributes to the rate-determining step in the rhomboid catalytic cycle.

Table 1.0: Summary of the impact on rate by mutations around the substrate gate measured using various TM and water-based substrates.

Mutant	Mutant location	Impact of activity (approximate fold enhancement)	Substrate (*)	Reference
L229V/F232V/W236V	α5	4-fold increase	Spitz	Baker, R. P. *et al.* 2007[51]
F153A/W236A	α2 & α5	10-fold increase	Spitz	Baker, R. P. *et al.* 2007[51]
		5-fold increase	TatA	Bohg, C. *et al.* 2023[53]
		40-fold increase	Gurken	Xue, Y. & Ha, Y. 2013[36]
		15-fold increase	Spitz	Baker, R. P. & Urban, S. 2012[54]
		2-fold increase	Casein	Sherratt, A. R. *et. al.* 2012[39]
		3-fold increase	TatA	Dickey, S. W. et al. 2013[55]
W157A/F232A	α2 & α5	7-fold increase	Spitz	Baker, R. P. *et al.* 2007[51]
F153A	α2	2.5-fold increase	TatA	Bohg, C. *et al.* 2023[53]
		20-fold increase	Gurken	Xue, Y. & Ha, Y. 2013[36]
		5-fold increase	Spitz	Baker, R. P. & Urban, S. 2012[54]
W236A	α5	2.5-fold increase	TatA	Bohg, C. *et al.* 2023[53]
		5-fold increase	Spitz	Baker, R. P. & Urban, S. 2012[54]

W157A	$\alpha 2$	2-fold increase	Gurken	Xue, Y. & Ha, Y. 2013[36]
F153A/W157A	$\alpha 2$	5-fold increase	Spitz	Baker, R. P. & Urban, S. 2012[54]
F232A/W236A	$\alpha 5$ & $\alpha 5$	5-fold increase	Spitz	Baker, R. P. & Urban, S. 2012[54]
L229A/F232A/W236A	$\alpha 5$	5-fold increase	Spitz	Baker, R. P. & Urban, S. 2012[54]
F153A/W157A/F232A/W236A	$\alpha 2$ & $\alpha 5$	5-fold increase	Spitz	Baker, R. P. & Urban, S. 2012[54]
W157C/F232C	$\alpha 2$ & $\alpha 5$	7-fold decease	TatA	Bohg, C. *et al.* 2023[53]
		4-fold decrease	Spitz	Baker, R. P. *et al.* 2007[51]
Y160C/L229C	$\alpha 2$ & $\alpha 5$	2-fold decrease	TatA	Bohg, C. *et al.* 2023[53]
		7-fold decrease	Spitz	Baker, R. P. *et al.* 2007[51]
F153C/W236C	$\alpha 2$ & $\alpha 5$	8-fold decrease	TatA	Bohg, C. *et al.* 2023[53]

* All substrates are TM segments from known rhomboid substrates except for casein (water-soluble model substrate).

In 2006, a crystal structure of GlpG revealed an open conformation with the C-terminal half of $\alpha 5$ bent $\sim 35°$ away from the enzyme core (Fig. 1.6A). This opened a path for substrate entry into the active site that was accompanied by concomitant movement of the L5 cap.[35] Although this gave rise to the proposal that this represented a physiologically relevant open state for GlpG, this structure was only captured in detergent micelles, and was not reproduced in any of the numerous other structures of *ec*GlpG raising the possibility that this conformation was an artifact induced by crystal packing interactions. A subsequent crystal structure of *ec*GlpG covalently bound to the irreversible inhibitor diisopropylfluorophosphonate (DFP) revealed a small lateral shift of $\alpha 5$ by $\sim 1\text{Å}$ which gave rise to an alternate model for $\alpha 5$ gating[38] (Fig. 1.6B). Later crystal structures provided more evidence for this lateral model, with GlpG covalently modified by a peptide aldehyde derived from a substrate sequence showing a lateral displacement of $\alpha 5$ by $\sim 5\text{Å}$ away from $\alpha 2$. Although a promising result, this structure was only captured with the inactive Y205F mutant and in a membrane bicelle environment. Attempts to capture this structure with wild-type samples and in detergents micelles failed to reproduce the same shift.[33] However, time-resolved crystal structures captured for wild-type *ec*GlpG in a bicelle with a 13-mer peptide substrate showed a loss of electron density across the entire $\alpha 5$ helix and most of the L5 loop, suggesting a more disordered state that may be required for gate opening[48].

Together, these studies represent a growing body of evidence in support of a vital role for α5 in substrate gating. However, the structural basis of substrate gating and the physiological role of the open conformation are not well understood and will be explored further in this thesis.

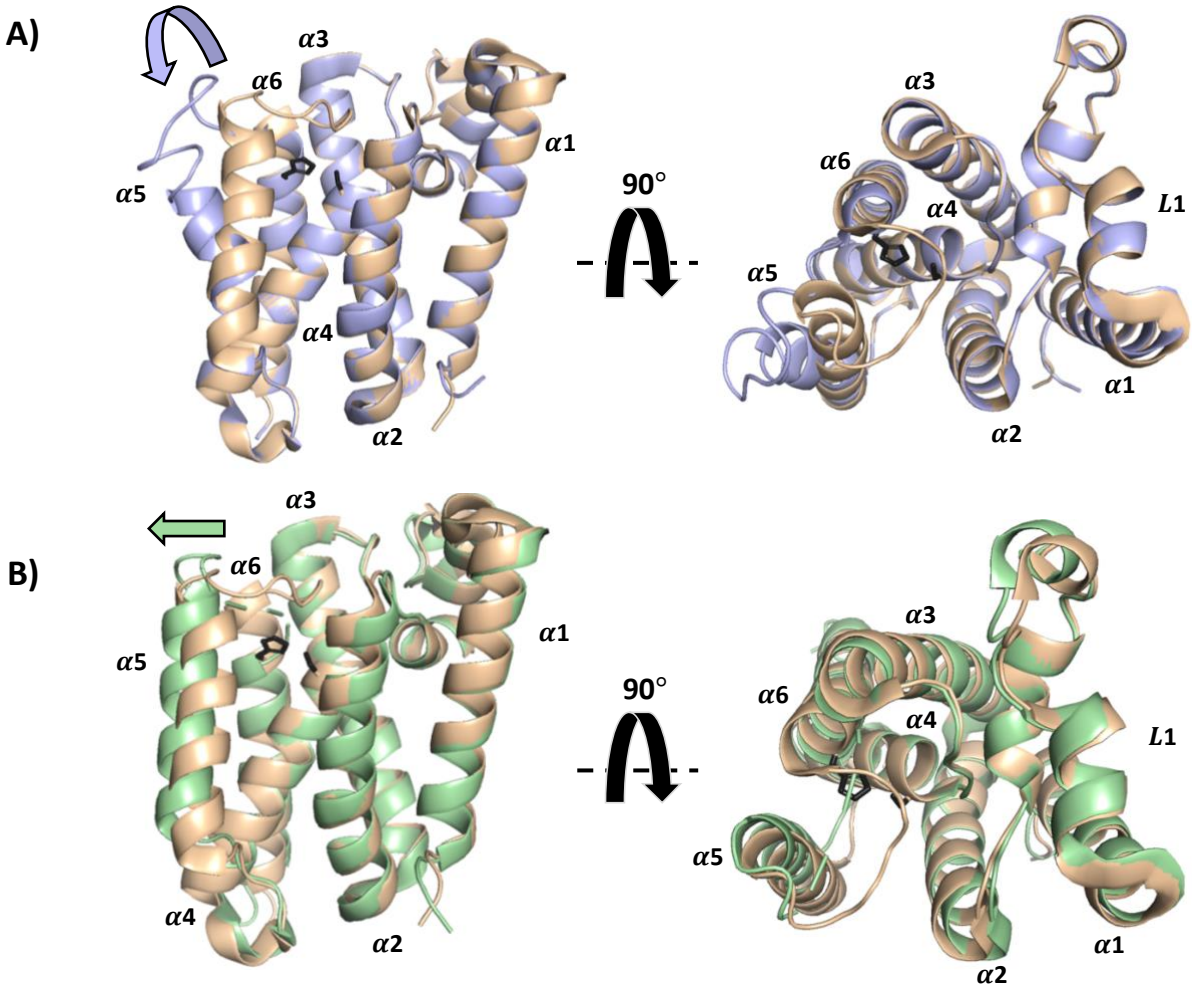


Figure 1.6: **Proposed *ec*GlpG gate-open models (A)** *ec*GlpG gate-open conformation showing α5 in a bent conformation (PDB 2NRF[35]) and **(B)** laterally displaced (PDB 5F5D[34]) shown in purple and green respectively, superimposed on the closed conformation of *ec*GlpG shown in beige (PDB 2IC8[3]). The catalytic dyad is shown in black.

1.5 Enzyme kinetics

Like its water-soluble counterparts, rhomboid protease activity is characterized using Michaelis-Menten kinetics. The general kinetic scheme that this model is based on (Equation 1.0) is a two-step process where the substrate (S) first binds reversibly to enzyme (E) to form an enzyme-substrate complex (ES) that then reacts to form free enzyme and product (P) in an irreversible manner.[56] The use of this model requires that the free enzyme (E) and substrate (S) can diffuse freely in solution, the concentration of substrate is in excess to that of the enzyme, and that product inhibition does not occur. To satisfy these conditions, it is necessary to measure initial reaction rates (v_0).[57,58]

(1.0) $E + S \leftrightharpoons ES \longrightarrow E + P$

From this model the Michaelis-Menten equation (Equation 1.1) can be derived which relates the initial rate of the reaction to the concentration of the substrate ([S]).

(1.1) $$v_0 = \frac{Vmax[S]0}{Km+[S]0}$$

According to this equation, a plot of initial reaction rate versus substrate concentration yields a hyperbolic curve as shown in Figure 1.7. The plateau of this graph represents the maximum velocity of the reaction where all substrate-binding sites in the enzyme are fully saturated by substrate (V_{max}). The substrate concentration required to reach half this maximum velocity represents K_M – the Michaelis constant.

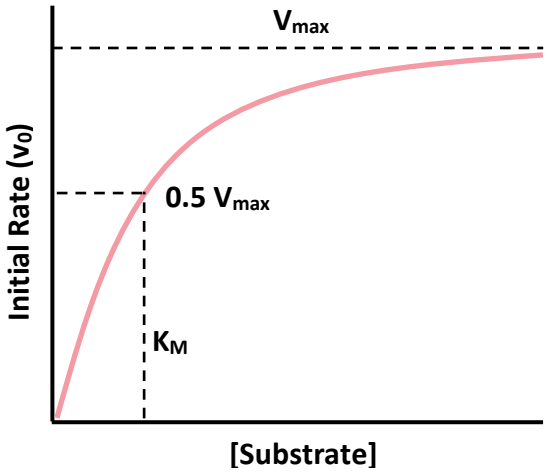


Figure 1.7: Schematic example of a plot of initial rate versus substrate concentration for an enzyme that follows the Michaelis-Menten kinetics.

Conventional methods make use of discontinuous assays using TM model substrates derived from physiological rhomboid substrates like TatA (*Providencia*), Gurken (*Drosophia*) or Spitz (*Drosophila*) where proteolysis rates can be quantified using the amount of cleavage product produced over a given time.[18,40,41,59] However, the challenge associated with measuring Michaelis-Menten kinetics using TM model substrates is that the diffusion of substrate to enzyme may be restrictive, which leads to difficulties in reporting accurate proteolysis rates using this model. To circumvent this, continuous assays using a water-soluble substrate was employed in this thesis. *ec*GlpG is able to cleave a commercially available casein protein substrate that can be monitored in a continuous assay to determine proteolysis rates approximated by the Michaelis-Menten model. This water-soluble substrate is labelled with multiple copies of a self-quenching fluorophore, where proteolytic cleavage can release the fluorophore from these quenching interactions to produce fluorescence.[39,60] According to the rhomboid substrate consensus motif described in 1.3.2, *ec*GlpG has the potential to cleave multiple sites in the β-casein sequence. However, previous SDS-page analysis revealed just one cleavage product, suggesting casein cleavage by *ec*GlpG to occur within the middle of its sequence at one of two possible cut sites[3,61]. Although casein is not a native rhomboid

substrate, it has been a valuable reporter of rhomboid proteolytic activity allowing for the acquisition of kinetic profiles for a range of rhomboid homologues including *ec*GlpG[59,61,62].

1.6 NMR spectroscopy

This thesis utilizes solution-state nuclear magnetic resonance (NMR) spectroscopy to probe structure and dynamics of *ec*GlpG. Solution-state NMR is a powerful technique that can be used to obtain information on conformational dynamics in proteins, allowing it to complement the high-resolution structural information provided by X-ray crystallography. In this thesis, 1D ^{19}F NMR was used to monitor the structure and dynamics of targeted regions within the α5 substrate gate of *ec*GlpG.

1.6.1 Basic principles of NMR spectroscopy

Solution NMR spectroscopy can be a helpful tool to characterize protein structure and dynamics by providing information on the local chemical environment of NMR-active nuclei within a molecule. Typically, nuclei experience a local magnetic field caused by the magnetic fields of electrons within their immediate chemical environment. Depending on the surrounding electron currents, this results in the nucleus being shielded or deshielded.[63] NMR involves detection of the magnetic signal from nuclei when placed within an external magnetic field (B_0). Nuclei typically used in protein NMR experiments such as ^{1}H, ^{15}N and ^{13}C all possess intrinsic spin angular momentum. When these nuclei are exposed to the external magnetic field, a dipole moment is induced, allowing the nuclei to align themselves in parallel or antiparallel spin states relative to B_0 in the z-axis. The overall summation of the individual spins in these states makes up the bulk magnetization vector along the z-axis. A transverse radio frequency (RF) pulse with energy equal to the difference between the spin states is then applied, where absorption of this energy leads to transitions between the spin states (Fig. 1.8).

This gives rise to a bulk magnetization vector in the transverse plane that is detected by the NMR spectrometer.

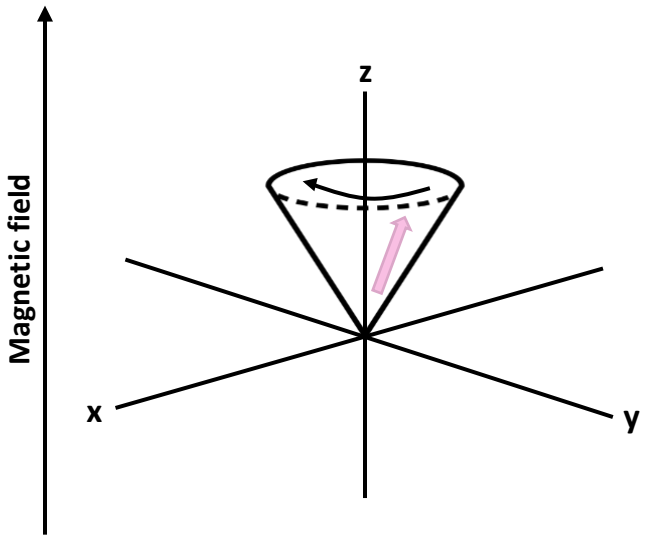


Figure 1.8: Schematic illustration showing the precession of a spin (pink) in a magnetic field on the surface of a cone in the xy-plane at the Larmor frequency specific to its nucleus.

This phenomenon is described by equation 1.3 where ω_0 is the Larmor frequency and represents the rate of precession of the spins governed by the energy difference between spin states, B_{eff} is the strength of the effective magnetic field experienced by the nuclei and γ is the gyromagnetic ratio of the nucleus of interest:

$$\textbf{\textit{(1.3)}}\ \omega_0 = \gamma\ B_{eff}$$

The local chemical environment around the nucleus influences the frequency of radiation absorbed and is reported as a chemical shift value, from a known standard, in units of parts per million (ppm).

1.6.2 Heteronuclear Single Quantum Coherence (HSQC) Spectroscopy

Although the 1D ^{1}H NMR experiment is the most widely used NMR experiment, it is not commonly employed to study proteins as its spectrum gives rise to an envelope of overlapping peaks due to the hundreds of proton resonances in proteins. To help resolve and assign these peaks, multidimensional NMR experiments are used, which reduces spectral overlap while providing correlations between nuclei that are coupled directly through bonds or dipolar interactions. The most common multidimensional NMR experiment used for proteins is the 2D ^{1}H-^{15}N heteronuclear single quantum coherence (HSQC) experiment. Peaks in the 2D ^{1}H – ^{15}N HSQC spectrum correlate the chemical shift of a ^{1}H in the one dimension with the chemical shift of its directly bonded ^{15}N nucleus in the second dimension. The backbone amide group gives rise to a peak for each non-proline residue from the polypeptide backbone. Additionally, this spectrum may also contain peaks from NH groups of Trp, Asn, Gln, Lys, His and Arg side chains. Specifically, tryptophan indole N-H groups have distinctive ^{1}H chemical shifts in the 9-11ppm range. This enables them to be easily identified and resolved from the main cluster of backbone and other side-chain amide peaks. Since each protein gives rise to a unique pattern of peaks, the ^{1}H – ^{15}N HSQC spectrum is widely regarded as a “fingerprint” spectrum. It is often used as a starting point to assess the structural integrity of protein samples giving information on the folded nature of the protein by evaluating the dispersion of the expected number of peaks.[64]

1.6.3 ^{19}F NMR

A particularly useful NMR-active nucleus is the ^{19}F isotope. This isotope exists in ~100% natural abundance, has a large gyromagnetic ratio (83% the sensitivity of ^{1}H) and a large chemical shift dispersion making it highly sensitive for NMR spectroscopy.[65,67] The utility of ^{19}F NMR in the study of protein structure and function is particularly informative when paired with structural information by other methods.[66-68] Since ^{19}F is rarely found in biological systems, there are no complications from background signals. Instead, the incorporation of ^{19}F labels at

targeted sites in a protein can provide information on the local chemical environment of residues, including those that may undergo conformational changes.[66] ^{19}F NMR has previously been used with great success to monitor conformational dynamics in the GPCR protein family[66, 69-74]. For instance, ^{19}F chemical shift changes were used to discern conformational changes between an inactive state *I* and an activated state *A* upon agonist binding to the β2-adrenergic receptor[73, 74]. These studies demonstrated the potential of solution ^{19}F NMR to monitor conformational dynamics in larger systems making it an attractive tool for the study of α5 gating dynamics in *ec*GlpG.

1.6.4 ^{19}F NMR to monitor conformational exchange

^{19}F NMR spectroscopy can be used to detect dynamic processes such as conformational exchange. The incorporation of ^{19}F probes at sites predicted to undergo a change in conformation are also likely to experience a change in local chemical environment which gives rise to different chemical shifts. As shown in Figure 1.11, if a spin is sampling two different states (e.g., open and closed), one peak will arise for each state, when the exchange rate (k_{ex}) is considerably smaller than the frequency difference between the states ($|\Delta v|$). In this situation, known as slow exchange on the NMR time scale, relative peak intensities will reflect the populations of the two states. As the exchange rate increases peak linewidths will broaden and shift towards the average chemical shift. When the exchange rate approaches the frequency difference between the states ($k_{ex} \approx |\Delta v|$), the system is in the intermediate time scale exchange regime. This can give rise to a single broad peak centered on the weighted average chemical shift of the two states. In the fast-exchange regime, the exchange rate is much faster than the frequency difference between states ($k_{ex} >> |\Delta v|$), yielding a single peak at the population-weight average chemical shift between both states.[80] Typically, slow exchange processes have exchange rates on the millisecond (ms) to second (s) timescale (i.e., $k_{ex} < 1/s^{-1}$), whereas fast exchange processes occur on microsecond (μs) timescale and faster (i.e, ., $k_{ex} > 10^5/ s^{-1}$). However, because $|\Delta v|$ is directly related to B_0, and k_{ex} is independent of B_0, the effect

of exchange on the quality of an NMR spectrum can be modulated by using a spectrometer of different field strength.

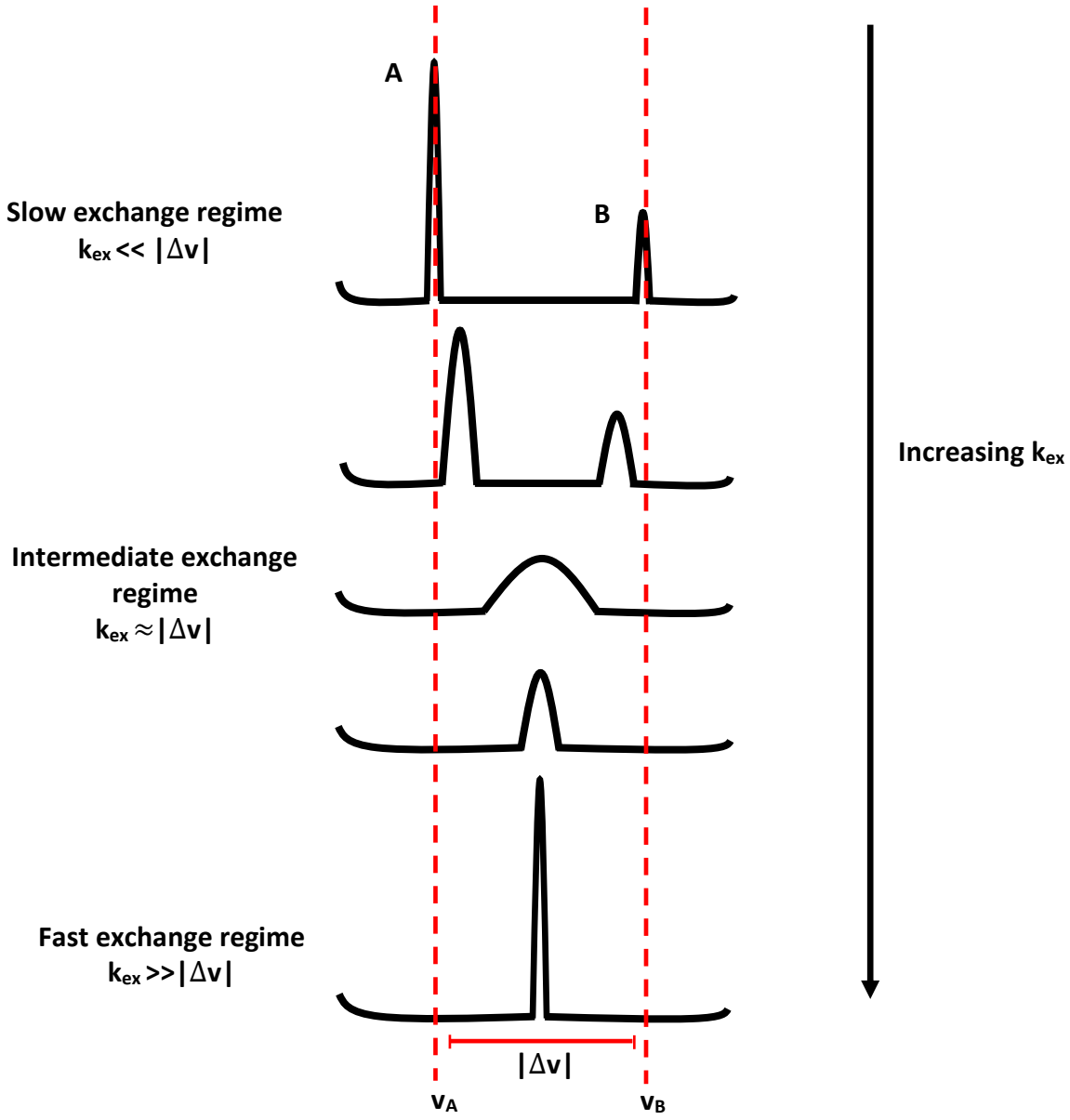


Figure 1.9: Schematic diagram illustrating the effect of exchange on a 1D NMR spectrum. Values v_A and v_B represent the frequencies for the two exchanging species (A and B) in units of Hz, while $|\Delta v|$ represents the difference between v_A and v_B and k_{ex} is the rate constant for exchange between the two states.

1.6.5 ^{19}F-labeling by covalent modification

^{19}F probes can be strategically incorporated into proteins using various methods, all of which require the use of extrinsic labels. ^{19}F can be incorporated by biosynthetic methods wherein the expression medium is supplemented with fluorinated amino acid analogs[68, 81-83] or sequence-specific incorporation of fluorinated amino acids at amber suppressor codons by the introduction of engineered tRNAs/aminoacyl-tRNA synthetases into genetically encoded positions[85-88]. Chemical modification of amino acid side chains is also frequently used to introduce ^{19}F probes into specific sites[66,67,70, 75-79]. In this method a fluorinated reagent is used to covalently modify specific protein sidechains. A common target for this covalent modification is cysteine as its sidechain thiol group is a nucleophilic residue[68]. Additionally, the low frequency of cysteines in endogenous protein sequences facilitates straightforward spectral assignment through the introduction of ^{19}F probes at desired sites[89,90].

In order to probe conformational dynamics in GlpG by ^{19}F NMR, we sought to label TMD81 samples at targeted sites using fluorinated agents that could react specifically with the thiol groups of cysteine residues[66,67]. Conveniently, there is only one endogenous cysteine residue (C104) in TMD81, and this residue appears to be buried within the hydrophobic core and should be less accessible to chemical conjugation by thiol-reactive agents (Fig. 1.9). Through site-directed mutagenesis other cysteine residues can be strategically engineered at useful locations around the substrate gating region.

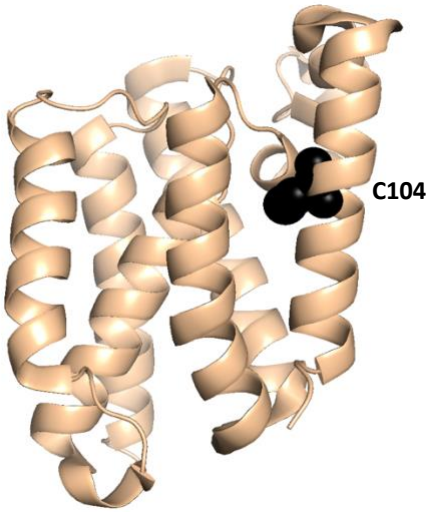


Figure 1.10: Structure of TMD81 (wheat, PDB 2IC8) showing the position of its native core cysteine – C104 (black).

A common group of fluorinated tags used to label cysteine residues are trifluoromethyl-containing probes. These probes typically exhibit a higher signal-to-noise ratio as there are three equivalent fluorine nuclei that undergo rapid methyl rotation, reducing the chemical shift anisotropy (CSA) that can broaden peaks thereby giving rise to narrower line widths[68,76]. In our studies, we employed 2,2,2-trifluoroethanethiol (TET) and 3,5-bis(trifluoromethyl)benzyl bromide (TFMB) to monitor conformational dynamics in the gating region (Fig. 1.10). The TET probe is commonly used in studies of membrane proteins as it is a highly selective reagent for cysteine, is commercially available, and is postulated to minimally perturb protein structure due to its relatively small size[66,81]. Labelling with TET typically begins with an activation step whereby cysteine thiol(s) groups are activated through reaction with 2,2'-dipyridyl disulfide (DPS). In a second step, a disulfide linkage is formed between the trifluoromethyl group of TET and the protein[66,76,91]. The TFMB probe has the potential to give even higher signal-to-noise spectra as it has the ability to introduce six equivalent fluorine atoms into a protein via a single step chemical conjugation reaction to form an irreversible covalent bond. This may allow ^{19}F NMR spectra to be obtained from lower protein concentrations[77].

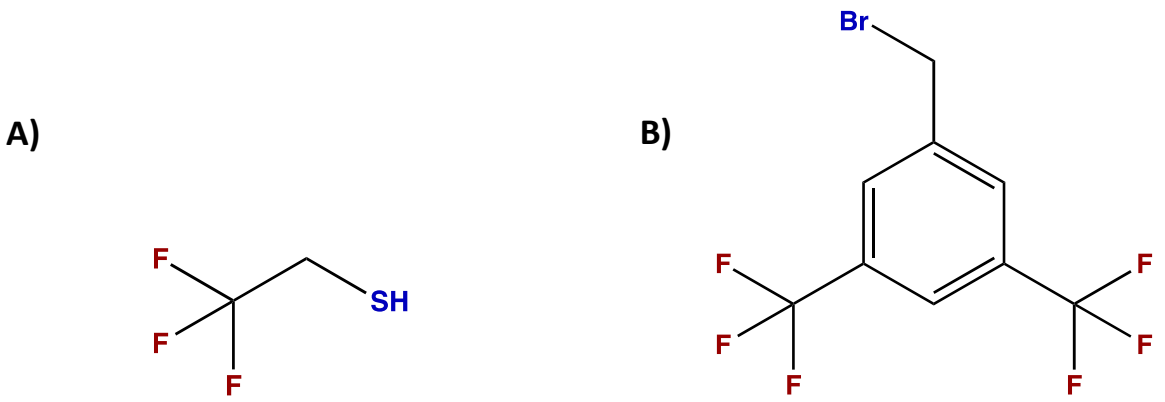


Figure 1.11: **Chemical structures for trifluoromethyl-containing probes employed to study α5 gating dynamics in *ec*GlpG: A)** 2,2,2,-Trifluoroethanethiol (TET) and **B)** 3,5 Bis(Trifluoromethyl)benzyl bromide (TFMB).

1.7 Solution NMR of membrane proteins

1.7.1 Detergent micelles and other membrane mimetic systems

The application of solution state NMR to membrane proteins like the rhomboid protease requires that pure samples be extracted from the native lipid membrane and subsequently solubilized in a membrane-mimetic system.[93] This is done to maintain protein structure and function while maintaining it in the liquid state required for solution-phase NMR.

In this thesis detergent micelles were employed to perform solution-state ^{19}F NMR studies of TMD81. Due to their relatively small size, detergent micelles are the most commonly used solubilization agents for membrane protein solution-state NMR studies[94] and have successfully been used to characterize a range of integral membrane proteins[95-98]. Micelles form due to the spontaneous self-assembly of detergent into spherical complexes with the hydrophobic regions directed inwards and hydrophilic headgroups facing the aqueous environment. Micelle formation occurs when the amount of detergent exceeds the critical micelle concentration (CMC). Membrane proteins can be solubilized by micelles through the

formation of a protein detergent complex (PDC) whereby the detergent interacts with the hydrophobic region of the membrane protein to protect it from the aqueous environment (Fig. 1.12)

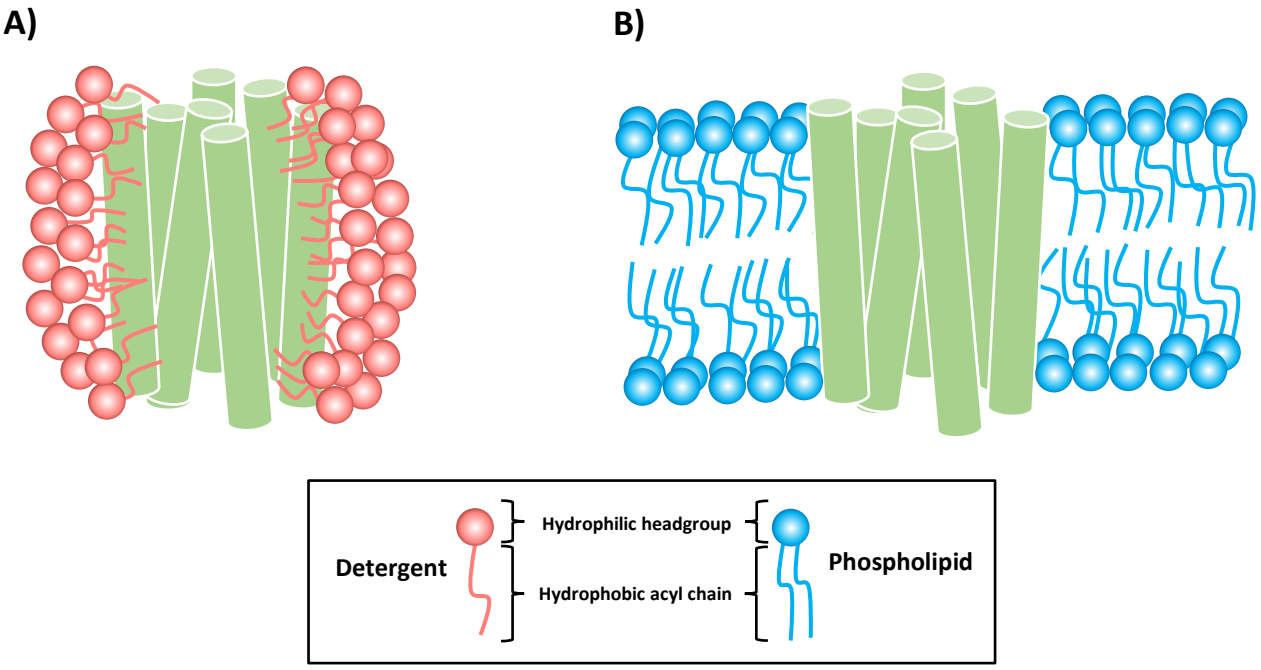


Figure 1.12: **Schematic illustration of the solubilization of membrane proteins** (green) **(A)** in detergent micelles (pink), and **(B)** within their native environment – the phospholipid bilayer (blue).

While micelles can mimic the general features of phospholipid membranes, there are nonetheless significant differences that may affect protein structure. Detergents commonly used for NMR studies have a single hydrophobic alkyl chain, which can cause a high degree of surface curvature in the micelle[99]. Additionally, micelles are more dynamic than phospholipid bilayers due to packing defects, leading to the entry of water molecules into the hydrophobic core.[6,100-102] This reduces the lateral pressure exerted on a detergent-solubilized protein which could affect its stability and conformational dynamics[103]. For these reasons, alternate membrane mimetic systems for IMPs have been developed, including phospholipid bicelles and nanodiscs. Bicelles are formed by mixtures of long-chain lipids and detergents that form bilayer

discs capped by the detergent[104,105]. Similarly, nanodiscs are also comprised of bilayers in a disc shaped complex, except that the edges of the lipid bilayer disc are capped by a membrane protein scaffold (MSP), thereby avoiding the need for detergents altogether[106]. Lipid discs can also be made using amphipols in place of MSP, with the polymers being composed of an amphipathic backbone with hydrophobic side chains that cap the edges of the bilayer disks to protect them from the aqueous phase[106]. All of these alternative membrane mimetic systems have been successfully used in a range of applications, however, the resulting complexes they form are significantly larger than those generated using detergent micelles. This can become an issue for solution NMR of membrane proteins as more rapid relaxation of the NMR signal from larger complexes decreases sensitivity and resolution, giving rise to significant line broadening of peaks[107]. Contributing to this loss of sensitivity is the slower molecular tumbling rates that can result from the use of high concentrations of membrane mimetics, further broadening signals in the NMR spectrum[108]. For these reasons our work on the *ec*GlpG has continued with the study of TMD81 in detergent micelles.

1.7.2 Solution NMR of *ec*GlpG

Attempts to study α5 gating in *ec*GlpG using solution NMR have predominantly centered around the use of 2-dimensional ^{1}H-^{15}N TROSY spectra with the use of Fos-choline-12 detergent micelles.[41] These spectra showed approximately 67% of the expected correlations in addition to enhanced signal intensity compared to spectra acquired using other detergents. Unfortunately, the quality of these spectra were not sufficient for backbone chemical shift assignments to be made. Instead, a series of tryptophan knockout mutations was done in an attempt to assign signals from the 11 well distributed tryptophan residues in the TMD81 sequence. NMR spectra of mutants made in the α5 and α2 interface, specifically W236A, W157F, F153A, and I230A showed chemical shift perturbations in these tryptophan residues that suggested changes to the TMD81 structure that may have been associated with the substrate gate. However, as the changes in the spectra were complex, and the assignment of tryptophan peaks tentative, it was

not possible to gain structural insight into the substrate gate from this data. Therefore, questions still remain concerning the extent to which $\alpha 5$ is displaced and whether this displacement is accompanied by rearrangements in other regions of TMD81.

To simplify the 2-dimensional spectrum to 1-dimension, 1D ^{19}F NMR spectra were acquired for TMD81 samples biosynthetically labelled with monofluorinated aromatic amino acid analogues.[109] In these experiments expression media was supplemented with fluorinated tryptophan analogues such as 5-fluorotryptophan (5FW) and 7-fluoroindole (7FI). Unfortunately, these 1D spectra suffered from low signal-to-noise and difficulties in assignment due to low peak dispersion and multiple spectral changes being induced by single tryptophan knock out mutants.

1.8 Thesis objectives

Although a growing body of evidence is present in support of a role for $\alpha 5$ dynamics in substrate entry into the active site, the details regarding the conformational states and their link to measured rates remain to be elucidated. In this thesis, we employed labeling of a site at the interface of $\alpha 2$ and $\alpha 5$ to probe gating dynamics. This was done using trifluoromethyl-labeled TMD81 probes that were monitored by ^{19}F NMR. The presence of more than one state in slow exchange was detected in some cases, consistent with conformational dynamics. However, interpretation of spectra was complicated by the presence of additional species and some loss of activity from the introduction of mutations needed for these studies. Together, results from this thesis suggest that further work to optimize label selection and mutation sites will be needed to elucidate the role of conformational dynamics in TMD81 function.

2 Materials and methods

2.1 Mutagenesis

A C-terminal hexahistidine-tagged construct encompassing residues 81 – 290 of *E. coli* GlpG rhomboid (TMD81) in a pET25b vector was used as the template to make all mutants. Primers (Table 2.0 and 2.3) were synthesized by Sigma Aldrich or Eurofins Genomics.

The cysteine knockout mutant, C104S was constructed by Dr. James A Davey using splicing by overlap extension (SOE)[110] followed by circular polymerase extension cloning (CPEC)[111]. To make the C104S insert and vector, touch-down PCR reactions were carried out using the Q5 High-Fidelity PCR kit (New England BioLabs). For the SOE reactions, the reaction mixtures each contained 1X Q5 reaction buffer, 0.2mM dNTPs, 2μM of the forward and reverse primers, 10ng of template DNA and 0.01U/μL of Q5 polymerase in a total reaction volume of 50μL. PCR tubes were placed in a S1000 Thermal Cycler (Bio-Rad) or a MasterCycler Personal (Eppendorf) running the program shown in Table 2.1. PCR products of the vector reaction were run on a 0.7% agarose gel and PCR products of the insert reaction was run on a 2% agarose gel both containing 1X SYBR Safe DNA gel stain (Invitrogen). Electrophoresis was carried out in SB buffer (36mM boric acid and 10mM NaOH) for a total of 75 minutes at 120V on an OWL EasyCast gel electrophoresis system (Thermo Scientific). PCR products were identified by UV light exposure using a ChemiDoc XRS+ imaging system (Bio-Rad). Reactions containing the PCR products were gel extracted using the E.Z.N.A. Gel Extraction Kit (Omega Bio-Tek) in accordance with the manufacturer's protocols. CPEC was then done to incorporate the insert into the vector using 1X Q5 reaction buffer, 0.2mM dNTPs, 90ng of insert, 150ng of vector and 0.01U/μL of Q5 polymerase. The PCR program shown in Table 2.2 was run and PCR reactions were visualized on a 0.7% agarose gel. Reactions containing the PCR products were cleaned using the E.Z.N.A. Cycle Pure Kit (Omega Bio-Tek). Plasmids were sequenced at the StemCore DNA Sequencing facility at the Ottawa Hospital Research Institute.

Table 2.0: DNA primer sequences used to generate knockout cysteine mutant; C104S

Primer type	Sequence
Vector	Forward: 5'- CGCGAAAACGAAAAGCTAGCCTCG-3' Reverse: 5'- GGCAATAAACACCACtACCGAtGCtATCATCATCACC-3'
Insert	Forward: 5'- GGTGATGATGATaGCaTCGGTaGTGGTGTTTATTGCC-3' Reverse: 5'- CGAGGCTAGCTTTTCGTTTTCGCG-3'

Table 2.1: PCR thermocycler program to generate C104S vector and insert by SOE

Step	Temperature (°C)	Time (sec)
1	98	30
2	98	10
3	72, subtract 1°C/cycle	20
4	72	168 (for vector) or 17 (for insert)
5	Go to step 2, (20 cycles)	
6	72	2 mins
7	4	**HOLD**

Table 2.2: PCR thermocycler program for CPEC of C104S

Step	Temperature (°C)	Time (sec)
1	98	30
2	98	30
3	66	20
4	72	168
5	Go to step 2, (20 cycles)	
6	72	2 mins
7	4	**HOLD**

All cysteine mutants were made using the Quick-change site-directed mutagenesis protocol (Agilent, Stratagene). Mutagenesis reactions of cysteine mutants were carried out using the KOD Xtreme Hot Start polymerase kit (Novagen). Individual reaction mixtures contained 1X Xtreme buffer, 0.4mM dNTPs, 0.3μM of forward and reverse primers, 10ng of template DNA and 0.02U/μL of KOD Xtreme Hot Start DNA polymerase to a total reaction

volume of 50μL. The PCR program shown in Table 2.4 was run and PCR reactions were visualized on a 0.7% agarose gel containing 1X SYBR Safe. PCR products were incubated with DPNI (NEB) for 60 minutes at 37°C followed by heat inactivation at 80°C for an additional 15 minutes. PCR products were then cleaned using the Cycle Pure Kit (Omega Bio-Tek) and plasmids were sent for sequencing as previously described.

Table 2.3: DNA primer sequences used to generate cysteine mutants and the W236A gate-open mutant

Mutants	Sequence
F153C	Forward: 5'-GCTGATGCATATCCTC**TGT**AACCTGCTCTGGTGGT-3' Reverse: 5'-ACCACCAGAGCAGGTT**ACA**GAGGATATGCATCAGC-3'
W236A	Forward: 5'-ATTATCTTTGCGCTGATC**GCG**ATTGTCGCCGGATGGTTTGATTTG-3' Reverse: 5'-CAAATCAAACCATCCGGCGACAAT**CGC**GATCAGCGCAAAGATAAT-3'

Table 2.4: PCR thermocycler program to generate cysteine mutants

Step	Temperature ((°C C)	Time (sec)
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

Chemically competent *E. coli* C43 (DE3) cells (Lucigen) and DH5α cells (Novagen) were individually divided into 50μL aliquots and flash-frozen in liquid nitrogen. Aliquots were stored at -80C.

10 ng of plasmid DNA was transformed into one aliquot of DH5α cells or C43 (DE3) cells for mutagenesis and protein expression respectively. Aliquots were mixed by flicking and allowed to incubate on ice for 30 minutes. Following incubation, cells were heat-shocked for 45 seconds at 42°C on a Standard Heatblock (VWR) and immediately returned to chill on ice for 2 minutes. 900μL of heated sterile LB broth (Lennox, Bioshop) was then added to the aliquot and placed into a MaxQ 4000 Benchtop Orbital Shaker (Thermo Fisher Scientific) to shake for 60 minutes at 37°C and 300 rpm. Cells were then plated on an ampicillin-selective LB agar plate (Bioshop) and left to grow overnight (12-16 hours) at 37°C.

2.3 Plasmid DNA preparation

An isolated colony of DH5α cells transformed with DNA plasmid was used to inoculate a 30mL culture of LB media in a 250mL flask with 100μg/mL of ampicillin. Cultures were shaken overnight (16-18 hours) at 37°C and 220 rpm. The following day, cells were harvested by centrifugation using a Spinchron 15 Tabletop Centrifuge (Beckman Coulter) with a S4180 swinging-bucket rotor for 10 minutes at 25°C and 5,000xg. Following centrifugation, plasmid DNA samples were purified from their pellets using the E.Z.N.A. Plasmid DNA Mini Kit II (Omega Bio-Tek). Plasmid DNA stocks were stored at -20°C until use.

2.4 Overexpression of TMD81

A pET-25b(+) expression vector encoding TMD81 (residues 81-276 of *ec*GlpG) was transformed into *E. coli* C43 (DE3) cells. This construct was overexpressed using previously published protocols. To summarize, a transformed colony was inoculated in a 100mL culture of LB media with 100 μg/mL of ampicillin and left to incubate overnight (16 hours) at 37°C and 220 rpm. The following day, overnight cultures were centrifuged, and pellets were resuspended in 1L culture flasks of either LB media (20g/L LB) or M9 minimal media (47.90mM Na_2HPO_4, 22.04mM of KH_2PO_4, 8.56mM of NaCl, 1mM of $MgSO_4$, 0.3% (w/v) of D-glucose, 0.1% (w/v) of NH_4Cl, 0.2g/L of LB, 1μM of $CaCl_2$, 100μg/mL of ampicillin). For generation of ^{15}N labelled TMD81, 10% w/v $^{15}NH_4Cl$ (Aldrich) was prepared in ddH_2O, sterilized by a 0.22μM filter used to prepare the M9 media in place of the natural abundance NH_4Cl.

Culture flasks were incubated at 37°C and 220 rpm on a MaxQ 5000 refrigerated incubator shaker (Thermo Fisher Scientific) and cell growth was monitored periodically by measuring the optical density (OD) of the cells at 600nm using an Ultrospec 2100 *pro* UV/Visible spectrophotometer (GE healthcare). As the optical density reached between 0.5-0.8, protein expression was induced with 1mM of isopropyl-β-D-thiogalatopyranoside (IPTG, BioShop). After induction, the cultures were allowed to incubate another 3 hours. Cells were then harvested by centrifugation using an Avanti J-E centrifuge (Beckham Coulter) with a JLA-16.250 fixed angle rotor for 10 minutes at 4°C and 6,000xg. Pellets were scrapped and stored at -20°C for purification.

2.5 Purification of *ec*GlpG in detergent micelles

Frozen cell pellets obtained from 1L of culture were thawed on ice and then mechanically resuspended in Harvest Buffer (pH 7.4, containing 20mM HEPES, 100mM NaCl, 5mM of MgCl, 10% glycerol and 1X protease inhibitor cocktail – 1mg/mL 4-(2-aminoethyl)benzenesulfonyl fluoride HCL (AEBSF), 2ug/mL pepstatin, 1.2ug/mL E-64, 1ug/mL

bestatin, 1.8 ug/mL phosphoramidon). To physically disrupt the cell membrane, cells were lysed by sonification on ice using a Sonic Dismembrator Model 500 (Fisher Scientific) at 50% intensity, 1 minute total run time with one-second pulses. This lysate was then centrifuged in an Avanti J-E centrifuge (Beckham Coulter) using a JA-25.50 fixed angle rotor for 1 hour at 4°C and 48,500xg to separate the insoluble membrane fraction from the rest of the cellular debris. The supernatant was discarded, and the pellet was mechanically resuspended for 1.5 hours in 1X solubilization buffer (pH. 7.4, containing 0.05M HEPES, 0.2M NaCl, 5mM imidazole, 10% glycerol, 1X protease inhibitor cocktail and 1% n-dodecyl-β-D-maltopyranoside (DDM)). This resuspended lysate was centrifuged another time for 30 minutes at 4°C and 48,500xg. The supernatant now containing the detergent-solubilized protein was applied and allowed to incubate for 10 minutes with 2.5mL nickel NTA agarose affinity resin (Ni-NTA, Qiagen), previously equilibrated in 2X solubilization buffer (pH 7.4, 0.1M HEPES, 0.4M NaCl, 0.01M imidazole and 20% glycerol). Impurities were removed through 2 successive wash steps using wash buffer 1 (pH 7.4, containing 0.05M HEPES, 0.5M NaCl, 0.03M imidazole, and 10% glycerol) and wash buffer 2 (pH 7.4, containing 0.05 HEPES, 0.2M NaCl, 30mM imidazole, and 10% glycerol). Finally, the protein was eluted using Elution Buffer (pH 7.4, containing 50mM HEPES, 0.2mM NaCl, 0.25mM imidazole, and 10% glycerol). The eluted protein sample was concentrated down to 1mL by centrifugation at 5,000xg and 4°C using a 30kDa molecular weight (MW) cut-off centrifugal filter device (Amicon). Samples were then subjected to size exclusion chromatography (SEC) on a Superdex-200 Increase 10/300 GL column (Cytiva) run on an AKTA FPLC system (GE Healthcare) equilibrated with 25mL of GlpG Buffer (pH 7.4, containing 25mM of Tris-HCl, 25mM NaCl, 0.05% DPC and 0.05mM EDTA). The eluted protein sample was monitored by absorbance at 280nm, and 0.5mL fractions of the main protein peak were collected for downstream applications.

2.6 SDS-PAGE analysis

SDS-PAGE analysis was performed following the Laemmli procedure to assess the purity of the protein samples following purification. In brief, 0.75mm gels were cast with a 12% resolving phase and a 5% stacking phase (containing either 12% or 5% acrylamide mix, 0.1% SDS, 0.1% ammonium persulfate, 0.06% TEMED and 1M Tris pH 8.8 or 1.5M Tris pH 6.8). Aliquots of the main protein peak following size-exclusion chromatography as well as aliquots of sample fractions at various stages during purification were mixed with 2X loading buffer (containing 125mM Tris pH 6.8, 20% (v/v) glycerol, 4% (w/v) SDS, 0.01% (w/v) bromophenol blue, and 0.02% (v/v) 2-mercaptoethanol) in a 1:1 ratio. Samples were incubated for 1 minute in boiling water before loading 10μL of sample and 3μL of EZ-Run Prestained Rec Protein Ladder (Fisher BioReagents) onto the gel. Gels were run in a Mini-PROTEAN electrophoresis cell (Bio-Rad) for 75 minutes at 150V in running buffer (containing 0.3% (w/v) Tris pH 8.5, 11.44% (w/v) glycine and 0.1% (w/v) SDS). Gels were immersed in Coomassie brilliant blue (CBB) R-250 stain solution (containing 0.1% Coomassie brilliant blue R-250, 50% methanol, 10% acetic acid) heated for 30 seconds in the microwave and then left on a rocking platform for 60 minutes. For destaining, excess stain was rinsed off and gels were heated for 30 seconds with a destaining solution (containing 50% methanol and 10% acetic acid) and left to incubate for 60 minutes on a rocking platform. When necessary, this destaining step was repeated with fresh solution to allow for more thorough destaining. Gels were stored in ddH_2O until imaging.

2.7 Protein concentration determination

A 1mg/mL stock of bovine serum albumin (BSA, Fisher) was used to prepare standards for use in the determination of TMD81 protein concentrations. Dilutions of 25, 50, 100, 200, 400, 500, 750 and 1000μg/mL BSA were prepared in ddH_2O as were protein sample dilutions. Samples were run in SDS-PAGE gels and stained using Coomassie brilliant blue R-250. The staining intensity of the CBB bands were captured with a ChemiDoc XRS+ System (Bio-Rad) and

analyzed using ImageJ processing software. Band intensities using BSA standards were used to generate a standard curve which was then used to determine protein concentrations for TMD81 samples run on the same gel.

2.8 Site-specific fluorine labelling of cysteine residues in *ec*GlpG

TMD81 was labelled with 2,2,2,-trifluoroethanethiol (TET, Sigma-Aldrich), by first activating purified protein samples with 2,2'-dipyridyl disulfide (DPS, ThermoFisher Scientific) overnight (12-16 hours) at 4°C. Unreacted DPS was removed the next day by buffer exchange using a Superdex-200 10/300 GL size exclusion column on an AKTA FPLC equilibrated with GlpG buffer. A 10-fold molar excess of 2,2,2-trifluoroethanethiol was then incubated with DPS-activated TMD81 overnight (12-16 hours) at 4°C. Unreacted label was removed the following day, again by buffer exchange using size exclusion chromatography.

Prior to labelling with 3,5-Bis(trifluoromethyl)benzyl bromide (TFMB, ThermoFisher Scientific), purified protein samples were first reduced with 1mM of tris(2-carboxyethyl)phosphine (TCEP, Sigma-Aldrich) for 1 hour at room temperature. These samples were then reacted with a 10-fold molar excess of 3,5-Bis(Trifluoromethyl)benzyl bromide and allowed to incubate with the protein overnight (12-16 hours) at 4°C. Unreacted label was removed the following day by buffer exchange using size exclusion chromatography.

2.9 Circular dichroism spectroscopy

Circular dichroism (CD) spectra were collected on 5μM of purified rhomboid samples solubilized in GlpG buffer. Spectra were acquired using a Jasco J-815 spectrometer. Spectra were recorded at 37°C and monitored from 250 to 200nm. A total of eight accumulations were collected and averaged to produce a single spectrum for each sample using a data pitch of 0.2nm and a scan rate of 20nm/min.

To access the effect of the cysteine knockout on the thermostability of *ec*GlpG, purified samples were heated from 35°C to 97°C at a rate of 0.3°C/min. The ellipticity at 222nm was monitored and fit to a two-state Boltzmann curve to determine the melting temperatures (T_m).

2.10 TMD81 activity assays

The EnzChek Protease Assay Kit (Invitrogen) was used to monitor *ec*GlpG activity against a water-soluble BODIPY-casein substrate. Casein substrate stocks of 0.0625, 0.5, 1, 2 and 3μM were prepared and incubated with 0.25μM of purified rhomboid samples in a clear flat-bottom black 96-well plate (Greiner, VWR). Using a Synergy H1 multimode microplate reader (BioTek, Agilent), fluorescence (RFU) was monitored for 2 hours at 37°C at a fixed gain of 75 measuring excitation at 485nm and emission at 530nm. A buffer blank with the varying concentrations of the BODIPY-casein substrate were subtracted from the resulting fluorescence readings. Initial rates of hydrolysis were calculated using the slope of relative fluorescence units (RFU) over 30 minutes after a 10-minute equilibration period and fit to the Michaelis-Menten equation to determine values of K_m and k_{cat} (Appendix 7.3).

2.11 ^{19}F NMR spectroscopy

^{19}F labelled protein samples were concentrated to 250μL using a molecular weight cut-off ultrafiltration centrifugation unit (Amicon, Millipore-Sigma). The NMR sample was prepared by adding D_2O to 10% (v/v) to the concentrated labelled protein sample (40-100 μM of protein) to give a final volume of 380μL. The samples were inserted into a 5mm Shigemi tube, with 8mm bottom (Bruker) and secured in place by a plunger. 1D ^{19}F experiments were performed at various temperatures using a Bruker 500MHz Avance spectrometer equipped with a liquids broadband probe with Z gradient. Experiments were acquired overnight (14-16 hours) using

30,000-36,000 scans with an acquisition time of 699ms and a 1s interscan delay time. Spectra were deconvoluted and analyzed using TopSpin 4.1.3.

NMR titrations were carried out by acquiring 1D ^{19}F NMR spectra on a TMD81 samples with of increasing amounts of peptide (Appendix 7.0 and 7.1). The peptide sequence was derived from the Drosophila Gurken rhomboid substrate (sequence RKVRMA, purity 98.57%) and were synthesized by CanPeptide (Montreal) with N-terminal acetylation and C-terminal amidation. Lyophyllized peptides were dissolved in DMSO to produce a ~40mM stock solution.

2.12 ^{1}H-^{15}N HSQC NMR spectroscopy

To obtain the fingerprint spectrum on the cysteine knockout mutant (C104S), ^{1}H-^{15}N HSQC spectra were acquired and compared to that of WTTMD81. ^{15}N labelled protein samples were concentrated down to 250μL using a molecular weight cut-off centrifugal filter. The NMR sample was prepared by combining the concentrated labelled protein with GlpG buffer and 10% (v/v) D_2O to a final volume of 300μL. The samples were inserted into a 0.25mm shaped tube (Bruker).

NMR experiments were performed at 318K on a Bruker Avance III HD 600 equipped with a liquids 5 mm triple resonance H/C/N autotuning cryoprobe. Experiments were acquired overnight (14-16 hours) using 200 scans. Spectra were processed by NMRPipe[112] and analyzed with NMRViewJ[113].

3 Results

3.1 Overexpression and purification of TMD81

Wild-type (WT) TMD81 was overexpressed with a C-terminal hexahistadine tag in *E. coli* C43(DE3) competent cells and purified by nickel affinity chromatography and size exclusion chromatography. Coomassie-stained SDS-PAGE gels of sample aliquots were taken at various stages during the purification process to monitor the effectiveness of the purification (Fig. 3.0). The presence of an intense band corresponding to TMD81 around the expected molecular weight of 24kDa begins to appear in the detergent soluble fraction (DSF). Some of this species may be present in the nickel affinity chromatography flow-through and wash steps (W1 and W2) but appears at higher levels in the elution fraction containing 250 mM of imidazole.

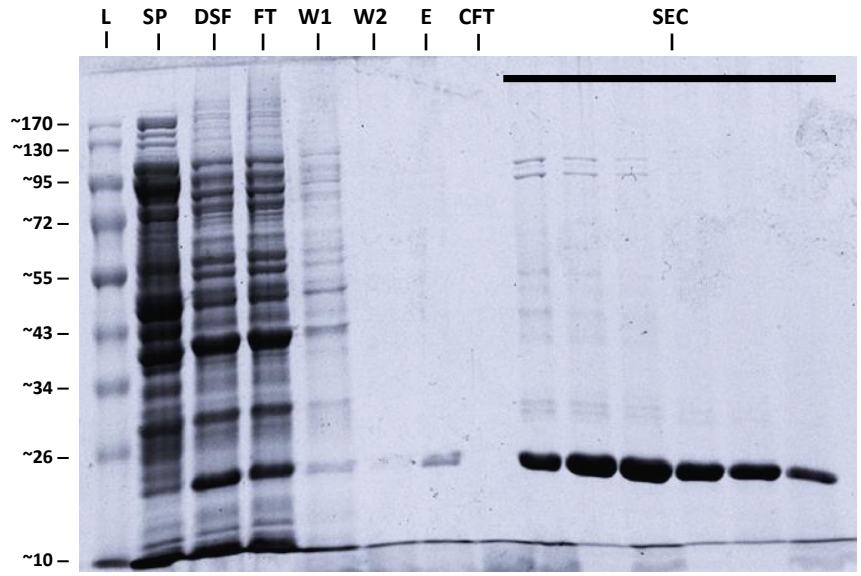


Figure 3.0: Coomassie-stained SDS-PAGE gel of fractions taken during the various stages of the purification of wild-type TMD81. **L**: molecular weight ladder in KDa, **SP**: soluble protein fraction, **DSF**: detergent soluble fraction. From the nickel affinity chromatography, **FT**: flow-through fraction, **W1** and **W2**: wash steps 1 and 2, and **E**: elution fraction. **CFT**: concentrator flow-through, **SEC**: size exclusion chromatography fractions containing purified wild-type TMD81.

Following nickel affinity chromatography, TMD81 samples were applied to a Superdex-200 10/300 GL size exclusion chromatography column for the final purification step (Fig. 3.1). Elution was monitored by absorbance at 280nm and found to give rise to a major peak corresponding to TMD81 as confirmed by SDS-PAGE.

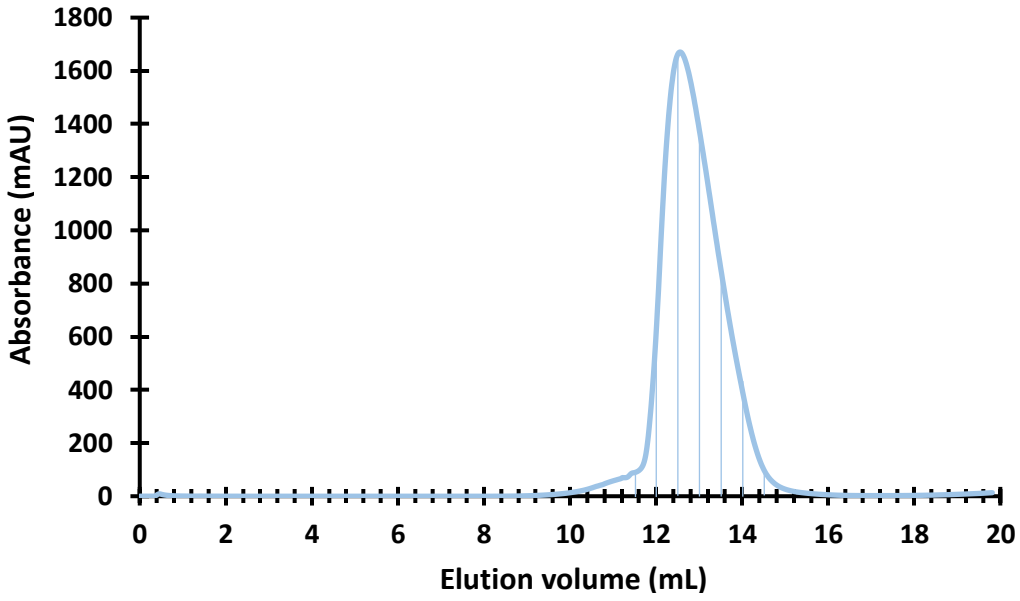


Figure 3.1: **Representative size exclusion chromatography profile** of wild-type TMD81 (blue) after purification by nickel-affinity chromatography monitored by absorbance at 280nm. Fractions spanning 11-15mL were collected for SDS-PAGE analysis.

3.2 ^{19}F NMR of TMD81

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

To determine if the native cysteine (C104) in TMD81 was accessible for labeling with sulfhydryl reactive probes, wild-type TMD81 was subjected to a labeling reaction with TET and ^{19}F spectra were acquired (Fig 3.2). This spectrum contained a single broad peak at -66.80 ppm which is slightly different from the chemical shift of free TET label in solution which showed a single narrow peak at -67.20 ppm. Although the appearance of this peak indicates that the

native C104 site is accessible to derivatization by TET, we hypothesized that it should be possible to deconvolute this peak from those produced by an additional label introduced at one other site. Therefore, mutations were made to introduce another cysteine residue in proximity to the gate to monitor dynamics around this region.

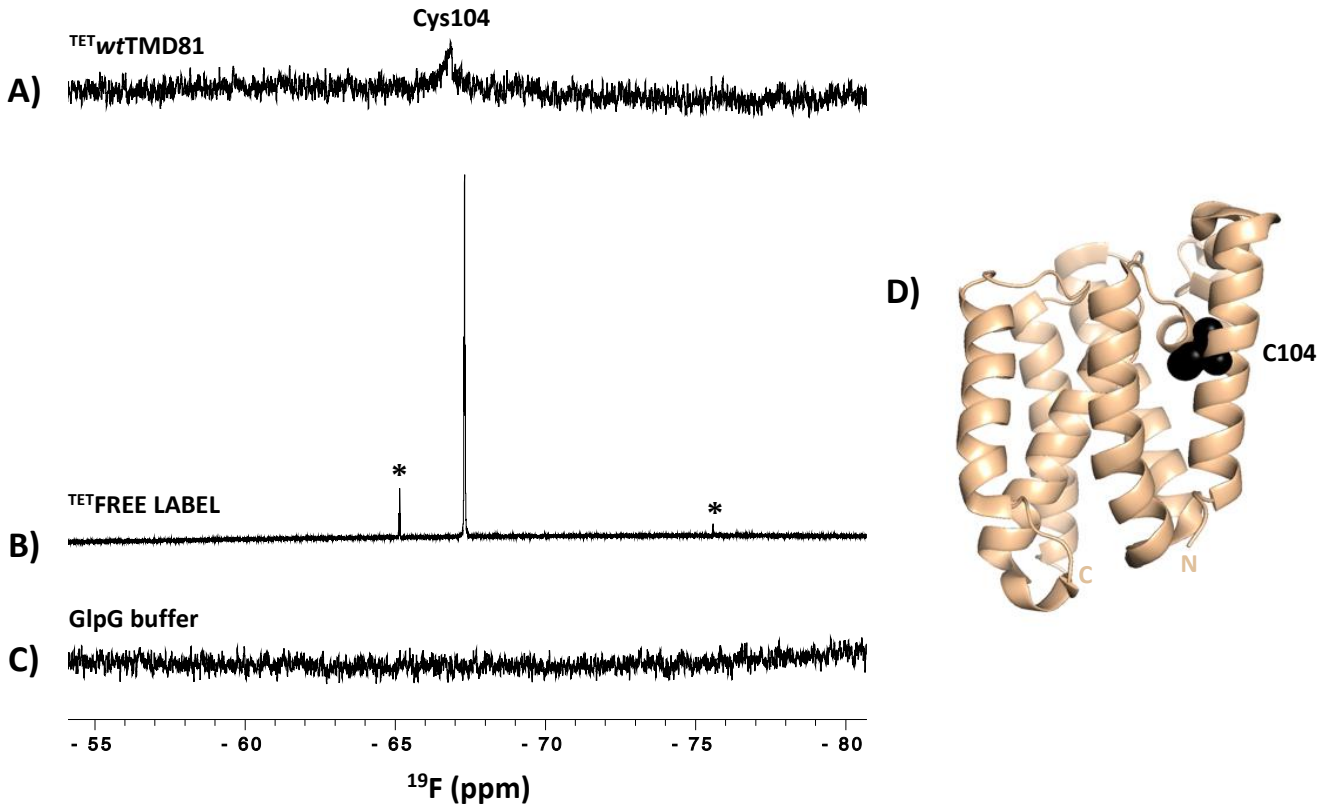


Figure 3.2: **(A)** ^{19}F NMR spectrum of 40μM wild-type TMD81 labelled with TET collected at 25°C. **(B)** ^{19}F NMR spectrum of free TET label in GlpG buffer collected at 25°C. **(C)** GlpG buffer containing no ^{19}F TET label. Note that peaks corresponding to contaminants are indicated by an asterisk (*). **(D)** Structure of *ec*GlpG (wheat) showing the position of the native Cys104 (black).

The first site chosen to introduce a non-native cysteine residue was Phe153 on helix α2, since this is found at the interface of the gating region and interacts with W236A from α5[51] (Fig. 3.3). As shown in Figure 3.4, the ^{19}F spectrum of TET-labelled F153C showed two distinct peaks with the peak at -66.90ppm (labeled peak 2), likely originating from the native C104 site and the other peak at -66.20ppm being attributed to the probe at residue 153 (peak 1).

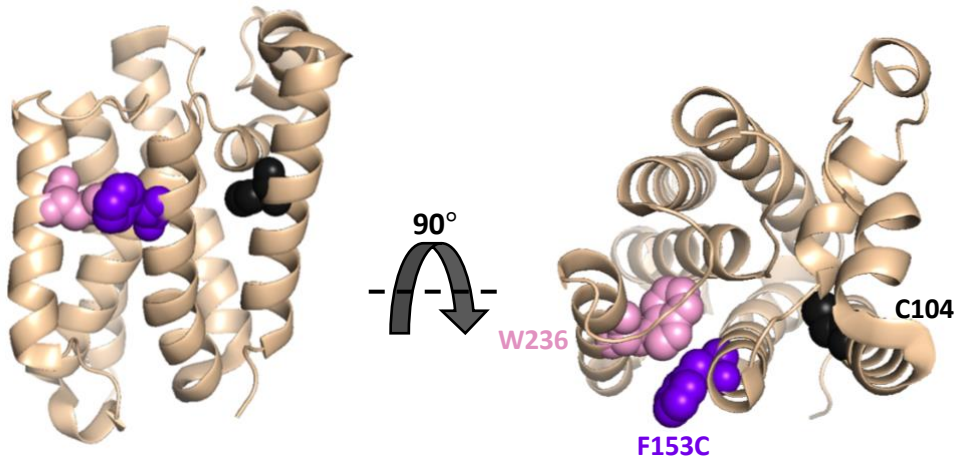


Figure 3.3: Structure of *ec*GlpG (wheat) showing the position of Phe153 (purple) relative to Trp236 (pink) and the native Cys104 (black).

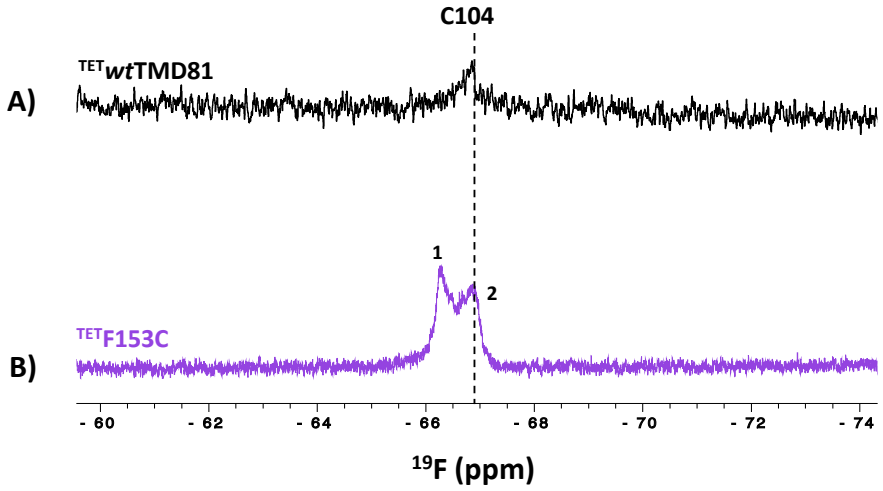


Figure 3.4: **A)** ^{19}F NMR spectrum of 40μM wild-type TMD81 labelled with TET collected at 25°C. **B)** ^{19}F NMR spectrum of 100μM F153C labelled with TET collected at 25°C.

To determine whether the two peaks in the F153C spectrum were due to distinct labels or arose from chemical exchange of one probe sampling more than one state, spectra were acquired on the same F153C sample over a range of temperatures, from 8°C to 45°C (Fig. 3.5). As expected, the relative intensity of the two species was similar for all spectra throughout the series, suggesting that each peak corresponded to a separate label, with peak 2 likely being the probe at Cys104. As expected, peaks became narrower as the temperature increased due to the

faster tumbling of the complexes at higher temperature. Some of this narrowing may also reflect an increase in the dynamics of the labeled site with increasing temperature, with the more solvent-accessible side chain of residue 153 showing a larger effect. It is important to note that throughout the duration of this temperature series, some of the TET label dissociated from the protein, possibly due to reduction of the disulfide bond, giving rise to sharp peaks in the NMR spectrum. Loss of TET label has also been observed at elevated temperatures for other proteins labelled in this manner[66,76,88].

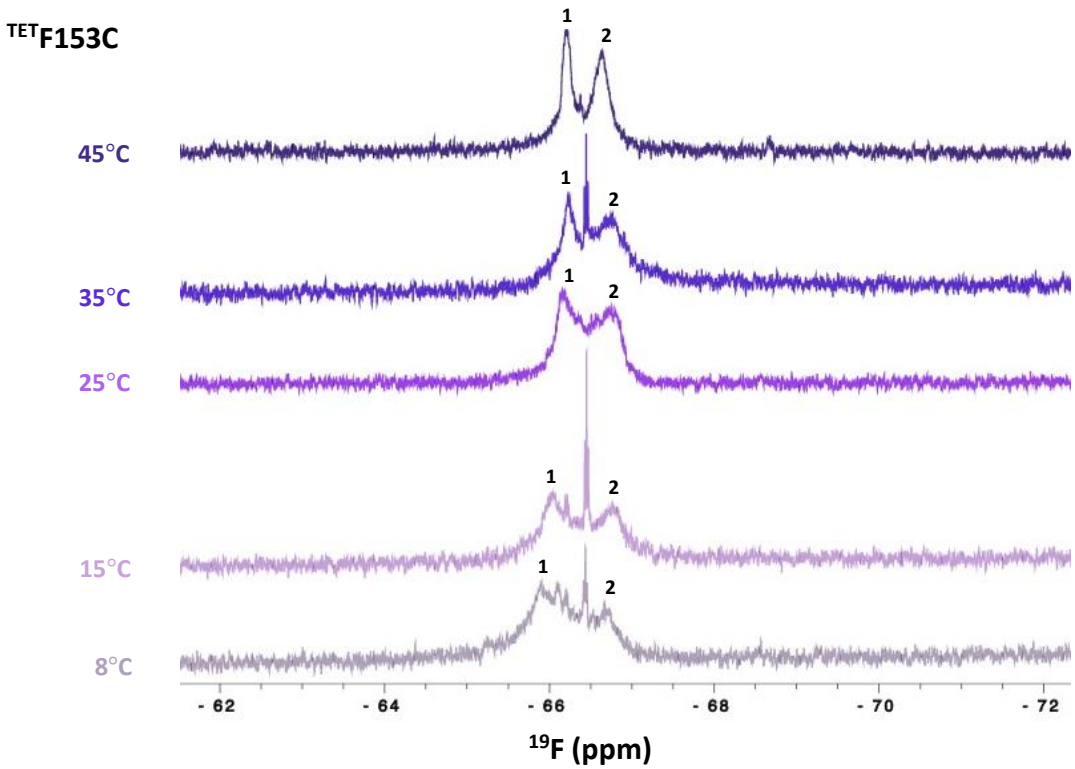


Figure 3.5: ^{19}F NMR spectrum of 100μM F153C labelled with TET collected at various temperatures throughout the variable temperature series.

To determine if the open-gate form of TMD81 would give rise to a change in the ^{TET}F153C spectrum, a W236A substitution mutation was introduced into the F153C mutant. This alanine substitution was previously found to weaken the interactions between α2 and α5, significantly enhancing enzymatic activity[36,37,39,53,54] and potentially promoting the opening of

the lateral α5 gate. To investigate this phenomenon, spectra were acquired on TET-labelled F153C/W236A and compared to that of TET-labelled F153C collected at 25°C (Fig. 3.6). This spectrum was comparable to that of ^{TET}F153C, with peak 2 likely corresponding to the native C104, however, both peaks in the ^{TET}F153C/W236A spectrum appear sharper relative to those in the spectrum of the single mutant.

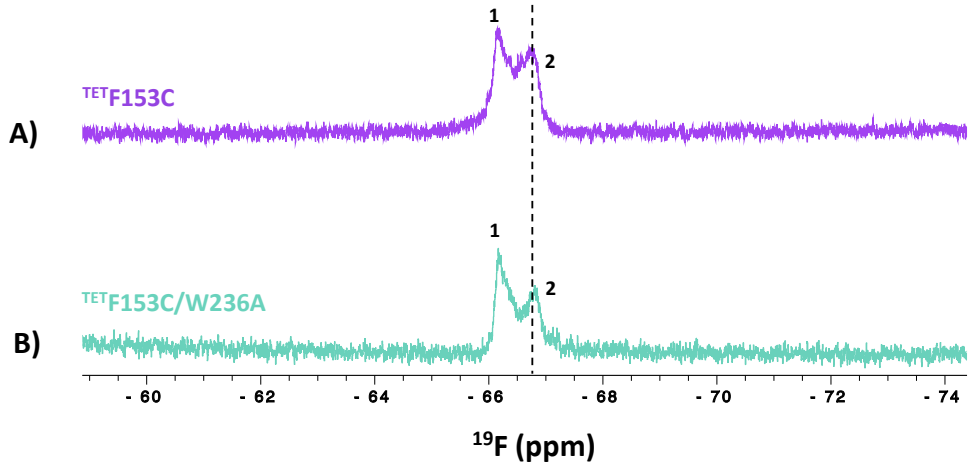


Figure 3.6: **A)** ^{19}F NMR spectrum of 100μM F153C labelled with TET collected at 25°C. **B)** ^{19}F NMR spectrum of 80μM F153C/W236A labelled with TET collected at 25°C.

The influence of temperature was also investigated on this sample (Fig 3.7.). Again, it was observed that peak 2, belonging to the native C104 site experienced slight sharpening as temperatures increased. Similarly, peak 1 also become narrower with increased temperatures. Interestingly, when the temperature was reduced a new peak arose in the spectrum around -66.58ppm, labelled peak 3. This peak is most likely the result of conformational exchange, possibly from an alternate rotameric state of the F153C site made available by the W236A mutation since the alanine residue is less sterically hindered at this interface.

Peaks in the TET-labelled F153C and TET-labelled F153C/W236A spectra appear to be similar at higher temperatures (25°C and above). Additionally, the chemical shift of peak 1 in the spectrum of open-gate F153C/W236A mutant does not change much with temperature but varies slightly in the F153C spectrum with fluctuating temperatures. This may indicate that the probe experiences the same environment at all temperatures in the open-gate F153C/W236A

mutant. These results suggest that there is potential for some exchange process to be detected at this site, although with the confounding presence of the C104 peak it is not possible to identify its origins at this time.

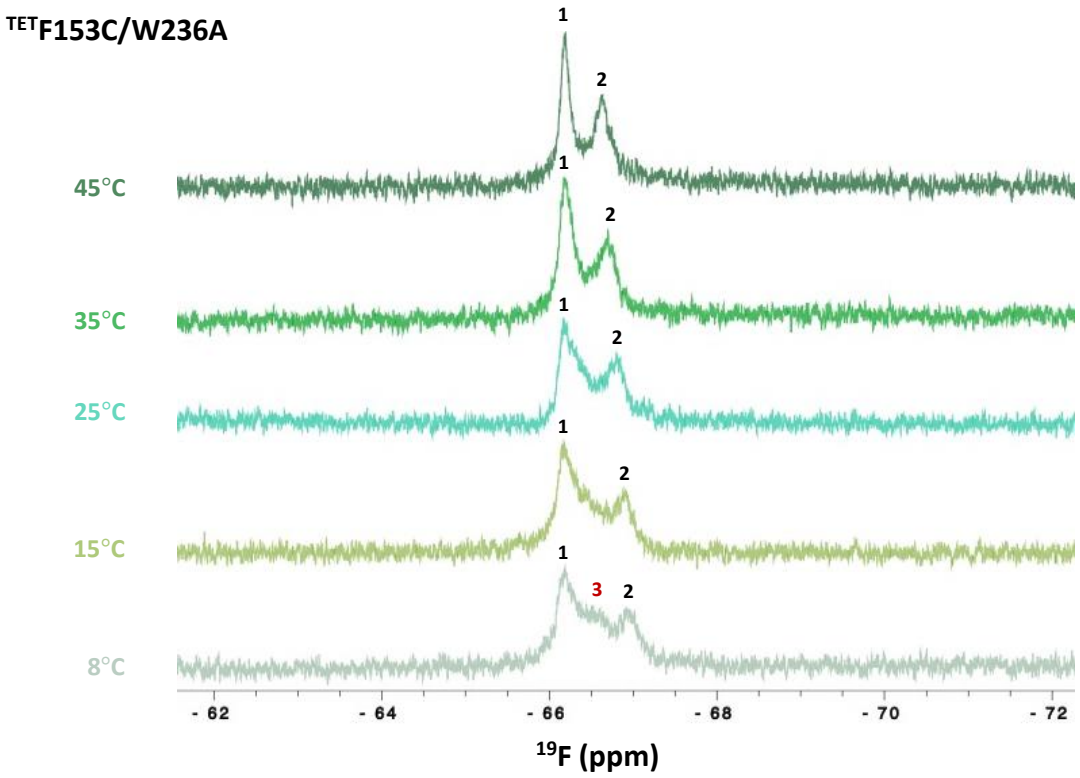


Figure 3.7: ^{19}F NMR spectrum of 80μM F153C/W236A labelled with TET collected at various temperatures throughout the variable temperature series.

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

We also investigated the utility of TFMB as a ^{19}F probe for the study of TMD81 dynamics. Its large size may replace the lost van der Waals interactions made by the mutation of Phe to Cys while the six ^{19}F nuclei offer greater NMR signal-to-noise. The ^{19}F NMR spectrum of wild-type TMD81 derivatized with TFMB was compared to that of free TFMB label in solution with and without DPC detergent micelles (Fig. 3.8). Unexpectedly, TFMB-labelled wild-type

TMD81 showed 2 distinct peaks, peak 1 at -62.50ppm and peak 2 at -63.30ppm both within the vicinity of the TFMB free label peaks. To explore whether temperature could impact the relative intensity of these peaks, spectra were collected on the same wild-type TMD81 sample at various temperatures. As seen in Figure 3.9, the relative populations of both peak 1 and peak 2 remained the same throughout the varying temperature gradient. Both peaks experienced narrowing at high temperatures and broadening at low temperatures, consistent with slower rotational correlation times due to increased sample viscosity at these temperatures[114]. This suggests that there is no significant exchange process occurring between the 2 peaks in the TFMB-labeled wild-type TMD81 sample, although it is not possible to discern the origins of these peaks at this time. Nevertheless, we decided to investigate the gating dynamics of the cysteine mutants using this probe, as it may be possible to distinguish these peaks from those arising from the F153C site.

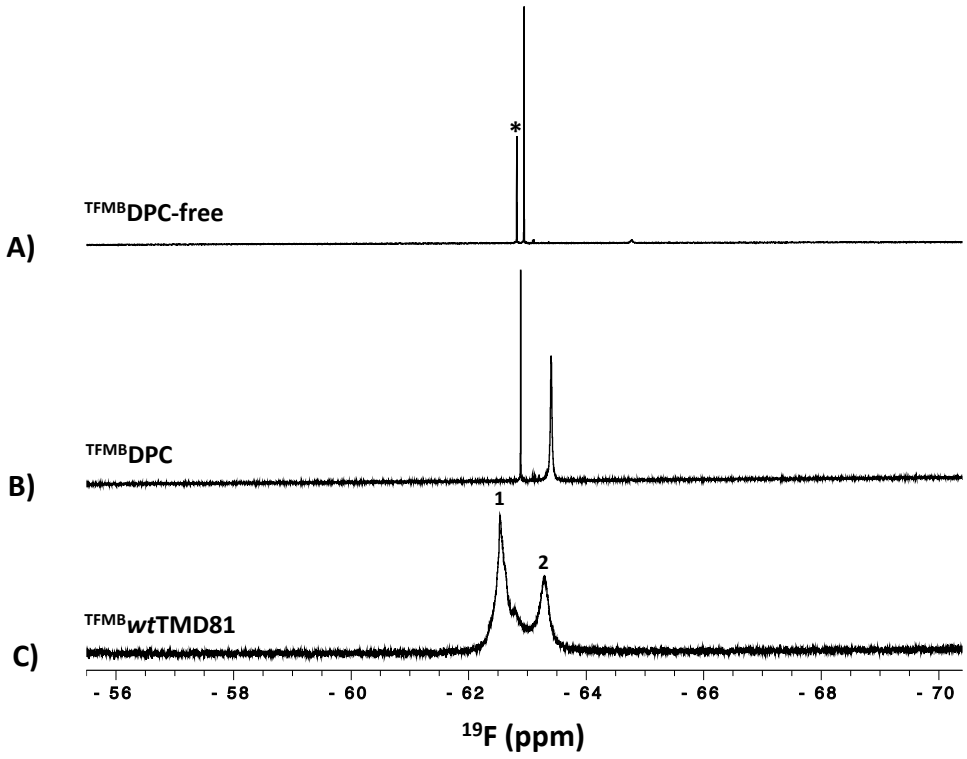


Figure 3.8: **(A)** ^{19}F NMR spectrum of free TFMB label in aqueous buffer collected at 25°C. **(B)** ^{19}F NMR spectrum of free TFMB label in 0.1% w/v DPC at 25°C. The broader peak at -63.5ppm may represent the population of the TFMB probe that has partitioned into the detergent micelle. **(C)** ^{19}F NMR spectrum of 100μM wild-type TMD81 labelled with TFMB 25°C. Peaks corresponding

to contaminants in the commercial label are indicated by an asterisk (*). These contaminants are likely removed during the purification process after the labelling reaction.

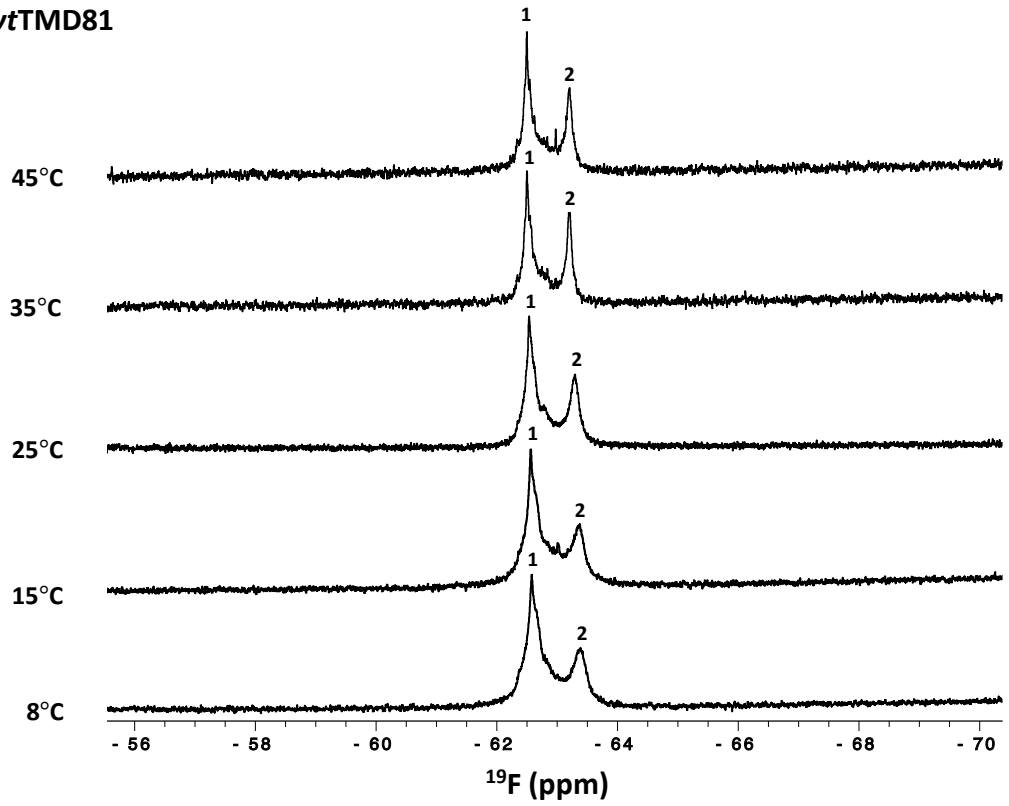
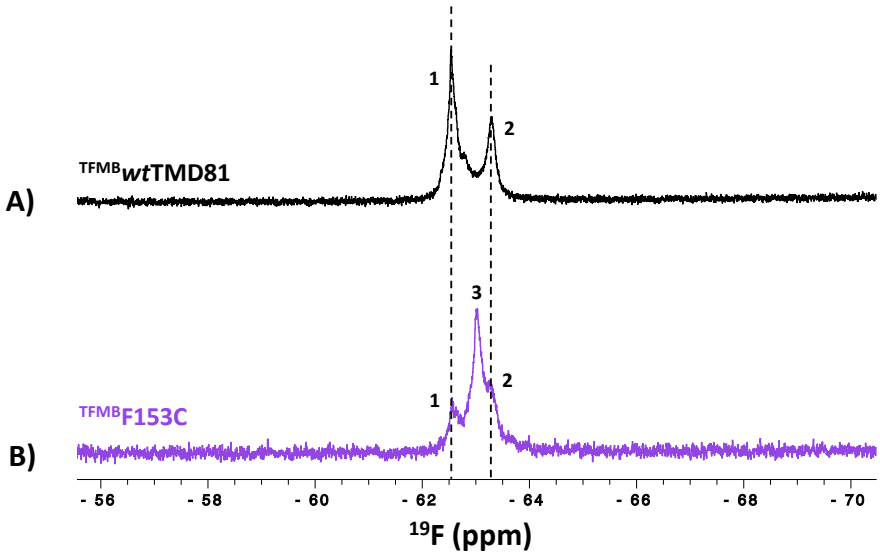


Figure 3.9: ^{19}F NMR spectra of 100μM *wt*TMD81 labelled with TFMB collected at various temperatures

To investigate the ability of this probe to study gating dynamics within TMD81, the F153C mutant was labelled with TFMB and its ^{19}F spectrum was collected and compared to that of wild-type TMD81 at 25°C (Fig. 3.10). TFMB-labelled F153C showed 3 peaks, with peaks 1 and 2 of F153C being aligned with those from the wild-type TMD81 at -62.50ppm and -63.30ppm. An additional peak was also observed (peak 3) that is likely due to the label at site Phe153, as this peak is absent in the wild-type spectrum. This peak is narrower and more intense than the other peaks, representing increased mobility of the label at site 153. The influence of temperature using the TFMB probe was also investigated on the F153C mutant (Fig. 3.10C). It was observed that at increased temperatures, the relative intensities for all three peaks were

similar. Interestingly however, peak 3 became broader suggesting an exchange process may be occurring that is slower on the NMR time scale at the low temperature.



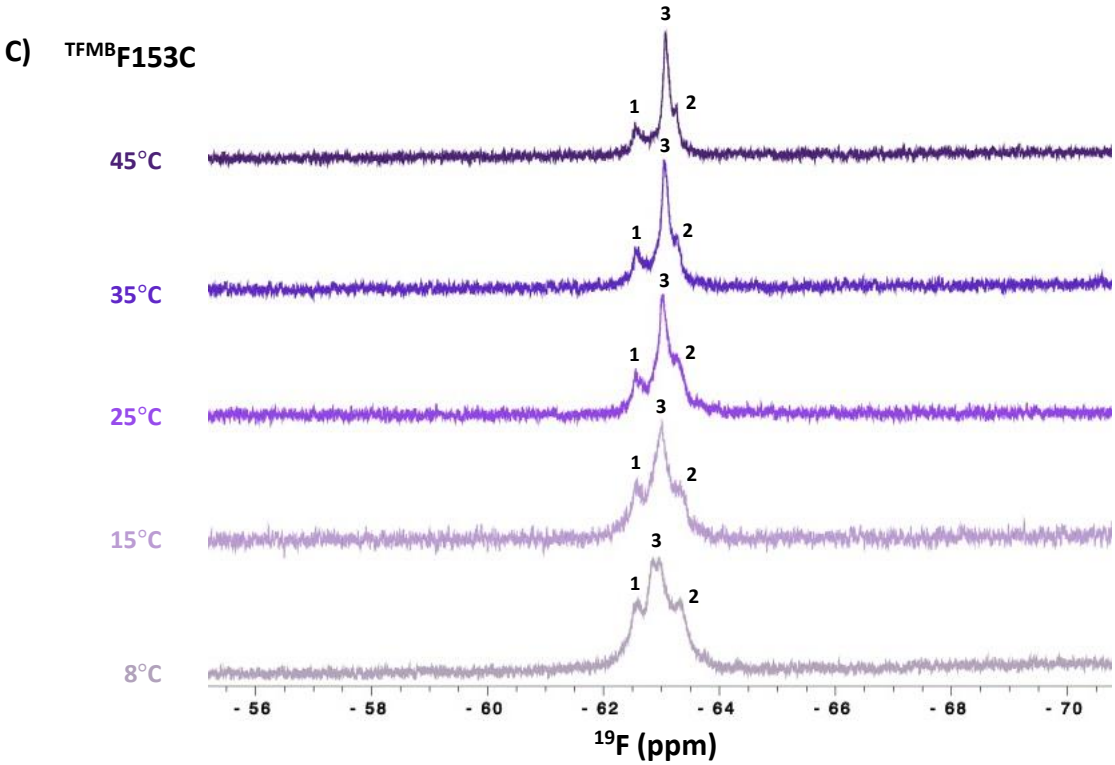
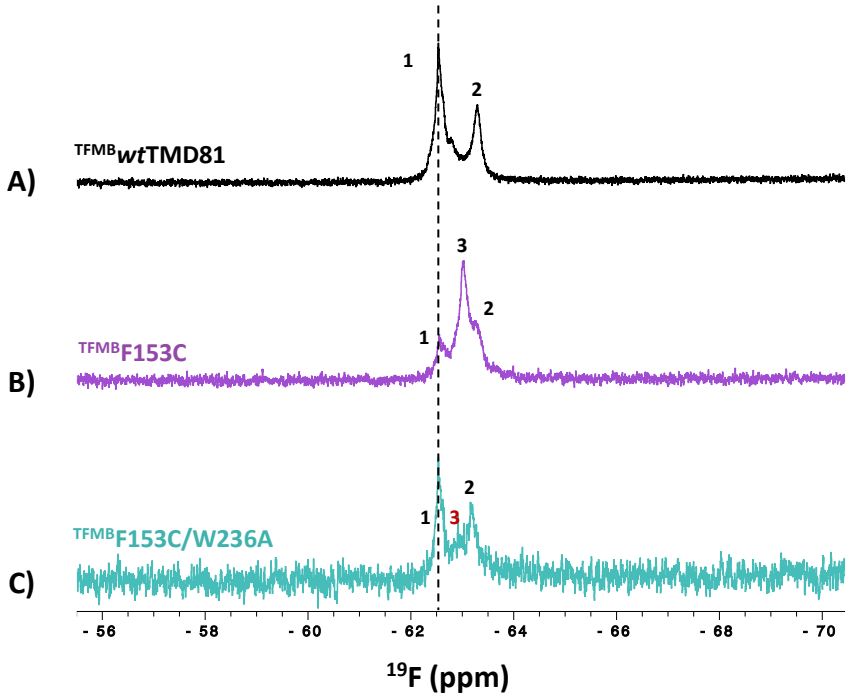


Figure 3.10: **(A)** ^{19}F NMR spectrum of 100μM wild-type TMD81 labelled with TFMB collected at 25°C. **B)** ^{19}F NMR spectrum of 70μM F153C labelled with TFMB collected at 25°C. **(C)** ^{19}F NMR spectra of F153C labelled with TFMB collected at various temperatures

To investigate the effect of a change in gate dynamics on this spectrum, the F153C/W236A mutant was labelled with TFMB and its ^{19}F spectrum was collected and compared to that of TFMB labelled F153C and wild-type TMD81 at 25°C (Fig. 3.11). The TFMB-labelled F153C/W236A spectrum exhibited low signal-to-noise relative to the wild-type and single mutant spectra, even though the concentration was similar between all three samples. It was possible to identify three peaks in this spectrum, with Peak 1 of F153C/W236A being the same as peak 1 in F153C and wild-type TMD81 at -62.50ppm. Peaks 2 and 3 of F153C/W236A have similar, although not identical, chemical shifts to peaks 2 and 3 of TFMB-labelled F153C. The significant loss in intensity for peak 3 in F153C/W236A may reflect intermediate timescale dynamics giving rise to significant broadening. Interestingly, at lower temperatures, peak 3 showed a small increase in intensity while peak 2 showed a decrease in intensity, (Fig. 3.11D) suggesting that these two species were undergoing exchange that became slow on the NMR timescale when the temperature decreased.



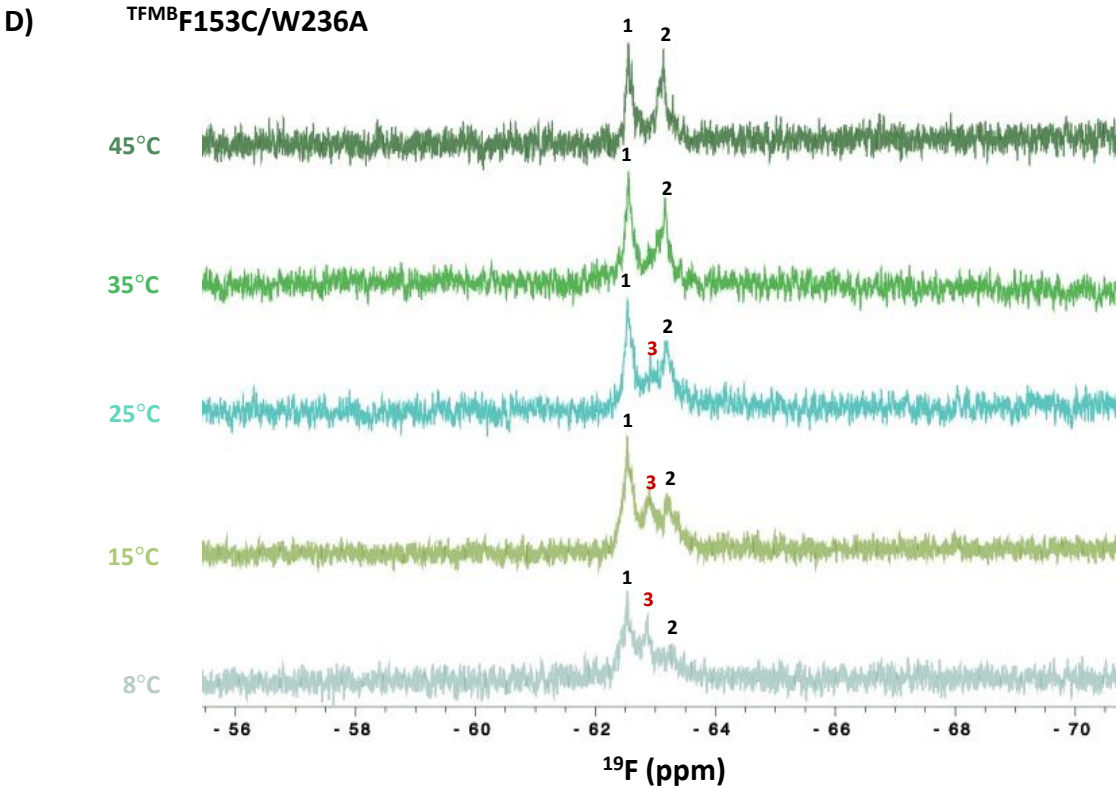


Figure 3.11: **(A)** ^{19}F NMR spectrum of 100μM wild-type TMD81 labelled with TFMB collected at 25°C. **(B)** ^{19}F NMR spectrum of 70μM F153C labelled with TFMB collected at 25°C. **(C)** ^{19}F NMR spectrum of 50μM F153C/W236A labelled with TFMB collected at 25°C. **(D)** ^{19}F NMR spectra of 50μM F153C/W236A labelled with TFMB collected at various temperatures.

3.3 Investigation of cysteine knockout C104S

Since the accessibility of Cys104 to reaction with the trifluoromethyl probes complicated the interpretation of ^{19}F spectra, this residue was mutated to serine to make C104S.

3.3.1 Structural analysis of C104S

3.3.1.1 ^{1}H-^{15}N HSQC NMR of C104S

To investigate the impact of the C104S mutation on the structure of TMD81, ^{1}H – ^{15}N HSQC spectra were acquired on ^{15}N labelled C104S samples and compared to that of wild-type TMD81 (Fig. 3.12). Most peaks in the C104S spectrum had chemical shifts that were similar to those in the wild-type TMD81 spectrum. In general, the C104S spectrum showed more broadening relative to the WT spectrum, as well as some changes in chemical shifts in both the backbone (Fig. 3.12A) and tryptophan side chain regions (Fig. 3.12B) of the spectrum. This suggests that there may be small changes in structure and dynamics introduced by the mutation.

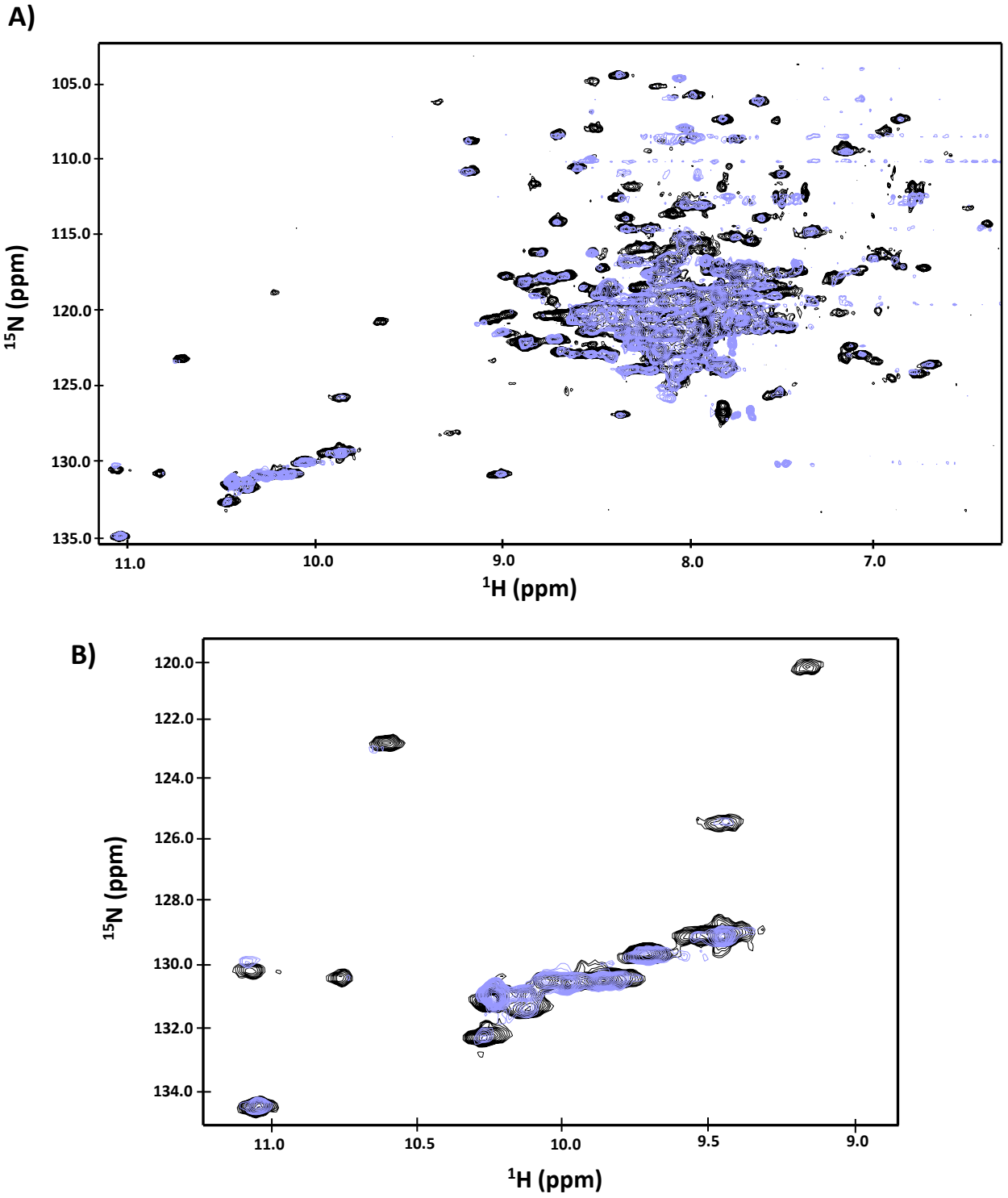


Figure 3.12. (A) ^{1}H – ^{15}N HSQC spectrum of uniformly ^{15}N-labelled C104S (purple) superimposed on the wild-type TMD81 spectrum (black). **(B)** Expanded view of the tryptophan side chain region. Spectra were recorded at 45°C, in 0.1% w/v DPC detergent, pH 7.4 with concentrations of 100μM (wild-type TMD81) and 70μM (C104S).

3.3.1.2 Circular dichroism studies on C104S

To assess the secondary structure content of the cysteine knockout mutant, circular dichroism (CD) spectra of C104S was compared to that of wild-type TMD81. Both protein samples produced spectra consistent with the high α-helical content expected for folded TMD81, showing characteristic minima at 208 and 220 nm. As seen in Figure 3.13, the CD spectrum of C104S (purple) shows agreement with the wild-type TMD81 spectrum (black) although there is a loss in relative intensity at both minima for this mutant. This may indicate that the C104S mutation produced a slight loss in the α-helical content of TMD81.

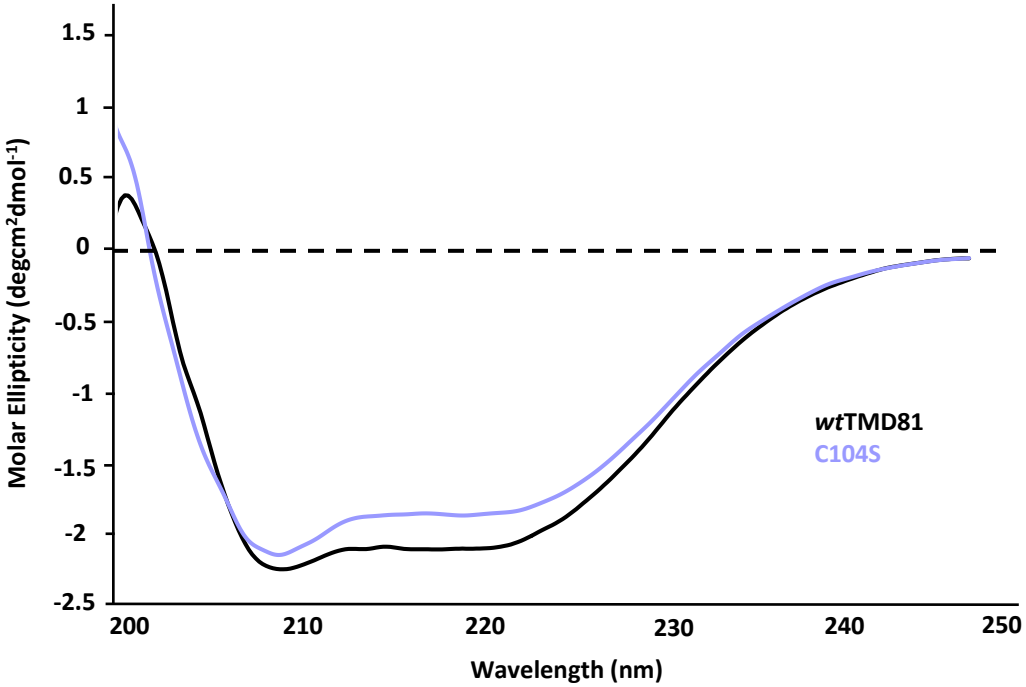


Figure 3.13. Representative circular dichroism spectra of 5μM C104S (purple) superimposed on the spectrum of 5μM wild-type TMD81 (black). Spectra were recorded at 37°C. Eight accumulations were collected for each sample using a data pitch of 0.2 nm and a scan rate of 20nm/min. Accumulations were averaged to produce a single spectrum.

3.3.1.3 Thermal denaturation studies on C104S

To assess the thermal stability of the cysteine knockout mutant, thermal denaturation studies were carried out (Fig. 3.14) on both WT TMD81 and C104S. A single phase transition

was observed in denaturation profiles, with melting temperatures of 76.1 $\pm$ 0.42°C and 73.9 $\pm$ 0.24°C for the WT and mutant respectively, similar to melting temperatures previously observed for *ec*GlpG[54]. The C104S mutation does seem to have a slightly lower melting temperature with a larger change in signal for its denatured state when compared to that of wild-type TMD81. This suggests that thermal stabilities between wild-type and mutant are not completely innocuous however the overall structural integrity and thermostability is not significantly altered by the C104S mutation.

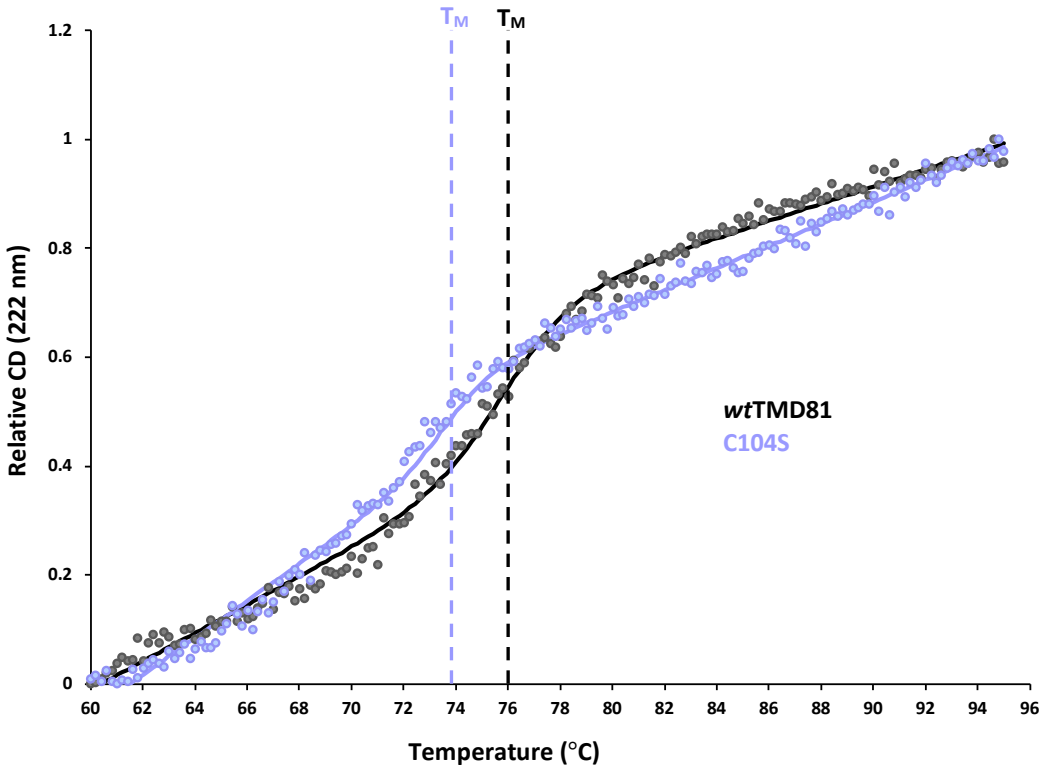


Figure 3.14. Thermal denaturation fitted curves[115,116] for 5μM of wild-type TMD81 (black) and C104S (purple). Spectra are obtained through monitoring the loss of CD signal intensity at 222nm from 60°C – 95°C at a rate of 0.3°C/min.

3.3.2 Functional analysis of C104S

To investigate the impact of the Cys104 mutation on the function of TMD81, an activity assay was conducted using a water-soluble BODIPY-labeled casein model substrate as described

previously[39]. Initial rates of proteolysis were found from the slope of the initial phase of the fluorescence versus time graph (Fig. 3.15A) and plotted as a function of substrate concentration (Fig. 3.15B). Both the wild-type TMD81 and C104S samples gave rise to Michaelis-Menten kinetic profiles where both samples showed similar observed activities. A significant and reproducible reduction in specific activity was observed for the C104S mutant over the range of tested substrate concentrations, indicating that this substitution may cause a loss in TMD81 activity (~2-fold decrease).

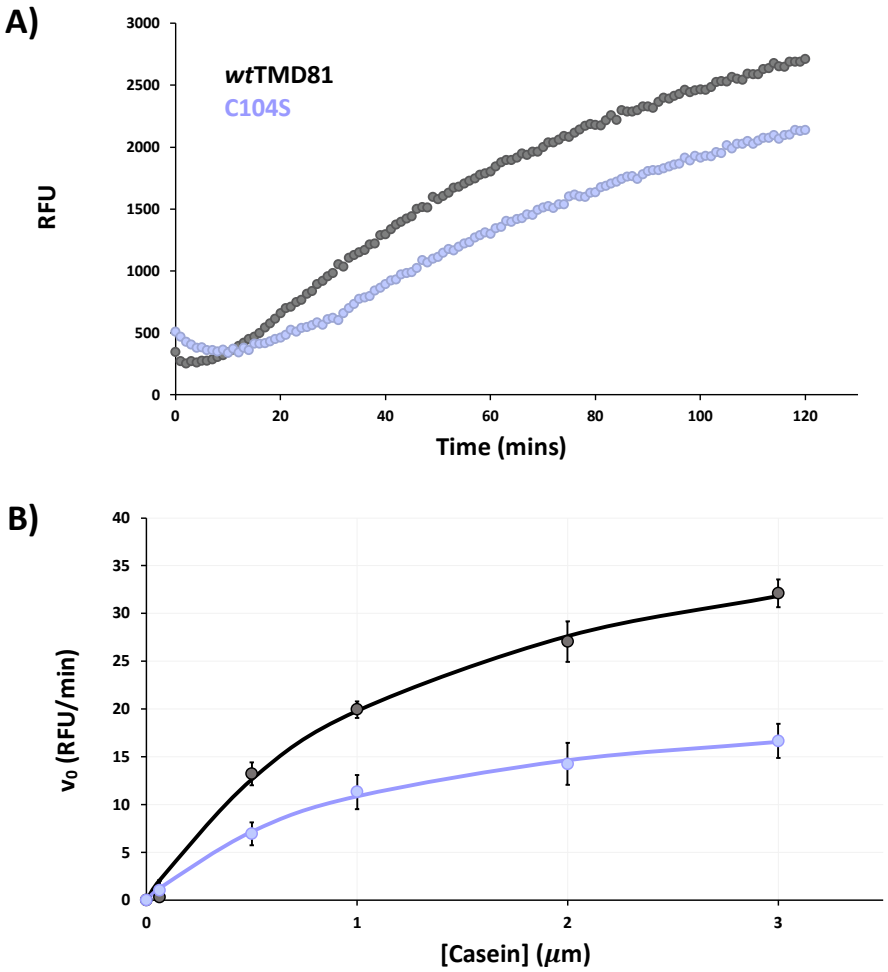


Figure 3.15. (A) Fluorescence profiles from BODIPY-labeled casein recorded with 0.5 μM wild-type TMD81 (black) and C104S (purple). **(B)** Michaelis-Menten plots showing initial rates (v_0) vs. casein substrate concentrations for wild-type TMD81 (black) and C104S (purple). K_M values were found to be 1.31 $\pm$ 0.1μM and 1.07 $\pm$ 0.2μM while k_{cat} values were found to be 1.52$\pm$0.14 RFU/seconds/μM and 0.74$\pm$0.1 RFU/seconds/μM for wild-type TMD81 and C104S respectively.

3.4 ^{19}F NMR of TET and TFMB-labelled TMD81 C104S

Although structural and functional studies on the C104S mutant showed some differences relative to wild-type TMD81, we were interested in investigating ^{19}F NMR spectra of labeled samples in the absence of the native cysteine residue as these spectra were anticipated to be more straightforward to interpret. To confirm that the single site of reactivity was no longer present in TMD81, a C104S sample was subjected to the same TET labeling protocol and a ^{19}F-NMR spectrum was recorded. As shown in Figure 3.16, no peak was detected, confirming that no other TET-susceptible sites were present.

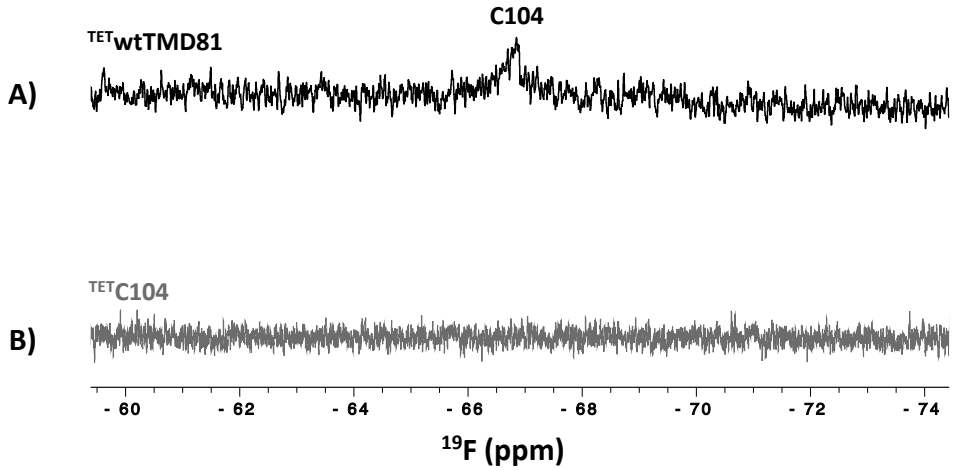


Figure 3.16: **A)** ^{19}F NMR spectrum of 40μM C104S labelled with TET collected at 25°C. **B)** ^{19}F NMR spectrum of 40μM wild-type TMD81 labelled with TET collected at 25°C.

These experiments were also repeated for C104S labeled with TFMB, and its ^{19}F NMR spectrum was compared to that of wild-type TMD81. As shown in Figure 3.17, a peak was still observed in the C104S spectrum with the same chemical shift as the species we had designated as peak 1 in the wild-type spectrum. Therefore, there appears to be an additional site that is being labelled by the TFMB probe, although it is not possible to know what this might be at this time. Since this peak was also present in the ^{19}F NMR spectra of TFMB-labeled F153C and F153C/W236A mutants, it appears that an analogous species/site was labelled in all TMD81 samples. This suggests that peak 2 in the TFMB-labelled wild-type TMD81 ^{19}F NMR spectrum belongs to the native C104 site as this peak is missing in the negative control C104S spectrum.

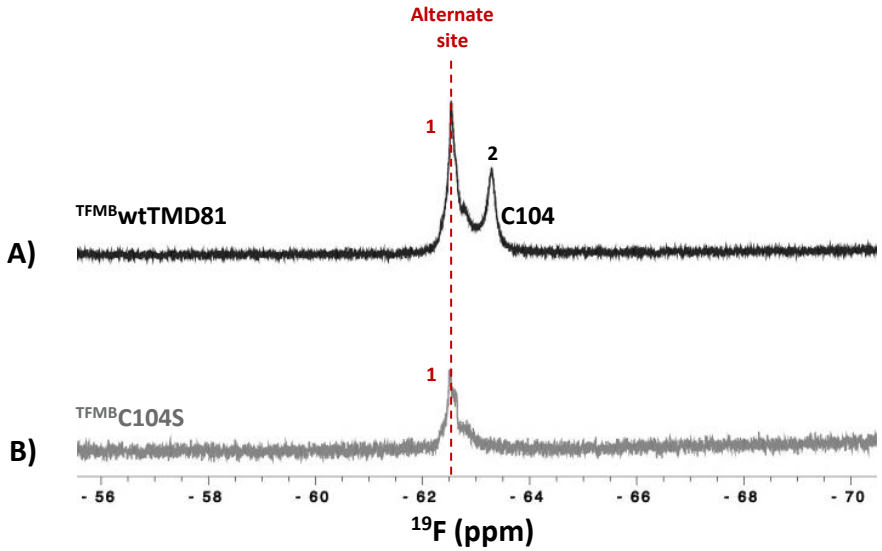


Figure 3.17: **A)** ^{19}F NMR spectrum of 100μM wild-type TMD81 labelled with TFMB collected at 25°C. **B)** ^{19}F NMR spectrum of 70μM C104S labelled with TFMB collected at 25°C.

3.5 ^{19}F NMR of TMD81 C104S mutants

3.5.1 Solution state ^{19}F NMR of TET-labelled TMD81 C104S mutants

Since the removal of the native C104 cysteine simplifies interpretation of the ^{19}F spectrum, TET-labelled C104S/F153C was prepared and compared to that of TET-labelled F153C (Fig. 3.18). Although poor signal-to-noise complicated spectral resolution, 2 distinct peaks can be resolved for C104S/F153C. Peak 1 of C104S/F153C has a chemical shift that is the same as that of peak 1 in the F153C spectrum. Interestingly, an additional peak in the spectrum is also apparent with higher intensity than peak 1, labeled peak 2. It is possible that its chemical shift could corresponds to a feature in the F153C spectrum which may correspond to the same species although spectral crowding could be obscuring this phenomenon. This suggests that the ^{19}F-methyl group of TET labeled F153C in both the WT and Cys-knockout mutant where peaks were undergoing slow exchange between two states. This is supported by the temperature series recorded for the F153C/C104S mutant since peak 2 became sharper with increasing temperature, and broader at lower temperatures (Fig. 3.18C). While it is possible that peak 1

was also undergoing temperature-dependent changes in intensity, the poor signal to noise for this species made it difficult to evaluate this trend.

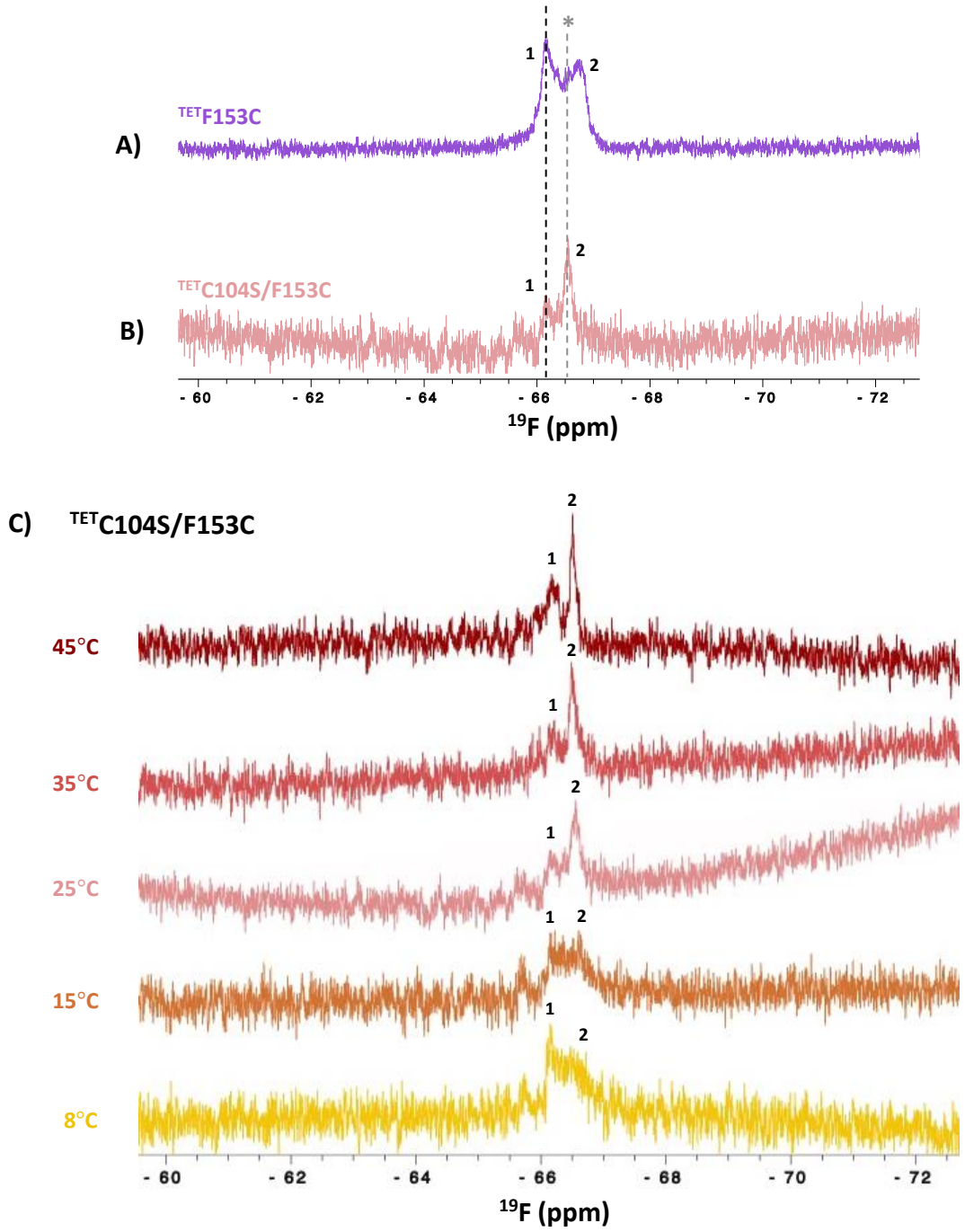


Figure 3.18: **A)** ^{19}F NMR spectra of 100μM F153C labelled with TET collected at 25°C. **B)** ^{19}F NMR spectra of 40μM C104S/F153C labelled with TET collected at 25°C. The additional similar

feature is marked with a grey asterisk (*). **C)** ^{19}F NMR spectra of 40μM C104S/F153C labelled with TET collected at various temperatures.

We also repeated these experiments after also introducing the W236A mutation that is designed to increase activity. As shown in Figure 3.19, the ^{19}F spectrum of this triple mutant (C104S/F153C/W236A) labelled with TET showed 2 peaks with similar positions and intensities to those seen in the spectrum of C104S/F153C. The signal to noise was higher in the triple mutant, which at least partly reflects the higher concentration of this sample. It is also possible that the higher peak intensities may reflect faster dynamics in the triple mutant relative to the double mutant. The temperature-dependent change in the populations of peaks in the triple mutant were also very similar to observations found with TET-labelled C104S/F153C (Fig. 3.19C). This suggests that while there may be changes in gating dynamics introduced by the W236A mutation, given the small size of the TET probe, it is possible that this probe may have changed the nature of the interactions in the gating interface.

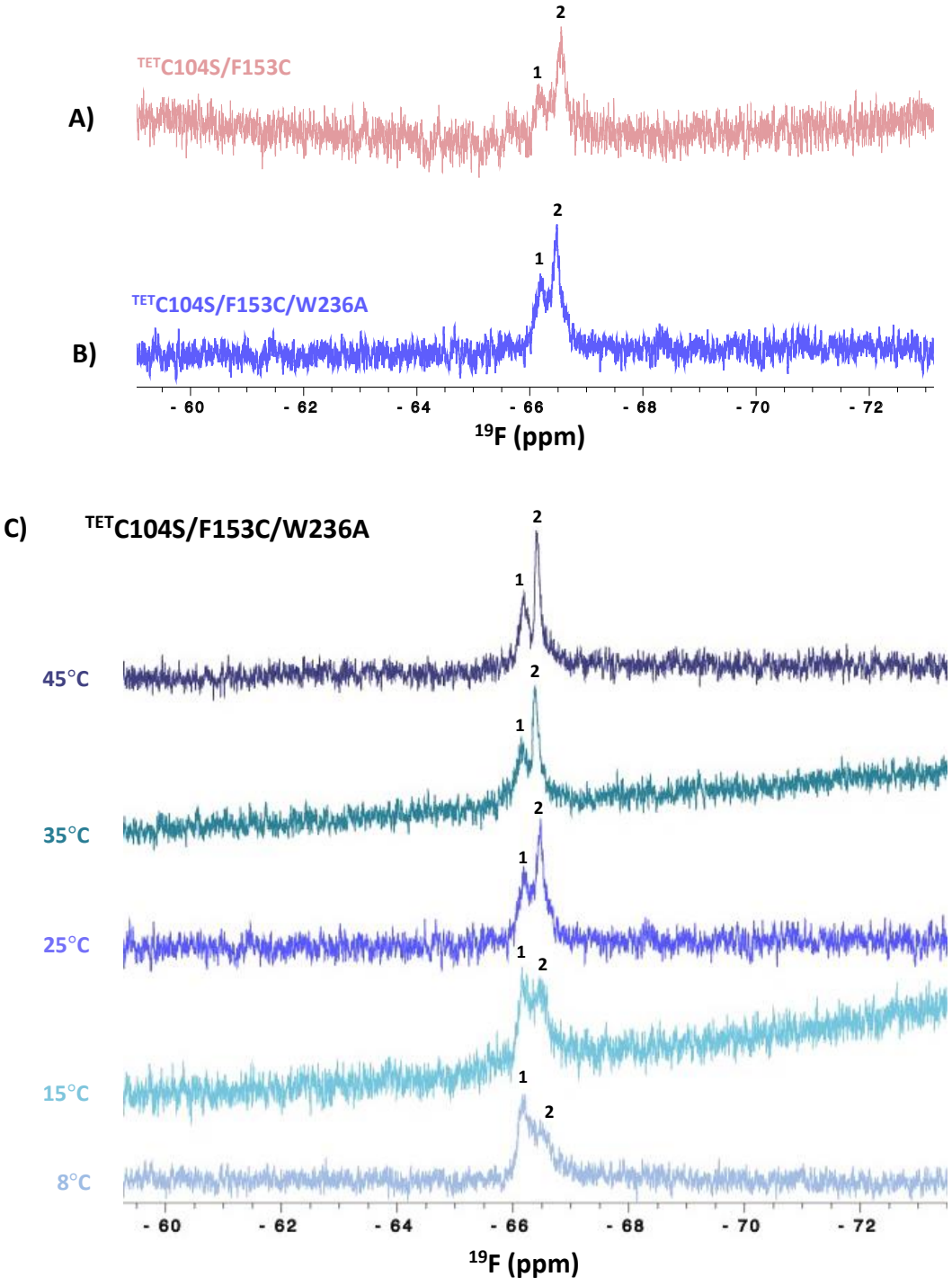


Figure 3.19: **A)** ^{19}F NMR spectra of 40μM C104S/F153C labelled with TET collected at 25°C. **B)** ^{19}F NMR spectra of 70μM C104S/F153C/W236A labelled with TET collected at 25°C. **C)** ^{19}F NMR spectra of 70μM C104S/F153C/W236A labelled with TET collected at various temperatures.

3.5.2 Solution state ^{19}F NMR of TFMB-labeled TDM81 C104S mutants

We also studied the native cysteine knockout mutants with the larger TFMB probe with the same series of Cys-knockout mutants. The double mutant C104S/F153C was labelled with TFMB and its ^{19}F NMR spectrum was compared to that of wild-type TMD81 and F153C (Fig.3.20). Unlike the F153C single mutant spectrum that had 3 peaks, the double mutant only showed 2 peaks, consistent with the loss of labeling at C104. One of the peaks corresponds to peak 1 in the TMD81 spectrum that appears to be due to labelling at an alternate site. The other peak in C104S/F153C corresponds to peak 3 in the F153C spectrum, likely arising from the TFMB label at site Phe153. Similar to the temperature series with the single mutant (Fig. 3.10), as temperatures are decreased, the peak arising from labeled F153C begins to broaden and at 8°C, a new peak is resolved (Fig. 3.21), suggesting a dynamic process that has become slow enough to show two distinct states at this lower temperature.

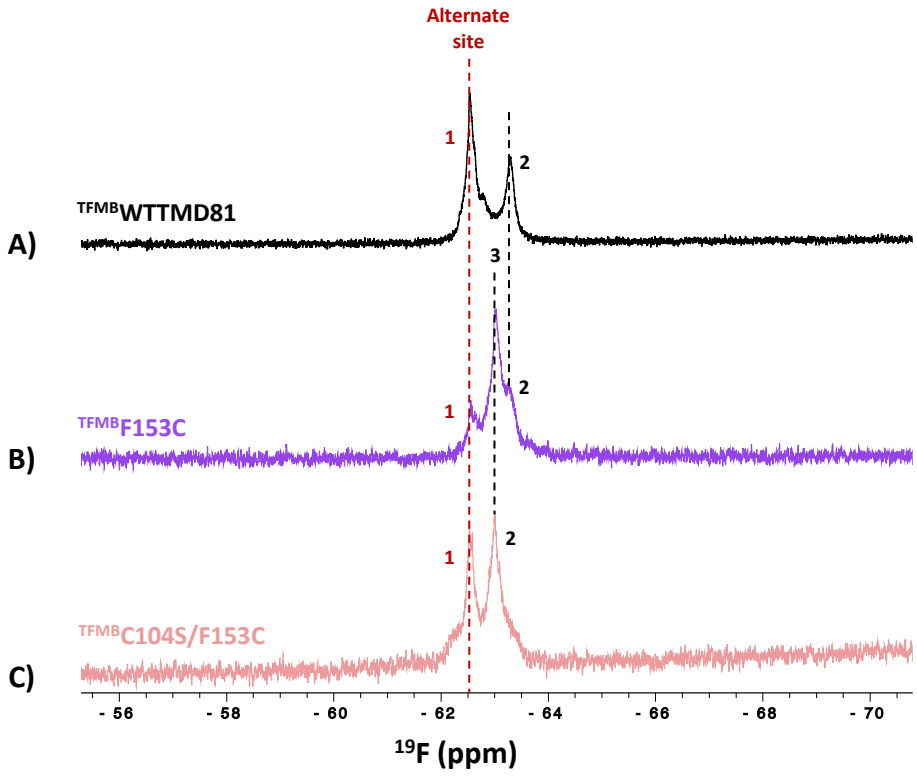


Figure 3.20: **(A)** ^{19}F NMR spectrum of 100μM wild-type TMD81 labelled with TFMB collected at 25°C. **(B)** ^{19}F NMR spectrum of 70μM F153C labelled with TFMB collected at 25°C. **(C)** ^{19}F NMR spectra of 100μM C104S/F153C labelled with TFMB collected at 25°C.

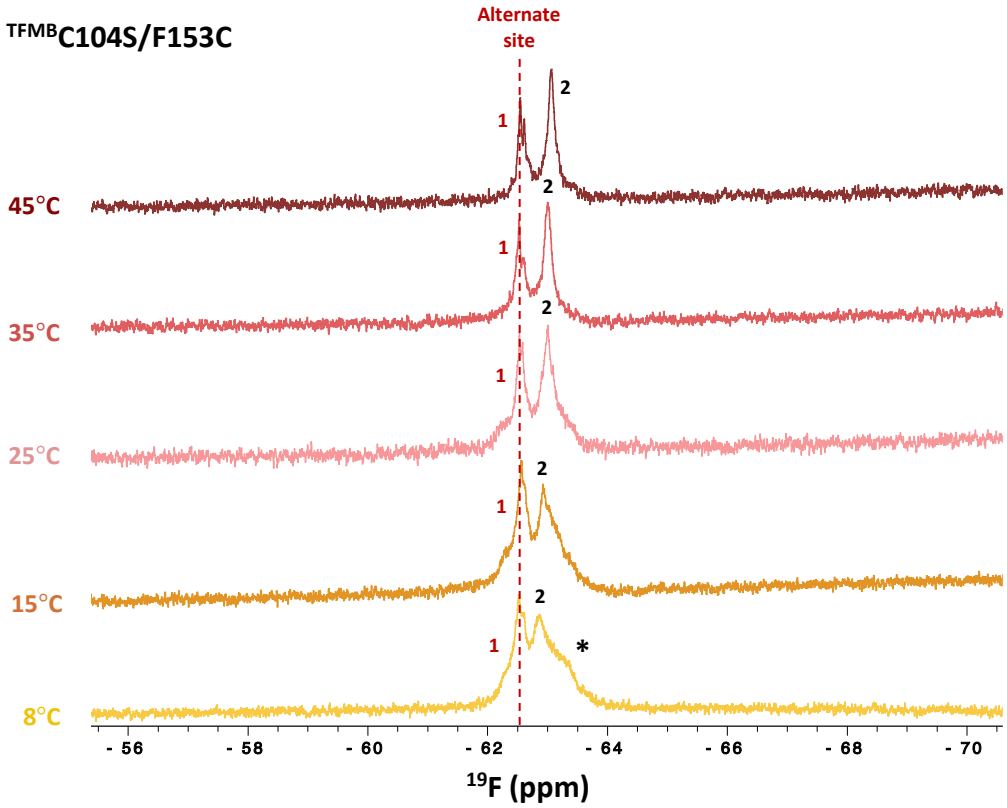


Figure 3.21: ^{19}F NMR spectra of 100μM C104S/F153C labelled with TFMB collected at various temperatures. The formation of the new peak is indicated with an asterisk (*).

We also evaluated the impact of the gate-opening mutation W236A on the spectrum obtained with the TFMB probe. As shown in Figure 3.22, the triple mutant (C104S/F153C/W236A) showed 3 identifiable peaks at 25°C. As expected, the peak corresponding to the alternate site appeared at the same chemical shift as the other mutants (peak 1). A lower intensity peak (labelled peak 3) also appeared at the same chemical shift as peak 2 in the F153C Cys knockout mutant. In addition, another peak was present at the same position as the most upfield peak in the F153C/W236A mutant (labelled peak 2). The temperature series for the labeled triple mutant (Fig. 3.23) showed both peaks are sharper at higher temperatures, and broader at lower temperatures, with a shift in the relative intensity of the peaks accompanying this change. Again, this is suggestive of exchange between two states, potentially due to dynamics around the gate.

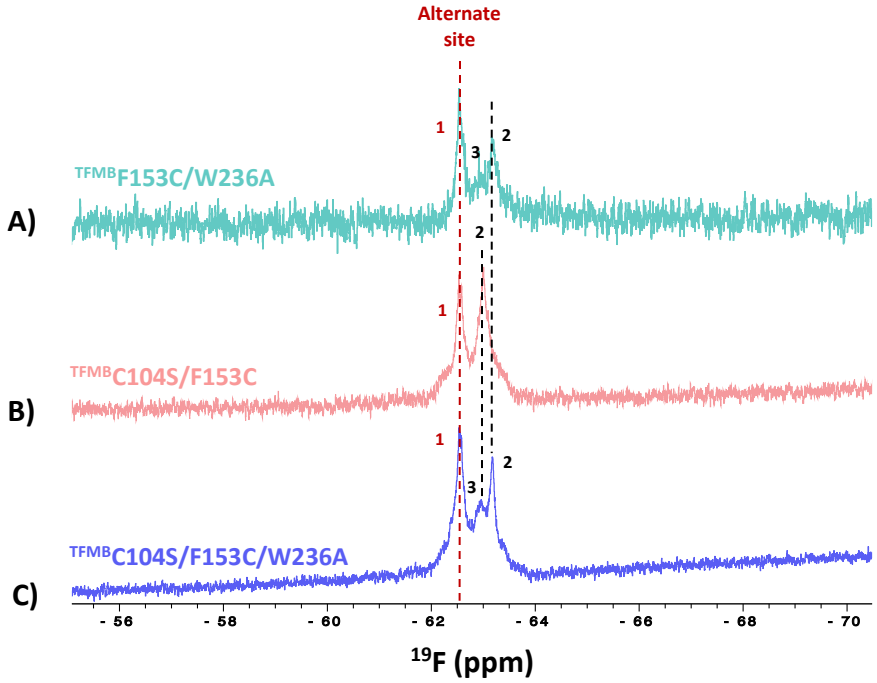


Figure 3.22: **(A)** ^{19}F NMR spectrum of 50μM F153C/W236A labelled with TFMB collected at 25°C. **(B)** ^{19}F NMR spectrum of 100μM C104S/F153C labelled with TFMB collected at 25°C. **(C)** ^{19}F NMR spectra of 90μM C104S/F153C/W236A labelled with TFMB collected at 25°C.

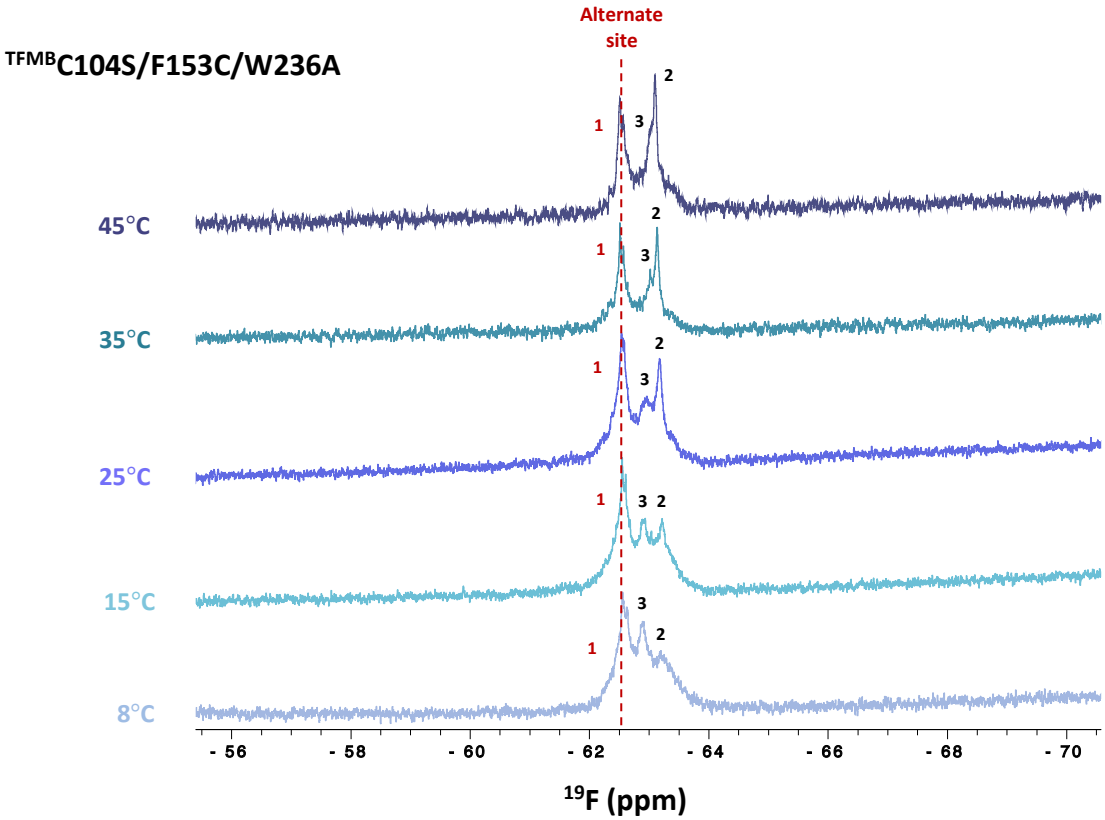


Figure 3.23: ^{19}F NMR spectra of 90μM C104S/F153C/W236A labelled with TFMB collected at various temperatures

3.6 Activity analysis of TMD81 cysteine mutants

To determine the functional implications of the various mutations around the α5-gate, kinetic measurements were performed using fluorogenic casein as the substrate (Fig. 3.24A). Initial rates were measured and plotted as a function of substrate concentration. Since it was not possible in all cases to reach saturating concentrations of substrate due to solubility limitations, we chose to use the slope to the initial region of this curve to give an estimate of k_{cat}/K_M. which is the catalytic efficiency of the enzyme. While all mutants retained significant levels of catalytic activity against the water-soluble substrate, the C104S mutation did cause a small but significant decrease in activity in all mutants harbouring the cysteine knock out.

The influence of the trifluoromethyl probes on the function of wild-type TMD81 was also evaluated (Fig. 3.24B), and showed that both probes significantly decreased catalytic efficiency, with TET labeling inducing a 5-fold decrease and TFMB labeling giving rise to a 30-fold decrease.

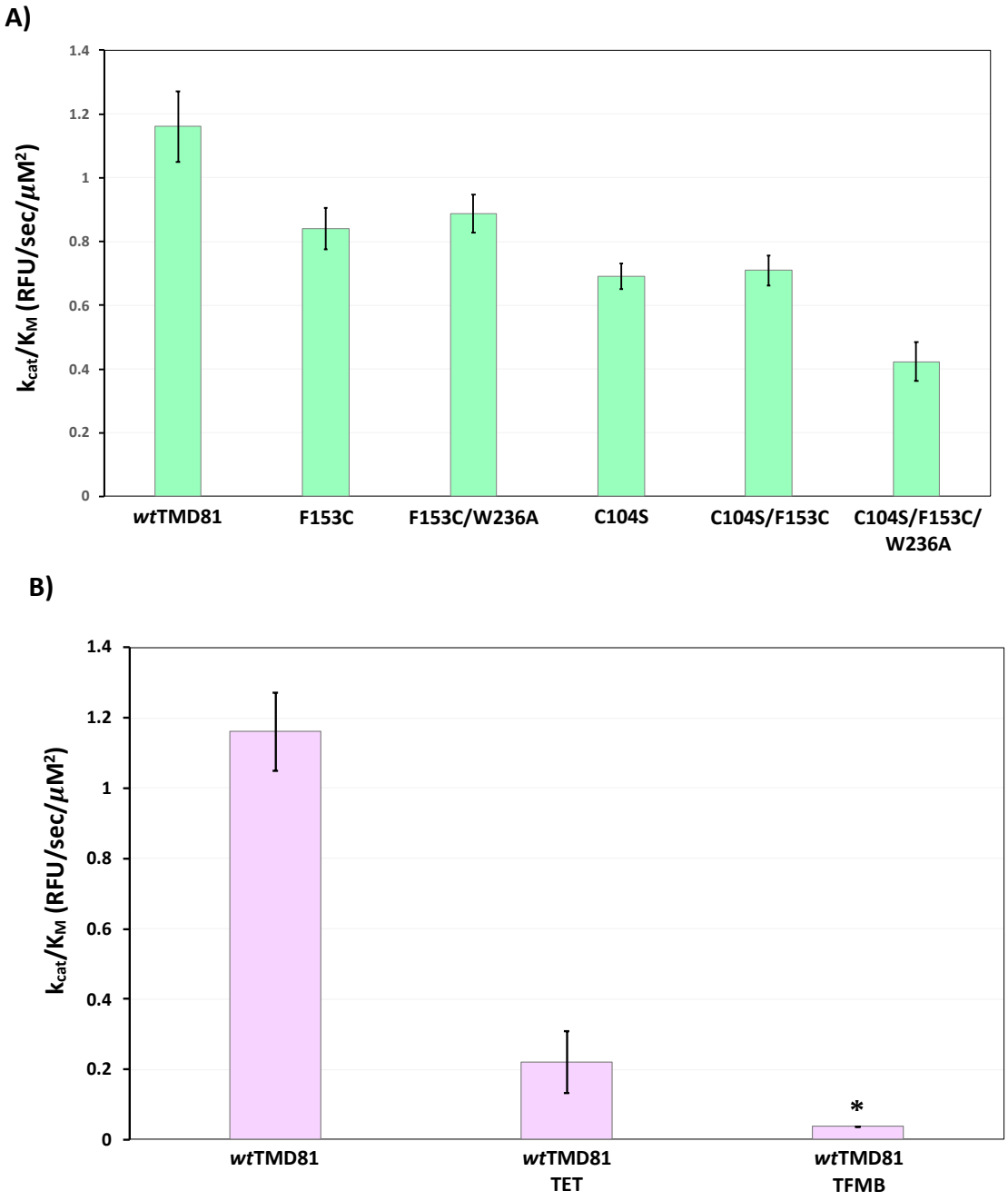


Figure 3.24: **(A)** Effect of various mutations on TMD81 activity monitored using fluorogenic casein as the substrate to obtain the catalytic efficiency k_{cat}/K_M for each indicated mutant. **(B)** Effect of labeling with trifluoromethyl containing probes on the activity of wild-type TMD81. Note that samples with near-zero values indicated with an asterisk (*) have no errors bars showing undetectable activity under the conditions employed (duplicate trials).

4 Discussion

4.1 Evidence for exchange in ^{19}F spectra

Spectra from both TET and TFMB-labelled TMD81 displayed evidence of an exchange process that was on the slow to intermediate NMR timescale. Initially the spectrum of TET-labelled F153C was thought to contain only 2 peaks, with the broader peak corresponding to labeled C104. However, the spectrum of this same mutant that lacked the native cysteine demonstrated that the TET-labeled F153C actually gave rise to two peaks. The peak that had been overlapped by signal from the labeled native cysteine residue was narrow, suggesting free mobility of the probe in this state. The downfield species (peak 1) of TET-labeled C104S/F153C was much broader, but appeared to gain intensity at lower temperatures, while the upfield peak (peak 2) broadened under these conditions. It is likely that the two peaks were undergoing chemical exchange, while it is possible that the mobility of the probe also changes with temperature, which further complicates the analysis. These same two species were observed for the TET-labeled C104S/F153C/W236A mutant that is designed to increase mobility around the gate, with very similar peak widths relative to the double mutant. The only difference was that peak 1 appears to be higher intensity and/or has a more narrow linewidth relative to the spectrum of the double mutant. Temperature dependent loss of intensity in peak 2 relative to peak 1 as the temperature decreased is consistent with two-state exchange between these species. Since van der Waals contacts can significantly affect the shielding of fluorine nuclei[67,108], the more deshielded chemical shift of peak 1 would be consistent with a more buried state of the probe in this state.

Based on the chemical shift difference between exchanging species (~0.5 ppm) in the 500 MHz spectrometer, it is possible to estimate an upper limit on the exchange rate using the Larmour frequency of 470 MHz for the ^{19}F nucleus. This turns out to be ~200 s^{-1}, which suggests lifetimes of each state to be on the millisecond time scale or longer. This is consistent with conformational exchange around the substrate gate, since it has been proposed to be a rate-determining step in the rhomboid catalytic cycle. However, if this is the case it is puzzling

that the impact of the supposed gate-opening mutation W236A on the NMR spectrum was not more significant. This may reflect a perturbation of the dynamics around the gate by the introduction of label at position 153. Due to the relatively small size and flexible nature of the TET probe, it is possible that it was not able to interact with the Trp236 residue in the way that Phe153 does (Fig. 4.0). Interestingly, it was previously found that the point mutation of F153A alone increased *ec*GlpG activity when proteolysis was performed using TM substrates by potentially promoting the α5 open gate conformational state[36,37,51,59]. However, this increase was not observed with the F153C mutation as it displayed equivalent activity to wild-type TMD81, although this was measured with the water-soluble substrate model casein. Likewise, the open-gate double mutant F153A/W236A was previously shown to significantly enhance activity[36,51,54] whereas the F153C/W236A mutant displayed similar activity to the wild-type in our study. These discrepancies may explain why a larger effect on the NMR spectrum was not observed for the W236A substitution mutation.

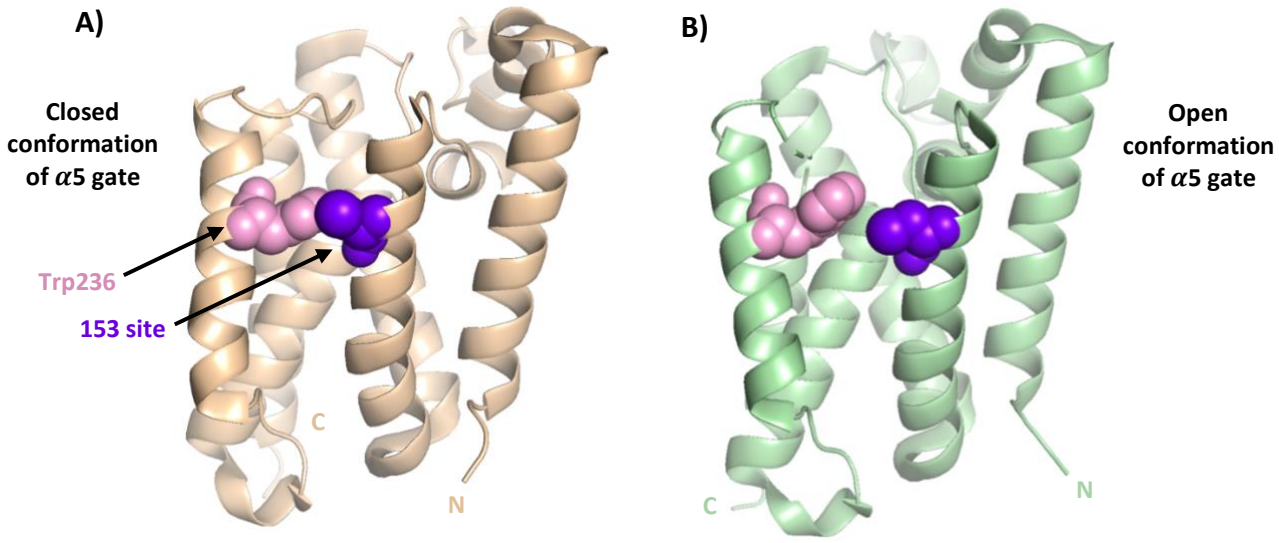


Figure 4.0: Structure of *ec*GlpG (wheat) showing the proximity of the 153 site (purple) to Trp236 (pink) in the **(A)** closed and **(B)** laterally displaced open-gate conformation of the α5 gate.

Although spectra of TFMB labeled TMD81 samples were complicated by labeling of a site other than cysteine residues, these spectra also showed evidence of an exchange process.

TFMB-labeled C104S/F153C showed 1 peak for the Phe153 site at higher temperatures with the emergence of a 2nd peak as temperatures were dropped. Unlike the TET probe, TFMB being a larger probe could have affected the conformational equilibrium we were trying to study and may have retained interactions with the Trp236 side chain. In contrast, when Trp236 was mutated to Ala, the ^{19}F NMR spectra showed 2 peaks for the Phe153-labeled site. Again, temperature dependent changes in this spectrum were difficult to interpret because there was a change in intensity and peakwidth that may reflect a change in populations of the two states and/or a change in probe mobility due to changes in probe interactions. It is tempting to speculate that the more broad, less dynamic peak that appeared downfield in the 25°C spectrum represents the closed state of TMD81 with the aromatic ring of this probe interacting with Trp236 while the sharp, more dynamic upfield peak may represent the open state of the α5 gate, although more experiments would be needed to determine if this is the case. For example, paramagnetic species could be introduced via addition of lipophilic probes containing paramagnetic relaxation enhancement (PRE) agents to selectively broaden peaks of residues/regions that are more solvent exposed[117-119].

Although background labeling of Cys104 did occur in our experiments when this residue was not mutated out, it was possible to compare these spectra to those obtained with knockout mutants to identify signals coming from the F153C site. Although small probes like TET are usually non-perturbing[64,66], it is possible that its relatively small size may reduce its ability to detect smaller-scale conformational changes in proteins[120]. Additionally, while the larger TFMB probe may have the potential to restore lost interactions between residues upon mutation in TMD81, we were surprised to see that the cysteine knockout mutant, C104S was still labeled by TFMB in spite of the absence of cysteine residues. Although the origins of this background signal are not known, it is possible that the TFMB label reacted with a nucleophilic residue in TMD81 that is solvent exposed. One possibility is that the active site Ser201 reacts with this probe since this residue is activated by its interactions with His254 to make it more nucleophilic. This is supported by the functional data obtained in this thesis using fluorogenic casein since TFMB-labeled wild-type TMD81 essentially appeared to be inactive. In contrast,

TET labeled wild-type TMD81 showed reduced, but significant levels of activity. Although we did try to work around this background reactivity, the non-functional state of TFMB-labeled TMD81 significantly reduces its utility for the study of GlpG dynamics.

4.2 Mutation of Cys104 affects *ec*GlpG structure and function

Somewhat unexpectedly, the conservative mutation C104S was shown to have a small but significant impact on the functional and structural integrity of TMD81. For example, the 2D ^{1}H-^{15}N HSQC of this mutant indicated that the mutation may destabilize the protein structure as there was some peak broadening throughout the spectrum. These results are in agreement with the slight decrease in secondary structure and thermal stability observed in CD spectra and thermal denaturation studies respectively, while functional analysis also showed a decrease in activity relative to wild-type TMD81. This was also seen for the cysteine mutants which contained the Cys104 knockout as they also displayed a small decrease in activity when compared to the same mutants that contained the native C104.

Examination of the literature for similar mutants showed some similarities to our results with the cysteine knockout mutant. In particular, the C104A mutant showed a decrease to GlpG T_M by 3.8°C[54], mirroring results obtained in this thesis. However, a subsequent study on a C104V mutant showed similar unfolding transitions to the wild type protein as monitored by Trp fluorescence, and comparable folding rates as determined by stopped-flow kinetic measurements[121]. These results suggest that a valine substitution may have been less disruptive on TMD81 structure and function given that this non-polar residue could be considered to be more isosteric to cysteine than serine. In general, mutation of buried cysteine residues has a greater tendency to disrupt the folding and stability of proteins relative to other residues[122-124].

4.3 Other labeling strategies and towards solid-state NMR of TMD81

The results obtained in this thesis demonstrate that slow-exchanging states consistent with an open-closed system potentially correlated to α5 gating dynamics have potential to be monitored by ^{19}F NMR spectra using trifluoromethyl tags. However, the difficulties associated with structural perturbations introduced by these tags, and the potential of these tags to alter the dynamic processes under study suggest that it may be more appropriate to use other types of labels that do not require the incorporation of non-native labels. For example, selective labeling of TDM81 samples using ^{13}C-methyl bearing enriched precursors for specific amino acids can provide sensitive probes to monitor dynamics without the need for covalent labels. A simple type of ^{13}C label that has been used in other systems is ^{13}C-methyl-labeled methionine residues[125]. In GlpG the methionine residues Met208 from α4 and Met247 from α6 lie just above and below Trp236 which could make them effective reporters on α5 gating dynamics (Fig. 4.1).

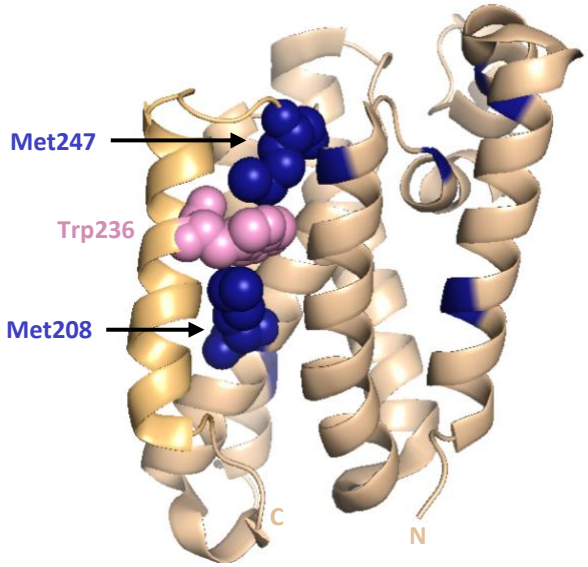


Figure 4.1: Structure of *ec*GlpG (wheat) showing the position of the Trp236 (pink) relative to Met208 and Met247 (dark blue).

Additionally, recent developments in solid-state NMR (ssNMR) have expanded the range of proteins that can be studied at high resolution[126]. During the writing of this thesis, a ssNMR

study on *ec*GlpG reconstituted into liposomes was published where triple resonance experiments allowed chemical shift assignment of backbone atoms and relaxation measurements to be performed[53]. Alanine mutations designed to reduce interactions in the $\alpha 2/\alpha 5$ interface showed increased conformational heterogeneity in spectra, both in the mutated regions and the entire protein while cross-linked cysteine residues using a maleimide-maleimide (M2M) crosslink showed slightly lower R_1 relaxation rates when compared to wild-type GlpG. This study also acquired ^{15}N relaxation dispersion data which revealed the presence of a major and a minor state of the $\alpha 5$ helix that undergoes conformational exchange on a time scale of ~40 μs.[126] These results suggest that dynamic measurements of $\alpha 5$ gating are possible on the slow exchange timescale as was postulated in this thesis.

5 Conclusion and outlook

In this thesis, we attempted to monitor the structure and dynamics of the *ec*GlpG TMD, specifically in the of $\alpha 2/\alpha 5$ gating interface using solution ^{19}F NMR. Site-specific labelling of TMD81 with cysteine reactive trifluoromethyl probes gave rise to spectra which facilitated assignments. While the C104S knockout mutation was necessary for straightforward peak assignment at a single site in the gating region, its removal unexpectedly imposed structural and functional restrains on the overall integrity of TMD81. A temperature-dependent change in the population between 2 states – one postulated to be the open state of the $\alpha 5$ gate and the other predicted to be closed state was observed. This change in population may be consistent with slow exchange (on the millisecond timescale) between open and closed states where the open state is observed at higher temperatures, and the closed-gate state favoured at lower temperatures. However, little difference was observed upon introduction of the W236A mutant designed to open the gate, which may reflect perturbations introduced by the mutation of F153 to Cys and its labelling, a hypothesis supported by functional studies showing decreased activity. For these reasons, in the future alternate labeling strategies like ^{13}C-methyl labeling of methionine residues may be more appropriate to report on slow exchange in the $\alpha 5$ gating region.

6 References

1. Deamer, D. W. & Bramhall, J. Permeability of lipid bilayers to water and ionic solutes. *Chemistry and Physics of Lipids* **40**(2-4), 167–188 (1986).
2. Zhou, Y., Moin, M. S., Urban, S., & Zhang, Y. An internal water-retention site in the rhomboid intramembrane protease GlpG ensures catalytic efficiency. *Structure* **20**(7), 1255-1263 (2012).
3. Wang, Y., Zhang, Y. & Ha, Y. Crystal structure of a rhomboid family intramembrane protease. *Nature* **444**(7116), 179–183 (2006).
4. Wolfe, M. S. Intramembrane-cleaving Proteases, *Journal of Biological Chemistry* **284**(21), 13969-13973 (2009).
5. Urban, S. & Freeman. M., Intramembrane proteolysis controls diverse signalling pathways throughout evolution. *Current Opinion in Genetics & Development* **12**(5), 512-518 (2002).
6. Moin, S. M. & Urban, S. Membrane immersion allows rhomboid proteases to achieve specificity by reading transmembrane segment dynamics. *eLife* **1**, 1–16 (2012).
7. Beard, A. B., et al. Discovery of cellular roles in intramembrane proteases. *ACS Chemical Biology* **14**(11), 2372-2388 (2019).
8. Kühnle, N., Dederer, V., & Lemberg, M. K. Intramembrane proteolysis at a glance: from signalling to protein degradation. *Journal of Cell Science* **132**(16), (2019).
9. Auld, S. D. Zinc catalysis in metalloproteases. *Structure and Bonding* **89**, 29-50 (1997).
10. Polgár, L. The mechanism of action of aspartic proteases involves 'push-pull' catalysis. *FEBS Letters*, **219**(1), 1–4 (1987).
11. Hedstrom, L. Serine protease mechanism and specificity. *Chemical Reviews* **102**(12), 4501-4524 (2002).
12. Fujinaga, M., *et al*. The molecular structure and catalytic mechanism of a novel carboxyl peptidase from *Scytalidium lignicolum*. *Proceedings of the National Academy of Sciences of the United States of America* **101**(10), 3364-3369 (2004).[3]
13. Wolf, V. E. & Verhelst, L, H. S. Inhibitors of rhomboid proteases. *Biochimie* **122**, 38-47 (2016).
14. 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**(3) (2003).
15. Dogga, K. S. & Soldati-Favre, D. Biology of rhomboid proteases in infectious diseases. *Seminars in Cell & Developmental Biology* **60**, 38-45 (2016).
16. Urban, S., Lee, J. R. & Freeman, M. Drosophila rhomboid-1 defines a family of putative intramembrane serine proteases. *Cell* **107**(2), 173-182 (2001).
17. 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**(3), 1003-1008 (2007).
18. Ochsner, U. A., *et al.* 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**(12), 8312-8317 (2002).

19. Bolhuis, A *et al.* Subunit interactions in the twin-arginine translocase complex of Escherichia coli. *FEBS Letters* **472**(1), 88-92 (2000).
20. 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**(30), 10482-10486 (2005).
21. Fleig, L. *et al.* Ubiquitin-dependent intramembrane rhomboid protease promotes ERAD of membrane proteins. Molecular Cell **47**(4), 558-569 (2012).
22. 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**(10), 979-981 (2012).
23. Paschkowsky, S. *et al.* Alternative processing of the amyloid precursor protein family by rhomboid protease RHBDL4. *The Journal of Biological Chemistry* **291**(42), 21903-21912 (2016).
24. 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**(11), 4146–4151 (2005).
25. Conseil, V., Soête, M. & Dubremetz, J. F. Serine protease inhibitors block invasion of host cells by toxoplasma gondii. *Antimicrobial Agents and Chemotherapy* **43**(6), 1358–1361 (1999).
26. 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**(10), 0922–0932 (2006).
27. Zeng G. *et al*. Repressor for the sn-glycerol 2-phosphate regulon of Escherichia coli K-12: primary structure and identification of the DNA-binding domain. *American Society for Microbiology* **178**(24)**,** 7080–7089 (1996).
28. Maegawa, S., Ito, K. & Akiyama, Y. Proteolytic action of GlpG, a rhomboid protease in the Escherichia coli cytoplasmic membrane. *Biochemistry* **44**(41)**,** 13543– 13552 (2005).
29. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology* **2,** 2006.0008 (2006).
30. 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**(6), 1–15 (2017).
31. Xue, Y. & Ha, Y. Catalytic mechanism of rhomboid protease GlpG probed by 3,4-dichloroisocoumarin and diisopropyl fluorophosphonate. *Journal of Biological Chemistry* **287**(5), 3099-3107 (2012).
32. Vinothkumar, K. R. *et al*. Structure of rhomboid protease in complex with β-lactam inhibitors defines the S2' cavity. *Structure* **21**(6), 1051-1058 (2013).
33. 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**(3), 329-340 (2016).
34. Vinothkumar, K. R. *et al*. The structural basis for catalysis and substrate specificity of a rhomboid protease. *The EMBO Journal* **29**(22), 3797-3809 (2012).

35. 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**(12), 1084-1091 (2006).
36. 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).
37. Dickey, S. W. et al. Proteolysis inside the membrane is a rate-governed reaction not driven by substrate affinity. *Cell* **155**(6), 1270-1281 (2013).
38. Xue, Y. & Ha, Y. Catalytic mechanism of rhomboid protease GlpG probed by 3,4-dichloroisocoumarin and diisopropyl fluorophosphonate. *Journal of Biological Chemistry* **287**(5), 3099-3107 (2012).
39. Sherratt, A. R. *et al.* Activity-based protein profiling of the Escherichia coli GlpG rhomboid protein delineates the catalytic core. *Biochemistry* **51**(39), 7794-803 (2012).
40. Ghasriani, H., Kwok, J. K., Sherratt, A. R., Foo, A. C., Qureshi, T., & Goto, N. K. Micelle-catalyzed domain swapping in the GlpG rhomboid protease cytoplasmic domain. *Biochemistry* **53**(37), 5907–5915 (2014).
41. Foo, A.C. *New Insights into the Role of Membrane Interactions and Conformational Dynamics in Intramembrane Proteolysis by GlpG Rhomboid* [Doctoral thesis, University of Ottawa] (2017).
42. Zeng G. *et al*. Repressor for the sn-glycerol 2-phosphate regulon of Escherichia coli K-12: primary structure and identification of the DNA-binding domain. *American Society for Microbiology* **178**(24)**,** 7080–7089 (1996).
43. Baker, R. P. & Urban, S. Cytosolic extensions directly regulate a rhomboid protease by modulating substrate gating. *Nature* **523**, 101–105 (2015).
44. Lohi, O., Urban, S. & Freeman, M. Diverse substrate recognition mechanisms for rhomboids; thrombomodulin is cleaved by mammalian rhomboids. *Current Biology* **14**(3), 236-241 (2004).
45. Urban, S., Schlieper, D. & Freeman, M. Conservation of intramembrane proteolytic activity and substrate specificity in prokaryotic and eukaryotic rhomboids. *Current Biology* **12**(17), 1507-1512 (2002).
46. Vinothkumar, K. R. & Freeman, M. Intramembrane proteolysis by rhomboids: catalytic mechanisms and regulatory principles. *Current Opinion in Structural Biology* **23**(6), 851–858 (2013).
47. Strisovsky, K., Sharpe, H. J. & Freeman, M. Sequence-specific intramembrane proteolysis: identification of a recognition motif in rhomboid substrates. *Molecular Cell* **36**(6) 1048-1059 (2009).
48. 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**(10), 910–918 (2019).
49. Akiyama, Y., & Maegawa, S. Sequence features of substrates required for cleavage by GlpG, an Escherichia coli rhomboid protease. *Molecular Microbiology* **64**(4), 1028–1037 (2007).

50. Brooks, C. L., & Lemieux, M. J. Untangling structure-function relationships in the rhomboid family of intramembrane protease. *Biochimica et Biophysica acta* **1828**(12), 2862-28872 (2013).
51. Baker, R. P. *et al.* 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**(20), 8257-8262 (2007).
52. 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).
53. Bohg, C. *et al.* The opening dynamics of the lateral gate regulates the activity of rhomboid proteases. *Science Advances* **9**(29), (2023).
54. Baker, R. P. & Urban, S. Architectural and thermodynamic principles underlying intramembrane protease function. *Nature Chemical Biology* **8**(9), 759-768 (2012).
55. Dickey, S. W. et al. Proteolysis inside the membrane is a rate-governed reaction not driven by substrate affinity. *Cell* **155**(6), 1270-1281 (2013).
56. Briggs, G. E. & Haldane J. B. S. H. A note on the kinetics of enzyme action. *Biochemical Journal* **19**(2), 338-339 (1925).
57. Rami Tzafriri, A. & Edelman, E. R. Quasi-steady-state kinetics at enzyme and substrate concentrations in excess of the Michaelis-Menten constant. *Journal of Theoretical Biology* **245**(4), 737-748 (2007).
58. International Union of Biochemistry and Molecular Biology. Symbolism and terminology in enzyme kinetics. *Journal of Biochemistry* **128**(2-3), 281-291 (1982).
59. Urban, S. & Wolfe, M. S. Reconstitution of intramembrane proteolysis in vitro reveals that pure rhomboid is sufficient for catalysis and specificity. *Proceedings of the National Academy of Sciences of the United States of America* **102**(6), 1883-1888 (2005).
60. Maegawa, A., Ito, K. & Akiyama, Y. Proteolytic action of GlpG, a rhomboid protease in the Escherichia coli cytoplasmic membrane. *Biochemistry* **44**(41), 13543-13552 (2005).
61. 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**(7)**,** 1127–42 (2013).
62. Foo, A. C. Y., Harvey, B. G. R., Metz, J. J. & Goto, N. K. Influence of hydrophobic mismatch on the catalytic activity of *E. coli* GlpG rhomboid protease. *Protein Science* **24**(4), 464–473 (2015).
63. Tognarelli, J. M. et al. Magnetic Resonance Spectroscopy: Principles and Techniques: Lessons for Clinicians. *Journal of Clinical and Experimental Hepatology*, **5**(4), 320-328 (2015).
64. Diaz-Parga, P. & de Alba, E. Chapter fourteen – protein interactions of the inflammasome adapter ASC by solution NMR in Methods in enzymology. *Academic Press* **625**, 223-252 (2019).
65. Overbeck, J. H., Kremer, W., & Sprangers, R. A. suite of ^{19}F based relaxation dispersion experiments to assess biomolecular motions. *Journal of Biomolecular NMR* **74**(12), 753-766 (2020).

66. Didenko, T. *et al*. Fluorine-19 NMR of integral membrane proteins illustrated with studies of GPCRs. *Current Opinion in Structural Biology* **23**(5), 740–747 (2013).
67. Danielson, M. A. & Falke, J. J. Use of ^{19}F NMR to probe protein structure and conformational changes. *Annual Review of Biophysics and Biomolecular Structure* **25**, 163– 195 (1996).
68. Kitevski-LeBlanc, J. L. & Prosser, R. S. Current applications of ^{19}F NMR to studies of protein structure and dynamics. *Progress in Nuclear Magnetic Resonance Spectroscopy* **62**, 1–33 (2012).
69. 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**(25), 9465–9474 (2013).
70. Manglik, A. *et al.* Structural insights into the dynamic process of β2-adrenergic receptor signaling. *Cell* **162**(6), 1101–1111 (2015).
71. Picard, LP. & Prosser, S. R. Advances in the study of GPCRs by ^{19}F NMR. *Current opinion in Structural Biology.* **69**, 169-176 (2021).
72. Ye, L. *et al*. 19F NMR: A promising tool for dynamic conformational studies of G protein-coupled receptors. *Structure* **30**(10), 1372-1384 (2022).
73. Liu, J. J., et al. Biased signaling pathways in β2-adrenergic receptor characterized by ^{19}F-NMR. *Science* **335**(6072), 1106-1110 (2012).
74. Horst, R., Liu, J. J., Stevens, R. C. & Wüthrich, K. β2-adrenergic receptor activation by agonists studied with ^{19}F NMR spectroscopy. *Angewandte Chemie - International Edition* **52**(41), 10762–10765 (2013).
75. Loewen, M. C. *et al.* Solution ^{19}F 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**(9), 4888–4892 (2001).
76. Ye, L. et al. A comparison of chemical shift sensitivity of trifluoromethyl tags: optimizing resolution in ^{19}F NMR studies of proteins. *Journal of Biomolecular NMR* **62**(1), 97-103 (2015).
77. Somlyay, M., Ledolter, K., Kitzler, M., Sandford, G., Cobb, S. L., & Konrat, R. ^{19} F NMR spectroscopy tagging and paramagnetic relaxation enhancement-based conformation analysis of intrinsically disordered protein complexes. *Chembiochem: a European Journal of Chemical Biology*, **21(5)**, 696–701 (2020).
78. Gimenez, D. et al. ^{19}F NMR as a tool in chemical biology. *Beilstein Journal of Organic Chemistry* **17**, 293-318 (2021).
79. Xin, L. *et al*. Paramagnetic perturbation of the ^{19}F NMR chemical shift in fluorinated cysteine by O_2: a theoretical study. *Journal of Physical Chemistry B* **113**(31), 10916-10922 (2009).
80. Kleckner, I. R. & Foster, M. P. An introduction to NMR-based approaches for measuring protein dynamics. *Biochimica et Biophysica acta - Proteins and Proteomics* **1814**(8), 942–968 (2011).
81. Frieden, C., Hoeltzli, S. D. & Bann, J. G. The preparation of ^{19}F-labeled proteins for NMR studies. *Methods in Enzymology* **380**, 400–415 (2004).

82. 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**(1-2), 34–36 (1990).
83. Crowley, P. B., Kyne, C. & Monteith, W. B. Simple and inexpensive incorporation of ^{19}F-tryptophan for protein NMR spectroscopy. *Chemical Communications* **48**, 10681–10683 (2012).
84. Jackson, J. C., Hammill, J. T. & Mehl, R. A. Site-specific incorporation of a ^{19}F-amino acid into proteins as an NMR probe for characterizing protein structure and reactivity. *Journal of the American Chemical Society* **129**(5), 1160–1166 (2007).
85. Furter, R. Expansion of the genetic code: site-directed p-fluoro-phenylalanine incorporation in Escherichia coli. *Protein Science* **7(2)**, 419–426 (1998).
86. Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the genetic code of Escherichia coli. *Science* **292**(5516), 498–500 (2001).
87. Xie, J. *et al.* The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. *Nature Biotechnology* **22**(10), 1297–1301 (2004).
88. Xie, J. & Schultz, P. G. An expanding genetic code. *Methods* **36**(3), 227–238 (2005).
89. Bak, W. D. et al. Cysteine reactivity across the subcellular universe. *Current Opinion in Chemical Biology* **48**, 96-105 (2019).
90. Luchette, P. A., Prosser, R. S. & Sanders, C. R. Oxygen as a paramagnetic probe of membrane protein structure by cysteine mutagenesis and ^{19}F NMR spectroscopy. *Journal of the American Chemical Society* **124**(8), 1778–1781 (2002).
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 ^{19}F NMR. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 13744–13749 (1999).
92. E. Y. Lau and J. T. Gerig. Origins of Fluorine NMR Chemical Shifts in Fluorine-Containing Proteins. *Journal of the American Chemical Society* **122**(18) 4408-4417 (2000).
93. Ratkeviciute, G., Cooper, F. B. & Knowles, J. T. Methods for the solubilisation of membrane proteins: the micelle-aneous world of membrane protein solubilisation. *Biochemical Society Transactions* **49**(4), 1763–1777 (2021).
94. 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**(12), 3098–3106 (2007).
95. Fernández, C. & Wüthrich, K. NMR solution structure determination of membrane proteins reconstituted in detergent micelles. *FEBS Letters* **555**(1), 144–150 (2003).
96. Chill, J. H., Louis, J. M., Miller, C., & Bax, A. NMR study of the tetrameric KcsA potassium channel in detergent micelles. *Protein science: a Publication of the Protein Society* **15**(4), 684–698 (2006).
97. Vinogradova, O., Sönnichsen, F. & Sanders, C.R. On choosing a detergent for solution NMR studies of membrane proteins. *Journal of Biomolecular NMR* **11**(4), 381–386 (1998).

98. Ng, E. Y. *et al.* Expression, purification of zika virus membrane protein-NS2B in detergent micelles for NMR studies. *Protein Expression and Purification* **154**, 1–6 (2019).
99. 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**(44), 15914–15931 (2019).
100. 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**(1-2), 275–288 (2004).
101. Garavito, R. M. & Ferguson-Miller, S. Detergents as tools in membrane biochemistry. *Journal of Biological Chemistry* **276**(35), 32403–32406 (2001).
102. Morrison, E. A. & Henzler-Wildman, K. A. Reconstitution of integral membrane proteins into isotropic bicelles with improved sample stability and expanded lipid composition profile. *Biochimica et Biophysica acta* **1818**(3)**,** 814–820 (2012).
103. Cross, T. A., Murray, D. T., & Watts, A. Helical membrane protein conformations and their environment. *European Biophysics Journal: EBJ* **42**(10), 731–755 (2013).
104. Caldwell, T. A. *et al*. Low- q bicelles Are mixed micelles. *The Journal of Physical Chemistry Letters* **9**(15), 4469–4473 (2018).
105. Dürr, U. H., Soong, R., & Ramamoorthy, A. When detergent meets bilayer: birth and coming of age of lipid bicelles. *Progress in Nuclear Magnetic Resonance Spectroscopy*, **69** 1–22 (2013).
106. Warschawski, D. E. *et al.* Choosing membrane mimetics for NMR structural studies of transmembrane proteins. *Biochimica et Biophysica acta* **1808**(8)**,** 1957–1974 (2011).
107. Van Ingen, H., & Bonvin, A. M. Information-driven modeling of large macromolecular assemblies using NMR data. *Journal of Magnetic Resonance* **241**, 103–114 (2014).
108. Mineev, K. & Nadezhdin, K. Membrane mimetics for solution NMR studies of membrane proteins. *Nanotechnology Reviews*, **6**(1), 15-32 (2017).
109. Hassan, I.A. *Development of ^{19}F NMR methods for the GlpG rhomboid protease in detergents and lipid nanoparticle systems* [Masters thesis, University of Ottawa] (2021).
110. Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., & Pease, L. R. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* **77**(1), 51–59 (1989).
111. Quan, J., & Tian, J. Circular polymerase extension cloning of complex gene libraries and pathways. *PloS One* **4**(7), e6441 (2009).
112. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., & Bax, A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *Journal of Biomolecular NMR* **6**(3), 277–293 (1995).
113. Johnson, B. A., & Blevins, R. A. NMR View: A computer program for the visualization and analysis of NMR data. *Journal of Biomolecular NMR*, **4**(5), 603–614 (1994).
114. Sugiki, T., Kobayashi, N., & Fujiwara, T. Modern technologies of solution nuclear magnetic resonance spectroscopy for three-dimensional structure determination of proteins open avenues for life scientists. *Computational and Structural Biotechnology Journal* **15**, 328–339 (2017).

115. Clarke, J., & Fersht, A. R. Engineered disulfide bonds as probes of the folding pathway of barnase: increasing the stability of proteins against the rate of denaturation. *Biochemistry* **32**(16), 4322–4329 (1993).
116. Santoro, M. M., & Bolen, D. W. Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry* **27**(21), 8063-8068 (1988).
117. 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**(13), 4389–4397 (2006).
118. Iwahara, J. & Clore, G. M. Detecting transient intermediates in macromolecular binding by paramagnetic NMR. *Nature* **440**(7088)**,** 1227–1230 (2006).
119. MacRaild, C. A., Howlett, G. J. & Gooley, P. R. The structure and interactions of human apolipoprotein C-II in dodecyl phosphocholine. *Biochemistry* **43**(25), 8084–8093 (2004).
120. Marsh, E. N., & Suzuki, Y. Using ^{19}F NMR to probe biological interactions of proteins and peptides. *ACS Chemical Biology* **9**(6), 1242–1250 (2014).
121. Otzen, D. E. *et al*. Cys-labeling kinetics of membrane protein GlpG: a role for specific SDS binding and micelle changes?. *Biophysical Journal **120***(18), 4115–4128 (2012).
122. Marintchev, A., Frueh, D., & Wagner, G. NMR methods for studying protein-protein interactions involved in translation initiation. *Methods in Enzymology*, **430**, 283–331 (2007).
123. Xia, X., Longo, L. M., & Blaber, M. Mutation choice to eliminate buried free cysteines in protein therapeutics. *Journal of Pharmaceutical Sciences* **104**(2), 566–576 (2015).
124. Bhopatkar, A. A., Uversky, V. N., & Rangachari, V. Disorder and cysteines in proteins: a design for orchestration of conformational see-saw and modulatory functions. *Progress in Molecular Biology and Translational Science* **174**, 331–373 (2020).
125. Fischer, M. *et al.* Synthesis of a ^{13}C-methyl-group-labeled methionine precursor as a useful tool for simplifying protein structural analysis by NMR spectroscopy. *ChemBioChem* **8**(6)**,** 610–612 (2007).
126. Bohg, C. *et al*. A combination of solid-state NMR and MD simulations reveals the binding mode of a rhomboid protease inhibitor. *Chemical Science*, **12**(38), 12754-12762 (2021).

7 Appendix

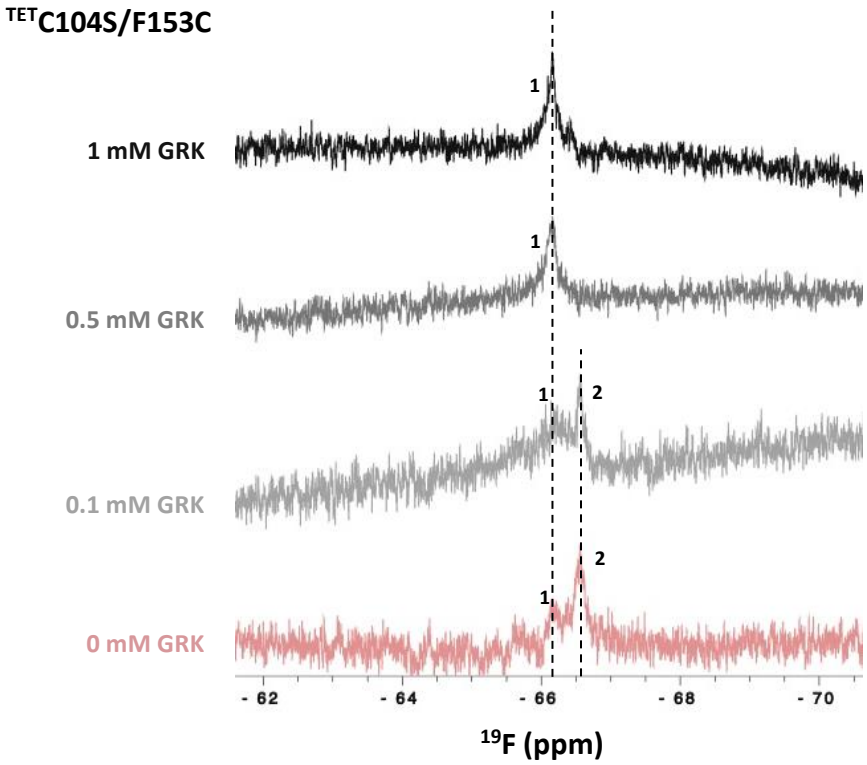


Figure 7.0: **Preliminary studies on the effect of peptide binding on TMD81.** ^{19}F NMR spectra of 40μM of TET-labeled C104S/F153C in the presence of Gurken-derived peptide fragment collected at 25°C. As the peptide was titrated into solution, the intensity of peak 2 disappears completely while peak 1 begins to initially broaden and then intensify drastically.

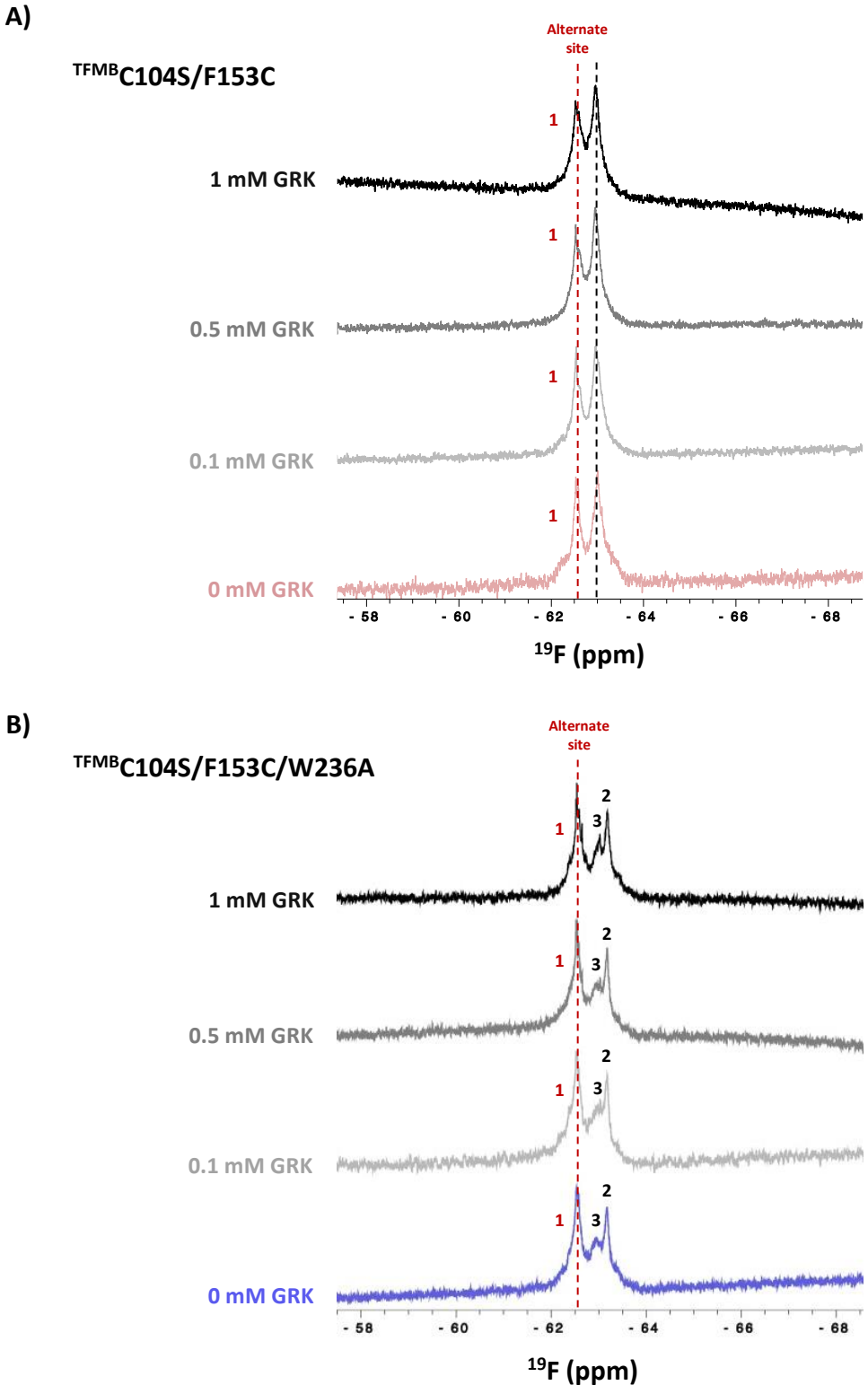
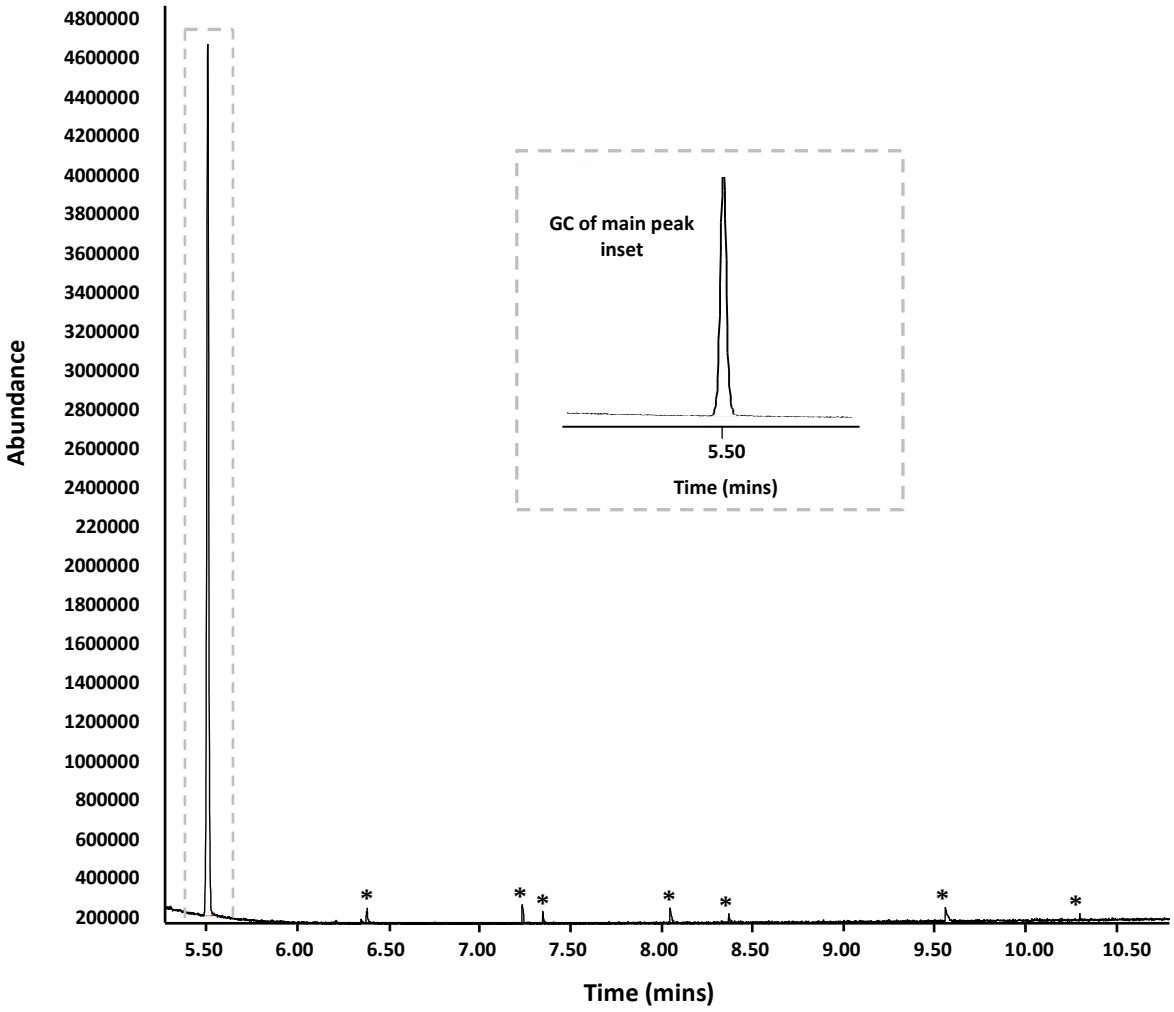


Figure 7.1: **Preliminary studies on the effect of peptide binding on TMD81**. **(A)** ^{19}F NMR spectra of 100μM C104S/F153C labelled with TFMB in the presence of Gurken-derived peptide

fragment collect at 25°C. **(B)** ^{19}F NMR spectra of 90μM C104S/F153C/W236A labelled with TFMB in the presence of Gurken-derived peptide fragment collect at 25°C. No significant changes in peak intensities are observed for either spectrum likely due to sample inactivity caused by labeling with TFMB.

A)



B)

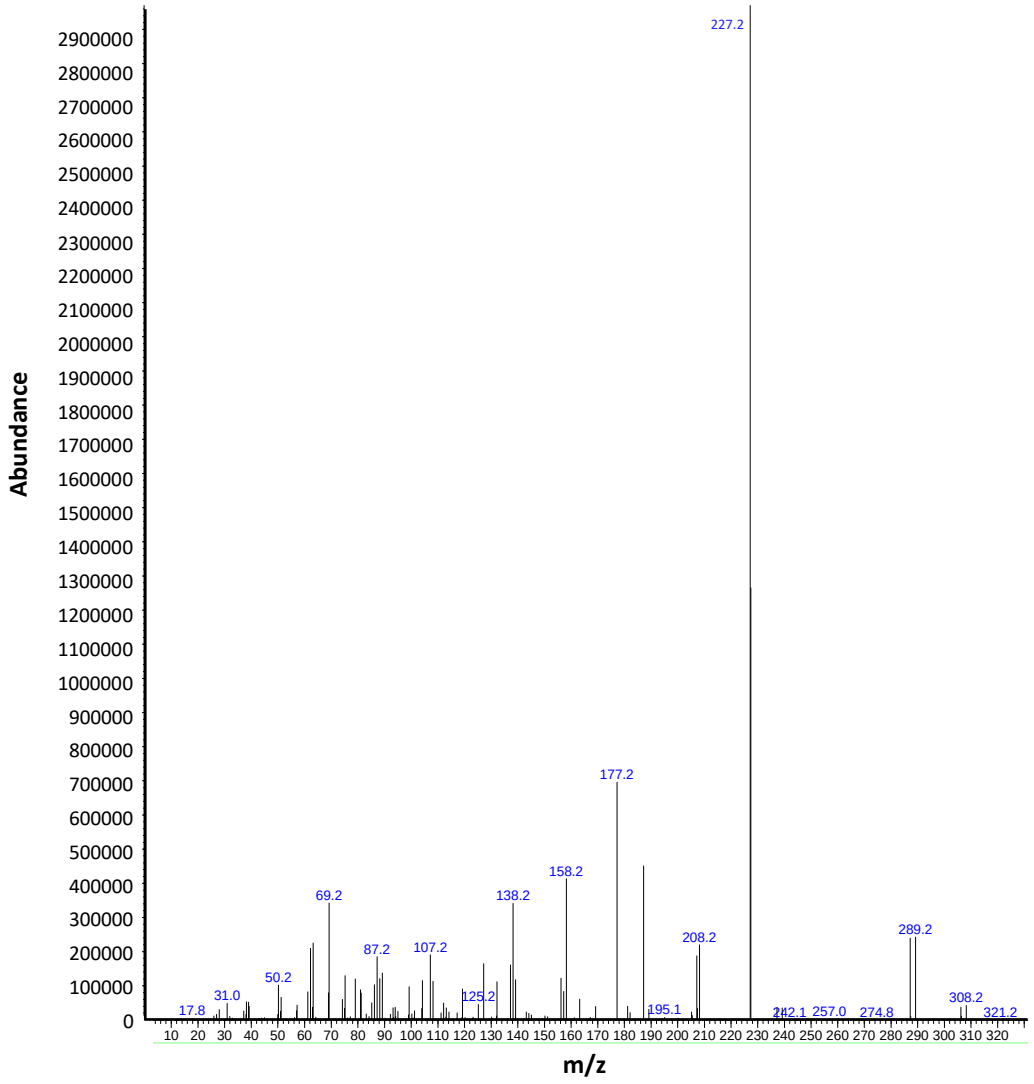


Figure 7.2: GC:MS findings for the TFMB label (A) Gas chromatograph of TFMB showing the main peak in the inset from time 5.45-5.55 minutes. A dilution of 1μL/mL of TFMB in ethyl acetate was used for the run. Note that impurity peaks are marked with as asterisk (*). **(B)** Mass spectrum of the main peak showing an M+ peak at 308.2 which corresponds with the molecular weight of TFMB at 307.02.

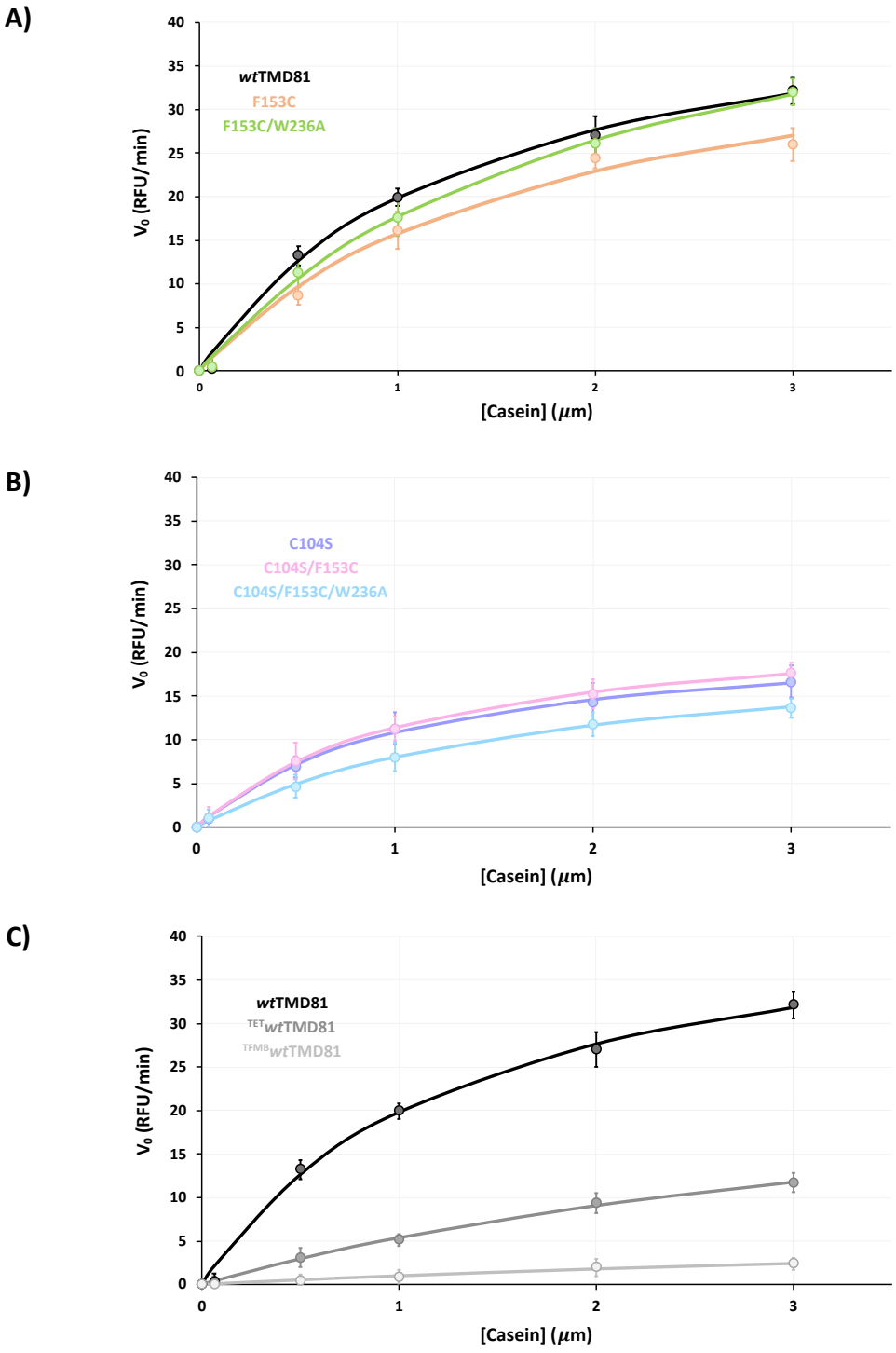


Figure 7.3: Michaelis-Menten plots showing initial rates (v_0) vs. casein substrate concentration. (A) Plots for wild-type TMD81 (black), F153C (beige) and F153C/W236A (green) producing K_M values of 1.31±0.1, 1.68±0.1, and 1.99±0.2μM as well as k_{cat} values of

1.52±0.14, 1.42±0.2, and 1.77±0.13 RFU/seconds/μM respectively. **(B)** Plots for C104S (purple), C104S/F153C (pink) and C104S/F153C/W236A (blue) producing K_M values of 1.07±0.2, 1.12±0.13, and 1.7±0.14μM as well as k_{cat} values of 0.74±0.1, 0.8±0.2, and 0.72±0.15 RFU/seconds/μM respectively. **(C)** Plots for unlabeled wild-type TMD81 (black), TET-labeled wild-type TMD81 (grey) and TFMB-labeled wild-type TMD81 (light grey) producing K_M values 1.31±0.1, 4.43±0.22, and 7.37±0.25 μM as well as k_{cat} values of 1.52±0.14, 0.98±0.1, and 0.28±0.01 RFU/seconds/μM respectively.