Beyond the Active Site of the Bacterial Rhomboid Protease: Novel Interactions at the Membrane to Modulate Function.

Allison Rae Sherratt

Thesis submitted to the
Faculty of Graduate and Postdoctoral Studies
In partial fulfillment of the requirements for the
PhD degree in Biochemistry

Department of Biochemistry, Microbiology & Immunology
Faculty of Medicine
University of Ottawa

Abstract

Rhomboids are unique membrane proteins that use a serine protease hydrolysis mechanism to cleave a transmembrane substrate within the lipid bilayer. This remarkable proteolytic activity is achieved by a core domain comprised of 6 transmembrane segments that form a hydrophilic cavity submerged in the membrane. In addition to this core domain, many rhomboids also possess aqueous domains of varying sizes at the N- and/or C-terminus, the sequences of which tend to be rhomboid-type specific. The functional role of these extramembranous domains is generally not well understood, although it is thought that they may be involved in regulation of rhomboid activity and specificity. While extramembranous domains may be important for rhomboid activity, they are absent in all x-ray crystal structures available. For this reason, we have focused on uncovering the structural and functional relationship between the rhomboid cytoplasmic domain and its catalytic transmembrane core.

To investigate the structure and function of the bacterial rhomboid cytoplasmic domain, full-length rhomboids from *Escherichia coli* and *Pseudomonas aeruginosa* were studied using solution nuclear magnetic resonance (NMR) spectroscopy, mutation and activity assays. The *P. aeruginosa* rhomboid was purified in a range of membrane-mimetic media, evaluated for its functional status *in vitro* and investigated for its NMR spectroscopic properties. Results from this study suggested that an activity-modulating interaction might occur between the catalytic core transmembrane domain and the cytoplasmic domain. Further investigation of this hypothesis with the *E. coli* rhomboid revealed that protease activity relies on a short but critical sequence N-terminal to the first transmembrane

segment. This sequence was found to have a direct impact on the rhomboid active site, and should be included in future structural studies of this catalytic domain.

The structure of the cytoplasmic domain from the *E. coli* rhomboid was also determined by solution NMR. We found that it forms slowly-exchanging dimers through an exchange of secondary structure elements between subunits, commonly known as three-dimensional domain swapping. Beyond this rare example of domain swapping in a membrane protein extramembranous domain, we found that the rate of exchange between monomeric and dimeric states could be accelerated by transient interactions with large detergent micelles with a phosphocholine headgroup, but not by exposure to other weakly denaturing conditions. This novel example of micelle-catalyzed domain swapping interactions raises the possibility that domain swapping interactions might be induced by similar interactions *in vivo*. Overall, the results of this thesis have identified detergent conditions that preserve the highest level of activity for bacterial rhomboids, defined the minimal functional unit beyond what had been identified in available x-ray crystal structures, and characterized a novel micelle-catalyzed domain-swapping interaction by the cytoplasmic domain.

Acknowledgements

First and foremost I would like to thank Dr. Natalie Goto, my supervisor, for accepting me into her lab and providing years of mentorship and guidance in my research. I knew from our first meeting that she would be a great match not only on a professional level, but on a personal level as well – it is rare to find a professor with a great sense of humor and a soft spot for SpongeBob SquarePants! Beyond the thought, support, patience and encouragement that she has provided me with respect to my project, I can not thank Natalie enough for all the time she has spent preparing me for presentations and providing feedback on written documents (especially this thesis). 'Da Boss' truly cares about the success of all her students, and I feel honored to be her first PhD graduate.

I would also like to express my gratitude for my thesis advisory committee members and colleagues that have helped me in this project. Dr. John Baenziger and Dr. Ajoy Basak posed excellent questions and suggestions over the course of this thesis. As a post-doc of the Goto lab, Dr. Thierry Ducat was a patient and thorough teacher with regards to anything NMR-related. He is also a master of lab technique, and I basically tried to conduct myself in such a way that would avoid Thierry's disapproving, "Just bad!" commentary. Dr. Houman Ghasriani, another post-doc of the Goto lab, was also instrumental in my ability to solve a protein structure using NMR. His talents regarding data processing and structure calculation were always impressive and key to my learning. I owe deep thanks to Dr. John Pezacki for the rhomboid activity-based labeling idea, and to Dr. David Blais for cheerfully conducting all the labeling assays – we were a great team! I would also like to thank the undergraduate students that have contributed to this work: Mike Braganza, Elizabeth Nguyen, Josephine Ta and Ben Chung. They were all wonderful to work with and greatly appreciated!

In addition, I am very thankful for all the indirect contributions people have made during my studies. I can not count the number of times Pierre Pilon and Marc Fredette of the Undergraduate Biochemistry Lab have helped me out: everything from loaning equipment to providing molecular biology advice. I am also grateful for access to the Scaiano lab microplate reader as well as the Boddy lab MALDI mass spectrometer, and of course for all the assistance Dr. Glenn Facey and Cheryl McDowell have provided from the NMR facility. I would also like to thank both past and present members of the Goto lab for many good memories. Some highlights include: practical jokes using ransom letters or our muscular mascot; pumpkin carving; the Peanut Lover incident; Bon Jovi sing-alongs; parties and group lunches; and all our trips including Wonderland, Cedar Point, Montreal, San Diego and Costa Rica. I am very grateful to have made such amazing friends! A very warm shout-out to the original gang: Lisette, Dennis, Jenny, Steve, Andrew and Tatjana ☺

I owe my deepest thanks to my family. My parents, Karen and Glen, and sister, Angela, have provided unwavering support and encouragement throughout my life. There was never pressure to pursue one path or another, and all they cared about was my own happiness. I am lucky to have such a great family! My in-laws, Pat and Paul, are my second family and have provided so much support from my first year of undergrad to present day. They have treated me like their own daughter, and have definitely kept me well-fed all these years! And finally, my husband Paul has given me the strength to keep going and has always believed in me, regardless of how I felt. Words cannot express how happy I am to have him in my life. I owe so much to him, and will never forget how much he has done for me especially in these last few months. Paul has been extremely patient and accommodating during all my studies, and it is difficult to imagine how I could achieve this without him.

Table of Contents

Abstract…….....ii

Acknowledgements……..... iv

List of Abbreviations…….....x

List of Figures……..xiv

List of Tables……...xvii

Chapter 1: Introduction…….....1

1.1 Intramembrane Proteolysis……...2

1.1.1 Proteases……...2

1.1.2 The three intramembrane protease families…….....5

1.2 The rhomboid protease…….....7

1.2.1 Biological function……...8

1.2.1.1 Involvement in cell signaling……...10

1.2.1.2 Role in parasite pathogenicity…….....12

1.2.1.3 Function of the mitochondrial rhomboid…….....13

1.2.2 Substrate specificity……....14

1.2.3 Structure…….....16

1.2.3.1 General features that support hydrolysis within the membrane…….....16

1.2.3.2 The rhomboid active site……..16

1.2.3.3 Structural elements important for rhomboid activity……...19

1.2.3.4 Models for substrate entry into the active site……..21

1.2.4 Structure beyond the catalytic transmembrane core……...23

1.2.4.1 General features of rhomboid extramembranous domains...23

1.2.4.2 Role of the RHBDL2 rhomboid cytoplasmic domain……..25

1.2.4.3 The *Pseudomonas aeruginosa* rhomboid cytoplasmic domain…….....25

1.2.5 Regulation……...28

1.2.5.1 Modulation of PARL function……..29

1.2.5.2 Regulation through expression and/or compartmentalization……...29

1.2.5.3 Prokaryotic rhomboid regulation……..30
1.2.6 Research focus and objectives……31
1.3 Nuclear Magnetic Resonance Spectroscopy (NMR)……..32
1.3.1 Basic principles of solution NMR……..32
1.3.2 The heteronuclear single quantum coherence (HSQC) experiment...35
1.3.3 Chemical shift assignment……..37
1.3.4 Structure determination by NMR……...42
1.3.5 Membrane proteins and NMR spectroscopy……..46
1.3.5.1 The detergent micelle……46
1.3.5.2 Small isotropic bicelles……...48
1.3.5.3 Advances in membrane protein solution NMR……..49
1.4 Thesis overview…….....50
Chapter 2: Insights into the effect of detergents on the full-length rhomboid protease from Pseudomonas aeruginosa and its cytosolic domain……....53
2.1 Introduction…….54
2.2 Materials & Methods……..57
2.2.1 Plasmids……..57
2.2.2 Protein expression……...59
2.2.3 Purification of detergent-solubilized membrane protein samples…..59
2.2.4 Transfer of purified *pa*GlpG or Bla-Spitz$_{TM}$-MBP into bicelle solutions……60
2.2.5 Cytoplasmic domain purification……...61
2.2.6 NMR spectroscopy……62
2.2.7 *In vitro* protease activity assay……...62
2.2.8 Cytoplasmic domain titration experiments……63
2.3 Results……....64
2.3.1 *pa*GlpG purification……....64
2.3.2 Rhomboid activity in different membrane mimetics……..66
2.3.3 NMR spectroscopy of detergent-solubilized *pa*GlpG……....69
2.3.4 Investigation of the detergent-binding properties of the isolated cytoplasmic domain……....71

2.3.5 Does the *pa*CytD bind lipid bilayers……..75
2.4 Discussion……..79
2.4.1 Influence of various membrane-mimetic media on *pa*GlpG structure and function…….....79
2.4.2 Functional role of the *pa*GlpG cytoplasmic domain……...81
Chapter 3: ***Novel interactions at the N-terminus of the E. coli rhomboid protease***…..85
3.1 Introduction…….86
3.2 Materials & Methods……..86
3.2.1 Plasmids……..86
3.2.2 Protein expression and purification…….....87
3.2.3 MALDI-MS…….....90
3.2.4 Rhomboid protease assay using fluorogenic casein substrate…….....92
3.2.5 Active rhomboid labeling with FP-PEG-rhodamine…….....92
3.3 Results…….93
3.3.1 Isolation and activity of *ec*GlpG TMD core……...93
3.3.2 Interactions of the TMD cytoplasmic face…….96
3.3.3 Active site labeling of purified rhomboid protease……98
3.3.4 Active site labeling of the rhomboid in crude membrane extracts...100
3.4 Discussion……102
3.4.1 Role of the critical N-terminal sequence……..102
3.4.2 Interactions of the TMD cytoplasmic face……...104
3.4.3 Casein as a water-soluble rhomboid substrate……..107
3.4.4 Activity-based profiling of the rhomboid protease – an insightful assay……..108
Chapter 4: ***Three-dimensional domain swapping in the E. coli rhomboid cytoplasmic domain***……..111
4.1 Introduction……...112
4.2 Materials & Methods……....112
4.2.1 *ec*CytD expression and purification……..112
4.2.2 Measurement of interconversion rates between monomeric and domain-swapped dimeric *ec*CytD…….114

4.2.3 Preparation of a ^{15}N, ^{13}C-/^{14}N, ^{12}C-labeled heterodimeric ecCytD sample……...116
4.2.4 NMR spectroscopy and monomeric *ec*CytD structure calculation...116
4.2.5 Purification of full-length *ec*GlpG and isolation of TMD core…...118
4.3 Results……...118
4.3.1 Identification of dimeric *ec*CytD……..118
4.3.2 Energetics of dimerization……120
4.3.3 Structural characterization of *ec*CytD by solution NMR…….120
4.3.4 Effect of detergent micelles on the *ec*CytD……..129
4.3.5 Effects of detergent size and charge on dimerization……...132
4.3.6 Dimerization induced by incubation with a denaturant.....135
4.3.7 *ec*CytD-mediated dimerization in the full-length GlpG protein.....135
4.4 Discussion…….138
4.4.1 *ec*CytD dimerizes through 3D domain swapping…….138
4.4.2 The detergent micelle as a catalyst for domain swapping.....140
4.4.3 How does the detergent micelle catalyze domain-swapping? …….140
4.4.4 Micelle properties important for catalysis…….....142
Chapter 5: Discussion and future directions……....145
5.1 Effect of membrane mimetics on the full-length rhomboid protease……...146
5.2 The effect of detergent monomers on aqueous regions of membrane proteins…….150
5.3 Interactions at the membrane in the *E. coli* rhomboid……..152
5.4 Dimerization of the *ec*GlpG CytD through 3D domain-swapping……...155
5.5 Catalytic mechanism of 3D domain swapping by detergent micelles……..157
5.6 Biological implications from micellar catalysis of CytD domain swapping……..162
5.7 Future directions……...163
5.8 Conclusion……164
References……....167
Appendix 1……...187
Curriculum Vitae……192

List of Abbreviations

3D	Three-dimensional
4-HCCA	alpha-cyano-4-hydroxycinnamic acid
AarA	*Providencia stuartii* Rhomboid
Aβ	amyloid β peptide
ADAM	A Disintegrin And Metalloproteinase
AMS	4-acetamido-4′-maleimidylstilbene-2,2′-disulfonic acid
APP	Amyloid Precursor Protein
Bla	β-lactamase
BODIPY	Boron-Dipyrromethene
βOG	β-octylglucoside
BS-RNAse	Bovine Seminal Ribonuclease A
c-di-GMP	Cyclic Dimeric Guanosine Monophosphate
CSA	Chemical Shift Anisotropy
CHAPS	3-[(3-Cholamidopropyl)dimethylammonio]-1-propanesulfonate
CMC	Critical Micelle Concentration
CytD	Cytoplasmic Domain
COS	CV-1 (simian) in Origin, carrying SV40 genome
COSY	Correlation Spectroscopy
DAGK	Diacyl Glycerol Kinase
DCI	3,4-Dichloroisocoumarin
DDM	n-Dodecyl-β-D-maltoside
DHPC	1,2-dihexanoyl-sn-glycero-3-phosphocholine
DMPC	1,2-dimyristoyl-sn-glycero-3-phosphocholine
DMSO	Dimethyl Sulfoxide
DPC	Dodecylphosphocholine
DT	*Diphtheria* Toxin
DsbB	*Escherichia coli* Disulfide bond formation protein B
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
EhROM1	*Entamoeba histolytica* Rhomboid 1

ER	Endoplasmic Reticulum
*ec*GlpG	*Escherichia coli* Rhomboid
EDTA	Ethylenediaminetetraacetic Acid
Fos10	Decylphosphocholine
Fos12	Dodecylphosphocholine
Fos16	Hexadecylphosphocoline
FPLC	Fast Protein Liquid Chromatography
FP-PEG-rhodamine	Fluorophosphonate-Polyethylene Glycol-rhodamine
g3p	Glycerol-3-phosphate
GB1	Immunoglobulin G binding protein B1 domain from *Streptococcus*
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
*hi*GlpG	*Haemophilus influenzae* Rhomboid
His_6	Hexahistidine tag
HSQC	Heteronuclear Single Quantum Coherence
IPTG	Isopropyl B-D-Thiogalactopyranoside
iRHOM	Inactive Rhomboid
K_d	Dissociation constant
kDa	Kilodalton
KdpD	*Escherichia coli* histidine kinase K+ sensor
LDAO	Lauryldimethylamine-oxide
L1	Loop 1
L5	Loop 5
LacYTM2	LacY Transmembrane Helix 2
LMPC	1-myristoyl-2-hydroxy-sn-glycero-3-phosphocholine
LMPG	1-myristoyl-2-hydroxy-sn-glycero-3-[phospho-rac-(1-glycerol)]
LPPG	1-palmitoyl-2-hydroxy-sn-glycero-3-[phospho-rac-(1-glycerol)]
MALDI	Matrix Assisted Laser Desorption/Ionization
MALDI-TOF	Matrix Assisted Laser Desorption/Ionization Time Of Flight
MBP	Maltose Binding Protein
mRNA	Messenger Ribonucleic Acid
MRS	Mitochondrial Retrograde Signaling

MS	Mass Spectrometry
MWCO	Molecular Weight Cut Off
NG	*n*-nonyl-β-D-glucoside
Ni-NTA	Nickel-Nitrilotriacetic acid
NOE	Nuclear Overhauser Effect
NOESY	Nuclear Overhauser Effect Spectroscopy
NMR	Nuclear Magnetic Resonance
OmpA	Outer Membrane Protein A
p13suc1	Suppressor of Cyclin Dependent Kinase 1
*pa*GlpG	*Pseudomonas aeruginosa* Rhomboid
PARL	Presenilin-associated rhomboid like
Pcp1	*Schizosaccharomyces pombe* mitochondrial rhomboid
PDC	Protein Detergent Complex
PfROM4	*Plasmodium faciparum* Rhomboid 4
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
POPE	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine
PPM	Parts Per Million
PROD	PARLl-Rhomboid-Domain
pSRII	*Natronomonas pharaonis* Sensory rhodopsin II
RHBDL2	*Homo sapiens* Rhomboid 2
Rho-1	*Drosophila melanogaster* Rhomboid 1
Rho-2	*Drosophila melanogaster* Rhomboid 2
Rho-3	*Drosophila melanogaster* Rhomboid 3
Rho-7	*Drosophila melanogaster* Mitochondrial Rhomboid 7
RNAseA	Ribonuclease A
SDS	Sodium Dodecyl Sulfate
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SREBP	Sterol Regulatory Binding Element Protein
SPP	Signal Peptide Peptidase
TatA	Twin Arginine Translocase
TEV	Tobacco Etch Virus

TGFα	Transforming Growth Factor
TgROM5	*Toxoplasma gondii* Rhomboid 5
TM	Transmembrane
TMD	Transmembrane Domain
TMH	Transmembrane Helix
TOCSY	Total Correlation Spectroscopy
TROSY	Transverse Relaxation Optimized Spectroscopy
TPCK	Tosyl Phenylalanyl Chloromethyl Ketone
WT	Wild-type
YqgP	*Bacillus subtilis* Rhomboid

List of Figures

Figure 1.1 Mechanism of the first step of peptide hydrolysis in the four general protease families……...3

Figure 1.2 The serine protease mechanism……..…….....4

Figure 1.3 Schematic representation of representatives from the three intramembrane protease families, along with their substrates……..6

Figure 1.4 Examples of rhomboid function in biological systems……....9

Figure 1.5 X-ray crystal structure of the *ec*GlpG rhomboid protease…….18

Figure 1.6 Active site of *ec*GlpG co-crystalized with the 7-amino-4-chloro-3-methoxy-isocoumarin inhibitor……..20

Figure 1.7 Proposed model of substrate entry to the active site…….....24

Figure 1.8 Proposed membrane interaction of the *pa*GlpG CytD……...27

Figure 1.9 Energy level splitting of spin ½ nuclei in an external magnetic field (B_0)....34

Figure 1.10 The ^{1}H-^{15}N Heteronuclear Single Quantum Coherence (HSQC) experiment…….....36

Figure 1.11 The effect of chemical exchange in NMR spectra…….38

Figure 1.12 Correlations provided by 3D NMR experiments used in this thesis…….....40

Figure 1.13 Overview of the HNCO experiment……...41

Figure 1.14 Sequential assignment of the amide backbone using the HNCACB experiment……..43

Figure 1.15 Through-space distances between nuclei are determined using the ^{15}N-NOESY-HSQC……...45

Figure 1.16 Schematic diagram of micelle, bicelle and vesicle environments……..47

Figure 2.1 Comparison of sequences from the two rhomboids for which core catalytic domain crystal structures are available (EC, *E. coli*, HI, *H. influenza*) with the *P. aeruginosa* (PA) rhomboid sequence……..55

Figure 2.2 Coomassie-stained SDS-PAGE gel of fractions taken at various stages in the purification of *pa*GlpG……65

Figure 2.3 Activity of *pa*GlpG in different membrane mimetic environments……...67

Figure 2.4 ^{1}H-^{15}N HSQC spectra of full-length *pa*GlpG purified in DDM, DPC, LMPC and LPPG……...70

Figure 2.5 The assigned ^{1}H-^{15}N HSQC spectrum of the CytD, superimposed on that from full-length *pa*GlpG solubilized in LMPC……..72

Figure 2.6 ^{1}H-^{15}N HSQC spectra of a 175 μM solution of the CytD in the absence of detergent superimposed with spectra obtained in the presence of DPC…….74

Figure 2.7 Average backbone amide chemical shift differences for *pa*CytD residues induced by addition of a 2% (w/v) solution of DHPC/DMPC bicelles with q=0.44 or CHAPS/DMPC bicelles with q=0.5……...77

Figure 2.8 Effects of bicelles and lipid vesicles on the *pa*CytD…….....78

Figure 3.1 Purification and limited trypsin digestion of *ec*GlpG…….....91

Figure 3.2 Schematic diagram of activity-based profiling approach for monitoring the activity of *ec*GlpG in crude membrane extracts…….....94

Figure 3.3 Full-length GlpG is more active than trypsin-generated TMD…….....95

Figure 3.4 Identification of functionally important residues at the cytoplasmic face of the TMD…….....97

Figure 3.5 MALDI mass spectra of full-length GlpG compared to the trypsin-isolated TMD……...99

Figure 3.6 Rhomboid activity in purified and crude extracts can be evaluated using a fluorophosphonate fluorescent probe……...101

Figure 3.7 Interactions with the lipid bilayer and/or cytoplasmic face of the TMD may modulate rhomboid activity……..105

Figure 4.1 Schematic representation of domain swapping……....113

Figure 4.2 Evidence of *ec*CytD dimerization…….....……....119

Figure 4.3 Interconversion between monomer and dimer occurs at elevated temperatures……..121

Figure 4.4 Kinetic profiles of *ec*CytD dimerization shown for 45°C, 50°C, 55°C and 60°C……...122

Figure 4.5 Energetics of *ec*CytD dimerization…….....123

Figure 4.6 The assigned ^{1}H-^{15}N HSQC spectrum of monomeric *ec*CytD, recorded at pH 6.5, 25˚C…….....125

Figure 4.7 Solution NMR structure of the *E. coli* GlpG CytD, aligned with the *P. aeruginosa* CytD …….....126

Figure 4.8 Chemical shift differences between monomeric and dimeric *ec*CytD.…....127

Figure 4.9 Cα secondary shift analysis and model for domain swapping of the *ec*CytD……...128

Figure 4.10 Catalysis of *ec*CytD dimerization by the detergent micelle……....131

Figure 4.11 Effect of micelle size and charge on *ec*CytD dimerization rates.....…...134

Figure 4.12 The effect of 1M urea on the *ec*CytD dimerization rate…….....136

Figure 4.13 Size exclusion chromatography profiles suggest full-length, but not TMD *ec*GlpG also undergoes dimerization.…….....137

Figure 4.14 Energy diagram for the interconversion between monomer and dimer, assuming that it can be as a two-state system with a single transition state……..……..……..…….....143

Figure 5.1 Model of domain swapping for *ec*CytD with the proposed mechanism for detergent micelle catalysis……....160

Figure 5.2 Comparison of inactivating γ-cleavage of mammalian Parl with the critical N-terminal sequence of *ec*GlpG.…….....165

List of Tables

Table 1.1 Rhomboid crystal structures.....17

Table 1.2 Detergent chemical structures.....51

Table 2.1 Physical properties of detergents used in this chapter.....56

Table 2.2 Primer sequences for *pa*GlpG constructs.....58

Table 3.1 Primer sequences for truncation constructs.....88

Table 3.2 Mutagenic primers used for site-directed mutagenesis.....89

Table 3.3 Potential *ec*GlpG rhomboid cleavage sites of Bovine β-casein.....109

Table 4.1 Structural statistics for the final ensemble of 20 water-refined structures of monomeric *ec*CytD.....117

Chapter 1: Introduction

1.1 Intramembrane Proteolysis

1.1.1 Proteases

Proteases are enzymes that catalyze the hydrolysis of peptide bonds and are responsible for a vast number of biological processes. They have critical roles in every living cell and are also essential in viral life cycles (1). The biological roles played by proteases are highly diverse and range from general intra- or extracellular degradation to specific cleavage for activation or signaling. Progression through the cell cycle (2, 3), apoptosis (4) and regulation of blood coagulation (5) are just a few examples that highlight the importance of proteases in nature. A number of disease states have also been identified that are either caused or exacerbated by abnormal protease function, providing attractive targets for drug development (6).

Proteases are classified into four main groups: cysteine, serine, aspartyl and metallo proteases (7) (Fig. 1.1). The cysteine and serine proteases both contain the namesake residue in the active site, maintained in a nucleophilic state by interactions with an activating base (8, 9). As shown for the serine protease in Figure 1.2, nucleophilic attack on the peptide carbonyl carbon creates a negatively charged tetrahedral intermediate that is stabilized in the oxyanion-binding 'hole'. The covalent intermediate formed between the enzyme and C-terminal cleavage product is later released through hydrolysis with a water molecule (8). Aspartyl and metalloproteases, on the other hand, use activated water as the nucleophile to attack the carbonyl carbon and form the tetrahedral intermediate (10, 11). Protonation of the scissile amide occurs after rearrangement of this intermediate.

Until recently, all proteases were thought to be globular, aqueous-phase proteins, either in a free or membrane-tethered state. This notion was challenged with the surprising

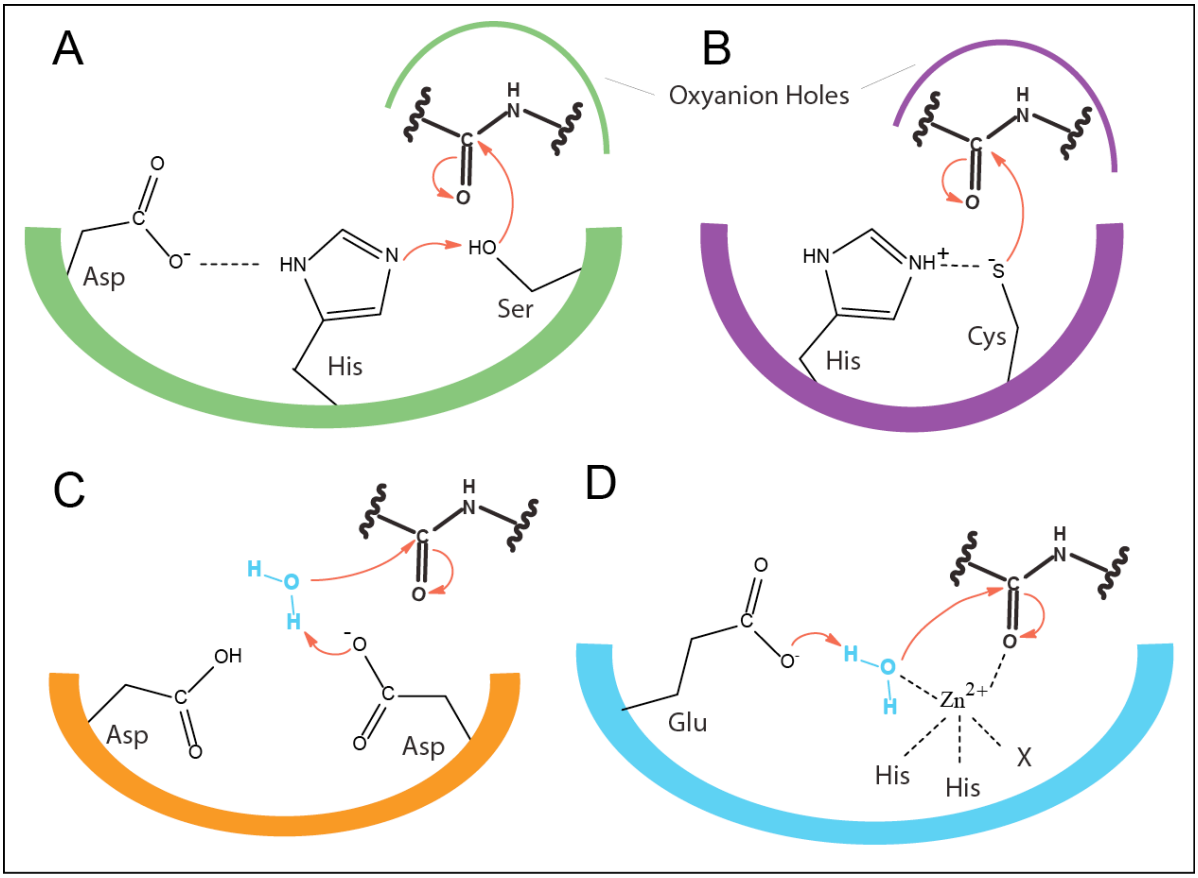


Figure 1.1 Mechanism of the first step of peptide hydrolysis in the four general protease families. (A) Serine and (B) cysteine proteases promote nucleophilic attack of the peptide carbonyl carbon by serine or cysteine, through activating interactions with a proximal histidine side chain. The oxyanion hole stabilizes the tetrahedral intermediate formed after initial nucleophilic attack of the peptide carbonyl, and during hydrolysis of the acyl enzyme intermediate. (C) The aspartyl protease coordinates a water molecule that is activated through abstraction of a proton while the metalloprotease (D) uses glutamate and a coordinated metal ion (often zinc) to activate a water molecule to attack the peptide carbonyl. Arrows indicate the movement of electrons and dotted lines indicate hydrogen bonding or electrostatic interactions. The peptide substrate is shown in bold.

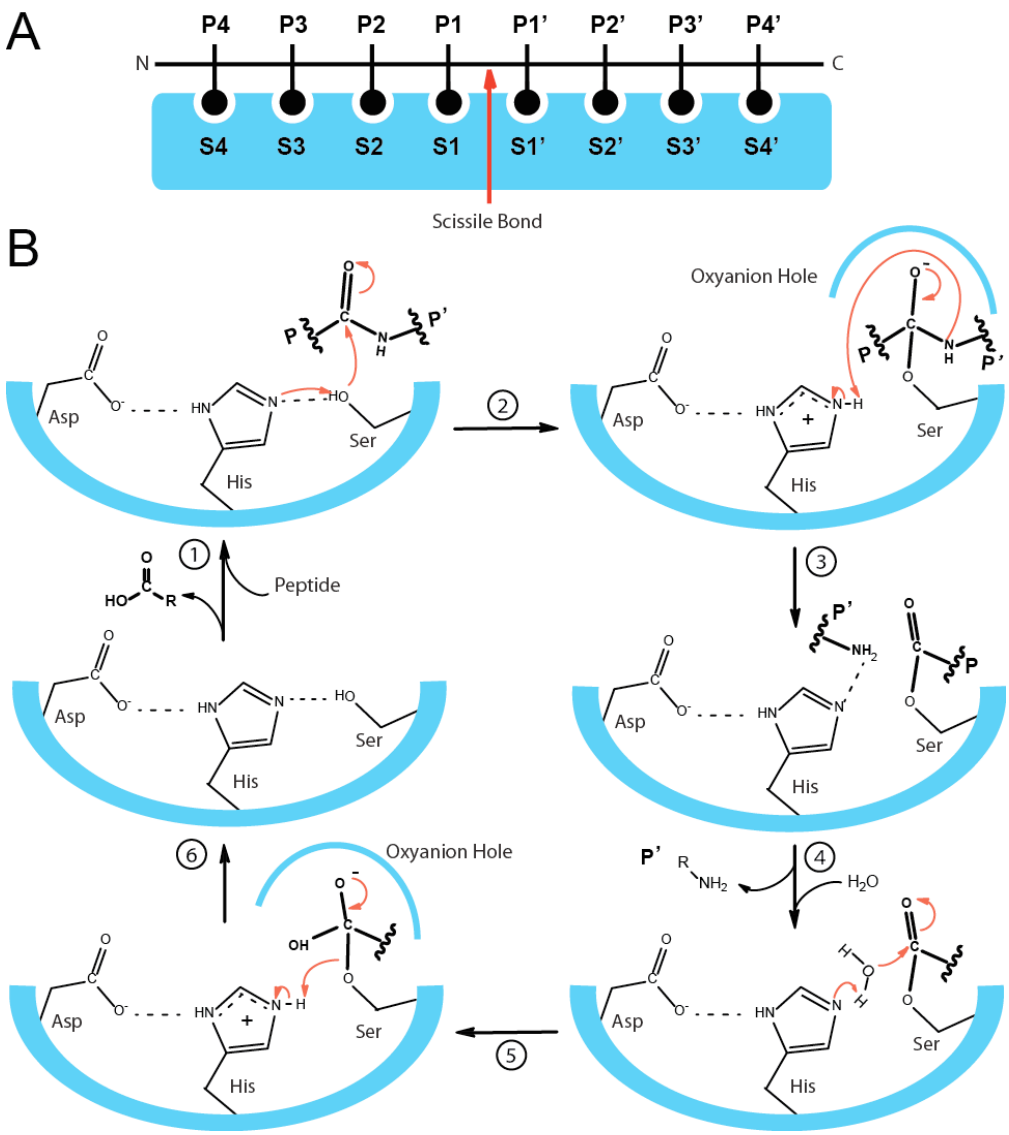


Figure 1.2 The serine protease mechanism. (A) General nomenclature of substrate cleavage site positions with respective protein binding sites. Interactions between the substrate sidechains (P_n) n residues away from the cleavage site, and enzyme binding pockets (S_n) responsible for binding to residue n determines specificity. (Note that residues C-terminal to the cleavage site are distinguished from N-terminal residues by the designation n'.) (B) The complete serine protease mechanism. (1) The substrate enters the active site and positions the P1 carbonyl close to the serine through interactions in (A). (2) Stabilized by aspartate, the histidine abstracts the hydroxyl proton to activate serine, generating the covalently-bound tetrahedral intermediate stabilized by the oxyanion hole. (3) Collapse of the tetrahedral intermediate leads to peptide bond cleavage and acyl enzyme intermediate formation. (4) Water is activated by His to attack the carbonyl carbon creating a tetrahedral intermediate (5) that is stabilized by the oxyanion hole. (6) The tetrahedral intermediate collapses, leading to dissociation of the C-terminal cleavage product from the enzyme.

discovery of integral membrane proteins with protease activity against other transmembrane proteins (12-14). These intramembrane proteases were considered to be unique for their ability to use a hydrolysis mechanism for peptide cleavage in the water-poor environment of the lipid bilayer. Since this initial discovery, three different classes of intramembrane protease have been identified, each of which acts via a distinct mechanism.

1.1.2 The three intramembrane protease families

The first intramembrane protease was identified in a study of human cholesterol metabolism (12, 15). The signaling factor responsible for turning on cholesterol and fatty acid synthesis genes, known as sterol regulatory binding element protein (SREBP), is processed and released from the membrane (15) by two cleavage events: first by the membrane-associated Site-1 protease, and the second by the intramembrane metalloprotease Site-2 protease at a site predicted to be buried in the lipid bilayer (16, 17) (Fig. 1.3a). The Site-2 protease is a multi-spanning helical membrane protein that contains a conserved HExxH sequence also found in water-soluble zinc metalloproteases, thought to form coordination bonds with zinc (16, 17). To gain access to the transmembrane location of the SREBP Site-2 cleavage site, an essential asparagine/proline pair 11 residues C-terminal to this site is required (18). These residues have a low propensity to form α-helices and their presence in the middle of the TM segment has been proposed to promote partial helix unfolding once the first processing event by Site-1 has occurred (18).

Since the initial discovery of the metalloprotease class of intramembrane proteases, an aspartyl protease class of intramembrane protease has also been described. One of the most well-characterized members of that family is the catalytic subunit of γ-secretase which

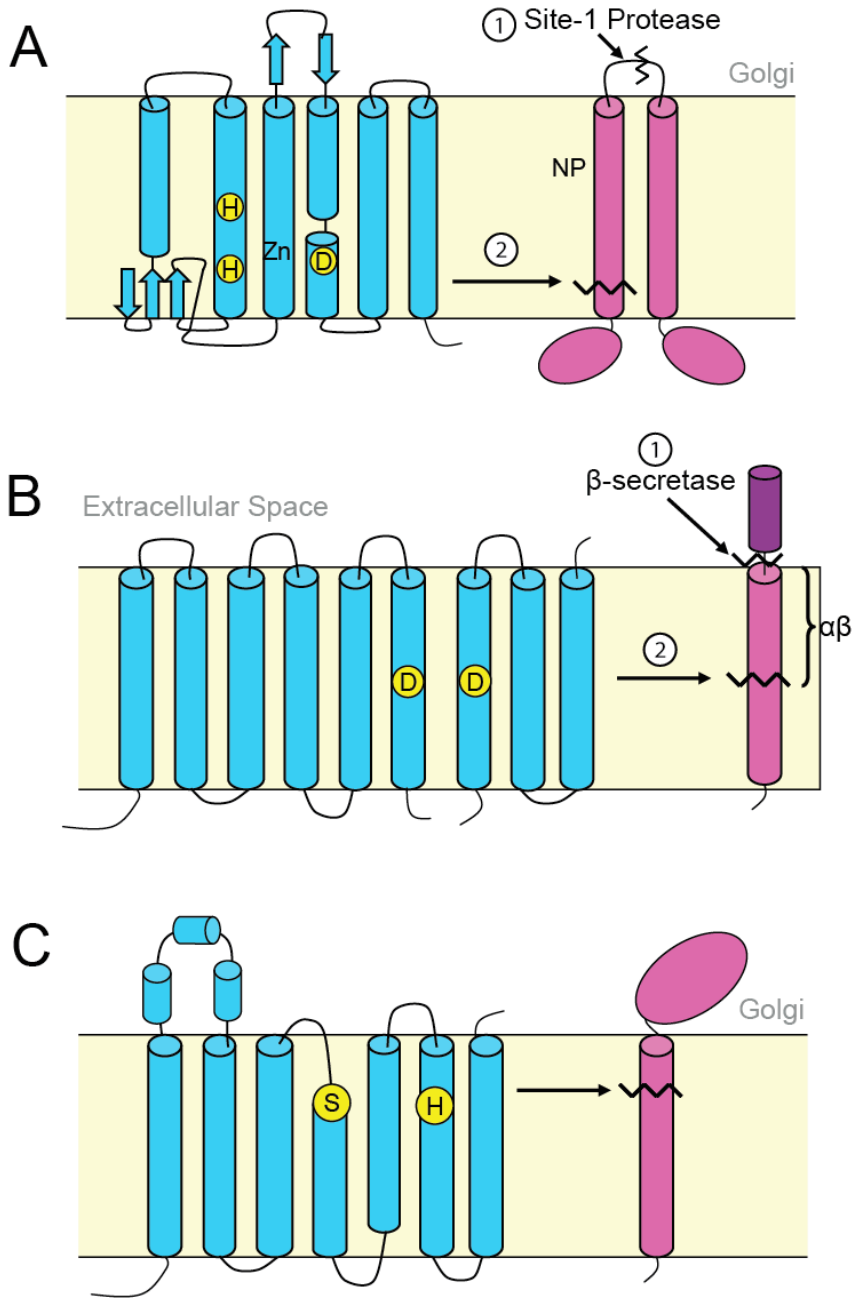


Figure 1.3 Schematic representation of representatives from the three intramembrane protease families (shown in blue), along with their substrates (pink). (A) Site-2 metalloprotease and its substrate SREBP. SREBP requires pre-processing by the membrane-associated Site-1 protease (zig-zagged line 1) to release the N-terminus before it can be cleaved by Site-2 protease (2). The appropriate position of the helix destabilizing asparagine/proline (NP) pair essential for Site-2 cleavage is indicated for SREBP. (B) Presenilin aspartyl protease and its APP substrate. Presenilin (in the γ-secretase complex) is processed into an N- and C-terminal fragment and cleaves APP once β-secretase (1) has released the extracellular domain. (C) *Drosophila* Rhomboid-1 serine protease and its EGFR-ligand Spitz substrate. Spitz is not processed prior to rhomboid cleavage.

cleaves the amyloid precursor protein (APP) to produce the amyloid β peptide (Aβ) (19) (Fig. 1.3b). Mutations in this subunit affect where APP is cleaved, generating variants of Aβ that aggregate and form the senile plaques that are thought to cause Alzheimer's disease (20, 21). Biochemical studies of this protein have been challenging, partly due to its large size and association with multiple proteins in the functional complex, limiting what is known about substrate specificity and mechanism. However, it has been shown that γ-secretase and other members of this family require preprocessing of their substrates by other proteases prior to intramembrane cleavage (16, 17).

The most recently discovered class of intramembrane protease, a member of which is the focus of this thesis, is known as the rhomboid family. The rhomboid comprises a highly conserved group of proteins that use a serine protease mechanism to carry out a wide variety of biological functions (22). The rhomboid consists of a 6 TMH core that contains the catalytic active site. In some cases an extra helix at either the N- or C-terminus is present, depending on its location in the cell (Fig. 1.3c). In contrast to the other two types of intramembrane proteases, the rhomboid does not require preprocessing of its substrate prior to cleavage (23), raising interesting questions regarding how this protease might act on a helical transmembrane substrate.

1.2 The rhomboid protease

Originally discovered as a mutation in *Drosophila* that caused the head to develop with a rhomboid shape (24), rhomboid function was initially not well understood. Early results suggested that the *Drosophila* Rhomboid-1 (Rho-1) had a role in development through the epidermal growth factor receptor (EGFR) pathway, and that it was active in the

signal-sending cell. In 2001, Rho-1 was shown to act as a serine protease that catalyzed the release of EGFR signaling proteins (13) (Figures 1.3c and 1.4a). In this landmark discovery, the *Drosophila* EGFR signaling protein Spitz was co-expressed in a number of eukaryotic cell lines with Rho-1, giving rise to rhomboid-specific Spitz cleavage. Mutational analysis of conserved amino acids identified a number of functionally important residues and highlighted the potential for Rho-1 to be the first identified intramembrane serine protease based on the requirement for three specific residues: asparagine, serine and histidine (Fig 1.3c). The critical serine residue was located in a motif that is highly conserved in soluble serine proteases (GxSxG), providing additional evidence that the rhomboid is a serine protease. Further support for this idea was provided by the ability of serine protease inhibitors to inhibit rhomboid activity, while other classes of protease inhibitors showed no activity against the rhomboid (13). This study, as well as an extensive phylogenetic analysis (22) revealed that the rhomboid is a highly conserved protein with at least one representative found in almost all sequenced genomes. Since the publication of these early results, development in the rhomboid field has been rapid with primary advances dramatically improving understanding of the catalytic mechanism and biological roles (reviewed in (25)).

1.2.1 Biological function

The biological role of a rhomboid protease depends on the system in which it is found, with a wide range of roles having been described (26-29). Beyond EGFR signaling in *Drosophila*, there is evidence that the mammalian rhomboid, RHBDL2, could be involved in activation of the EGFR through EGF release during development (30).

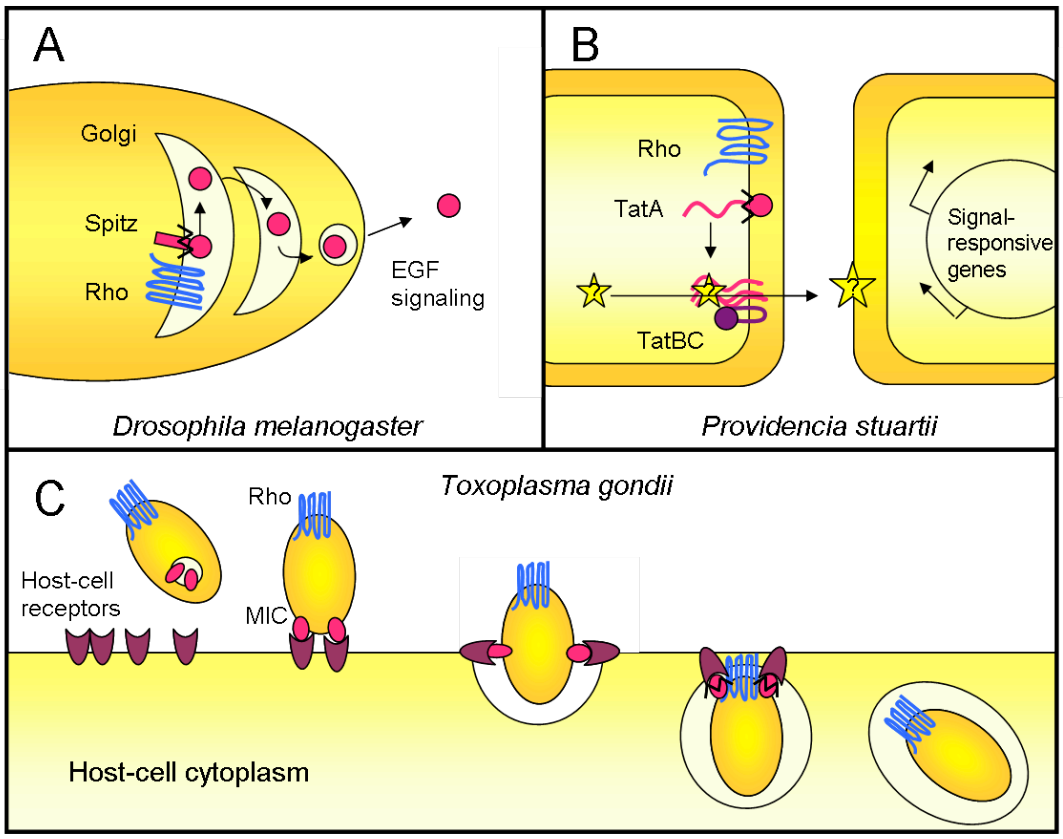


Figure 1.4 Examples of rhomboid function in biological systems. (A) In the *Drosophila* golgi, Rhomboid-1 (blue) cleaves the luminal side of the Spitz transmembrane domain (pink) for release of the extracellular domain to be exported out of the cell for subsequent epidermal growth factor (EGF) signaling. (B) The *P. stuartii* rhomboid, AarA (blue), is proposed to cleave TatA (pink) which can then bind the TatB/C complex (purple) to form a pore. This pore releases a signaling molecule that has not yet been characterized (shown as a yellow star), but is known to turn on signal-responsive genes for quorum sensing. C) Invasion of the protozoan parasite *T. gondii* into host cells is mediated by the TgRom5 (blue). *T. gondii* MIC adhesins (pink) bind to host-cell receptors (purple) and are cleaved during host-cell invasion to release the parasite.

In addition to a role in bacterial cell signaling (31), other roles that have been described for the rhomboid include host-cell invasion by intracellular parasites (32-34), immune system evasion by parasitic amoeba (35), and mitochondrial remodeling (36).

1.2.1.1 Involvement in cell signaling

Once the role of Rho-1 in release of Spitz from the membrane had been established, additional studies were performed to determine the extent of rhomboid involvement in the *Drosophila* EGFR signaling pathway. For this purpose, four *Drosophila* rhomboid proteins (including Rho-1) were assessed for their ability to cleave other EGFR ligands and activate the EGFR *in vivo* (37). Since all EGFR ligands tested were found to be cleaved by the four rhomboids, it was proposed that these rhomboids have distinct functions as EGFR regulators. Some overlap in function was also demonstrated, since Rho-3 could fulfill the EGFR signaling role during eye development (38), consistent with previous findings from a Rho-1 knockout that showed no change in fly head morphology (39). Rho-3 has also been shown to have EGFR signaling roles in embryogenesis (40) and apoptosis protection in epidermal cells (41). Finally, Rho-2 is expressed in *Drosophila* germ cells and appears to be involved in EGFR signaling to the surrounding somatic cells (42, 43). Based on these reports, there is a strong consensus that the rhomboid plays a key role in EGFR signaling in the development of *Drosophila*.

While it is generally accepted that the rhomboid is involved in EGFR-mediated cell development in *Drosophila*, only recently did evidence of a similar role in mammals become available. One reason for this delay is that membrane-tethered EGFR signaling factors are typically released by members of the ADAM (a disintegrin and metalloproteinase) family of proteases (44, 45). However, the Freeman group has now

identified EGF as a substrate of the mammalian RHBDL2 rhomboid since it could be processed by RHBDL2, and EGFR signaling can occur in a metalloprotease-independent manner (30). This suggests that RHBDL2-mediated release of EGF could be important at specific stages of embryogenesis, with cleavage by ADAM potentially occurring at different stages. Also worth noting is the homologous human rhomboid, RHBDD2, that has recently been shown to be overexpressed in carcinomas that are negative for the estrogen receptor (46). Most interestingly, breast cancer cell proliferation can be reduced by *in vitro* silencing of this rhomboid, raising the possibility that RHBDD2 could be acting in cell proliferation or survival signaling pathways.

A distinct role in signaling was described for the bacterial AarA rhomboid from *Providencia stuartii* that is required for quorum sensing (47). AarA activates a subunit of the twin arginine translocase (TatA) by removing an inhibitory sequence of TatA, allowing the transport of as yet unidentified factors in quorum sensing (Fig. 1.4b). This appears to be a unique mechanism for rhomboid-mediated inter-cellular signaling, since *P. stuartii* is currently the only known species that requires processing of the TatA protein. It is still unclear what the function of other bacterial rhomboids might be, even though most bacterial genomes contain one or more rhomboids. Although their functions remain unclear, bacterial rhomboids have turned out to be amenable to a range of structural and functional studies, and are consequently the best characterized in the rhomboid family.

The *Escherichia coli* rhomboid, *ec*GlpG, is currently the most extensively studied member of the rhomboid proteases, although its function has yet to be elucidated. A GlpG knockout *E. coli* strain (48, 49) has been shown to be viable, and does not display any obvious phenotypic abnormalities (49). However, the *glpG* gene is located in the glpEGR

operon of the glp regulon that contains genes responsible for glycerol-3-phosphate (g3p) catabolism (50). In the same operon as *glpG* is *glpE*, a gene that encodes a protein of unknown function, and *glpR* that encodes a DNA-binding repressor protein. GlpR negatively controls all glp operon transcription by binding to sequences close to or overlapping with *glp* operon promoters (50). This inhibition is relieved by GlpR binding to g3p, allowing g3p-dependent expression of glp genes. It was found that GlpE, GlpG and GlpR are co-transcribed since there are no internal terminators within the operon; however, the presence of multiple promoters raises the possibility of differential expression of these genes (51). Overall, little is known about the function of this protein other than its ability to cleave certain EGFR ligands (like Gurken) (23) or substitute for AarA in *P. stuartii* (52, 53).

1.2.1.2 Role in parasite pathogenicity

The rhomboid protease is gaining increasing attention for its role in invasion, growth and survival of parasites that infect a significant proportion of the human population, particularly in developing countries (reviewed in (28)). This was first discovered in *Toxoplasma gondii*, the organism that causes toxoplasmosis, with cell-surface adhesins being identified as substrates of the rhomboid protease that are cleaved during entry into the cell (33) (Figure 1.4c). Inhibition of this cleavage prevents host cell invasion without having a significant effect on parasite morphology, metabolism or gliding motility (54). Adhesins in *Plasmodium,* the causative agent of malaria, are also cleaved by a rhomboid, in this case PfROM4 (34). While it appears that this cleavage is not directly linked to invasion of the host cell, it instead seems to be required for normal parasite growth. A distinct function for rhomboids has also been described for *Entamoeba histolytica*, an extracellular parasite that can cause amoebic dysentery (35). In this case, the rhomboid EhROM1 seems

to be required for immune system evasion, cleaving opsinized antigens that would normally target the parasite for phagocytosis by white blood cells. Evidence for this role includes the observation that EhROM1 localizes to the same region where opsinized antigens are typically shed from the parasite, and its cleavage of these antigens in cell culture based assays. Overall, the discovery of critical roles for rhomboid proteases in parasite infection and survival has made them an attractive target for drug development.

1.2.1.3 Function of the mitochondrial rhomboid

A sub-class of rhomboid that differs from the *Drosophila* Rho-1, is found in the mitochondrial inner membrane. The presenilin-associated rhomboid like (PARL) family is characterized by the presence of an extra transmembrane helix at the N-terminus of the 6-TMH core. Members of the PARL family are present in a number of eukaryotic systems, with yeast, mammalian and insect members being the best characterized. As the functions of these PARL rhomboids were revealed, it became apparent that they do not associate with presenilin as was originally postulated from a yeast two-hybrid screen (55). Instead, the PARL rhomboid seems to be involved in the regulation of mitochondrial membrane dynamics. This is particularly true for the insect and yeast systems since deletions of the yeast pcp1 (56) and insect Rho-7 (57) yield a fragmented mitochondrial morphology. One substrate of the yeast pcp1 rhomboid is the Mgm1 GTPase, which needs to be released from the membrane to function in mitochondrial membrane fusion (36, 57). Mitochondrial membrane fusion is also regulated by the *Drosophila* Rho-7, which is thought to cleave the Mgm1 homologue, Opa. The mammalian system is more complicated, however, since mice with a PARL deletion do not display the abnormal mitochondrial morphology seen in yeast and insects, although they do have shorter life spans (58). Instead, PARL may play a role in

cytochrome c release and apoptosis, although further studies are required to understand the role of PARL in these events. Based on these eukaryotic systems, mitochondrial rhomboid function could be as diverse as non-mitochondrial rhomboids.

1.2.2 Substrate specificity

Rhomboids typically cleave single-pass type I transmembrane domains on the extracellular or luminal side, with the cleavage site usually being found near helix destabilizing residues. In a landmark study by the Freeman group (59), the activity of *Drosophila* Rho-1 was tested with a series of chimeric Spitz-based substrates, allowing the identification of a seven residue substrate sequence (ASIASGA) required for cleavage. Moreover, the insertion of a non-native GA or GG motif in the N-terminal region of a TMH was found to convert a non-substrate TMH into a cleavable sequence, showing the importance of helix destabilizing residues in the substrate. The presence of hydrophilic residues N-terminal to the cleavage site also appeared to be important, and was suggested to have a role in substrate entry to the active site (59). Based on these findings it was possible to identify a similar motif in *Toxoplasma* adhesins (59) that were subsequently demonstrated to serve as substrates for three unrelated rhomboids.

The idea that helix-destabilizing residues are required for substrate cleavage by rhomboids was further substantiated when the *E. coli* rhomboid, *ec*GlpG, was found to cleave a TM segment originally meant to serve as a negative control. Specifically, Maegawa and coworkers created a large fusion protein consisting of the second transmembrane helix of LacY (LacYTM2) with β-lactamase at the N-terminus and maltose binding protein at the C-terminus (49). Surprisingly, this substrate was cleaved by *ec*GlpG

in vivo and *in vitro* even though it was quite different from known rhomboid substrate sequences. The sequence features of LacYTM2 were then systematically analyzed in Δ*glpG E. coli* cells for their effects on cleavage. It was found that residues surrounding the cleavage site were only partially responsible for determining amount of substrate cleavage (60), and not completely dependent on a recognition motif. However, a three-fold increase in substrate cleavage could be obtained simply by the introduction of a helix-breaking Gln-Pro (QP) sequence into the middle of the TM segment, approximately seventeen residues away from the site of cleavage. Furthermore, it was possible to convert the Spitz TM that is normally not cleaved by this protease, into a GlpG substrate by introducing the QP sequence seventeen residues from the cleavage site. The presence of destabilizing residues around the cleavage site and midway through the transmembrane segment was hypothesized to bend the substrate into a favorable position for GlpG cleavage in the membrane (60).

To characterize the seemingly elusive rhomboid substrate specificity in better detail, the Freeman group tested the ability of an extensive series of TatA substrate mutants to be cleaved by the bacterial AarA rhomboid (61). While it was found that the presence of helix destabilizing residues in the TM substrate were an important determinant for cleavage *efficiency*, they did not seem to control *where* the substrate would be cleaved (61). Instead, amino acids adjacent to the cleavage site appear to be important for cleavage site specificity: specifically P4, P1 and P2'. In particular, P4 requires a large, hydrophobic side chain, P1 only accommodates a small side chain and P2' can accept any hydrophobic side chain regardless of size. These results were supported by analysis of other substrates that can be cleaved by AarA, and of substrates that are cleaved by divergent rhomboids. These results indicate that divergent rhomboids appear to recognize a general substrate recognition motif.

1.2.3 Structure

1.2.3.1 General features that support hydrolysis within the membrane

Early studies using topology prediction and biochemical analysis of conserved residues provided important insights into rhomboid structure and mechanism (13, 62, 63), however the development of protocols to generate purified samples suitable for crystal formation was a critical milestone for the rhomboid field (62, 63). This led to multiple rhomboid structures for the *E. coli* rhomboid *ec*GlpG (64-66) and the *Haemophilus influenzae* GlpG (*hi*GlpG) (67, 68). The structures (summarized in Table 1.1) revealed that the reaction takes place within a water-filled cavity formed by its 6 TM helices. This cavity is lined with hydrophilic residues that surround the active site serine approximately 10 Å below the membrane surface at the top of the short TM helix 4 (Figure 1.5). The conserved GxSG motif contains the active site serine, along with residues that form stabilizing interactions to keep the loop connecting TM3 and TM4 in place. These interactions ensure that the active site Ser will remain accessible to both water and substrate. There is also a conserved AHxxGxxxG motif found on TM6 that is responsible for ensuring that an active site histidine (underlined) remains within hydrogen bonding distance of the serine nucleophile. This includes a GxxxG motif often found in interacting TM helices (69), that in the rhomboid interacts with a matching motif in TM4, ensuring close contact between the active site histidine and serine.

1.2.3.2 The rhomboid active site

The active site of the rhomboid protease is similar to that of typical serine proteases with one exception: it uses a catalytic dyad rather than a triad. Early biochemical analyses of the rhomboid implicated a conserved asparagine as a potential catalytic triad residue

Table 1.1 Rhomboid crystal structures.

PDB ID	Resolution (Å)	Membrane Mimetic [a]	Distinguishing Feature	Ref
E. coli GlpG				
2IC8[a]	2.10	NG	First published intramembrane protease structure. TM5 is positioned close to TM2.	(65)
2IRV	2.30	LDAO/ DDM	Two antiparallel molecules in the asymmetric unit with TM5 positioned slightly farther from TM2 than in 2IC8.	(64)
2NRF	2.60	NG	Two antiparallel molecules in the asymmetric unit with a large variation in TM5 position relative to TM2 between the two structures.	(66)
2O7L	2.50	NG	Open cap conformation showing lifted Loop 5 and exposed active site S201.	(70)
3B44	1.70	NG	Conserved W136A mutant in Loop 1 showing narrow hydrophobic belt and unaltered structure compared to native enzyme.	(71)
3B45	1.90	NG	Native structure generated for comparison with W136A mutant.	(71)
2XOW	2.09	NG	Bound to isocoumarin inhibitor.	(72)
2XOV	1.65	NG	Native structure for comparison with isocoumarin complex.	(72)
2XTU	1.85	NG	S201T inactive mutant.	(73)
2XTV	1.70	DHPC/ CHAPS	S201T inactive mutant in bicelles.	(73)
H. influenzae GlpG				
2NR9	2.20	$C_{12}E_8$	First *H. influenzae* structure. Similar to *E. coli* GlpG, with an extended C-terminal helix and no N-terminal cytoplasmic domain.	(67)
3ODJ[b]	2.84	$C_{12}E_8$	Loop 4, helix 5 and loop 5 are disordered.	(68)

[a] Where NG is n-nonyl-β-D-glucoside, LDAO is lauryldimethylamine-oxide, DDM is n-dodecyl-β-D-maltoside, DHPC is dihexanoyl phosphatidylcholine, CHAPS is 3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate and $C_{12}E_8$ is octaethylene glycol monododecyl ether.

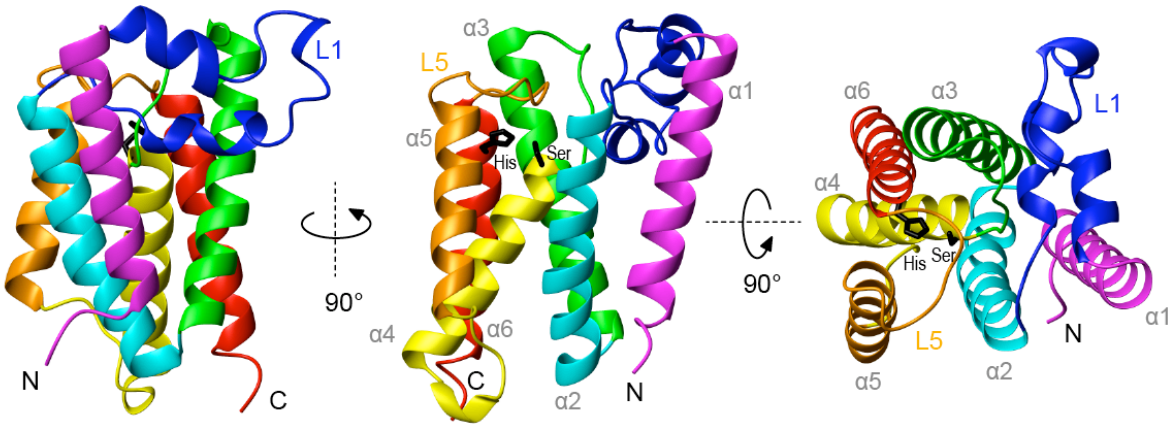


Figure 1.5 X-ray crystal structure of the *ec*GlpG rhomboid protease. The first published rhomboid structure is shown (PDB 2IC8) (65), consisting of the C-terminal fragment (residues 91 – 272) isolated by limited proteolysis with chymotrypsin. Active site dyad side chains are shown in black with the L1 loop and 6 transmembrane helices distinguished by different colours.

(13). Mutation of the candidate Asn residue in *Drosophila* Rho-1 suggested that this residue was essential for activity *in vivo* (13); however, *in vitro* activity assays of the *Bacillus subtilis* YqgP rhomboid provided conflicting evidence, suggesting that this Asn was not directly involved in catalysis (62). This issue was resolved by the *E. coli* GlpG rhomboid structures, since the conserved asparagine (N154) was not in the expected location proximal to H254 that would be expected for a catalytic triad residue. Instead N154 forms part of the oxyanion hole responsible for stabilizing the tetrahedral intermediate (72). Other parts of the oxyanion binding site are formed by the backbone amide of the active site serine (S201) and L200, and the side chain of H150 (Fig. 1.6).

1.2.3.3 Structural elements important for rhomboid activity

In addition to the water-exposed active site, the rhomboid crystal structures revealed a number of structural elements important for protease activity. First and most surprisingly, a lateral protrusion thought to be embedded in the membrane surface connects TM1 and TM2 (named L1), and appears to be a structural feature unique to the rhomboid family (Fig. 1.5). A number of residues in this loop are highly conserved, which is characterized by the consensus sequence E/QxWRxxS/T (25). Mutation of the arginine in this sequence was found to abolish activity (71, 74) while mutation of the tryptophan decreased activity in cell membranes, although not when reconstituted in detergents (75). The WR pair forms interactions within the L1 loop with the arginine alone forming five hydrogen bonds (65). These residues appear to stabilize the structure of this lateral protrusion while the surface facing the bilayer consists of hydrophobic residues that are thought to help position the rhomboid in the membrane. It was found that mutation of these hydrophobic residues to either serine or phenylalanine also decreased protease activity, further supporting a

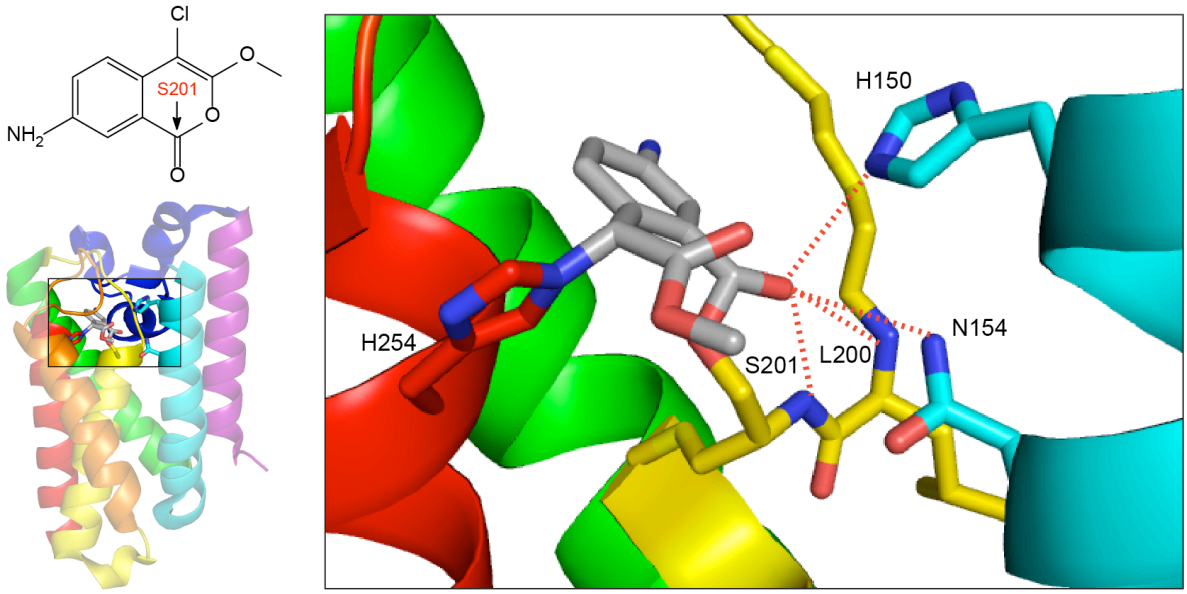


Figure 1.6 Active site of *ec*GlpG co-crystalized with the 7-amino-4-chloro-3-methoxy-isocoumarin inhibitor. The inhibitor is covalently bonded to the catalytic serine (S201) and histidine (H254) side chains. The inhibitor carbonyl (nucleophilic attack indicated by an arrow on the inhibitor, top left) is within hydrogen bonding distance of the S201 and L200 backbone amides as well as N154 and H150 side chains, forming the oxyanion binding pocket (hydrogen bonding interactions marked with dashed lines). TM helix colours are identical to Figure 1.5 with the inhibitor shown in grey for structure 2XOW (72).

membrane positioning role for L1 (74). The hydrophilic cavity is capped by loop 5 (L5), which seems to have flexibility based on a crystal structure with the core exposed to the aqueous environment through an "open" conformation of this loop (70). The structure of the isocoumarin-inhibited state also revealed significant displacement of the L5 cap away from the active site (72). It is expected that a similar conformational change is involved in polypeptide substrate entry, although this has yet to be shown experimentally.

Based on the crystal structure of *ec*GlpG rhomboid with an isocoumarin inhibitor and structural data from soluble serine proteases, Vinothkumar and coworkers modeled a substrate peptide into the GlpG structure to determine how it might interact (72). A model made with the TatA substrate tetrapeptide ($_{P2}$TA/AF$_{P2'}$) confirmed that residues in the P1 and P2' positions determined cleavage specificity. These side chains made extensive interactions with the rhomboid in the model, where the phenylalanine in particular fit well in a hydrophobic pocket of *ec*GlpG. In addition, side chains for amino acids in the P2 and P1' positions were oriented towards the solvent, explaining why residues at these positions do not influence cleavage efficiency. Similar results were obtained when a tetrapeptide representing the *Drosophila* Rho-1 substrate Gurken ($_{P2}$MAHI $_{P2'}$) was modeled into the *ec*GlpG active site, suggesting that divergent substrates bind rhomboids in a conserved manner.

1.2.3.4 Models for substrate entry into the active site

One issue that has not yet been resolved by the available library of rhomboid structures is the mode of substrate entry into the active site from its lipid bilayer environment. When the GlpG structure was initially published, it was suggested that L1 played a role in this process (65), although subsequent investigations on the effect of

mutations in this region did not support this idea (71, 74). More recent data suggests that substrate entry may occur through a gap between TM2 and TM5. This hypothesis arose from the variation in the position of TM5 in the various structures published of *ec*GlpG, with some structures suggestive of a "closed" conformation, and others of a more "open" conformation. In the most open state, TM5 was tilted 35° away from the core and was accompanied by changes in the active site cap formed by L5 (66). However, since these structures were solved in the absence of lipid membranes, these samples may have lacked the lateral compression that is normally imposed on a protein embedded in a bilayer environment (25). This would be expected to restrict the movement of TM5 *in vivo*, raising the possibility that the apparent flexibility of TM5 was simply an artifact of crystal packing. To address this issue, the Urban group created an extensive series of mutants in TM5, and evaluated them for function both *in vivo* and *in vitro*. They found that mutating large hydrophobic and aromatic residues at the TM2/TM5 interface into either valine or alanine led to increased rhomboid activity (74, 75). This increase in activity was most apparent for mutations closer to the side facing the active site. Conversely, double cysteine mutants in TM5 to TM2 that could be cross-linked between the two helices reduced the ability of these mutants to cleave a TM substrate when oxidized, with wild-type activity being restored after the addition of reducing agent (74). Taken together, this data provides strong evidence that lateral movement of TM5 facilitates substrate entry.

In addition to flexibility in TM5, substrate entry may be assisted by membrane distortions around the rhomboid. Early evidence for this model was provided by crystal structures showing detergent molecules in a hydrophobic belt around the rhomboid that was narrower than the hydrophobic core of a regular lipid bilayer (71). Molecular dynamics

simulations also suggest that the membrane bilayer directly surrounding the protein is about 10% thinner than the bulk membrane (76). This has given rise to a model for substrate entry where the TM substrate enters the distorted lipid bilayer immediately surrounding the rhomboid, destabilizing the TM helix as the end protrudes from the hydrophobic membrane core. This process would be facilitated by helix-destabilizing residues in the substrate sequence, allowing the peptide backbone to be exposed for nucleophilic attack following entry into the active site via the gap between TM2 and TM5 (Fig. 1.7).

1.2.4 Structure beyond the catalytic transmembrane core

1.2.4.1 General features of rhomboid extramembranous domains

While the rhomboid active site is located in its 6-TMH core domain, many rhomboids contain additional TM helices and/or soluble domains of various sizes at their N- and/or C-termini. The functional role of these soluble domains is generally not well understood, although it is thought that they may be involved in regulation of rhomboid activity and specificity. A thorough evolutionary analysis was published by the Pellegrini group who proposed that the near-universal presence of rhomboids in eukaryotes and prokaryotes reflects their origin in bacteria and subsequent spread to other species through horizontal gene transfer (22). Diversification in rhomboid domain structure would have occurred following these transfer events.

Phylogenetic analysis (22) also identified the presence of a number of soluble domains that appear to be rhomboid-type specific. Some rhomboids appear to contain domains that encode different types of zinc-fingers of unknown function. Diverse rhomboids of *Drosophila melanogaster*, *Anopheles gambiae* (mosquito) and *Homo sapiens*

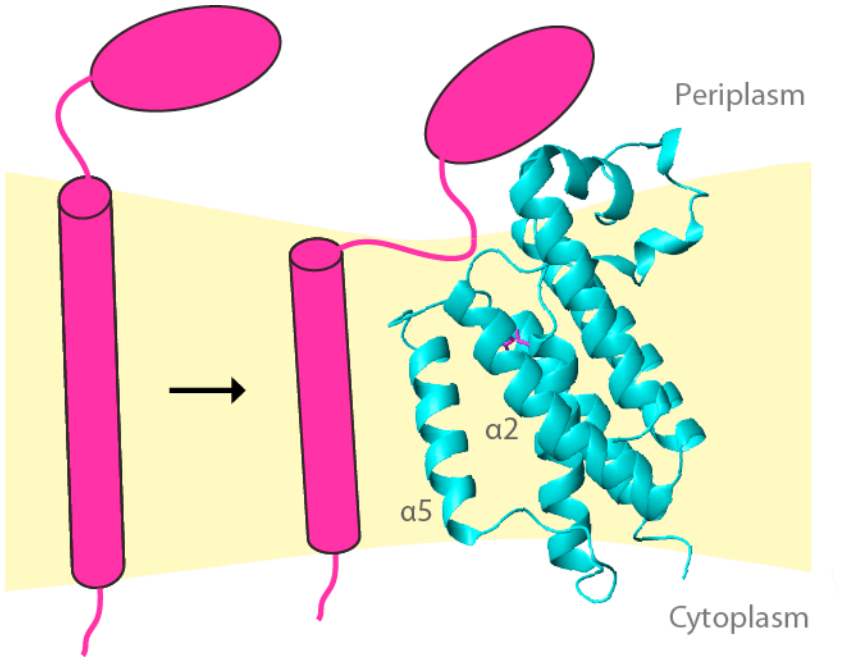


Figure 1.7 Proposed model of substrate entry to the active site. Analysis of detergent molecule localization in GlpG structures (71) and molecular dynamics simulations (76) suggest that the lipid bilayer is thinner in the region immediately surrounding the *ec*GlpG rhomboid protease (blue). Exposure of the end of the TM substrate (pink) to this distorted membrane could induce a conformational change that exposes the peptide backbone to proteolytic attack. Active site entry is believed to occur between TM2 and TM5.

were even found to contain one or two calcium binding EF hands, implying possible roles in signaling or calcium buffering. A rhomboid of the plant *Arabidopsis thaliana* is predicted to also contain a ubiquitin-binding domain, potentially reflecting a role in protein degradation. Although these examples provide some insight into the various soluble domains that can be found in rhomboids, most are still uncharacterized, having no sequence homologies to provide any insights into their functions.

1.2.4.2 Role of the RHBDL2 rhomboid cytoplasmic domain

From a functional viewpoint, the N-terminal cytoplasmic domain from the mammalian RHBDL2 rhomboid is one of the few rhomboid regulatory domains that has been characterized. Thrombomodulin, an anticoagulant cell-surface protein, was identified as a potential substrate for this rhomboid through analysis of the mouse genome (77) and *in vitro* assays of thrombomodulin cleavage by RHBDL2. This cleavage was found to be dependent on the presence of the thrombomodulin cytoplasmic domain. Moreover, fusion of this cytoplasmic domain to a non-substrate TM could convert it into a RHBDL2 substrate. In addition, deletion of the RHBDL2 cytoplasmic domain significantly reduced activity against thrombomodulin suggesting that the RHBDL2 cytoplasmic domain may interact with that from thrombomodulin. However, RHBDL2 is also involved in EGFR signaling (30), and can also cleave ephrin B3, another substrate involved in inter-cell communication (78). Since EGF and ephrin B3 do not possess anything homologous to the thrombomodulin cytoplasmic domain, this rhomboid may have multiple mechanisms for substrate specificity control.

1.2.4.3 The Pseudomonas aeruginosa rhomboid cytoplasmic domain

The only structural data that is available for a rhomboid domain outside of the

catalytic core comes from the bacterial *Pseudomonas aeruginosa* rhomboid (*pa*GlpG). This rhomboid has a structured N-terminal cytoplasmic domain (CytD) that was solved through solution nuclear magnetic resonance (NMR) spectroscopy (79) (Fig. 1.8a). The CytD forms a mixed β-strand/α-helix fold comprised of a three-stranded β-sheet and two α-helices. A number of bacterial rhomboids possess a similar N-terminal soluble domain, with no sequence homology to other known proteins. However, the NMR structure showed weak homology with the N-terminal domain of EscJ, an *E. coli* protein that helps form a sensory probe involved in pathogenesis (80). Covalent modification of EscJ with a lipid group leads to its localization to the bacterial inner membrane (81), prompting the Ghose group to hypothesize that the *pa*GlpG CytD also has a lipid binding role. To test this hypothesis, a lipid-like zwitterionic detergent, hexadecylphosphocholine (Fos16), was titrated into a solution containing the CytD, and the effect on the local chemical environment of CytD residues was monitored by NMR (79). Based on changes in the spectrum, Del Rio and coworkers identified a number of amino acids that appeared to interact with Fos16, all of which mapped to the β-sheet region of the CytD. Surprisingly, they also found that many residues involved in the interaction had buried side chains, suggesting that a significant conformational change was required for binding to occur. Based on these findings, it was proposed that the *pa*GlpG rhomboid CytD could bind to the lipid bilayer and position the transmembrane active site for optimal activity (Fig. 1.8b). However, if the L1 loop already plays a role in active site positioning in the membrane it is difficult to understand why an additional positioning domain would be required. In addition, the other rhomboid of known structure, *hi*GlpG of *H. influenzae* (67), does not contain a cytoplasmic domain in its sequence, suggesting that the CytD could have an alternate role in rhomboid function.

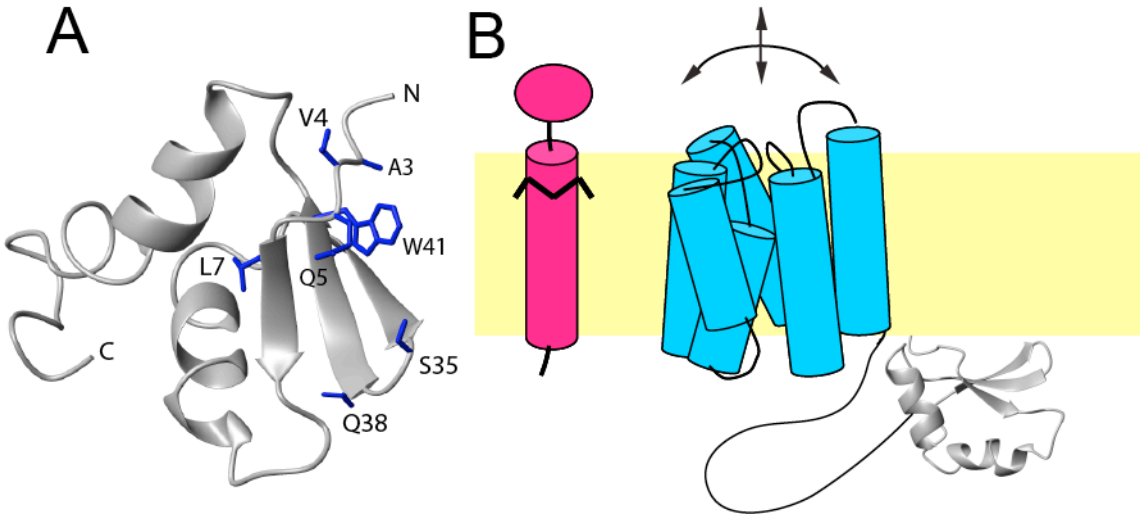


Figure 1.8 Proposed membrane interaction of the *pa*GlpG CytD. (A) Solution NMR structure of the CytD (PDB: 2GQC, (79)) with side chains from residues that interact with hexadecylphosphocholine (Fos16) indicated in blue (79). (B) Schematic diagram of the proposed membrane interaction of CytD. CytD could bind the membrane and position the active site within the bilayer to promote proteolysis of the substrate (pink). The structure of the *pa*GlpG TMD (blue) has not been solved and is represented as a model.

Since crystallization of the rhomboid was only possible if N-terminal cytoplasmic residues were removed by limited proteolysis, there is no structural data on a full-length rhomboid. However, during the development of crystallography samples it was found that the isolated TMD had lower activity compared to the full-length rhomboid (65). This raises questions regarding how the cytoplasmic domain could be involved in protease activity, particularly considering its cytoplasmic location far from the active site. Although the L1 loop is proposed to have an essential role in membrane positioning, is it possible that the cytoplasmic domain binds to the lipid bilayer for a similar type of membrane positioning role? Or could the cytoplasmic domain have an alternative regulatory role? Little is known about this domain, and in particular if it associates with the TMD (or other proteins), and for this reason is the primary focus of my thesis. More specifically, this thesis describes my investigation of the potential role that this domain may play to regulate the activity of this class of bacterial rhomboid.

1.2.5 Regulation

While a number of aspects of rhomboid structure and function have become well understood over a relatively short period of time, little is still known about how these proteases are regulated. In the case of soluble proteases, regulation is often achieved by expression of a proenzyme form that must be proteolytically processed for activation. Alternatively, aspartyl and metallo intramembrane proteases both require their substrates to be cleaved by another protein prior to recognition as a substrate (23). In contrast, rhomboids appear to be constitutively active, yet do not require any pre-processing of their substrates, raising questions about whether rhomboid activity is regulated in any way. Although a

unified theme of rhomboid regulation has not been identified, partly due to the diversity in rhomboid function, some ideas regarding regulation have been forwarded for a small number of systems.

1.2.5.1 Modulation of PARL function

In a unique form of regulation, mammalian PARL can undergo two processing events (α and β-cleavage) in a long loop that connects the extra N-terminal helix to the 6-TMH core. Specifically, α-cleavage is constitutive and occurs during import into the mitochondria to produce mature PARL, whereas β-cleavage removes part of the N-terminus and depends on PARL phosphorylation. This β-cleavage event activates PARL and results in altered mitochondrial morphology (82). When cells were transfected with truncated PARL with a sequence that started at the β-cleavage site rather than the full-length protein, it caused the cells to have a fragmented mitochondrial morphology. Surprisingly, the cleaved N-terminal peptide was also found to traverse the mitochondrial membranes and localize to the nucleus in what is described as mitochondrial retrograde signaling (MRS) (83). MRS allows the nucleus to monitor and coordinate structural and functional events in the mitochondria; the possibility that the rhomboid is involved in signaling through release of part of itself is a very interesting concept that appears to be unique to PARL. The role of PARL in MRS requires further investigation to determine the requirements for this type of signaling, but it appears to provide one role for an extra rhomboid structure element outside the catalytic core.

1.2.5.2 Regulation through expression and/or compartmentalization

The best characterized mechanism of rhomboid regulation occurs at the level of rhomboid expression and/or compartmentalization. Even before the rhomboid was

identified as an intramembrane protease, expression patterns of Rho-1 in *Drosophila* embryogenesis were found to be complex, revolving around the EGFR pathway (39, 84). Regulation through compartmentalization in the cell was later uncovered with Rho-1 localizing to the Golgi apparatus (Figure 1.4a) while its substrate, Spitz, is retained in the ER. Therefore, Spitz can only be processed when it is allowed to move from the ER to the Golgi, an event that requires the Star trafficking protein as an escort (85). Another example of regulation through compartmentalization is seen in *T. gondii* parasite invasion mediated by TgROM5. The substrates for this rhomboid are adhesins that are normally localized to internal organelles. When these organelles fuse with the cell surface, the adhesins localize to the plasma membrane where TgROM5 is also found, leading to the adhesin cleavage that is required for parasite invasion to proceed (33).

1.2.5.3 Prokaryotic rhomboid regulation

Although physical separation of enzyme and substrate could be a common form of regulation in eukaryotic organisms, this is unlikely to be the case for prokaryotes. A more plausible form of regulation would be at the transcriptional level (although this has not been shown experimentally for bacteria), or through changes in the membrane environment. In a study by Urban and Wolfe, four bacterial rhomboids were reconstituted in lipid vesicles of various compositions and assayed for activity (63). The lipid environment was found to have either a stimulatory or an inhibitory effect on rhomboid activity, depending on the protease assayed. For example, the *P. stuartii* AarA protease displayed low activity when assayed in the non-ionic detergent, dodecylmaltoside (DDM); however, reconstitution in thirteen different lipid environments resulted in a stimulatory effect for each. Conversely, the same lipid environments had an inhibitory effect on *ec*GlpG with the exception of

phosphotidylethanolamine, which seemed to maintain the same activity found in DDM. Reconstitution of *ec*GlpG in *E. coli* lipid extract, surprisingly, seemed to abolish protease activity altogether. These results imply that the lipid environment has an important role in bacterial rhomboid function (63), although it should be noted that the proportion of protein that was successfully reconstituted was not measured in this study. Consequently, some of the activity differences could reflect differences in reconstitution efficiencies. Nonetheless, if the lipid environment has a role in protease activity, it could help explain how a membrane-embedded active site could be regulated in prokaryotes.

1.2.6 Research focus and objectives

As mentioned above, one potential mode of rhomboid regulation could be through interactions between the catalytic TMD and the extramembranous domains of these proteins. An example may be provided by RHBDL2, whose cytoplasmic domain appears to have a role in activity against thrombomodulin (77). However, little is known about the function of N- and/or C-terminal soluble domains possessed by many rhomboid proteases. Structural insight into the role of these domains has also been limited since these domains are removed from the rhomboid to make them amenable for x-ray crystallography (64-66). Given the evidence that the *ec*GlpG cytoplasmic domain appears to be functionally important (65), and that little is known about how rhomboids are regulated in general, we have sought to elucidate the role of the cytoplasmic domain in bacterial rhomboid function through the following main objectives:

1) Assess the full length rhomboid for structural characterization in solution.
2) Solve the structure of the isolated rhomboid cytoplasmic domain.

3) Characterize interactions that involve the cytoplasmic domain and catalytic transmembrane domain.

This thesis describes my work toward these goals, using a combination of biochemical assays of rhomboid function along with structural studies. Part of this work relies heavily on structural analysis of the cytoplasmic domain using nuclear magnetic resonance spectroscopy, the theory and application of which will be described in the following sections.

1.3 Nuclear Magnetic Resonance Spectroscopy (NMR)

1.3.1 Basic principles of solution NMR

Nuclear magnetic resonance (NMR) spectroscopy is a powerful tool in structural biology with a unique ability to study both macromolecular structure and dynamics. NMR spectroscopy is most commonly considered to be an alternate method of structure determination for proteins that cannot be studied by x-ray crystallography (86, 87). Less appreciated is the fact that these two methods are highly complementary, and can reveal important features of a system when used together (87). For example, NMR can be used to uncover functionally relevant conformational dynamics of a macromolecule in solution, while x-ray crystallography can be used to determine its structure to high precision (86-88). The complementary nature of these methods is also exemplified by the *ec*GlpG rhomboid protease which could not be crystallized until the cytoplasmic domain was removed (64-66, 70, 73). However, the small size of the cytoplasmic domain makes it an excellent candidate for solution NMR spectroscopy as both an isolated domain and potentially as part of the full-length protease.

NMR is a technique that uses the intrinsic magnetic property of nuclei that have spin angular momentum (I) (89). Experiments typically utilized in protein structure determination use ^{1}H, ^{15}N, and ^{13}C, all of which have spin I = ½. When these isotopes are placed in an external magnetic field (B_0) as is provided by an NMR spectrometer, a dipole moment is induced. Each dipole adopts one of two orientations with respect to the external magnetic field, separated by a small difference in energy (Fig. 1.9). A radiofrequency pulse with energy that is equal to the difference between spin states, called the resonance or Larmor frequency, is the frequency at which absorption occurs and leads to transitions between states. This frequency reflects contributions from the local chemical environment surrounding the nucleus, and is reported as a chemical shift in parts per million (ppm) from a standard frequency. Transitions are detected for the small excess of nuclei populating the low energy state (as determined by the Boltzmann distribution). The size of this population difference is related to the energy difference between states; since the energy difference between these two states is small, the population difference is also small, making NMR an insensitive technique (90, 91).

Sensitivity is also affected by the size of the molecule being analyzed, with larger molecules having spectra with decreased signal-to-noise relative to their lower molecular weight counterparts (90, 91). This is due to slower molecular reorientation rates of larger molecules. Molecular tumbling creates an environment of fluctuating fields caused by the dipoles carried by these molecules. Some of these fluctuations will occur at frequencies that can induce transitions between states. The population of oscillators that have these transition-inducing frequencies depends on the molecular reorientation rates of these molecules. The larger the molecule, the larger the population of frequencies that will induce

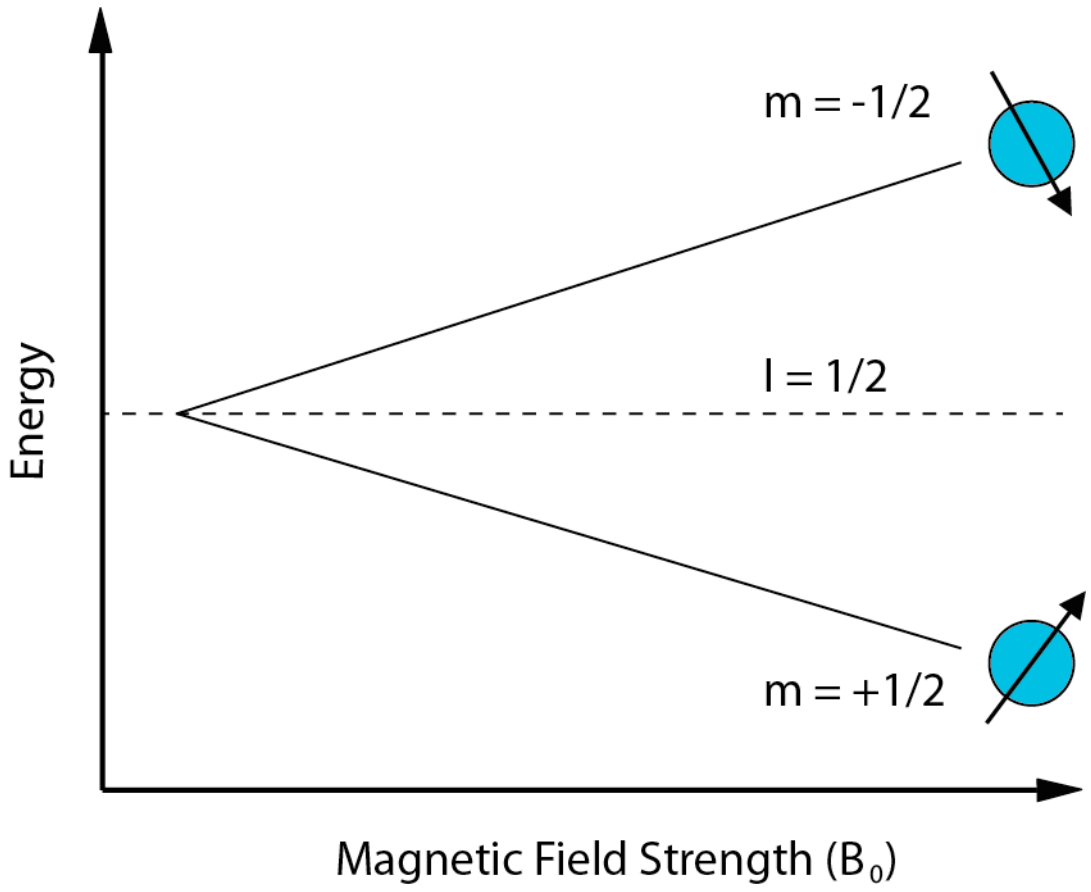


Figure 1.9 Energy level splitting of spin ½ nuclei in an external magnetic field (B_0). For proton spins, the orientation parallel to B_0 is lower in energy while anti-parallel to B_0 is higher in energy. The energy difference (ΔE) is proportional to the strength of B_0 and radiofrequency radiation (v) with $\Delta E = hv$, which can be used to stimulate transitions between states (h is Planck's constant). The population of nuclei in each spin state is defined by the Boltzmann distribution, $N_{high}/N_{low} = e^{-\Delta E/kT}$, where the population of the lower energy spins (N_{low}) slightly outnumbers that of the higher energy spins (N_{high}) at room temperature. (k is the Boltzmann constant, T is the temperature.)

transitions that dephase magnetization. Relaxation also reduces the efficiency of coherence transfer elements in multidimensional experiments, causing additional loss in signal. Consequently, the sensitivity and resolution of NMR spectra obtained from macromolecules such as proteins are significantly reduced compared to small organic compounds. For these reasons, high sample concentrations (~ 0.5 mM or more) are required for the acquisition of protein NMR spectra. The triple-resonance experiments routinely acquired for structure determination also require that protein samples be uniformly labeled with ^{15}N and ^{13}C, spin ½ nuclei (92).

Compared to simple organic molecules, proteins contain a large number of protons with resonances that cannot be entirely resolved in a one-dimensional proton NMR spectrum. To solve this problem, multidimensional NMR experiments have been created that improve the resolution of these resonances while providing informative chemical shift correlations between spins that are connected by one or more bonds, or are close together in space (93, 94). The most widely used type of NMR experiment that is used for this purpose is the two-dimensional heteronuclear single quantum coherence (HSQC) experiment (95).

1.3.2 The heteronuclear single quantum coherence (HSQC) experiment

The ^{15}N-HSQC spectrum correlates the amide proton chemical shift with that of its directly attached nitrogen (95) (Fig. 1.10). It is widely considered to be a protein "fingerprint" spectrum since each residue except for proline is represented by one backbone peak in the spectrum. Peaks are also observed from Trp, Asn and Gln side chain ^{1}H-^{15}N groups, sometimes along with Arg and Lys side chains, depending on the conditions being used. The ^{15}N-HSQC is also used to evaluate structural integrity since a well-folded protein

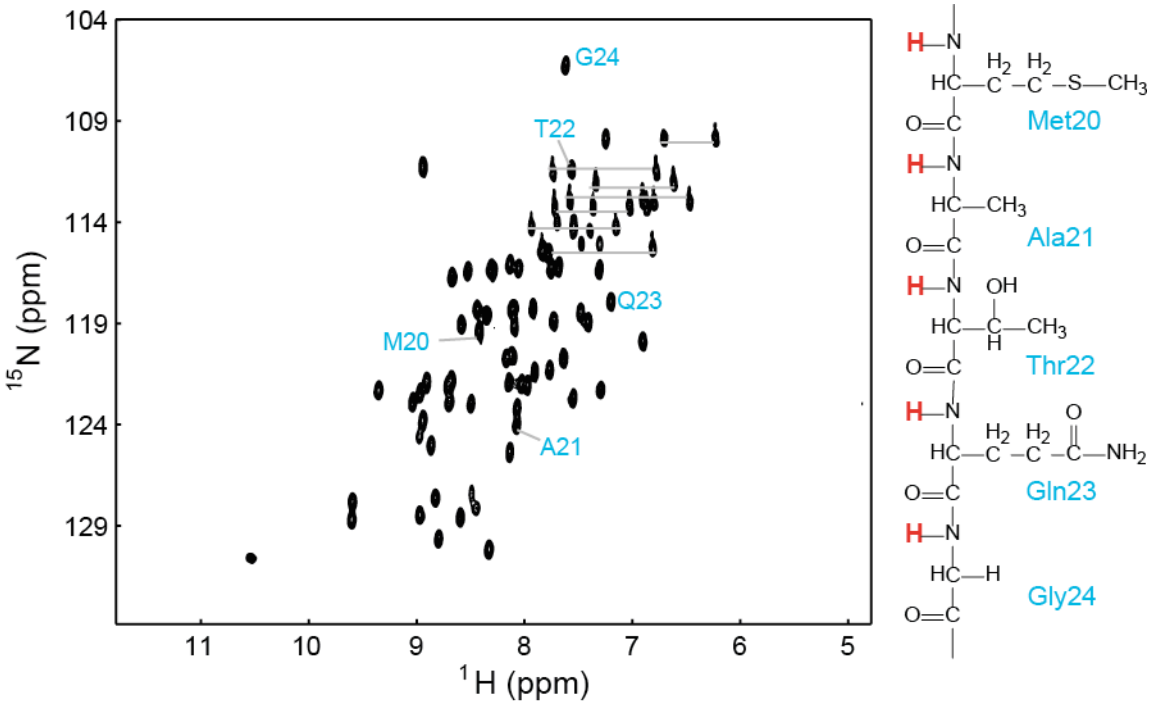


Figure 1.10 The ^{1}H-^{15}N Heteronuclear Single Quantum Coherence (HSQC) experiment. Resonances arising from the amide proton are separated in the nitrogen dimension to provide this characteristic 2D protein fingerprint spectrum. The example shown here is from the 8 kDa ^{15}N-labeled *ec*GlpG cytoplasmic domain with five amide backbone resonances indicated for simplicity. Amide side chain resonances are connected by the horizontal lines.

will display a wide chemical shift dispersion in the ^{1}H-dimension. In contrast, the backbone amide groups in an unfolded protein are highly solvent-exposed, giving rise to less variation in local chemical environment and therefore a narrower range of proton chemical shifts.

Residues involved in protein-protein or protein-ligand interactions can be identified using the ^{15}N-HSQC experiment in a process called chemical shift mapping (reviewed in (96)). Since the position of each peak reflects the local chemical environment experienced by an amide proton, these positions can change for amide protons that experience a change in local chemical environment upon interaction with an unlabeled binding partner (97). How these binding interactions affect the spectrum is also governed by the rate of complex dissociation, and how it compares to the frequency difference between peaks. As shown in Figure 1.11, an interaction with a slow rate of dissociation (ms - s) will give rise to two different peaks for each residue under sub-saturating conditions; one representing the bound state and the other the free state. In contrast, nuclei involved in a rapid exchange between free and bound states (ps - ns) will appear as an averaged signal where the chemical shift is weighted by populations in the bound and unbound states. If the interaction occurs on an intermediate time scale (i.e. the frequency difference between peaks is comparable to the frequency of exchange) then the peak will be broadened, appearing as a decrease in peak intensity. The slow timescale of exchange may have been relevant in the case of the *pa*GlpG cytoplasmic domain interaction with Fos16, since some peak intensities decreased as detergent was titrated into solution (79).

1.3.3 Chemical shift assignment

To perform chemical shift mapping or NMR structure determination it is necessary

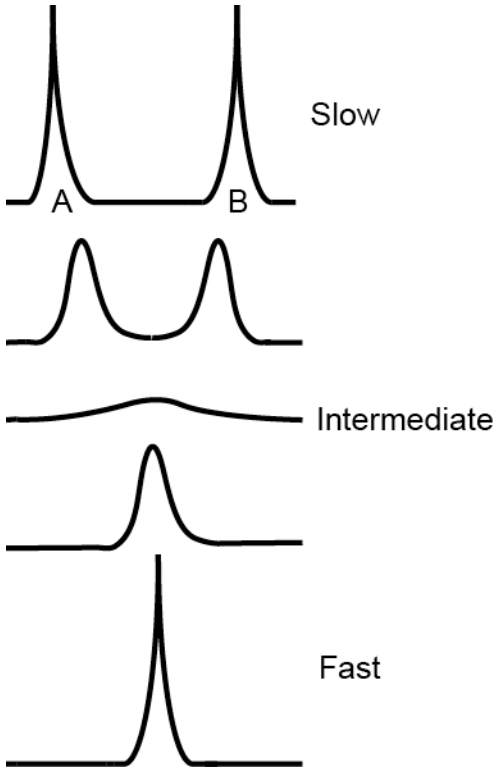


Figure 1.11 The effect of chemical exchange in NMR spectra. Slow exchange has a ms – s rate of interconversion, allowing both states (A and B) to be detected at distinct positions in the spectrum, while fast exchange is characterized by ps – ns interconversion times, giving rise to a single peak at a population weighted average of the chemical shift of the two species. Intermediate exchange is characterized by a rate of exchange that is comparable to the frequency difference separating the two states, and gives rise to peak broadening and coalescence of the peaks from the two exchanging species.

to assign chemical shifts using a series of three-dimensional experiments. One of the simplest of these experiments is called the HNCO which correlates the amide proton and nitrogen chemical shifts to that of the carbonyl carbon (CO) of the residue immediately preceding it (98) (Figure 1.12). This 3D experiment can be visualized as a cube where an end-on view of the ^{1}H-^{15}N dimension would resemble an ^{15}N-HSQC spectrum. However, each peak in this ^{15}N-HSQC will have a different position along the z-axis according to the chemical shift of the carbonyl carbon to which it is attached (Fig. 1.13). This experiment can help to determine if there are overlapping peaks in the ^{15}N-HSQC since they can be resolved in the carbon dimension. In addition to backbone assignment, carbonyl carbon chemical shifts are also important for secondary structure prediction (99).

Key experiments used in the process of protein backbone chemical shift assignments are the CBCA(CO)NH (100) and the HNCACB (101), where chemical shifts are recorded for the Cα, Cβ and amide ^{1}H and ^{15}N atoms (Fig. 1.12). However, in the CBCA(CO)NH experiment the amide proton and nitrogen shifts of each residue (position i), are only correlated with the alpha and beta carbon atoms of the preceding residue (i-1). The HNCACB adds an additional set of correlations by connecting the Cα and Cβ chemical shifts for both residues i (intra-residue correlation) and i-1 (inter-residue) to the amide shifts of residue i (Fig. 1.12). Consequently, for most residues the HNCACB will contain four peaks in the carbon dimension at each backbone amide ^{1}H – ^{15}N peak position, while the CBCA(CO)NH experiment will only contain two peaks at this position. Identification of an amide with inter-residue Cα and Cβ correlations that are identical to intra-residue Cα/Cβ correlations of another residue would mean that they are sequentially located in the protein sequence. The range of Cα and Cβ shifts characteristic of each amino acid (102) are then

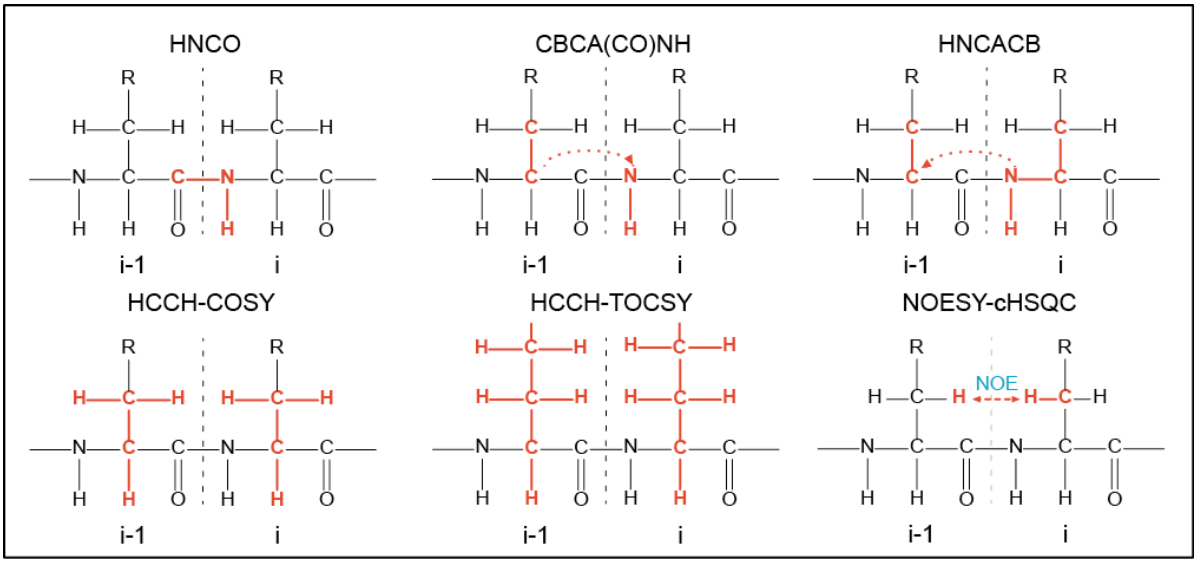


Figure 1.12 Correlations provided by 3D NMR experiments used in this thesis. The experiment name indicates the nuclei involved in magnetization transfer that would consequently be correlated in the experiment (shown on the structures in red). With respect to the HCCH-COSY, magnetization transfer occurs over one ^{13}C-^{13}C bond with correlations made to attached protons only, while the HCCH-TOCSY will detect correlations throughout the whole side chain.

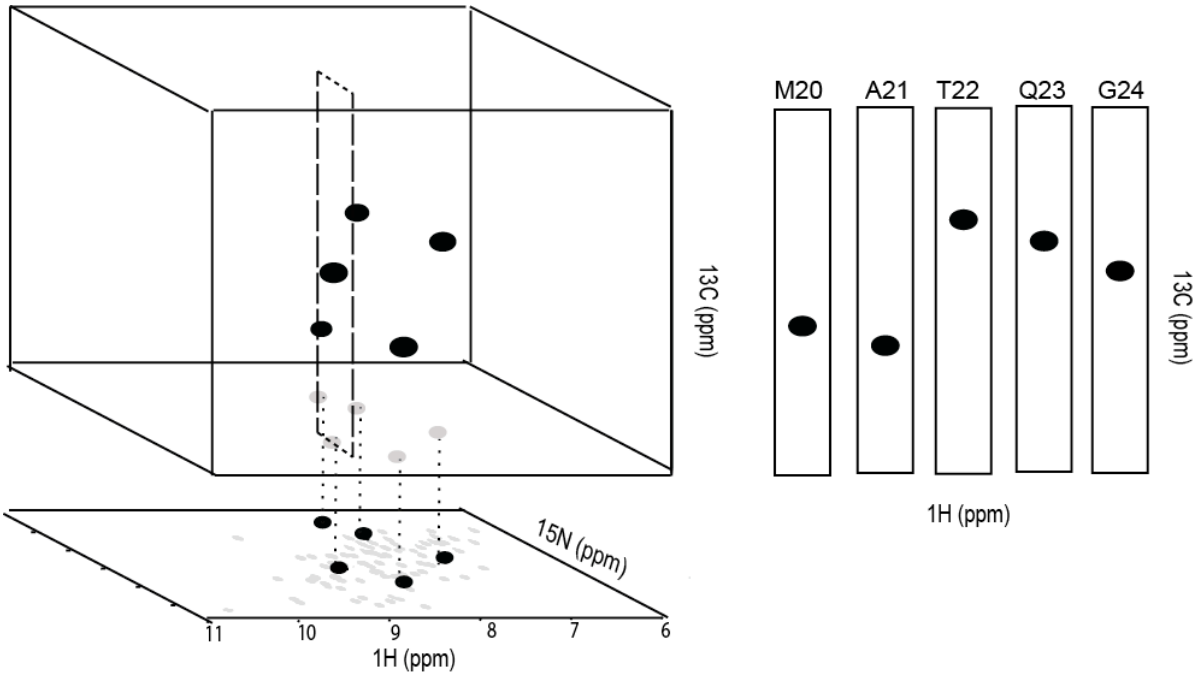


Figure 1.13 Overview of the HNCO experiment. Correlations are obtained between the amide proton, its directly attached nitrogen and the carbonyl carbon of the preceding (i-1) residue. Resonances of the corresponding HSQC are separated in the carbon dimension to obtain i-1 carbonyl chemical shifts. 3D experiments are analyzed as strips (right) taken from the position in the ^{15}N dimension where the peak is found.

used to determine the residue in the protein sequence that is associated with those chemical shifts. Using this process, usually with the assistance of specialized software (103-106) it is possible to identify the residues in the protein sequence that give rise to each backbone amide peak in the ^{15}N-HSQC (Fig. 1.14).

Assignment of side chain chemical shifts to the ^{1}H-^{13}C-HSQC spectrum follows a different approach that principally relies on two types of experiments: Total Correlation Spectroscopy (TOCSY, (107)) and Correlation Spectroscopy (COSY, (108)). In these experiments magnetization is transferred between aliphatic proton and carbon atoms along the amino acid side chain (Fig. 1.12). In the case of the TOCSY, each side chain proton is correlated to all other aliphatic protons or carbon atoms in the same side chain, depending on the version of experiment that is being run. Since this can often give rise to a crowded spectrum with significant peak overlap, the COSY must also be used to simplify analysis. This experiment correlates vicinal side chain protons or directly bonded carbon atoms (depending on the version being run); essentially a subset of the correlations observed in the TOCSY. Through analysis of chemical shifts and connectivities established in these complementary experiments, it is possible to assign peaks in the ^{13}C-HSQC to specific protons in the protein.

1.3.4 Structure determination by NMR

To determine a structure by NMR, chemical shift assignments must be used to identify atom pairs that undergo a Nuclear Overhauser effect (NOE) (109). This is a dipolar mechanism of magnetization transfer that can only occur between protons that are within ~5 Å of each other. These are measured in Nuclear Overhauser effect spectroscopy (NOESY)

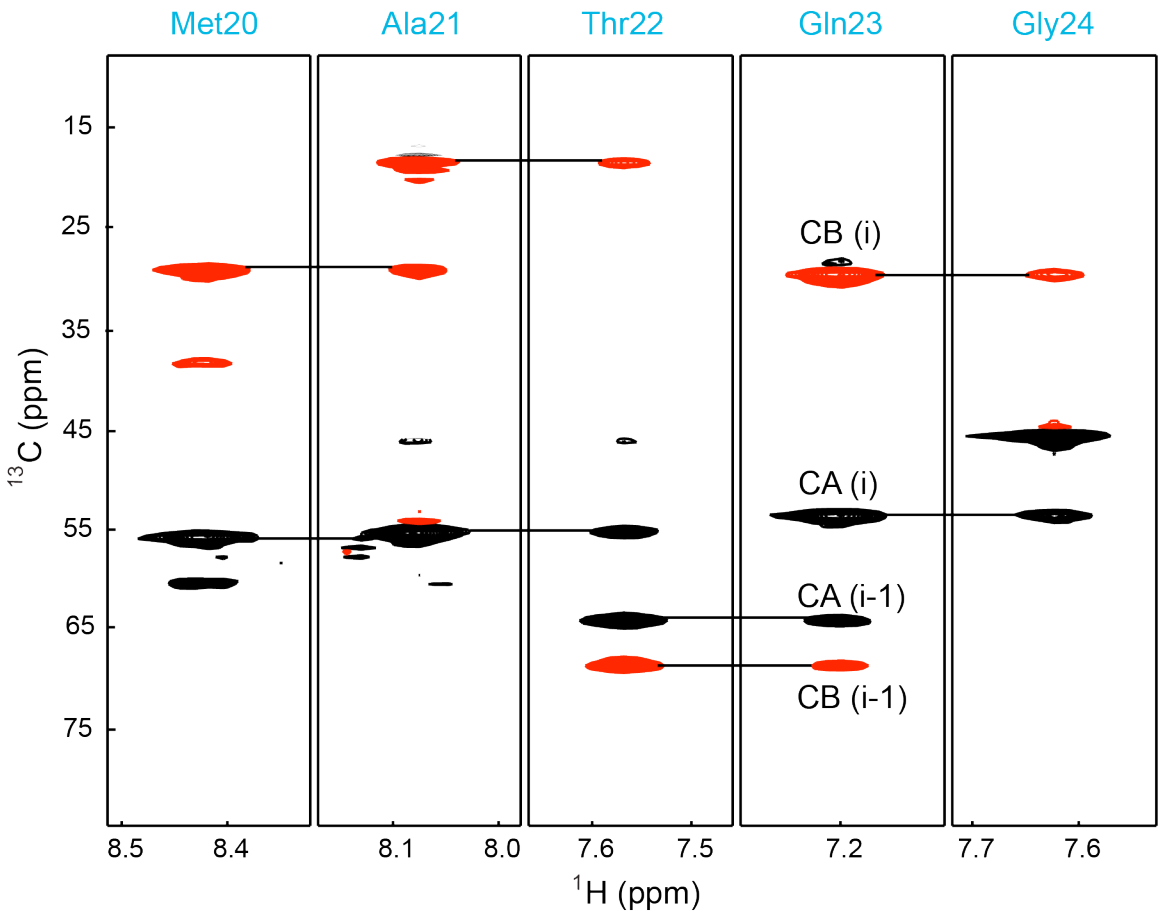


Figure 1.14 Sequential assignment of the amide backbone using the HNCACB experiment. Sample ^{13}C-^{1}H strips of the HNCACB spectrum centered about the amide proton and nitrogen chemical shifts of the residue indicated above each strip. Intra- and inter-residue correlations are shown for the peaks arising from Cα (black) and Cβ (red) atoms, with peaks from intra-residue correlations being more intense than those from inter-residue correlations. Matches between inter-residue peaks and intra-residue peaks in the strip from the preceding residue are indicated by black lines between strips.

experiments that can be filtered to selectively observe NOEs involving amide protons (^{15}N-NOESY-HSQC, Fig. 1.15) or aliphatic protons (^{13}C-NOESY-HSQC, Fig. 1.12) (110, 111). These interacting atom pairs are then used as one of the primary forms of input for programs that calculate structure. The program Cyana is particularly useful for this purpose since, in addition to structure calculation, it can also increase the list of assigned NOEs through an automated process that assigns NOEs that are consistent with the initial structure generated by the manual assignments (112) (Fig. 1.15). For structure calculation, Cyana uses torsion angle molecular dynamics in combination with simulated annealing – a process that subjects a random extended structure to a high temperature followed by gradual cooling of the system in order to computationally ease the structure into a global energy minimum. Since this calculation process can lead to local structural defects such as unsatisfied hydrogen bonding and non-optimal packing, 10% of the lowest energy conformers are subjected to additional refinement, often done in explicit water, typically using a program called XPLOR-NIH (113). This refinement process can significantly improve structure quality providing a more accurate representation of the structure.

The final group of refined lowest energy structures is generally referred to as the NMR ensemble. Members of this ensemble will agree well with each other when a sufficient number of structural restraints are included in the calculation. The rmsd value provides a good indication of ensemble precision where a well-defined structure should have a backbone and heavy atom rmsd of < 0.5 Å and < 1 Å, respectively (114). Structure quality analyses of the ensemble using programs like AQUA, PROCHECK (115, 116) and WHAT IF (117) can also give an indication of model accuracy based on distance and angle violations. When necessary, it is possible to improve the quality of a structure by using

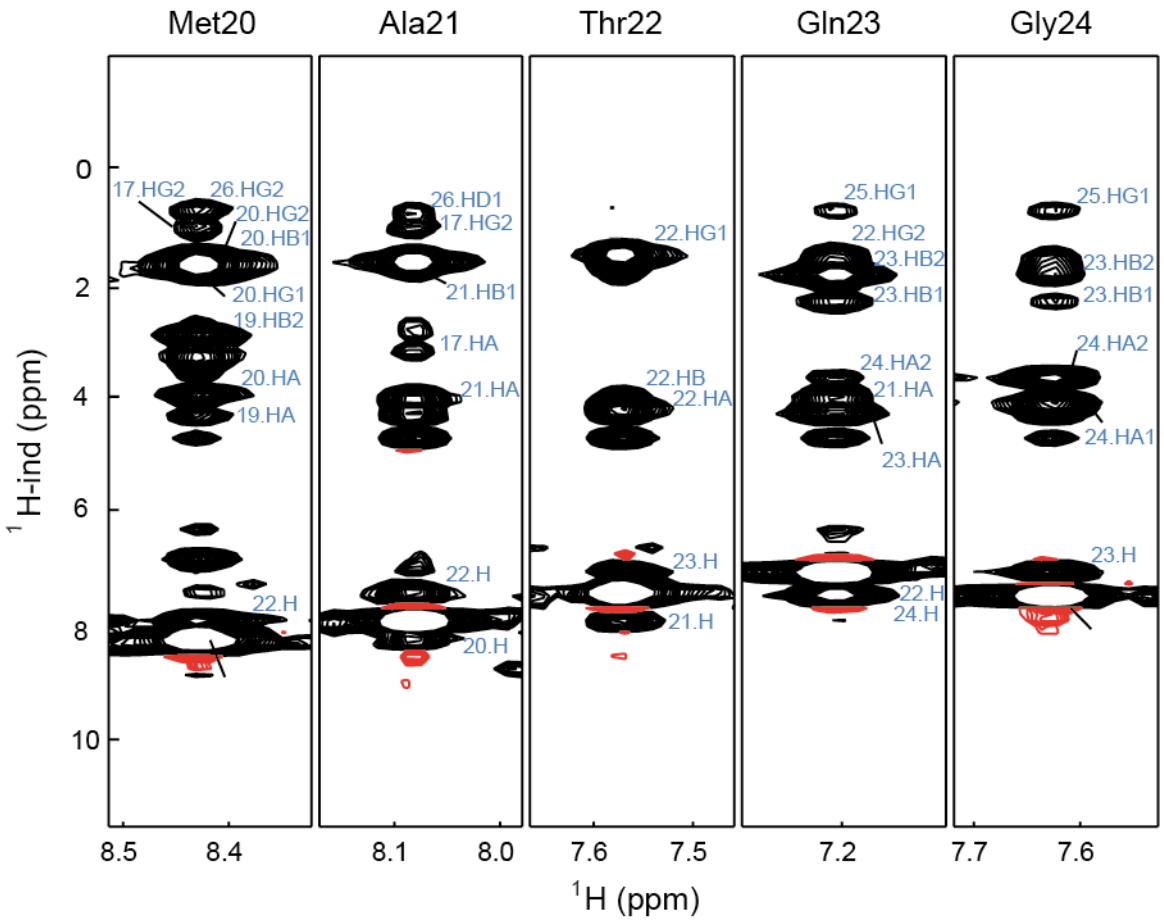


Figure 1.15 Through-space distances between nuclei are estimated using peak intensities in NOESY spectra (^{15}N-HSQC-edited version shown in this example). Once side chains have been assigned, these chemical shifts can be used to manually assign the NOESY spectrum. Additional NOEs assignments can be subsequently obtained with the structure calculation and automated NOE assignment program Cyana (112). Cross peak intensities are used to estimate the distance between each assigned nucleus and the amide proton from the corresponding strip.

additional types of restraints such as residual dipolar couplings (118, 119) and paramagnetic relaxation enhancement-based restraints (120, 121).

1.3.5 **Membrane proteins and NMR spectroscopy**

This thesis describes the structure as determined by solution NMR of the *E. coli* rhomboid cytoplasmic domain, to help understand its role in rhomboid activity. This domain is an excellent candidate for study by solution NMR spectroscopy because it is small (8 kDa) and highly soluble. However, there is also considerable interest in studying the structure of the full-length rhomboid, a prospect that was also explored in this thesis. However, NMR of the full-length rhomboid presents an additional challenge as it requires solubilization in a membrane mimetic detergent to ensure that the protein is properly folded and in an active conformation. This can give rise to a protein-detergent complex with very slow molecular reorientation rates and therefore low spectral sensitivity and resolution. Consequently, it is important to screen a number of membrane mimetic systems to identify conditions that give rise to the most favorable NMR spectral characteristics (122-125).

1.3.5.1 The detergent micelle

The detergent micelle (depicted in Figure 1.16) is the most commonly used membrane mimetic for NMR analysis of membrane proteins (124). Although the micelle is not comprised of phospholipids normally found in biological membranes, it is a convenient medium for solution NMR applications owing to its relatively small size. Micelles are a dynamic assembly of detergent molecules that organize to form roughly spherical complexes with polar headgroups facing the aqueous phase that help to minimize exposure of

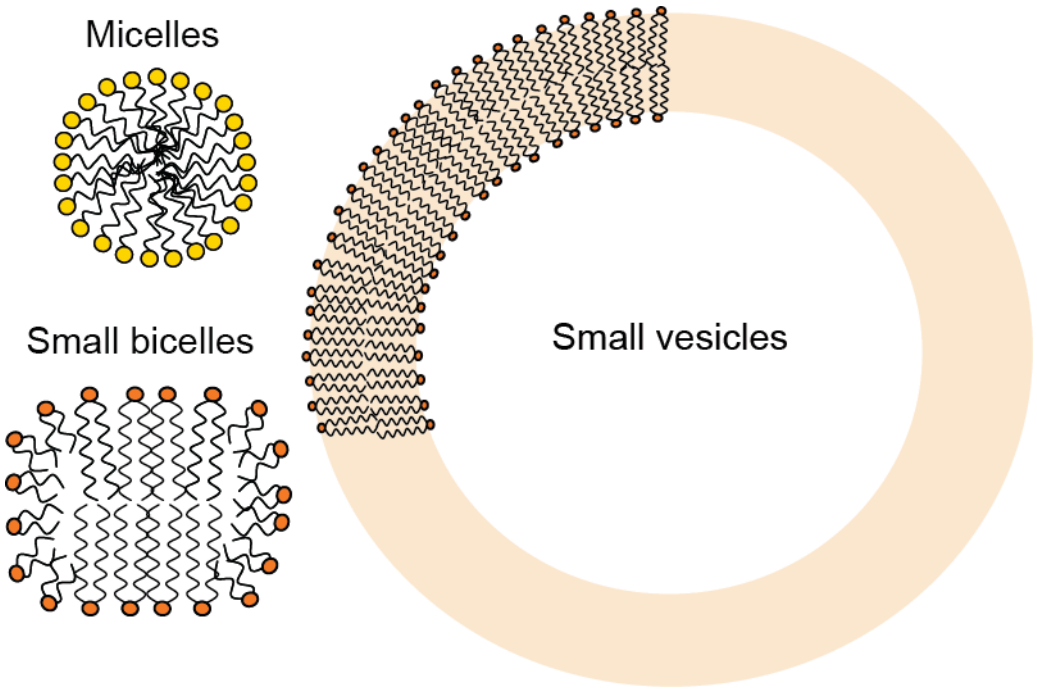


Figure 1.16 Schematic diagram of micelle, bicelle and vesicle environments. Acyl chains are shown as lines and hydrophilic headgroups are shown as yellow circles for detergents and orange circles for lipids. Figures are not to scale.

hydrophobic tails to water (126). In the presence of membrane proteins, it is thought that detergent molecules aggregate around the hydrophobic regions of the TM domain to reduce their exposure to the aqueous phase and create a soluble protein-detergent complex (PDC) (122).

Detergent selection is one of the most critical aspects of membrane protein study using solution NMR (123). Ideally, the detergent should be mild enough to maintain the properly folded conformation of the protein in a water-soluble form for the extended periods necessary to run NMR experiments. In addition to this requirement, it must support high protein concentrations, yet form relatively small complexes to avoid line broadening caused by slow tumbling in solution (127). Ionic detergents are typically thought of as excellent solubilizers but are also the most denaturing because they can disrupt protein-protein interactions (125, 128). Polar detergents, on the other hand, are considered to be the least denaturing and are often used to isolate active membrane proteins since they can disrupt protein-lipid interactions without disrupting protein-protein interactions (125, 128). The zwitterionic detergents appear to have intermediate qualities between the other two detergent types (125).

1.3.5.2 Small isotropic bicelles

Although detergent micelles have been successfully used for structure determination of many membrane proteins, many physical characteristics of the micelle are significantly different from that of a lipid bilayer. Specifically, micelles have strong curvature atypical of the membrane bilayer (129), and are more permeable to solvent water (122). If these differences significantly affect a membrane protein structure it is possible to use a more native-like membrane mimetic system that is still compatible with solution NMR known as a

small isotropic bicelles (130, 131) (Fig. 1.16). This type of bicelle is composed of long chain phospholipids that form a bilayer phase with shorter chain phospholipids/detergents coating its hydrophobic edge (132, 133). Bicelle size can be controlled by the ratio of long-chain and short-chain lipid, known as the q-value (where q = moles of long-chain lipid / moles of bicellar short-chain lipid). So long as the q-value is approximately < 1 the bicelles that are formed will tend to tumble isotropically in solution, allowing high resolution spectra to be obtained for bicelle-associated proteins (134).

Small bicelles are discoidal in shape (135), consistent with previous studies showing that the native-like bilayer phase is preserved (130, 136). This more native-like environment for membrane proteins can help maintain them in a stable, folded form. For example opsin, a protein that is unstable in detergent micelles, has been shown to be both correctly folded and active once purified in bicelles ($q=0.67$, $q_{eff}=0.98$) (137). The *Staphylococcus aureus* small multidrug-resistance pump is another protein that can be functionally reconstituted in small bicelles, allowing chemical shift assignments to be obtained for this state (124, 138).

1.3.5.3 Advances in membrane protein solution NMR

In spite of these advances in the development of suitable membrane mimetic systems, solution NMR of polytopic α-helical membrane proteins remains a challenging prospect. However, advances in spectrometer hardware (139), isotope labeling schemes (92, 140, 141), NMR pulse sequence development (142, 143) and data acquisition techniques (144) have made a large impact on the number of membrane protein structures solved through solution NMR. To date, NMR has been used to solve approximately 12% of the known integral membrane protein structures (145). In addition 24 of the 30 solved by NMR are α-helical with two or more TM helices. Although many of these are of homooligomers

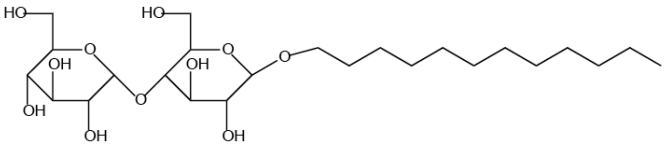
of subunits containing a single TM helix, there are a few examples of membrane proteins solved with four or more unique TM helices that would be more similar to that of the rhomboid protease. These include: the 21 kDa *E. coli* disulfide bond formation protein B (DsbB) in DPC micelles (146); the 11.4 kDa transmembrane domain of the *E. coli* histidine kinase K^+ sensor (KdpD) solved in LMPG micelles (147); and the impressive 26.7 kDa *Natronomonas pharaonis* sensory rhodopsin II (pSRII), with seven TM helices, solved in DHPC micelles (148). As noted earlier, there are many useful types of structure and dynamics data that can be provided by solution NMR, even when an x-ray structure has already been determined; consequently, these recent NMR landmarks raise the possibility that unique insights into rhomboid function might be accessible by solution NMR.

1.4 Thesis overview

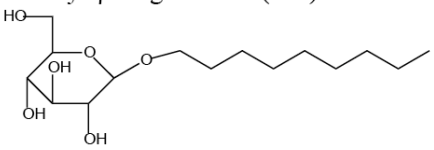
In this thesis, structural and functional insights into the cytoplasmic domain from the bacterial rhomboid are developed over three chapters. Chapter 2 examines the full-length rhomboid from *Pseudomonas aeruginosa*, and its functional and NMR spectroscopic properties in a range of membrane mimetic systems (Table 1.2). An interesting correlation was observed between the denaturing effects of certain detergents on the cytoplasmic domain structure, and loss of rhomboid activity. Furthermore, NMR data suggested that the cytoplasmic domain may interact with the transmembrane domain in a way that could regulate rhomboid activity. Chapter 3 describes experiments done to characterize the potential interaction between the TMD and cytoplasmic domain in *ec*GlpG. N-terminal truncation constructs and cytoplasmic TMD mutants identified regions in the cytoplasmic face of the TMD as well in part of the N-terminal cytoplasmic sequence that are required for

Table 1.2 Detergent chemical structures.

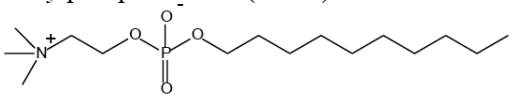
n-dodecyl-β-D-maltoside (DDM)



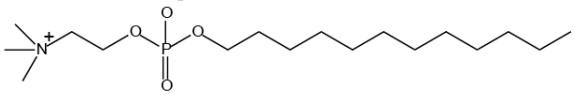
n-nonyl-β-D-glucoside (NG)



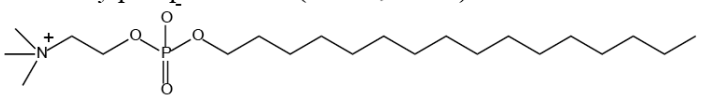
Decylphosphocholine (Fos10)



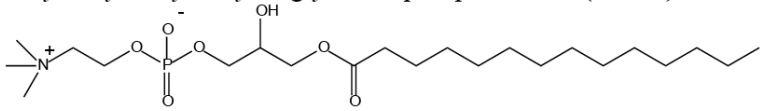
Dodecylphosphocholine (DPC, Fos12)



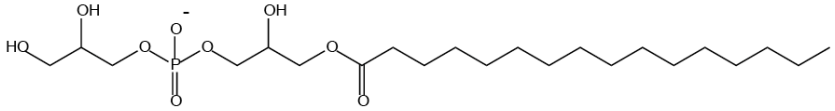
Hexadecylphosphocholine (HDPC, Fos16)



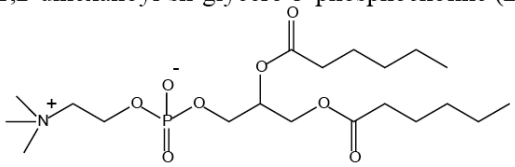
1-myristoyl-2-hydroxy-sn-glycero-3-phosphocholine (LMPC)



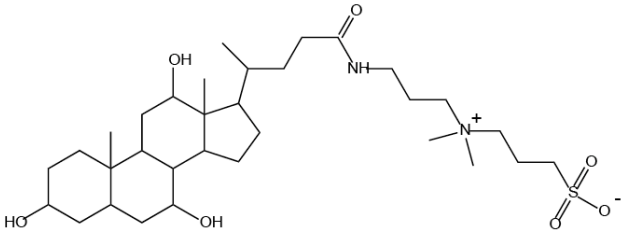
1-palmitoyl-2-hydroxy-sn-glycero-3-[phospho-rac-(1-glycerol)] (LPPG)



1,2-dihexanoyl-sn-glycero-3-phosphocholine (DHPC)



3-[(3-Cholamidopropyl) dimethylammonio]-1-propanesulfonate (CHAPS)



maximal function. In addition, a new rhomboid assay was developed that can probe the activation state of the active site serine without the need to detect substrate cleavage products. Finally, Chapter 4 reveals that the cytoplasmic domain can form a homodimer through three-dimensional domain swapping. Not only is this the first example of domain-swapping in an intramembrane protease, but we show that the detergent micelle plays a role in catalyzing this interaction. Taken together, this body of work reveals novel interactions that occur at the membrane that could modulate the function of the bacterial rhomboid protease.

Chapter 2: Insights into the effect of detergents on the full length rhomboid protease from Pseudomonas aeruginosa and its cytosolic domain.

Contributions of collaborators:

Results from this chapter have been published, and are reproduced with permission (License number 2763790754648, October 7 2011) from: **Sherratt AR**, Braganza MV, Nguyen E, Ducat T, Goto NK. Insights into the effect of detergents on the full-length rhomboid protease from Pseudomonas aeruginosa and its cytosolic domain. *Biochim Biophys Acta*. 2009 1788(11):2444-53; © 2009 Elsevier B.V.

Michael Braganza (University of Ottawa) and Elizabeth Nguyen (University of Ottawa) purified and assayed the *pa*GlpG rhomboid protease in DHPC/DMPC and CHAPS/DMPC bicelles, respectively. Thierry Ducat (University of Ottawa) aided in NMR data analyses.

2.1 Introduction

As noted in Chapter 1, many rhomboids contain N- and/or C-terminal extramembranous domains of unknown function. It has been postulated that some of these domains are required for specificity (77), or activity (79). In the case of the GlpG rhomboid from *E. coli* (*ec*GlpG), it has been shown that removal of the cytoplasmic domain reduces the ability of the core domain to hydrolyze a non-native water-soluble substrate (65). The mechanism by which the cytoplasmic domain modulates catalysis within the membrane-embedded active site is currently unknown. However, a potential clue was provided by the observation of detergent micelle binding properties for the cytoplasmic domain from a homologous rhomboid in *Pseudomonas aeruginosa* (79). This led to the proposal that in the full-length protein, this domain interacts specifically with the bilayer surface to help position the active site at a fixed point relative to the membrane surface. However, it is not yet known if the observed detergent-binding properties of this domain accurately reflect a native interaction between this domain and the surface of the lipid bilayer.

In order to gain more insight into the functional role of the GlpG cytoplasmic domain in the regulation of proteolytic activity, we sought to develop a bacterial rhomboid sample possessing a cytoplasmic domain that might be accessible to analysis by solution NMR. This chapter focuses on a rhomboid from *Pseudomonas aeruginosa* (*pa*GlpG) that is homologous to *ec*GlpG (Fig. 2.1), for which a solution NMR structure of the cytoplasmic domain (*pa*CytD) is available (79). In this chapter, the *in vitro* function of *pa*GlpG in a number of different detergents (detailed in Tables 1.2 and 2.1) is evaluated. These same samples were also examined by solution NMR to determine the extent of CytD association with the detergent-protein complex. In a separate set of experiments on the isolated

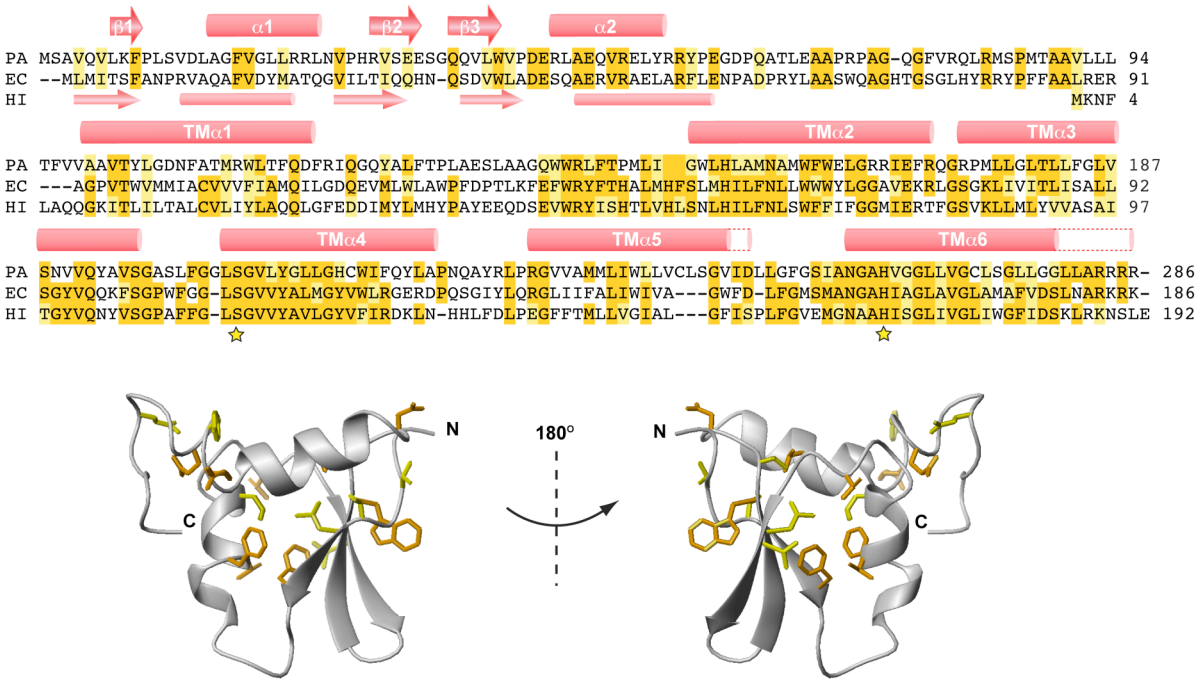


Figure 2.1 Comparison of sequences from the two rhomboids for which core catalytic domain crystal structures are available (EC, *E. coli*, HI, *H. influenza*) with the *P. aeruginosa* (PA) rhomboid sequence. *Top;* Core catalytic membrane domains from the 3 rhomboid sequences were aligned in ClustalW (149) and combined with the alignment for the CytD from PA and EC rhomboids. Identical residues are highlighted in orange and homologous residues in yellow (*i.e.* all hydrophobic, all aromatic, all polar, or all small (Ala/Ser/Gly)). Regions of secondary structure of the *E. coli* and *H. influenza* sequences in the x-ray crystal structures are indicated above the sequence with the longer helices found in the *H. influenza* structure (67) indicated with dashed extensions of helices 5 and 6. The catalytic His-Ser dyad residues are indicated with stars. Secondary structure elements in the CytD structure are also shown above the CytD sequence, with the secondary structure prediction from Jpred (150) for the EC sequence shown below. *Bottom;* Two views of the PA CytD structure (accession number 2GQC, (79)) with side chains displayed for conserved residues that are identical (orange) or similar (yellow) between the EC and PA sequences.

Table 2.1 Physical properties of detergents used in this chapter

Detergent	MW (Da)	CMC[a] (mM)	Aggregation Number[b]	~$MW_{micelle}$ (kDa)	Reference
DDM	511	0.1	78-149	58	AC[c]
NG	306	6	-	-	(151)
DPC	351	1.5	54	19	AC[c]
LMPC	468	0.07	122	57	(152, 153)
LPPG	507	0.02	125	63	(152, 154)
DHPC	453	14	19	16	(152)
CHAPS	615	8	10	6	(154)

a – Critical micelle concentration, above which micelles are spontaneously formed.

b – Average number of detergent molecules that associate to form a micelle.

c – Anatrace catalogue (2009-2010)

*pa*CytD, the intrinsic ability of this domain to interact with these same detergents was similarly probed. Overall, a model consistent with this work suggests that while the CytD is important for *pa*GlpG activity against transmembrane substrates, this domain does not act through an interaction with the membrane, and is likely interacting with some part of the core catalytic domain that has yet to be identified.

2.2 **Materials & Methods**

2.2.1 Plasmids.

A C-terminally hexahistadine (His_6) tagged version of the *P. aeruginosa* rhomboid PA3086 (*pa*GlpG) in a pET25b vector (62) was kindly provided by Dr. Christopher Koth's group in the Ontario Centre for Structural Genomics at the University of Toronto. This plasmid was also used to construct expression vectors containing the *pa*GlpG cytoplasmic domain (*pa*CytD) (residues 1-88) or C-terminal transmembrane domain (*pa*TMD) (residues 84-286), both with C-terminal His_6-tags separated from the rhomboid sequence with a TEV protease cleavage site, using primers synthesized by Integrated DNA Technologies Inc (Table 2.2). A plasmid encoding a fusion protein to the Spitz TM segment ($Spitz_{TM}$) with an N-terminal β-lactamase domain (Bla) and C-terminal His-tagged maltose binding protein (Bla-$Spitz_{TM}$-MBP) (49) was generously provided by the Akiyama group in the Institute for Virus Research at Kyoto University. Conveniently, the *E. coli* rhomboid does not cleave this substrate *in vivo*, circumventing any requirement to use GlpG knockout strains to express this substrate. All constructs were verified by sequencing performed at the Ontario Genomics Innovation Centre at the Ottawa Health Research Institute.

Table 2.2 Primer sequences for *pa*GlpG constructs.

Construct	Plasmid	Direction	Sequence[a]
*pa*CytD	pET25b	Forward	5' –GATATACATATGAGCGCGGTG– 3’
		Reverse	5' –CAGCGTCCTCGAGGGTCATCGGGCTCATCCG– 3’
*pa*TMD	pET25b	Forward	5’ –GCACGTCTCATATGAGCCCGATGACCGCCGCG– 3’
		Reverse	5’ –TTCTCTCTCGAGGCTAGCACG– 3’

[a] Restriction enzyme sites are underlined in the primer sequence. Forward primers contain the NdeI site and reverse primers contain the XhoI site.

2.2.2 Protein expression.

Expression vectors encoding the *pa*CytD, *pa*TMD or full length *pa*GlpG rhomboid were transformed into *E. coli* BL21(DE3) and grown at 37ºC in M9 minimal media with 100 µg/mL ampicillin containing 0.1% (w/v) ^{15}N-labeled ammonium chloride and 0.3% (w/v) glucose (U-^{13}C for the *pa*CytD) as the sole nitrogen and carbon sources, respectively. For *pa*TMD or full length *pa*GlpG, when the optical density of the culture at 600 nm reached 0.5-0.7 the temperature was reduced from 37ºC to 16ºC and expression induced with 1 mM isopropyl B-D-thiogalactopyranoside and allowed to proceed for 16-20 hours. In the case of the *pa*CytD, induction was carried out at 37ºC for four hours. Bla-Spitz$_{TM}$-MBP expression was done using the same protocol, except LB medium containing 50 µg/mL spectinomycin was used instead of minimal media with ampicillin, and expression was carried out at 37˚C for two hours.

2.2.3 Purification of detergent-solubilized membrane protein samples.

All constructs containing a membrane-associated domain (*i.e. pa*GlpG, *pa*TMD and Bla-Spitz$_{TM}$-MBP) were purified using a modified version of previously published protocols for *ec*GlpG (62). Briefly, the bacterial pellet from a 1 L culture was resuspended in 20 mL of 20 mM HEPES buffer pH 7.3 containing 100 mM NaCl, 5 mM $MgCl_2$, 10% glycerol and 0.5 mM benzamidine. Cells were lysed either by two passages through an Avestin homogenizer at 25 kpsi or by sonication (since we found that rhomboid activity in full-length and TMD *pa*GlpG was not affected by the method of lysis) and centrifuged using a Beckman Coulter Avanti centrifuge at 50,000xg for 1 h. The insoluble fraction was resuspended for 1 h in 50 mM HEPES buffer, pH 7.3, containing 1% (w/v) detergent (either dodecyl maltoside (DDM), dodecylphosphocholine (DPC), 1-myristoyl-2-hydroxy-sn-

glycero-3-phosphocholine (LMPC) or 1-palmitoyl-2-hydroxy-sn-glycero-3-[phospho-rac-(1-glycerol)] (LPPG)) or 2% (w/v) *n*-nonyl-β-D-glucoside (NG), 200 mM NaCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 10% glycerol, 5 mM imidazole and 0.5 mM benzamidine and then centrifuged at 50,000 x g for 30 min. (Note that a higher NG concentration was used as was done for previous purifications with this detergent for *ec*GlpG (66), since sample precipitation was observed with lower NG concentrations.) The supernatant containing detergent-solubilized protein was applied to a nickel affinity column equilibrated in the same buffer without detergent. The column was washed in 50 mM HEPES buffer, pH 7.3, containing 0.1% detergent, 500 mM NaCl, 1 mM $MgCl_2$, 1 mM $CaCl_2$, 10% (v/v) glycerol and 30 mM imidazole. A second wash of 50 mM HEPES buffer, pH 7.3, containing 0.1% (w/v) detergent, 200 mM NaCl and 30 mM imidazole was performed prior to elution in the same buffer supplemented with 0.2% (w/v) detergent (0.4% in the case of NG) and 250 mM imidazole. Elution fractions were concentrated using a 30 kDa molecular weight cut off (MWCO) ultrafiltration device (Amicon Ultra) for samples to undergo FPLC purification or a 50 kDa MWCO unit (Pall Macrosep/Microsep) for samples to be used for NMR experiments. Samples to be used for the activity assay were then applied to a Superdex-200 10/300 GL size exclusion column using an AKTA FPLC (GE Healthcare) equilibrated in 50 mM HEPES buffer pH 7.3 containing 0.2% (w/v) detergent, 100 mM NaCl, 0.5 mM benzamidine and 0.1 mM EDTA. Fractions of the main peak were then directly used in the protease activity assay.

2.2.4 Transfer of purified paGlpG or Bla-Spitz$_{TM}$-MBP into bicelle solutions.

Samples were purified using DDM as described above, and the 2 mL fraction containing the highest concentration of pure protein (typically ~ 20 - 30 μM) was rapidly

diluted into 50 mL of a 2% (w/v, total lipid) 1,2-dimyristoyl-sn-glycero-3-phosphocholine/1,2-dihexanoyl-sn-glycero-3-phosphocholine (DMPC/DHPC) bicelle solution with a DMPC:DHPC molar ratio (*q*) of 0.44 or a 2% DMPC/CHAPS solution with *q*=0.5 in 50 mM HEPES buffer, pH 7.3 and 100 mM NaCl. Bicelles were prepared as described (155) through multiple rounds of heating to 42ºC and cooling to room temperature, with vortexing in between until the solution became clear. Bicelle-solubilized rhomboid was applied to a Ni-NTA column equilibrated in the same buffer with no bicelles and eluted in 50 mM HEPES buffer pH 7.3, 100 mM NaCl, 2% (w/v) bicelles, 250 mM imidazole and 0.2 mM benzamidine. The 1 mL elution fraction that contained the highest concentration of *pa*GlpG or Bla-Spitz$_{TM}$-MBP was used directly in the activity assay or all elution fractions were combined and concentrated for analysis by NMR spectroscopy (DHPC bicelle sample only).

2.2.5 Cytoplasmic domain purification.

The bacterial pellet from a 1 L culture was resuspended in 20 mM HEPES pH 7.3, 200 mM NaCl, 1X EDTA-free protease inhibitor cocktail (Roche) and 10 mM imidazole and lysed by sonication. The supernatant was applied to a Ni-NTA column equilibrated in the same buffer and the column was washed with this same buffer supplemented with 30 mM imidazole before the *pa*CytD construct was eluted with 250 mM imidazole. Following concentration with a 5 kDa MWCO Amicon Ultra centrifugal filter unit, the *pa*CytD was further purified on a Superdex-75 10/300 GL size exclusion column in 50 mM HEPES buffer, pH 7.3, 200 mM NaCl, 0.5 mM benzamidine and 0.1 mM EDTA. MALDI-MS was performed on purified samples by the Protein Function Discovery Facility at Queen's

University and found to correspond to the expected molecular mass of the *pa*CytD construct, with approximately 60-70% of the sample missing the N-terminal Met.

2.2.6 NMR spectroscopy.

All NMR spectra were recorded at the University of Ottawa NMR Facility on a Varian Inova 500 spectrometer equipped with a triple resonance probe and pulsed field gradient unit. 2D ^{1}H-^{15}N HSQC spectra were recorded at 15 - 30ºC and data processed by NMRPipe (105) and analyzed with NMRView (156). Concentrations of *pa*GlpG were ~360 – 440 µM, with purity >90% determined by Coomassie-stained SDS-PAGE. Backbone ^{15}N and ^{1}HN chemical shift assignments for the *pa*CytD domain in 25 mM phosphate buffer, pH 6.5, 25 mM NaCl, 0.1 mM EDTA were made on a 1 mM ^{15}N, ^{13}C-labeled sample at 25ºC using standard triple resonance techniques (157) based on HNCACB, CBCA(CO)NH and HNCO experiments. The sample was then transferred into a 50 mM HEPES solution, pH 7.3 with 200 mM NaCl 0.5 mM benzamidine and 0.1 mM EDTA and assignments transferred from the low pH spectrum with the aid of an additional HNCACB spectrum recorded at 15ºC.

2.2.7 In vitro protease activity assay.

Proteolytic activity was assessed using an assay based on a previously described protocol using Bla-Spitz$_{TM}$-MBP as the substrate (49), purified under identical conditions to those used for each rhomboid sample. Briefly, 6 µM *pa*GlpG or *pa*TMD was mixed with 1 µM Bla-Spitz$_{TM}$-MBP in the same buffer as was used for FPLC purification. Aliquots were taken at time T=0 h and T=16 h and mixed with SDS-PAGE sample loading buffer. The amount of Bla-Spitz$_{TM}$-MBP remaining after incubation was assessed by Coomassie-stained SDS-PAGE by integrating protein band intensities and subtracting any non-specific intensity

reductions detected in control reactions performed in the absence of rhomboid using the general image processing program ImageJ (available at http://rsb.info.nih.gov/ij/index.html).

2.2.8 Cytoplasmic domain titration experiments.

^{1}H-^{15}N HSQC spectra of ~175 μM solutions of ^{15}N-labeled *pa*CytD in 50 mM HEPES buffer pH 7.3 with 200 mM NaCl, 10% (v/v) D_2O, at 15ºC were acquired in the absence and presence of DPC, LPPG, DDM, CHAPS or DHPC at a range of concentrations (typically 2 mM to 10 mM). ~400 μM samples of the *pa*CytD were also added directly to high concentrations of detergents (typically 250 mM) in the same buffer and found to give spectra that were identical to those acquired at endpoints of detergent-into-protein titrations. Peak intensities were monitored using the nlinLS module in nmrPipe. For the DPC and LPPG titrations, changes in peak intensities could only be monitored in the earlier stages of the titration. 2% (w/v) DHPC/DMPC (q=0.44) or CHAPS/DMPC (q=0.5) bicelles were also added to *pa*CytD samples in the same buffer and changes in peak positions determined in NMRView. Assignment of spectra recorded in the presence of bicelles was facilitated by following the smaller changes in peak positions that occurred during titrations of 2 mM to 10 mM bicelle solutions into a *pa*CytD sample. There was no significant difference in the average chemical shifts between the 10 mM and 40 mM bicelle samples, indicating that the sample was already close to, or at, saturation at 10 mM bicelle concentrations. Average amide chemical shift differences (Δδ) were calculated according to:

$$\Delta\delta = \sqrt{(\Delta\delta_{HN})^2 + (\Delta\delta_N/5)^2} \qquad [1]$$

where $\Delta\delta_{HN}$ and $\Delta\delta_{N}$ are the amide proton and nitrogen chemical shift differences, respectively.

^{1}H-^{15}N HSQC spectra were also recorded at 37ºC in the presence and absence of 2% (w/v) vesicles prepared from *E. coli* total lipid extracts (Avanti Polar Lipids) using established protocols (158). Briefly, a 25 mg/mL stock solution in chloroform was transferred into a glass test tube and the chloroform removed under a dry argon stream. The lipid film was resuspended in 50 mM HEPES buffer, pH 7.3, 200 mM NaCl, 0.1 mM EDTA and vortexed to make a suspension of large unilamellar vesicles. These were then sonicated with a probe sonicator at 20% power, 15 times for 2 seconds, creating an optically clear solution, and centrifuged to remove particulate matter.

2.3 Results

2.3.1 paGlpG Purification.

In order to probe the ability of relatively 'NMR-friendly' detergents to be used in the isolation of the *pa*GlpG rhomboid, bacterial membranes containing over-expressed *pa*GlpG were solubilized and subjected to purification in a range of detergent solvents. These included detergents that have previously been used for solution NMR of transmembrane domains from helical membrane proteins (namely DPC, LMPG and LPPG), and detergents known to support rhomboid activity *in vitro* (DDM and NG) (49, 62, 63, 65). Yields were generally similar for all detergents, with ~2 mg of >90% pure *pa*GlpG being typically isolated from 1 L of minimal media after nickel affinity chromatography (Fig. 2.2). However, in the case of purifications done with NG, which was the detergent used to produce a number of crystal structures of the *ec*GlpG transmembrane domain, precipitation

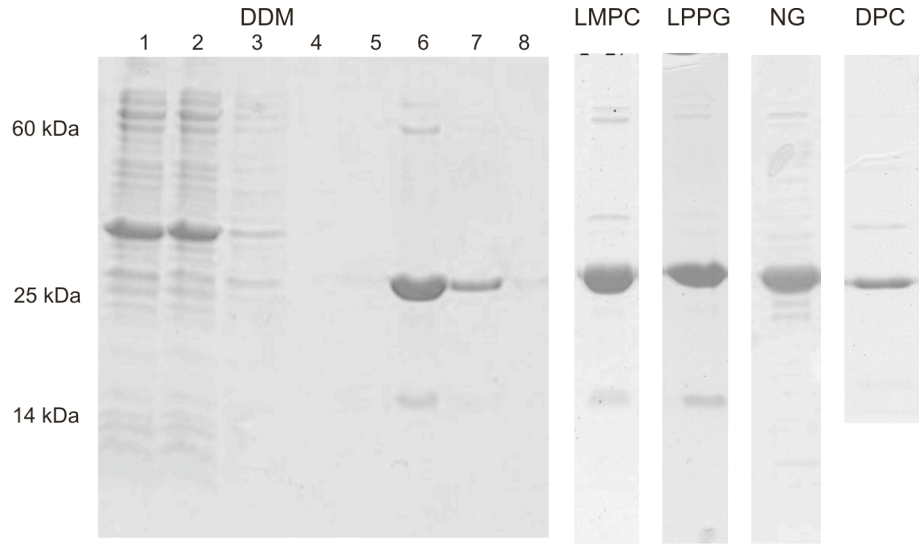


Figure 2.2 Coomassie-stained 15% SDS-PAGE gel of fractions taken at various stages in the purification of *pa*GlpG. Lane 1- Solubilized crude rhomboid in DDM; Lane 2 - Nickel affinity column flow through; Lane 3 - Flow-through after reapplication of solubilized membrane to the same column; Lanes 4 & 5 - Contaminants eluted after washing nickel affinity column with buffer containing 30 mM imidazole. Lanes 6 - 8; purified *pa*GlpG eluted with 250 mM imidazole. Representative elution fractions for purifications done in other detergents are also shown. The theoretical molecular mass of *pa*GlpG is 34.2 kDa.

reproducibly occurred after this step, even when no further manipulations were performed. Consequently, it was not possible to concentrate the NG sample for NMR analysis, although concentrations required for activity assays could be maintained under these conditions.

2.3.2 Rhomboid activity in different membrane mimetics.

In order to compare the functional status of *pa*GlpG in these different detergent solvents, the proteolytic activity of each sample was tested in vitro using Bla-Spitz$_{TM}$-MBP (49), a fusion protein containing the Spitz TM segment fused to globular proteins at the N- and C-termini (β-lactamase and maltose binding protein, respectively). The Spitz TM segment has previously been shown to be cleaved by *pa*GlpG, and since it is not a substrate for *ec*GlpG (49, 63), it was possible to express and purify it under the same range of detergent conditions that were used for *pa*GlpG. As shown in Figure 2.3, for purified *pa*GlpG and Bla-Spitz$_{TM}$-MBP in either DDM or NG, there is a decrease in intensity of the purified full-length fusion protein in Coomassie-stained gels that only occurs when *pa*GlpG is present. This change was largest for *pa*GlpG in DDM, with a 77% reduction in band intensity, whereas the NG-solubilized rhomboid showed a smaller decrease of 45%. This reduction in intensity could be correlated with the appearance of a band corresponding to the size of the C-terminal fragment that would be expected to arise if cleavage were to occur in the Spitz TM segment. No proteolysis was detected if the samples were also incubated with DCI, an inhibitor shown to be relatively specific for rhomboid proteases (13, 63).

In contrast with the activities obtained in the non-ionic detergents DDM and NG, significantly lower activities were found for *pa*GlpG in the ionic or zwitterionic detergents tested (Fig. 2.3). Based on Bla-Spitz$_{TM}$-MBP band intensity reductions for assays run under identical conditions, *pa*GlpG appeared to be ~4-times less active in LPPG, and ~10-times

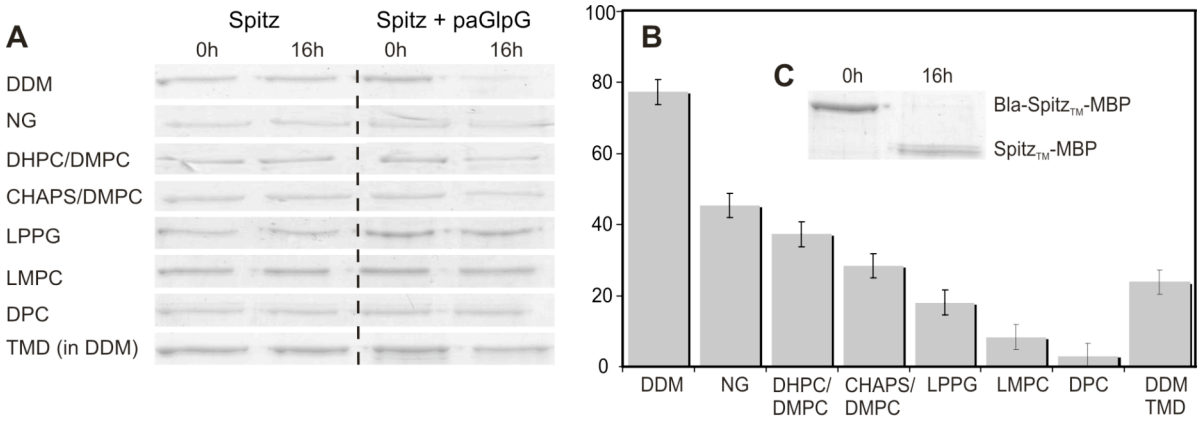


Figure 2.3 Activity of *pa*GlpG in different membrane mimetic environments. (A) Region of the Coomassie-stained SDS-PAGE gel showing the Bla-Spitz$_{TM}$-MBP substrate band before (0h) and after (16h) incubation at 37ºC in the presence (Spitz+*pa*GlpG) and absence (Spitz) of 6 μM *pa*GlpG in 0.2% (w/v) of the indicated detergent (except for NG, which was run in 0.4%), or 2% (w/v) bicelles. The last row shows assay results for just the TMD from *pa*GlpG purified in DDM. (B) Average percent reduction in Bla-Spitz$_{TM}$-MBP band intensity determined from assay gels. The error bars represent the maximum observed standard deviation obtained from three different purifications in a single detergent. (C) Portion of the Coomassie-stained gel showing the C-terminal product of degradation containing MBP and part of the Spitz TM segment (Spitz$_{TM}$-MBP).

less active in LMPC compared to its activity in DDM. Moreover, DPC did not support any rhomboid-specific proteolytic activity indicating that these detergents likely exert a detrimental effect on *pa*GlpG structure.

Since small DMPC-based bicelles can stabilize helical membrane proteins (155), and are also compatible with solution NMR studies (124, 130), *pa*GlpG was transferred from DDM micelles into small bicelles formed by DMPC and DHPC or CHAPS. Bicelle size is determined in part by the molar ratio of DMPC to short chain lipid or bile salt (commonly referred to as *q*) with values of 0.25 to 0.33 being previously used in solution NMR studies of membrane proteins with multiple TM helices (131, 159-161). However, it was shown for the 7-TM segment bacterial protein opsin that higher *q* ratios are required to stabilize this protein, with values of 0.4 to 0.6 being the minimum required to observe a significant effect (155). Therefore, we chose to test DMPC-based bicelles with q=0.50 for CHAPS and q=0.44 for DHPC. A rapid dilution technique previously used to transfer opsin (155) and the 3-TM helix trimer diacylglycerol kinase (162) was employed in this case, with nickel chromatography being used to complete the exchange into bicelles and reconcentrate the sample. Purified Bla-Spitz$_{TM}$-MBP in DDM was also subjected to this same bicelle transfer procedure.

The activity of *pa*GlpG transferred into either CHAPS/DMPC or DHPC/DMPC bicelles was similar to that obtained in NG, although this was approximately 40 - 50% of the activity obtained in DDM (Fig. 2.3). The ability of these bicelles to support higher proteolytic activity than that obtained in lysolipids suggests that the bicelles provided a more native-like environment for the rhomboid. In addition, since the mobility of the DMPC phase is lower than detergent mobility in micelles (123, 163), transfer of the Spitz substrate

into a rhomboid-containing bicelle may have been slower than the analogous transfer in detergent micelles, raising the possibility that the true activity of bicelle-bound *pa*GlpG could be higher than shown in this assay.

In addition to these studies on the full-length protein, we evaluated the activity of a truncated form of *pa*GlpG containing only the catalytic transmembrane domain (*pa*TMD, residues 84 – 288) in DDM since the influence of the CytD on proteolytic activity against transmembrane substrates has not been investigated. As shown in Figure 2.3, the *pa*TMD shows a ~70% reduction in activity relative to the full-length construct, demonstrating that the full-length rhomboid requires the CytD for maximal activity against transmembrane substrates.

2.3.3 NMR spectroscopy of detergent-solubilized paGlpG.

To determine whether the CytD was strongly associated with the detergent micelle complex in the range of solubilizing detergents tested, 2D ^{1}H-^{15}N HSQC NMR spectra were acquired for the full-length *pa*GlpG. Since the size of the rhomboid construct used in our study was 32.4 kDa, and the micelle could be expected to add 20 kDa or more to the complex size (Table 2.1), significant line-broadening in the NMR spectrum should be observed so long as all parts of the complex tumble as a single entity. However, any part of the protein that is not strongly associated with the detergent-protein complex will undergo faster rotational diffusion, and should give rise to more intense and narrow peaks in the NMR spectrum. As shown in Figure 2.4, this appeared to be the case when the three detergents that produced the least active forms of *pa*GlpG were used (*i.e.* DPC, LMPC and LPPG). Approximately one quarter of the expected number of backbone amide correlations could be observed in these samples. Comparison of these spectra to a spectrum acquired for

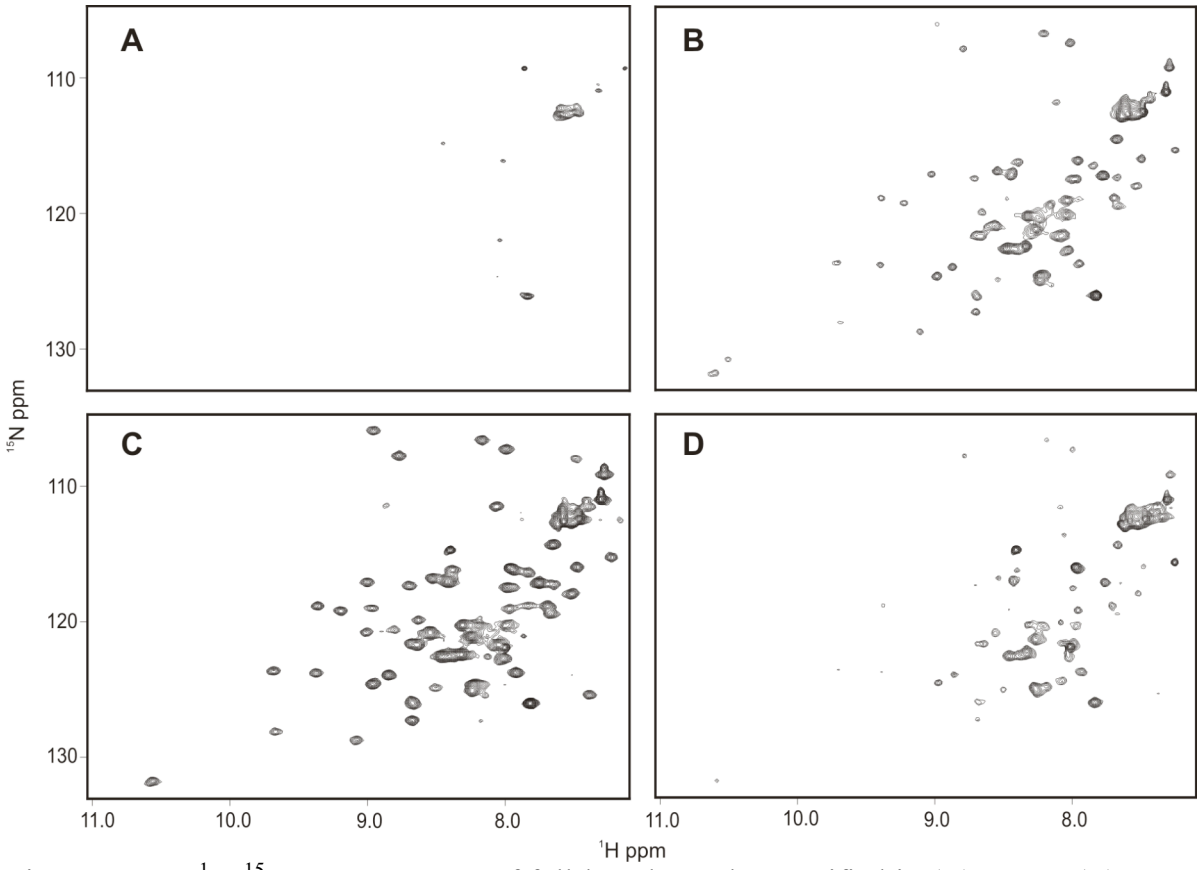


Figure 2.4 ^{1}H-^{15}N HSQC spectra of full-length *pa*GlpG purified in (A) DDM, (B) DPC, (C) LMPC and (D) LPPG. All spectra were recorded at pH 7.3, 15˚C and levels were adjusted to account for differences in protein concentration (0.36 - 0.44 mM).

the isolated *pa*CytD obtained under the same conditions in the absence of detergent demonstrated that almost all signals observed in the presence of DPC, LMPC or LPPG could be attributed to this domain. For example, the ^{1}H-^{15}N HSQC spectrum of the isolated *pa*CytD in aqueous buffer was virtually superimposable with that from the full-length rhomboid in LMPC (Fig. 2.5). The strong signals observed for the *pa*CytD in these detergents, along with the very high degree of similarity in amide chemical shifts for all of these samples, demonstrate that this domain is partly (DPC, LPPG), or almost completely (LMPC) dissociated from the bulk of the detergent-protein complex in these functionally compromised samples.

NMR spectra were also acquired on samples that showed a high degree of functionality, however very few peaks were observed in these cases (*e.g.* DDM Fig. 2.4a). Similarly, very few peaks were observed in spectra of the core catalytic domain lacking the cytoplasmic domain (data not shown), as would be expected for a detergent-protein complex exceeding 50 kDa in size. In fact, the approximate size of all protein-micelle complexes formed with *pa*GlpG in this study were at least 100 kDa or larger as determined by size exclusion chromatography. Therefore, the absence of signals from the CytD in DDM or in small bicelles is consistent with the possibility that the CytD remains associated with the large detergent-protein complex in these functional samples, imparting the same NMR spectroscopic properties to the CytD that prevents visualization of signals from the core domain.

2.3.4 Investigation of the detergent-binding properties of the isolated cytoplasmic domain.

Since NMR spectra of full-length rhomboid samples suggested that the *pa*CytD was associated with the detergent micelle-complex for functional samples, we were interested to

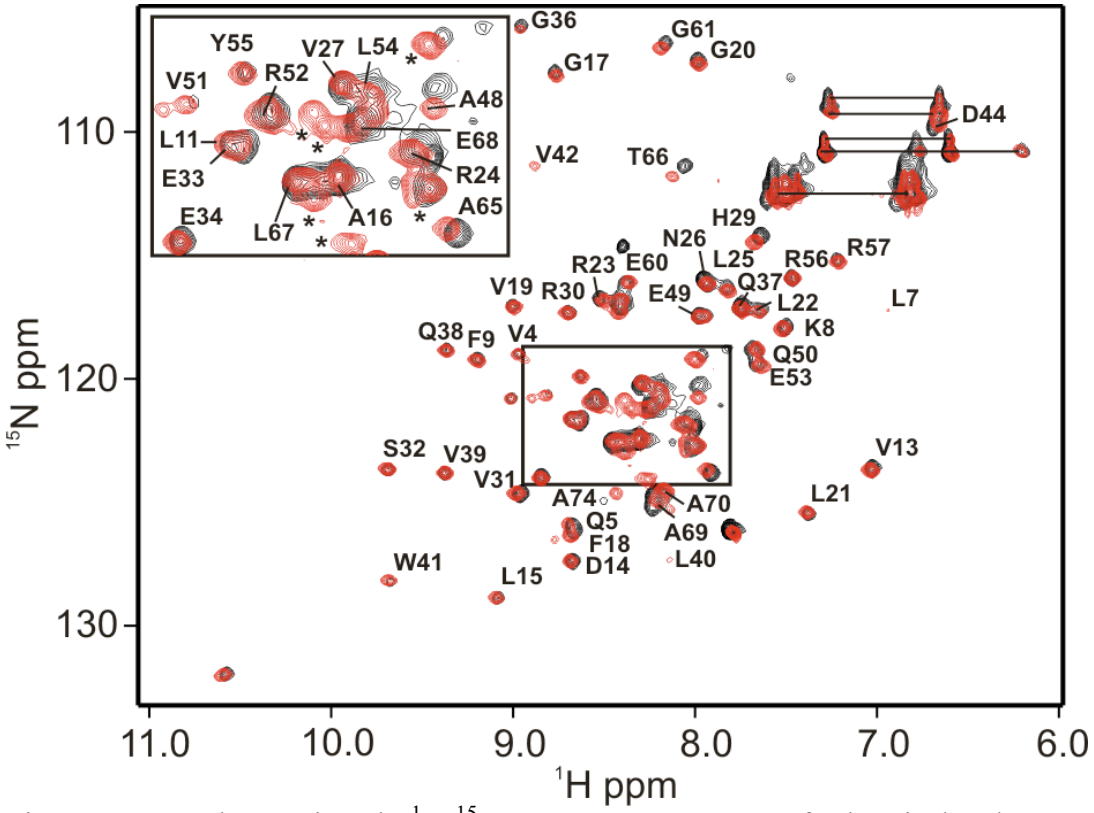


Figure 2.5 The assigned ^{1}H-^{15}N HSQC spectrum of the isolated CytD (red), superimposed on that from full-length *pa*GlpG solubilized in LMPC (black), recorded at pH 7.3, 15˚C. Peaks arising from the unstructured C-terminal region of the protein with ambiguous assignments are indicated with asterisks.

see to what extent this could be attributed to direct interactions between the detergent and *pa*CytD. For this purpose, ^{15}N-labeled samples of *pa*CytD were titrated with detergents that supported high (DDM), or low (DPC, LPPG), rhomboid activity, and monitored by ^{1}H-^{15}N HSQCs under conditions similar to those used for NMR spectroscopy of the full length protein.

As shown in Figure 2.6a, there was a significant reduction in peak intensity upon addition of increasing concentrations of micellar DPC. Sub-micellar concentrations of DPC did not give rise to any changes in the spectrum (data not shown), confirming that the interaction was micelle-specific. However, in contrast with previous studies using hexadecylphosphocholine (HDPC) (79), the loss of peak intensity that occurred upon introduction of DPC micelles occurred to a similar extent for all peaks in the spectrum, with no region of the *pa*CytD showing evidence for a preferential interaction with this detergent. In addition, at higher concentrations of micellar DPC a new series of peaks appeared in the central part of the spectrum that was characteristic of an unfolded protein (Fig. 2.6b). Only these peaks from the unfolded species were visible in the spectrum once the micelle:protein ratio was ~1:1 or greater. Therefore, although these results confirmed that the isolated *pa*CytD domain has a tendency to interact with phosphocholine-type detergent micelles, the NMR spectral changes suggest that this interaction leads to denaturation of *pa*CytD.

A similar type of denaturing interaction was also observed for micelles formed by the anionic lysolipid LPPG, with complete conversion into the unfolded form being observed when the micelle:CytD ratio was ~1:4 (data not shown). Again, it was not possible to identify the regions of the *pa*CytD that might have a preferential association with this detergent, since all regions of the protein showed similar peak intensity reductions at early

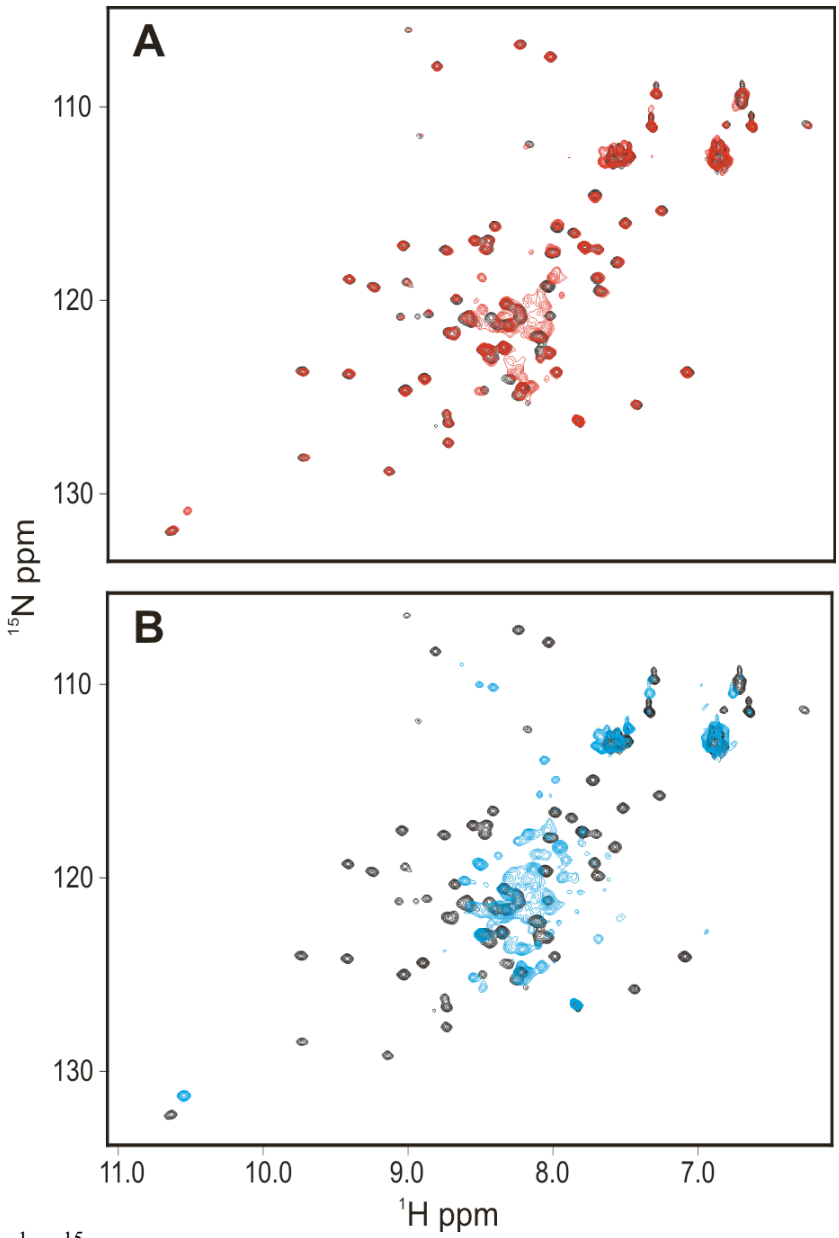


Figure 2.6 ^{1}H-^{15}N HSQC spectra of a 175 μM solution of the CytD in the absence of detergent (black) superimposed with spectra obtained in the presence of DPC at a concentration of (A) 2 mM (red) or (B) 10 mM (blue).

stages of the titration. These results indicate that denaturing *pa*CytD-micelle interactions are not specific for the phosphocholine headgroup. More significantly, since these same detergents give rise to NMR spectra for the full-length protein that show attenuated signals for the CytD (Fig. 2.4b and d), this denaturing interaction with the CytD may partly explain the low function observed for *pa*GlpG solubilized in these detergents.

Distinct from the lysolipid and DPC results, when DDM was added to solutions containing the *pa*CytD, no chemical shifts were detected over a wide range of detergent concentrations (data not shown). In addition, peak intensities were not attenuated, but remained constant for all residues in the spectrum at all tested concentrations of DDM. These results show that there is no interaction between the *pa*CytD and micelles formed by the detergent that retains the highest levels of rhomboid activity, demonstrating that CytD interactions with the membrane-mimetic environment are not required for a high level of *pa*GlpG activity.

2.3.5 Does the paCytD bind lipid bilayers?

Although our data to this point indicated that *pa*CytD interactions with detergent micelles do not promote rhomboid activity, it did not rule out the possibility that *pa*CytD specifically binds to more native-like phopholipid bilayers. To investigate this possibility, spectra were acquired for the *pa*CytD in the presence and absence of the same bicellar media that were used for the activity assays with full-length *pa*GlpG. For both DHPC and CHAPS bicelles made with DMPC, a subset of residues showed changes in chemical shifts that increased with increasing bicelle concentrations up to ~10 mM, characteristic of a system undergoing fast exchange. Shift changes incurred upon addition of either type of bicelle to a final concentration of ~40 mM (2% w/v) showed a similar pattern for both bicelle types,

with the magnitude of the shifts generally being larger for the CHAPS bicelles (Fig. 2.7). When residues showing the most significant shift changes were mapped onto the structure of the CytD, they localized to the same regions of the structure for both bicelle types; namely surface-exposed residues in α2, residues in the flexible region C-terminal to this helix and in an adjacent loop between α1 and β2 (Fig. 2.8).

Given that significant concentrations of free DHPC or CHAPS are known to exist in equilibrium with bicelles, we then investigated the possibility that the observed shifts with bicelles arose from interactions with the detergent monomers. As shown in Figure 2.8, spectra that strongly resembled those acquired in the presence of bicelles were obtained just through the addition of 6 mM DHPC or CHAPS. The concentration of submicellar detergent giving rise to these chemical shifts is comparable to previous estimates of the monomeric detergent concentrations that exist in solution with a 40 mM DHPC/DMPC or CHAPS/DMPC bicelle solution (~7 mM (133), and 4 mM - 10 mM (164), respectively). In addition, similar spectra were obtained for the *pa*CytD in the presence of 10 mM and 40 mM bicelles, which would be expected to have comparable amounts of free detergent in each solution. Taken together, these results demonstrate that the shifts induced by the bicelle solutions arise from an interaction with free short-chain lipid or CHAPS detergent and not with the bicelles themselves.

To test *pa*CytD membrane-binding affinity without the complication of additional interactions with free detergent, spectra of labeled *pa*CytD were also recorded in the presence of liposomes prepared from 2% (w/v) total *E. coli* lipid extracts. As shown in Figure 2.8d, the spectrum acquired in the presence of small unilamellar vesicles is superimposable with the lipid-free spectrum of the *pa*CytD, with no attenuation of signal

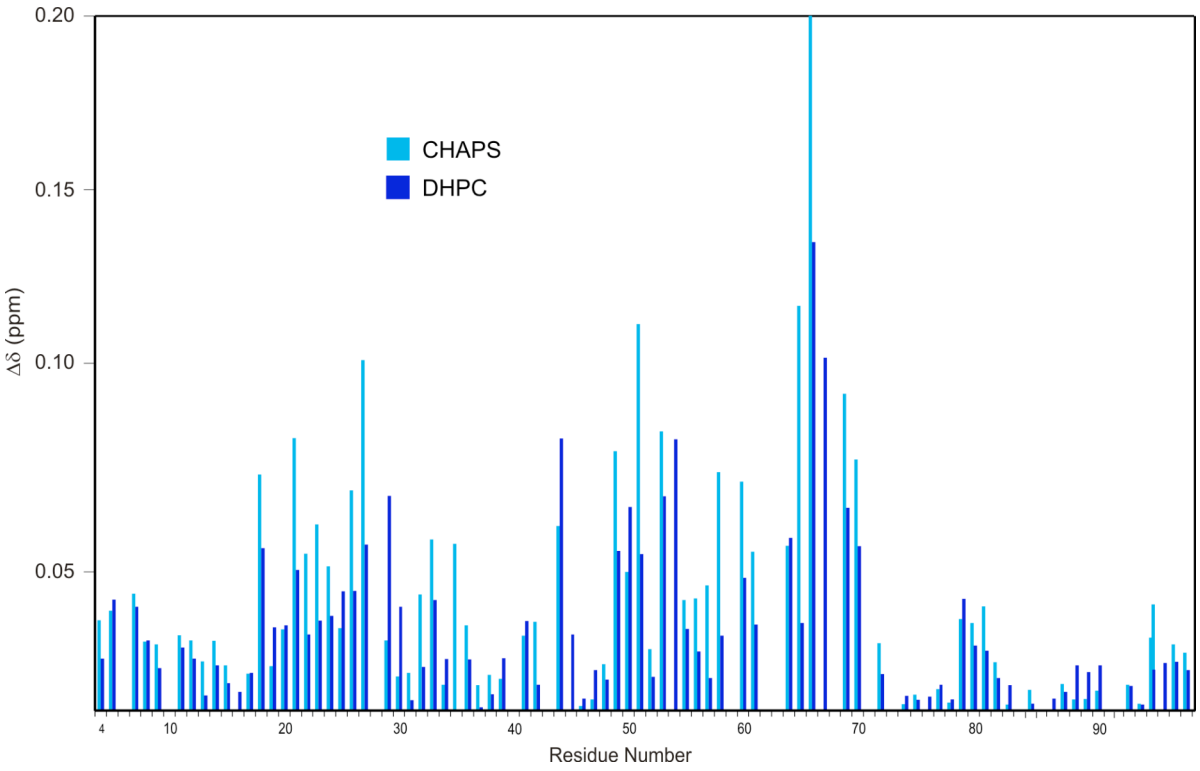


Figure 2.7 Average backbone amide chemical shift differences for *pa*CytD residues induced by addition of a 2% (w/v) solution of DHPC/DMPC bicelles with q=0.44 (dark blue) or CHAPS/DMPC bicelles with q=0.5 (light blue). Spectra were recorded at pH 7.2 to enable more *pa*CytD residues to be monitored, and to confirm that the non-native C-terminal residues from the TEV protease site and His-tag sequence (residues 89 - 106) did not participate in the interaction.

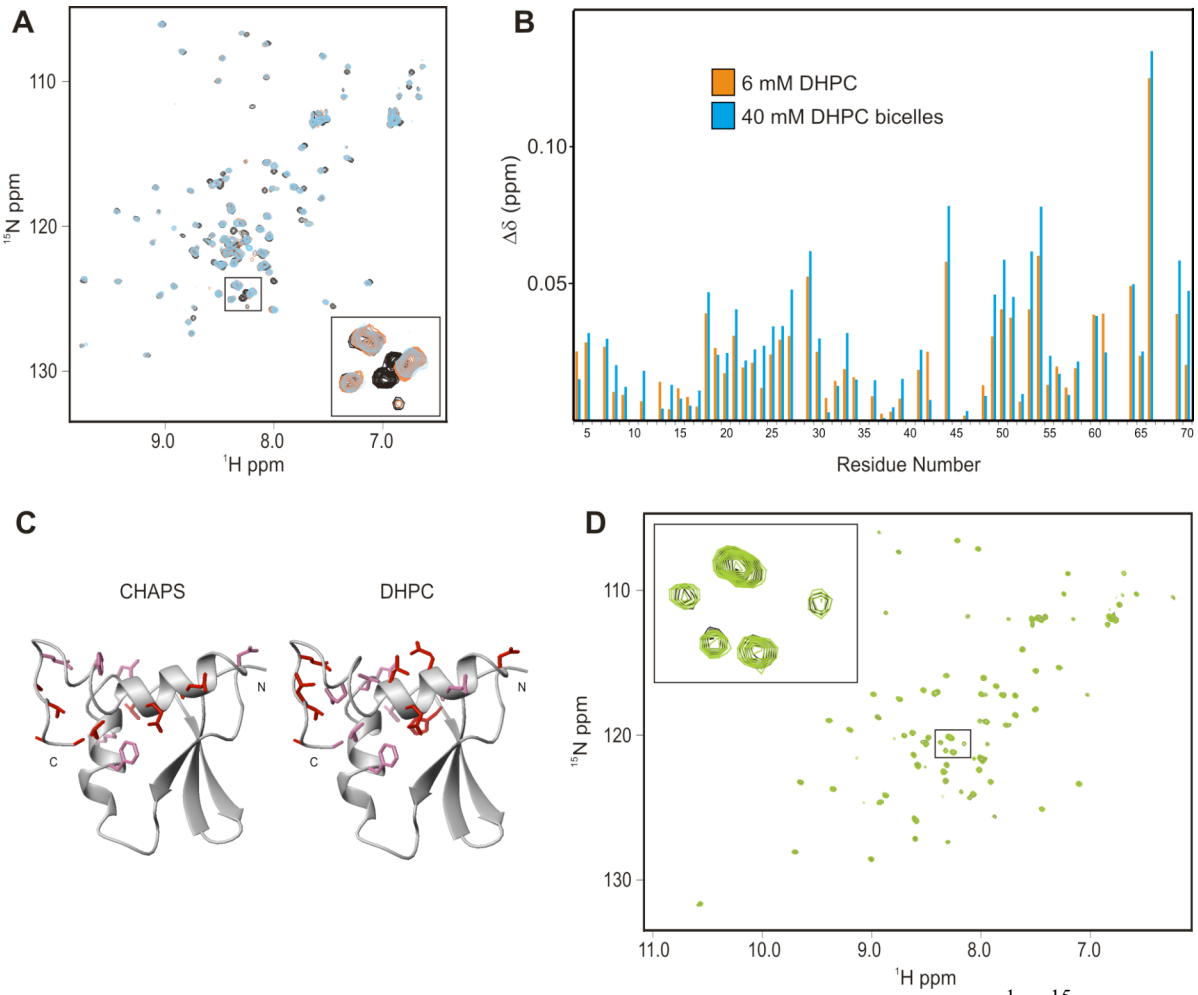


Figure 2.8 Effects of bicelles and lipid vesicles on the *pa*CytD. (A) ^{1}H-^{15}N HSQC spectra of a 350 μM solution of the *pa*CytD with no additive (black) superimposed with the spectrum acquired in the presence of 6 mM CHAPS detergent (orange) or 2% (w/v) CHAPS/DMPC bicelles (blue). A magnified portion of the spectrum is shown as an inset to illustrate the similarity between detergent and bicelle spectra. The pH of this solution was set to 7.2 to allow additional peaks from the unstructured C-terminal parts of the protein to also be monitored. (B) Average amide chemical shift differences (Δδ) measured for the *pa*CytD upon addition of either 6 mM DHPC detergent (orange) or 2% (w/v) DHPC/DMCP bicelles (blue) displayed as a function of residue number. (C) Residues experiencing the largest average amide shift changes upon addition of 2% (w/v) DMPC bicelles made with CHAPS or DHPC, mapped onto the *pa*CytD structure. Red side chains are shown for residues with shift changes that are larger than the average plus one standard deviation calculated for all observable residues, and pink sidechains for residues with shift changes larger than the average plus one standard deviation calculated without the residues shown in red. (D) ^{1}H-^{15}N HSQC spectra of a 240 μM sample in the absence (black) or presence (green) of 2% (w/v) small unilamellar vesicles made with *E. coli* phospholipids. A magnified portion of this spectrum is also shown in the inset.

intensity as would be expected from an interaction with lipid vesicles. Identical results were also obtained with large unilamellar vesicles, a bilayer medium that is closer to native conditions in structure due to the absence of significant membrane curvature (165). This demonstrates that the CytD from *pa*GlpG does not have an intrinsic affinity to bind lipid membranes and must therefore increase *pa*GlpG activity via an alternate mechanism.

2.4 Discussion

2.4.1 Influence of various membrane-mimetic media on paGlpG structure and function.

In the current study, we have shown that it is possible to purify *pa*GlpG in a range of membrane-mimetic environments that can, to varying degrees, support *in vitro* proteolytic activity against a fusion protein containing a central Spitz TM segment. Previous studies had only tested *pa*GlpG activity in Triton X-100-solubilized crude membrane preparations against a model Spitz substrate produced by *in vitro* translation (62). Since we were interested in comparing rhomboid activity for *pa*GlpG in a range of detergent conditions we chose to adopt an assay that had previously been developed for *ec*GlpG (49) that allowed a fusion protein substrate to be purified in the same range of detergent media. Although cleavage of this substrate is slower than rates obtained with *ec*GlpG using shorter substrates, cleavage rates in our study were similar to those obtained for *ec*GlpG in DDM against the same fusion protein containing a different TM segment (49, 71). The relatively slow cleavage rates are also in accord with previous observations with *ec*GlpG that shorter substrates are cleaved more efficiently (62), raising the possibility that larger aqueous domains can impede approach and/or entry of the substrate TM segment into the active site.

Results from the functional assay showed that DDM was the best detergent for preserving a functional state for *pa*GlpG, exceeding the activity obtained with NG, a detergent that had been used for crystallization of the *ec*GlpG core domain. Nonetheless, the lower activity we obtained in NG is similar to previous observations using the Spitz-cleaving rhomboid YqpG from *B. subtilitis* (63), indicating that the longer alkyl chain and/or the larger polar headgroup of DDM are important for maintaining a folded, functional state. In contrast, very low or no activity was observed in our assay when *pa*GlpG was purified in ionic or zwitterionic detergents that have been successfully used for solution NMR studies on smaller membrane proteins. NMR spectra of these samples indicate that the CytD did not tumble together with the main body of the protein-detergent complex, suggesting that destabilizing interactions between the CytD and these detergents could have disrupted a functionally important association with the complex. In DPC this effect seemed to be most severe, since no rhomboid activity could be detected, while some peaks from the unfolded CytD appear in the NMR spectrum.

This was the first study to incorporate a functional rhomboid protease into a bicelle medium, although reconstitution into lipid vesicles has been performed in the past (63, 166). Those studies indicated that, for some rhomboids, lipids may be required to preserve an active conformation (63). However, in the case of *ec*GlpG, activity against a model substrate was lower in most lipid vesicle types tested and virtually absent in *E. coli* phospholipid vesicles. It is possible that *pa*GlpG activity against a transmembrane substrate is also lower in the more constrained environment of a native lipid bilayer, however interactions with free detergent molecules in the bicelle solution could also explain the apparent reduction in *pa*GlpG protease activity. In fact, interactions between bicelle-

reconstituted membrane proteins and free detergent may be a more frequent occurrence than previously appreciated, particularly if hydrophobic mismatch exposes TM helix ends to the aqueous phase. For example, diacyl glycerol kinase (DAGK), a 3-TM helix trimer, may have been influenced by these types of interactions when reconstituted in DMPC bicelles, since its activity was lower than that in DMPC vesicles, or in bicelles composed of phospholipids with longer acyl chains (162). In the case of *pa*GlpG, the potential for CytD interactions with the significant amount of free detergent that exists in these solutions indicate that small bicelles are probably not suitable for functional and structural studies of the full-length rhomboid.

2.4.2 Functional role of the paGlpG cytoplasmic domain.

Previous experiments with the *pa*CytD in Fos16 micelle solutions showed selective attenuation of ^{1}H-^{15}N HSQC resonances from residues in the β-sheet, leading to the assignment of this region as a membrane-binding site (79). Since many of the side chains from these residues are not solvent accessible, and some side chain amides from buried residues also showed attenuation, it was proposed that a significant conformational change in the CytD was required for micelle binding. Our results with DPC and LPPG titrations are in line with these observations and indicate that the conformational change previously detected with Fos16 was likely due to unfolding of CytD. This idea is substantiated by the identification of slow-exchange timescale dynamics for the same part of the protein (79) that could reflect an equilibrium between folded and unfolded states. This equilibrium appears to be shifted toward the unfolded state through interactions with Fos16, with an even larger effect arising from interactions with DPC or LPPG micelles. The Trp41 indole amide peak in corresponding HSQC spectra also provides evidence that interactions with these

detergents similarly perturb the folding state of the CytD, since the chemical shift change induced by Fos16 binding is comparable to that seen with LPPG and DPC. Altogether the data indicates that the *pa*CytD β-sheet is a conformationally flexible region that does not comprise a micelle-binding surface itself, but instead is just one part of the protein that binds to the micelle in a denaturing interaction. It is unlikely that this interaction is physiologically relevant, especially since the isolated *pa*CytD did not bind DMPC bicelles or *E. coli* lipid vesicles. Instead, it appears that the loose packing of micellar detergent facilitates interactions with the small population of unfolded *pa*CytD, while closer packing interactions between lipid molecules in the lipid bilayer is enough to restrict access to the hydrophobic core.

Given that the *pa*CytD does not have an intrinsic affinity for DDM micelles or phospholipid bilayers, it is interesting that *pa*CytD signals from the full length protein were not observed in these media types. This raises the possibility that either; 1) dynamic processes are occurring on a timescale that is unfavorable for NMR (*i.e.* μs - ms), or; 2) the CytD is interacting with some other part of the rhomboid that is anchored to the micelle. The first scenario is less likely since the high quality of the ^{1}H-^{15}N HSQC spectrum obtained with the isolated *pa*CytD indicates that this domain does not have an intrinsic tendency to undergo spectroscopically unfavorable conformational exchange processes. Moreover, previous relaxation studies on the isolated CytD showed that only a small number of residues undergo millisecond to microsecond timescale dynamics (79). Therefore, it is more likely that the absence of signals from the CytD in NMR spectra of functional samples of full-length *pa*GlpG can be attributed to interactions between the CytD and some other part

of the protein in a way that would anchor it to the large detergent-protein complex and/or alter its dynamic properties.

The CytD interaction appears to be functionally relevant since, as we have shown with the model Spitz substrate, deletion of the CytD from *pa*GlpG reduces *in vitro* activity in DDM. This result is in agreement with previous observations with the TMD from *E. coli* GlpG, which was shown to have significantly reduced activity against a water-soluble casein substrate compared to the full-length protein (65). While this lower activity might be explained by the loss of a substrate-binding function for the CytD, only the TM segment from Spitz was included in the model substrate used in our assay, which may not be accessible for binding by the CytD. Also, the water-soluble substrate used with *ec*GlpG bore no resemblance to typical rhomboid substrates, making it unlikely that this substrate would be bound by the CytD. Therefore, the CytD may enhance catalysis by altering the structure and/or stability of the core catalytic domain, a function that would be most straightforwardly achieved through direct interactions between the CytD and some part(s) of the TMD.

Potential regions of the TMD that could be involved in catalytically important CytD interactions include cytoplasmic loops linking TM helices 2 and 3 and/or 4 and 5. The latter loop is particularly interesting given its proximity to TM helix 5, a region that has displayed conformational variability in different crystal structures and has been postulated to be important for gating substrate entry into the active site (66, 75). In addition to these loops, some TM helix ends could also serve as docking points for the CytD, since analysis of a recent high resolution *ec*GlpG structure indicated that the thickness of the hydrophobic core of the lipid bilayer surrounding the rhomboid may be only 20 Å (65). If that were the case

then a significant part of the C-terminal sides of the 2nd, 4th and 6th TM helices would be exposed to the aqueous phase, providing a larger potential surface for docking.

In summary, this study identified a functional role for the *pa*GlpG CytD that we hypothesize involves an interaction with the TMD. It will be interesting to characterize this interaction, although the absence of a solved structure for the *pa*GlpG TMD complicates this analysis. In contrast, the *ec*GlpG rhomboid has a known TMD structure, and also contains a homologous cytoplasmic domain (Fig. 2.1). For this reason our subsequent studies focused on identifying potential interactions between the TMD and CytD of the *ec*GlpG rhomboid. However, as will be revealed in the following chapters, different types of interactions at the membrane-aqueous interface were observed over the hypothesized TMD-CytD interaction, leading to the assignment of an alternate role for the CytD.

Chapter 3: Novel interactions at the N-terminus of the E. coli rhomboid protease

Contributions of collaborators:

All results obtained using the activity-based protein profiling probe were done in collaboration with the laboratory of Dr. John Pezacki (National Research Council of Canada). Dr. David Blais performed activity-based labeling and in-gel fluorescence on constructs of *ec*GlpG. The FP-PEG-rhodamine probe was synthesized by Shuqiong Lin.

3.1 Introduction

In this chapter, I evaluate the main hypothesis originating from the previous chapter. We propose that functionally important interactions occur between the cytoplasmic N-terminal region of the GlpG rhomboid and the cytoplasmic face of the TMD catalytic core. This hypothesis was evaluated through functional analysis of a number of *E. coli* GlpG (*ec*GlpG) mutants, made in both cytoplasmic regions of the TMD and the cytoplasmic domain, since the mutation of residues important for this type of interaction would be expected to lower protease activity. This work allowed us to identify residues in the cytoplasmic face of the TMD that play a role in protease activity. We also identified a critical sequence just outside the TMD that is absent in all x-ray crystal structures of GlpG, and was not previously known to play a role in rhomboid function. In addition, a novel application of an activity-based profiling probe was developed that provides a new way for rhomboid activity to be assessed in crude membrane extracts. In summary, this chapter identifies the minimal sequence of *ec*GlpG required for maximal function, while describing a new method for the study of rhomboid function.

3.2 Materials & Methods

3.2.1 Plasmids.

The C-terminally hexahistidine tagged *E. coli* GlpG rhomboid in a pET15b vector (62) was kindly provided by Dr. Christopher Koth's group of the Ontario Centre for Structural Genomics at the University of Toronto. This plasmid was used to create expression vectors for the C-terminal transmembrane domain consisting of residues 60-276 (TMD60), 81-276 (TMD81) and 91-276 (TMD91). TMD60 and TMD81 were cloned into

pET25b and TMD91 was cloned into pET30a using primers synthesized by Integrated DNA Technologies, Inc. The C-terminal His_6-tags of TMD60 and TMD81 is preceded by a TEV protease cleavage site, while the C-terminal His_6-tag of TMD91 is preceded by a non-native Leu-Glu sequence introduced by a XhoI restriction site required for sub-cloning. All mutants were generated using the Quik-Change site-directed mutagenesis protocol (Stratagene). Primers for truncation and mutant construction are shown in Tables 3.1 and 3.2, respectively. All constructs were verified by sequencing performed at the Ontario Genomics Innovation Centre at the Ottawa Health Research Institute.

3.2.2 Protein expression and purification.

Expression vectors were transformed into *E. coli* C43 (DE3) competent cells and grown at 37°C in LB media with 100 μg/mL ampicillin or 50 μg/mL kanamycin (TMD91 only). When the optical density of the culture at 600 nm reached 0.5-0.7 the temperature was reduced to 16°C, and expression was induced with 1 mM isopropyl β-D-thiogalactopyranoside (IPTG) and allowed to proceed for 16 to 20 h. Purification of detergent-solubilized protein was performed as described in Section 2.2.3, with some modifications. Briefly, the bacterial pellet from 500 mL of culture was resuspended in 20 mM HEPES buffer (pH 7.3) containing 100 mM NaCl, 5 mM $MgCl_2$, 10% glycerol and Complete™ EDTA-free protease inhibitor cocktail (Roche). Cells were lysed by sonication and centrifuged at 50,000 x g for 1 hour. The insoluble membrane fraction was resuspended in 50 mM HEPES buffer pH 7.3 containing 1% (w/v) DDM, 200 mM NaCl, 10 % glycerol, 5 mM imidazole and protease inhibitor cocktail for 1h and then centrifuged for 30 minutes at 50,000 x g. The DDM-solubilized protein in the supernatant was applied to a nickel affinity column equilibrated in the same buffer without detergent and then washed in 50 mM

Table 3.1 Primer sequences for truncation constructs.

Construct	Plasmid	Direction	Sequence[a]
TMD60	pET25b	Forward	5’ – GCACGTCTCATATGGCAGATCCGCGTTATCTGGC – 3’
		Reverse	5’ –TTTTCTCTCTCGAGGCTAGCTTTTC– 3’
TMD81	pET25b	Forward	5' –GCACGTCTCATATGCGCCGTTATCCTTTCTTTGCC– 3’
		Reverse	5' –TTTTCTCTCTCGAGGCTAGCTTTTC– 3’
TMD91	pET30a	Forward	5’ –GCACGTCTCATATGGAACGCGCAGGTCCGGTAAC– 3’
		Reverse	5’ –TTTTCTCTCTCGAGGCTAGCTTTTC– 3’

[a] Restriction enzyme sites are underlined in the primer sequence. Forward primers contain the NdeI site and reverse primers contain the XhoI site.

Table 3.2 Mutagenic primers used for site-directed mutagenesis.

Mutant[a]	Sequence [b]
N9A	5' –GATTACCTCTTTTGCT**GCC**CCCCGCGTGGCGCAG– 3'
R11A	5' –CTTTTGCTAACCCC**GCC**GTGGCGCAGGCGTTTG– 3'
D18A	5' –CGCAGGCGTTTGTT**GCT**TACATGGCGACGC– 3'
W38A	5' –CCAAAGCGATGTC**GCG**CTGGCGGATGAGTCC– 3'
F55D	5' –CGGAGCTGGCGCGT**GAT**CTCGAAAACCCGGC– 3'
K167A	5' –GGTGCGGTGGAA**GCA**CGCCTCGGTAGCG– 3'
S201A	5' –GTTTGGCGGGCTT**GCT**GGCGTGGTGTATG– 3'
E216A	5' –CTGGCTACGTGGC**GCA**CGCGATCCGCAAAG– 3'
R217A	5' –GGCTACGTGGCGAA**GCC**GATCCGCAAAGTGGC– 3'
D218A	5' –CGTGGCGAACGC**GCT**CCGCAAAGTGGC– 3'
Q220A	5' –GGCGAACGCGATCCG**GCA**AGTGGCATTTACCTGC– 3'
Y224A	5' –CCGCAAAGTGGCATT**GCC**CTGCAACGTGGGTTAATTATC– 3'
Q226A	5' –CGCAAAGTGGCATTTACCTG**GCA**CGTGGGTTAATTATCTTTGCG– 3'
R227A	5' –GTGGCATTTACCTGCAA**GCT**GGGTTAATTATCTTTGCGC– 3'

[a] All mutant *ec*GlpG sequences are in the pET15b plasmid.

[b] Only forward primers are shown since the reverse is complementary. Mutated nucleotides are shown in bold with residue codon underlined.

HEPES buffer pH 7.3 containing 0.1% DDM, 500 mM NaCl, 10% glycerol and 30 mM imidazole. A similar wash with 200 mM NaCl was performed prior to elution in 50 mM HEPES buffer 7.3 containing 0.1% DDM, 200 mM NaCl, 10% glycerol and 250 mM imidazole. Eluted protein was concentrated using a 30 kDa molecular weight cut off (MWCO) ultracentrifugation device (Amicon Ultra) and applied to a Superdex-200 10/300 GL size exclusion column using an AKTA FPLC (GE Healthcare) equilibrated in 50 mM HEPES buffer pH 7.3 containing 0.1% DDM, 200 mM NaCl, 10% glycerol and 100 μM EDTA. Limited proteolysis was performed as described (65) using trypsin (Sigma) in a 1:20 (by mass) ratio at 4ºC overnight, and the digest was directly reapplied to the size exclusion column to isolate the transmembrane domain. To minimize differences in sample purity and detergent composition between trypsin digested versus truncation or mutant constructs that did not undergo digestion, non-digested samples were also reapplied directly to the size exclusion column to perform one more round of purification (Fig. 3.1).

3.2.3 MALDI-MS

MALDI-MS was performed on purified samples using a Microflex MALDI-TOF mass spectrometer (Bruker) set to a m/z window between 5 and 40 kDa or between 10 and 100 kDa. The ultra thin layer method with alpha-cyano-4-hydroxycinnamic acid (4-HCCA) as the matrix was used for ionization of the membrane proteins (167). Briefly, a stainless steel MALDI sample plate was cleaned with methanol and deionized water and 30 μL of thin layer substrate was spread over the plate surface. The thin layer substrate consisted of 1 part saturated 4-HCCA (2 parts acetonitrile, 1 part water and 0.1% final Trifluoroacetic acid) mixed with 3 parts isopropanol. Once dried, the surface was wiped in one direction using a

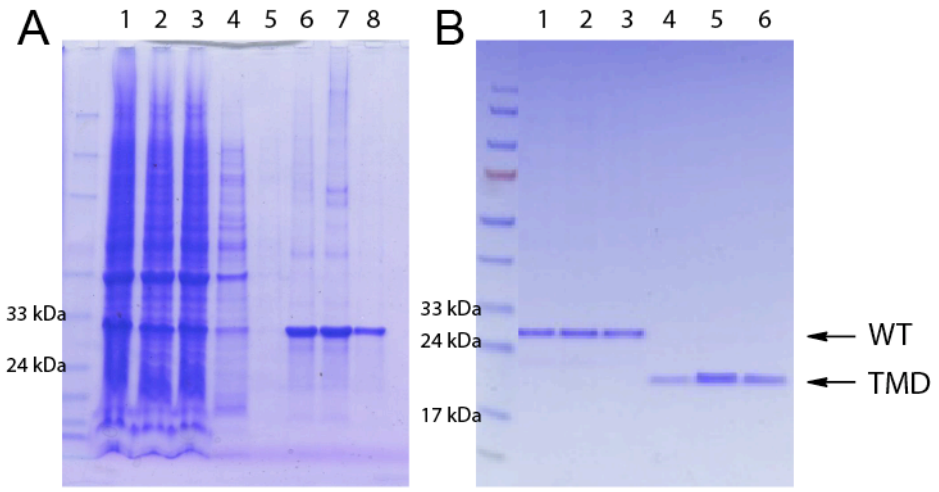


Figure 3.1 Purification and limited trypsin digestion of *ec*GlpG. (A) Gradient gel (4-20%, Pierce) of nickel affinity chromatography purification. Lane 1- Solubilized crude rhomboid in DDM; Lane 2 - Flow through indicating unbound rhomboid; Lane 3 - Flow-through after reapplication of solubilized membrane extract to the same column; Lanes 4 & 5 - Contaminants eluted after washing nickel affinity column with buffer containing 30 mM imidazole. Lanes 6 – 8: elution fractions with 250 mM imidazole. The predicted molecular mass of *ec*GlpG is 33.7 kDa. (B) 15% acrylamide gel containing eluted fractions of *ec*GlpG from the second round of size exclusion chromatography following incubation with or without trypsin. Lanes 1 – 3: Undigested *ec*GlpG FPLC fractions taken from the peak eluting between 11 and 12.5 mL; Lanes 4 – 6: FPLC fractions taken from the peak eluting between 12.5 and 14 mL obtained after overnight trypsin digestion of *ec*GlpG at 4ºC.

Kimwipe to create the ultra thin layer of substrate. The protein sample was diluted between 1:10 – 1:100 times in sample matrix (consisting of saturated 4-HCCA in 3 parts formic acid, 1 part deionized water and 2 parts isopropanol) and 0.5 μL was spotted on the plate. The precipitate was dried using vacuum suction then washed with 2 μL of ice cold trifluoroacetic acid (0.1%) in deionized water.

3.2.4 Rhomboid protease assay using fluorogenic casein substrate

Wild-type, truncated and mutant *ec*GlpG were assayed for activity using BODIPY-labeled casein (Invitrogen). This assay was based on previously described protocols (65), where 5 μg BODIPY FL casein was incubated with 0.8 μM *ec*GlpG in 50 mM HEPES buffer pH 7.3 containing 0.1% DDM, 200 mM NaCl, 10% glycerol and 100 μM EDTA at 37°C. Fluorescence excitation of a 200 μL volume was induced at 485 nm and emission monitored at 530 nm using a SpectraMax 5 microplate reader (Molecular Devices) in the top-reading mode. A black quartz microplate (Hellma) was set to mix for 5 seconds prior to the first reading, and each measurement consisted of 6 readings at 37ºC. All samples were assayed in duplicate against a buffer blank, with assays on each mutant being repeated with at least two different purified samples. Specific activities were calculated using the slope of relative fluorescence units (rfu) over 30 minutes after a 10 minute equilibration period.

3.2.5 Active rhomboid labeling with FP-PEG-rhodamine

Purified rhomboid (0.5 mg/ml) was incubated in the dark with 10 nM FP-PEG-rhodamine (from a 0.01 μM stock in 2% DMSO, final DMSO concentration of 0.2%) in a 20 μl volume for 2 minutes at 37°C in 50 mM HEPES buffer pH 7.3 containing 0.1% DDM, 200 mM NaCl, 10% glycerol. Crude membrane extracts were obtained (as described in Section 3.2.2) by resuspending the insoluble membrane fraction in 50 mM HEPES buffer

pH 7.3 containing 1% (w/v) DDM, 200 mM NaCl, 10 % glycerol, 5 mM imidazole and protease inhibitor cocktail. 2 mg/ml of this extract was incubated with 10 nM FP-PEG-rhodamine in 20 µl for 1 hour at 37ºC (Fig. 3.2). Reactions with purified rhomboid and crude membrane extracts were quenched by precipitating the protein with five volumes of ice-cold acetone at -80ºC for 1 hour to remove unreacted probe. After aspiration of acetone, the precipitate was air dried before resuspension in SDS-PAGE sample buffer and separation on a 15% gel (1.5 mm thickness) at constant voltage (100 V). The gels were scanned for fluorescence using the FMBIO III (Hitachi Solutions Ltd, Tsurumi-ku Yokohama, Japan) to visualize the active protein. The amount of FP-PEG-rhodamine labeling in crude membrane extracts was measured by integrating fluorescent band intensities and subtracting background intensity from these measurements. FP-PEG-rhodamine labeling was normalized for differential protein expression using relative intensities measured from coomassie-stained SDS-PAGE. Labeling of all constructs was performed with at least two different crude membrane preparations.

3.3 Results

3.3.1 Isolation and activity of the ecGlpG TMD core.

To confirm the functional importance of the *ec*GlpG cytoplasmic N-terminal domain we performed limited proteolysis with trypsin to isolate the TMD core as had been previously done for x-ray crystallography (64). Similar to previous observations using fluorogenic casein as a rhomboid substrate (65), this TMD sample displayed roughly half the specific activity of the full-length *ec*GlpG over the range substrate concentrations tested (Fig. 3.3a). Both the full-length and TMD samples gave rise to Michaelis-Menten kinetic

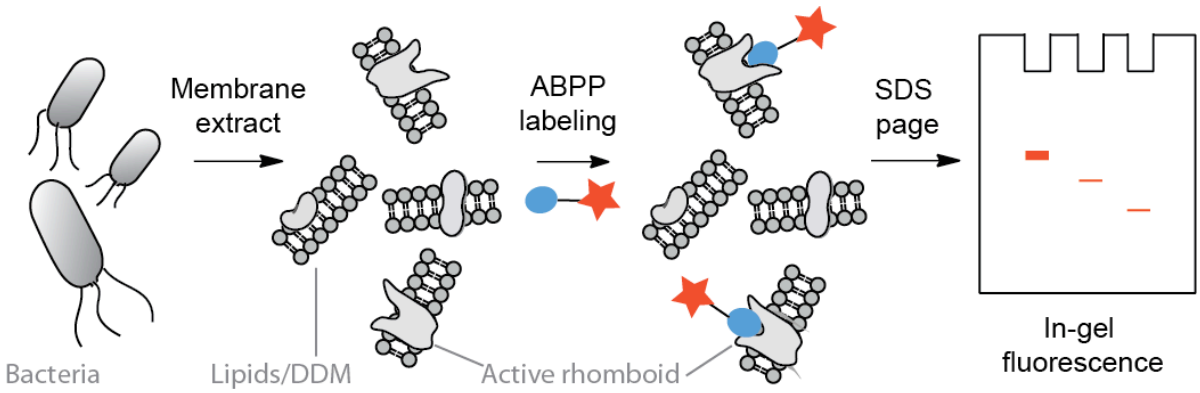


Figure 3.2 Schematic diagram of activity-based profiling approach for monitoring the activity of *ec*GlpG in crude membrane extracts. Bacterial membranes are disrupted and isolated in 1% DDM, then incubated with FP-PEG-rhodamine. Active protein is assessed by in-gel fluorescence.

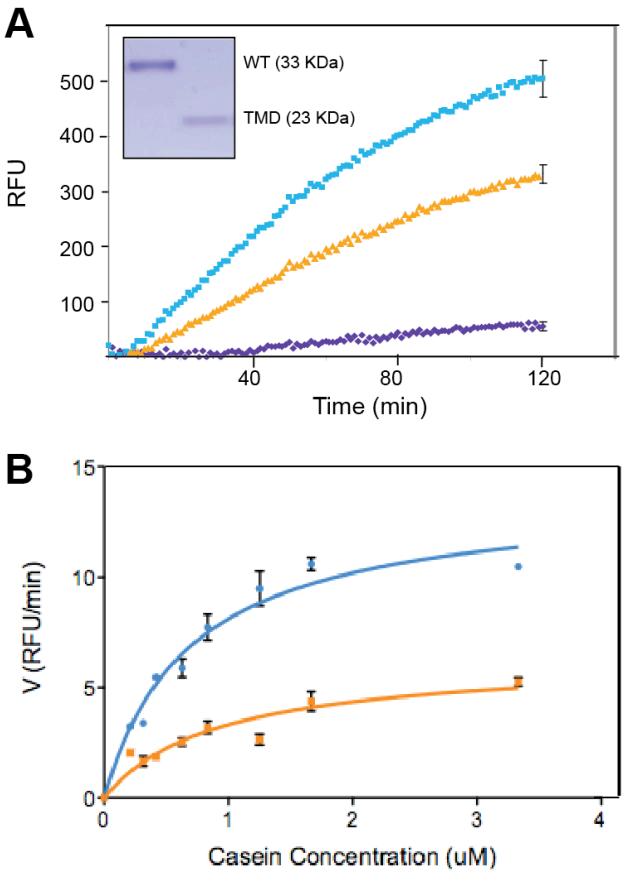


Figure 3.3 Full-length *ec*GlpG is more active than trypsin-generated TMD. (A) Rate profiles of full-length *ec*GlpG (blue), inactive S201A (purple) and trypsinized TMD (orange). A Coomassie-stained gel of purified full-length *ec*GlpG and TMD samples used in the assays is shown in the inset box. Casein substrate (5 μg) was incubated with 0.8 μM protein in 50 mM HEPES pH 7.3, 0.1% DDM, 200 mM NaCl, 10% glycerol and 100 μM EDTA for 2 hours at 37°C and the relative fluorescence (RFU) released by casein cleavage was measured. Initial rates were determined to be 0.13 ± 0.01, 0.067 ± 0.01 and 0.013 ± 0.007 RFU/s/μM for full-length, TMD and S201A, respectively. (B) Michaelis-Menten plot of initial rate (V) versus casein substrate concentration for the full-length (blue) and TMD (orange) variants of *ec*GlpG. The Vmax was calculated to be 0.13 ± 0.02 and 0.29 ± 0.02 RFU/s/μM for the TMD and full-length *ec*GlpG, respectively. The Km values were found to be 0.69 ± 0.1 and 0.90 ± 0.3 μM for TMD and full-length, respectively, and therefore not significantly different.

profiles, with the V_{max} for the TMD being roughly half that of the full-length protein, while K_m was not altered (Fig. 3.3b).

3.3.2 Interactions of the TMD cytoplasmic face.

One possible explanation for the reduction in protease activity observed upon removal of the cytoplasmic domain is that this domain may normally interact with the TMD in a way that augments rhomboid activity. To determine if these interactions exist, alanine mutations were made in the CytD and cytoplasmic face of the TMD at sites selected based on their sequence conservation in *Enterobacteriaceae* species (Appendix 1), since these rhomboids all contain a homologous cytoplasmic domain. The activity of these mutants was assessed using the same casein-based assay (Fig. 3.3). Two of these were found to have greatly reduced activity compared to wild-type *ec*GlpG, namely E216A and R227A (Fig. 3.4) with activity levels essentially the same as the active site serine knockout mutant S201A. A small but significant decrease in activity was also observed for K167A, D218A and Y224A, which showed 30-40% lower activity than wild-type *ec*GlpG. Mutations made in the CytD, however, did not significantly affect protease activity suggesting that an interaction with the TMD involving this domain may not be required for activity against this model substrate.

Since CytD mutations did not significantly alter rhomboid activity against casein, and limited proteolysis using trypsin removes most of the N-terminal extramembranous sequence (66), three TMD truncation constructs were created to delineate more precisely the functionally important cytoplasmic region. Specifically, we made TMD60 (*ec*GlpG residues 60-276) that lacks only the structured part of the cytoplasmic domain, TMD91 (residues 91-276) that omits all N-terminal residues not observed in the *ec*GlpG TMD x-ray structures,

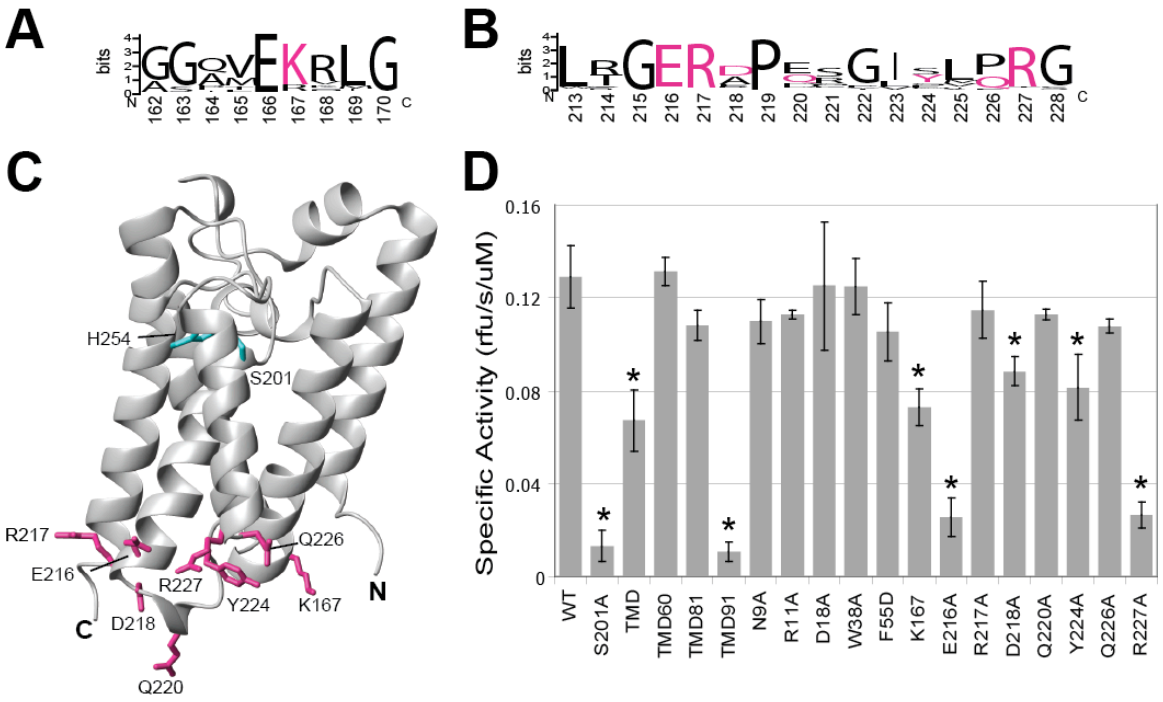


Figure 3.4 Identification of functionally important residues at the cytoplasmic face of the TMD. (A) and (B) depict sequence conservation for residues in the TMD on the cytoplasmic aqueous-exposed side of the protein among *Enterobacteriaceae* species using WebLogo (168). Residues that were chosen for mutation to alanine are shown in pink. (C) The location of mutations in the *ec*GlpG TMD (pdb: 2IC8) cytoplasmic face. Charged and polar residues chosen for mutation to alanine are highlighted in pink, and active site residues in blue. (D) Comparison of specific activities for trypsin-isolated TMD, *ec*GlpG truncation mutants and single-site mutants against the fluorogenic casein substrate. Activities that were significantly lower than WT ($p < 0.05$) are identified by asterisks.

and TMD81 (residues 81-276) which is slightly longer than the TMD obtained after chymotrypsin digestion (65). These constructs were expressed, purified and assayed in the same manner as done for the full-length protein. As shown in Figure 3.4d, the construct lacking the structured cytoplasmic domain (TMD60) displayed the same level of activity as the full-length protein, indicating that this part of the cytoplasmic region is not responsible for the decrease in activity found after trypsin digestion. Shortening the sequence by 20 residues to remove most of the N-terminal loop in TMD81 also appeared to have no significant effect on protease activity. Surprisingly, TMD91 was essentially inactive, even though *ec*GlpG x-ray structures begin at E91 or later in the sequence (64, 65, 72). This suggested that the additional N-terminal amino acids (residues 81 – 90) are essential for maintaining an active state for the TMD.

Although TMD91 showed no activity in the casein assay, the TMD product of limited trypsin proteolysis did show significant activity, albeit at lower levels than the full-length protein. However, the loop connecting the cytoplasmic domain to the TMD has multiple potential cleavage sites that could produce a mixture of active and inactive digestion products. In order to evaluate the homogeneity of the purified trypsin digest, MALDI-TOF mass spectrometry was performed on a purified TMD produced by trypsinolysis (Fig. 3.5). These spectra showed that this sample is comprised of a mixture of cleavage products and is consistent with the decrease in V_{max} and unchanged K_m that was observed relative to the full-length protein, since the trypsin digest produced a mix of active and inactive TMD.

3.3.2 Active site labeling of purified rhomboid protease

As a complementary assay to probe the activation state of the catalytic serine in these

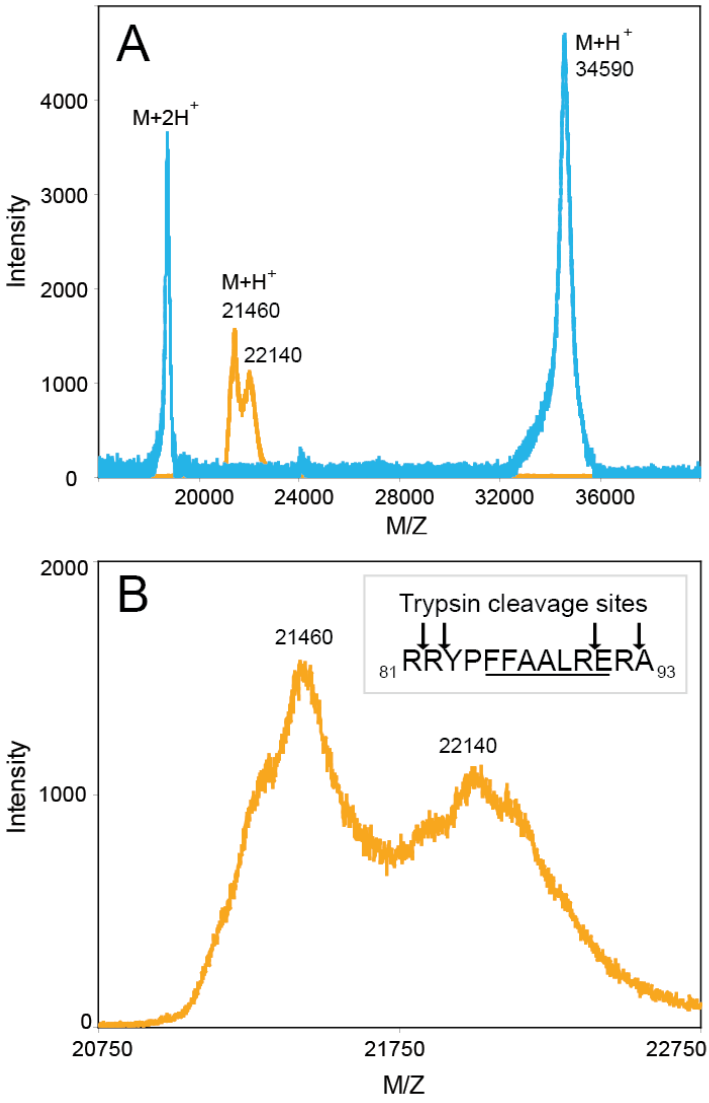


Figure 3.5 (A) MALDI mass spectra of full-length *ec*GlpG (blue) compared to the trypsin-isolated TMD (orange). (B) Expanded view of trypsin cleavage product MS spectrum. Trypsin cleavage sites of the N-terminal loop are indicated in the inset box and are accompanied by 5 additional cleavage sites at the C-terminus.

mutants, we also used a rhodamine-tagged fluorophosphonate inhibitor (FP-PEG-rhodamine, Figures 3.2 and 3.6a) known to be specific for serine hydrolases (169). This type of probe has been used in activity-based protein profiling of soluble serine hydrolases (170, 171) as well as the membrane-associated carboxylesterase 1 (172) and is expected to enter the rhomboid active site from the aqueous-phase. Covalent modification of purified wild-type *ec*GlpG and inactive S201A mutant was evaluated by measuring the fluorescence of the labeled rhomboid band in an SDS-PAGE gel of the precipitated reaction product. As shown in Figure 3.6b, purified wild-type *ec*GlpG was labeled with FP-PEG-rhodamine while the active-site serine mutant S201A mutant was not, demonstrating the feasibility of using the FP-PEG-rhodamine probe to directly probe the rhomboid activity site. Moreover, the truncation mutants showing activity in the casein assay were efficiently labeled while TMD91 was not, confirming that the active site serine was no longer in an activated state.

An interesting feature of this probe is its potential to reveal differences in activation states that may not be apparent in the casein assay. This was observed to be the case for E216A which, like R227A, was largely inactive in the assay with the bulky casein substrate. In contrast the FP-PEG-rhodamine activity-based probe was able to modify E216A, but not R227A (Fig. 3.6b) suggesting that the lower activity seen for the E216A mutant in the casein assay may arise from hindered entry of a larger polypeptide-based substrate.

3.3.4 Active site labeling of the rhomboid in crude membrane extracts.

A significant advantage offered by this serine protease activity-specific probe is the potential to perform this test on crude membrane extracts without the need for purification (Fig. 3.6c). Crude membrane extracts are expected to preserve protein-bound lipids that could have an effect on rhomboid function (62, 173). This was particularly useful for our

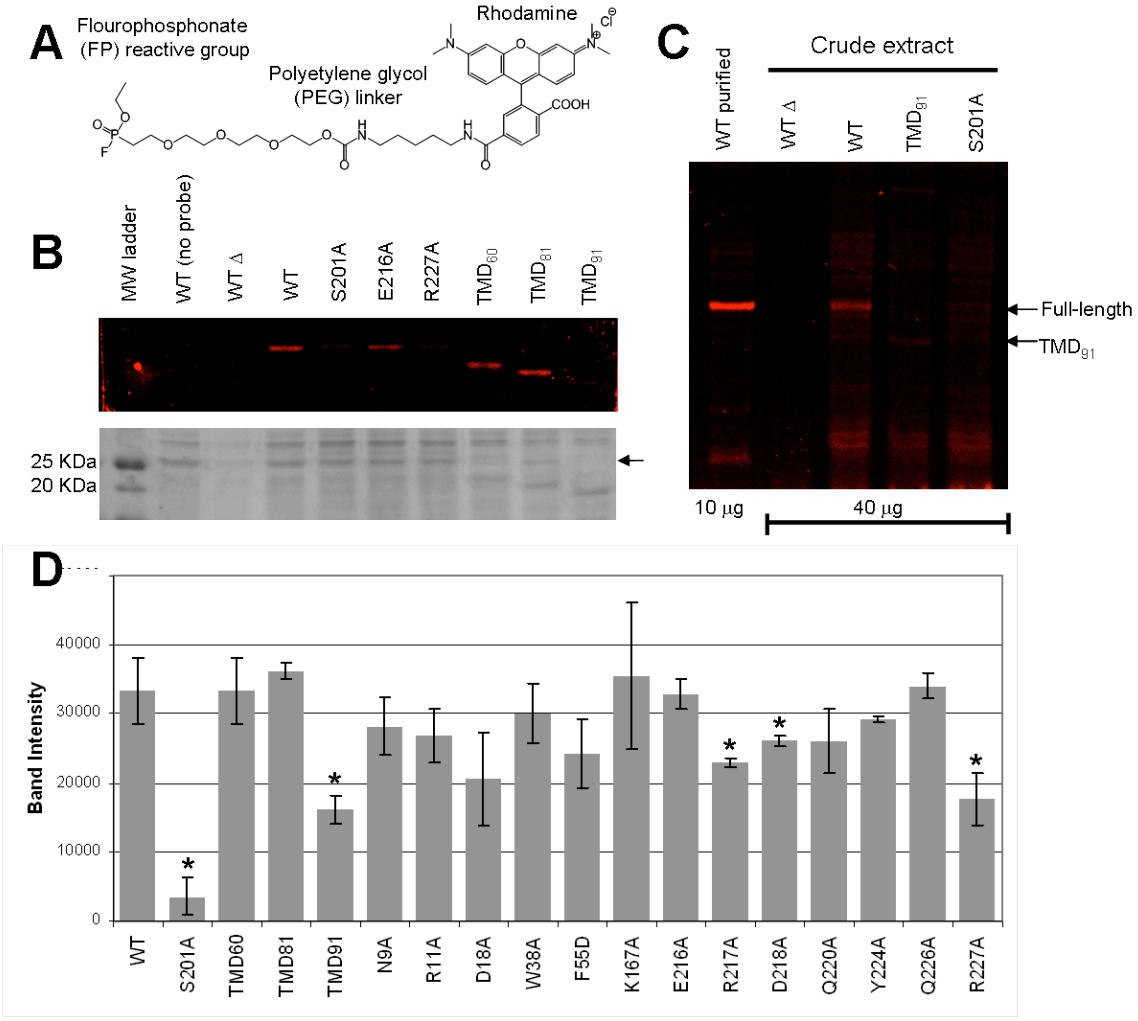


Figure 3.6 Rhomboid activity in purified and crude extracts can be evaluated using a fluorophosphonate fluorescent probe. (A) Chemical structure of the fluorophosphonate probe (FP-PEG-rhodamine). (B) Activity of wild-type, mutant and truncated versions of the GlpG rhomboid protease. Visualization was done by stimulation of rhodamine fluorescence (top), followed by silver stain shown below. Purified rhomboid samples were labeled for two minutes in 50 mM HEPES pH 7.3, 0.1% DDM, 200 mM NaCl, 10% glycerol and 100 μM EDTA with 10 nM FP-PEG-rhodamine (from a 0.5 μM stock in DMSO, final DMSO concentration of 0.2%) before precipitation with ice-cold acetone to isolate the protein fraction from unreacted probe. The expected mobility of full-length WT and mutant GlpG is indicated by the arrow. (C) Rhomboid activity can be probed in crude membrane extracts. Extracts were incubated with the probe for 1 hour and 40 ug was loaded onto the gel. (D) Comparison of gel band intensities measured for TMD truncation constructs and cytoplasmic face TMD mutants isolated in crude membrane extracts. Intensities were measured using ImageJ and normalized for variations in expression. Significant differences in labeling from WT levels are identified by asterisks.

studies since the fluorogenic casein substrate cannot be used to assay the activity of crude membrane extracts due to high background activity (data not shown). Figure 3.6c shows selective labeling of a band corresponding to full-length wild type *ec*GlpG, but not S201A *ec*GlpG, indicating that the labeling observed in purified extracts mirror the activity under more native-like conditions.

Since the more native-like environment offered by crude membrane extracts has the potential to alter the activity of rhomboid mutants in our screen, crude extract labeling was done for all mutants and compared to that for the wild-type *ec*GlpG (Fig. 3.6d). Active-site labeling in crude membrane extracts was found to be the same for most mutants tested, with only R227A and TMD91 showing large differences in labeling relative to wild-type. As was observed in the casein assay, D218A showed a small but statistically significant decrease in labeling. R217A also showed a similarly small decrease in labeling, suggesting that the E216 – D218 stretch of amino acids plays a role in maintaining a catalytically active structure.

3.4 Discussion

3.4.1 Role of the critical N-terminal sequence.

Work described in Chapter 2 on the *Pseudomonas* GlpG-like rhomboid suggested that the cytoplasmic and transmembrane domains might interact in a way to enhance activity. The limited proteolysis of *ec*GlpG that was used to isolate the TMD core for crystallization was also previously found to reduce activity by about 50% (65), as we have observed here (Fig. 3.3). Surprisingly, the structured domain of the CytD (residues 1 – 60) was not responsible for this decrease in activity, although it is still possible that this domain

plays a functional role for the currently unidentified *ec*GlpG substrate. Instead, we found that residues 81 to 91 are needed for maximal activity, with TMD91 showing almost no activity in the casein assay. In fact, this truncation mutant was even less active than the fragment produced by limited proteolysis with trypsin. However, this critical sequence contains four trypsin cleavage sites, potentially giving rise to a mixture of shorter inactive, and longer active cleavage products. MALDI-TOF mass spectrometry performed on the purified cleavage mixture confirmed the presence of multiple cleavage products, with at least 5 different species between ~20 – 21 kDa in this sample (Fig 3.5). This indicates that the observed decrease in activity arises from a heterogeneous sample containing both active and inactive cleavage products, consistent with the decrease in V_{max} and unaltered K_m.

This is the first study to implicate a functional role for residues R81-R90, a segment that is immediately adjacent to the N-terminus of TM1. While there is no structural information yet available for this part of the protein, secondary structure analysis suggests the potential for amphipathic helix formation between F85 and E91 (150). This could facilitate a direct interaction with the lipid bilayer that may provide additional membrane positioning function. In fact molecular dynamics simulations of the TMD in model membranes suggest that the greatest amount of rhomboid-induced membrane deformation occurs under loop 1 (L1) at the cytoplasmic side around the N-terminus (76), which might facilitate this interaction (174, 175). These simulations also suggest that any changes to the interaction between L1 and the membrane are transmitted to the active site, raising the possibility that changes in membrane structure induced by amphipathic helix binding on the cytoplasmic side could be transmitted to L1, and therefore the active site. However, a membrane-binding role for this segment does not provide a straightforward explanation for

the observed loss of activity in detergent-solubilized TMD91, since interactions with the micelle would be expected to differ from those with the native lipid bilayer.

An alternate possibility that is also consistent with these results is that this region is involved in a specific structural role, potentially involving interactions with functionally important residues identified in this study. In fact Y83 is absolutely conserved among *Enterobacteriaceae* species (Appendix 1), a level of conservation that would be more consistent with a specific structural role rather than a membrane-interacting role that is normally fulfilled by a wide range of hydrophobic residues. Residues in this N-terminal sequence could interact with functionally important residues K167, Y224 and/or R227 (Fig. 3.7a). For example, K167 on TM2 is proximal to the N-terminus and could form interactions with this N-terminal extension. This is a region of low variability among crystal structures, with the highly conserved E166 of TM2 forming extensive interactions with a number of residues in the N-terminus of TM1. However, since K167A retains wild-type activity in crude membrane extracts, stabilization of this region may be more important when the rhomboid is solubilized in detergent micelles. Meanwhile it is also possible that residues in the N-terminal R81-R90 sequence could form an extended structure to interact with more remote regions of the TMD such as Y224 and/or R227 (Fig. 3.7a).

3.4.2 Interactions of the TMD cytoplasmic face.

Another region of the cytoplasmic side of *ec*GlpG that appears to be important for maintaining a functional state is the group of charged residues at the C-terminus of TM4 E216, R217 and D218. This helix contains the precisely positioned active site serine at its N-terminal side, which could be altered by these mutations. For example, R217 appears to hydrogen bond with S269 on TM6 (Fig. 3.7b), an interaction that would be important for

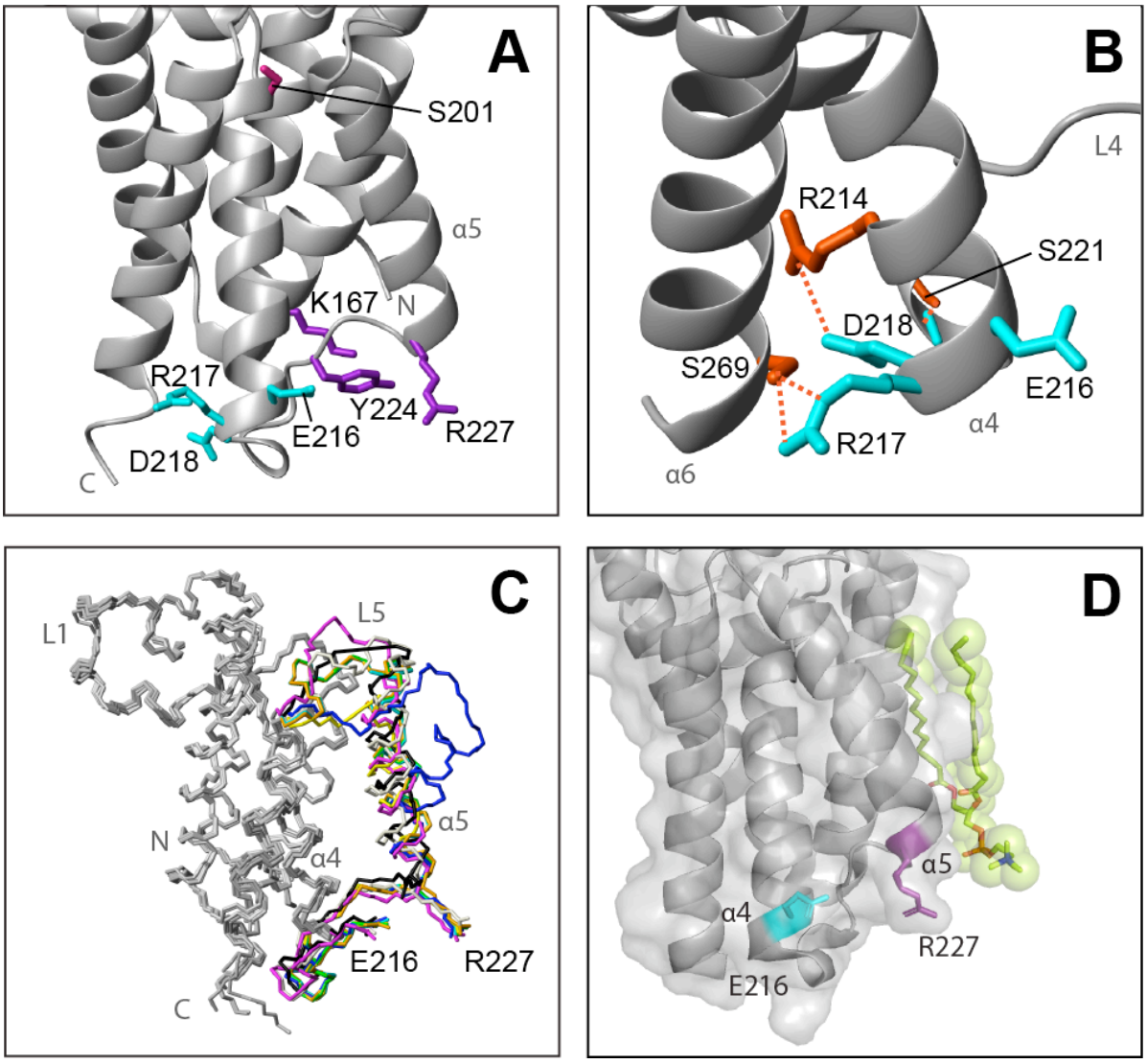


Figure 3.7 Interactions with the lipid bilayer and/or cytoplasmic face of the TMD may modulate rhomboid activity. (A) Residues with significant reduction in activity in DDM micelles and/or crude membrane extracts. Functionally important cytoplasmic-facing residues are: K167, E216, D218, Y224 and R227 in detergent micelles; and R217, D218 and R227 in crude membrane extracts. Residues in cyan are proposed to be involved in TM4 positioning and residues in purple may interact with the short sequence N-terminal to TM1. The active site serine is shown in pink. (B) Interactions involved in TM4 stabilization between residues identified in this study (cyan) and bonding partners (orange) using the structure in 2XOV (72) (C) Alignment of *ec*GlpG TMD structures reveals variation in TM5 and L5. Structures are shown for 2IC8 (65) (cyan), 2NRF (66) molecules A (blue) and B (yellow), 2IRV(64) molecules A (beige) and B (black), 3B45 (71) (green), 2XOV (72) (orange) and the acylated 2XOW (72) (pink). (D) Location of a DMPC molecule (green) in the cleft between TM2 and TM5 with its choline group close to R227, in the rhomboid structure obtained in a bicelle environment (2XTV) (73).

maintaining the relative position of the active site histidine (on TM6) and serine (on TM4). Similarly, D218 appears to engage in electrostatic interactions with S221, disruption of which could affect the tight bend at the bottom of TM4 that connects the loop to TM5.

In contrast to R217 and D218, E216A was found to have a very large reduction in activity, particularly in purified samples. Coincidently, this residue was found to be proximal to R227, a residue in TM5 that showed a similar level of activity reduction when mutated to Ala. Although the orientation of these side chains in the structures might suggest the presence of an electrostatic interaction (Fig. 3.7), the average distance between the E216 carboxylate and the R227 guanidinium group is 9.3 ± 0.7Å for ten published structures while a salt bridge interaction typically requires that this distance be less than 4 Å (176). However, there is one x-ray structure (2IRV) that shows TM5 shifted closer to TM4, giving rise to a distance of 4.8 Å between charged atoms in these side chains. This indicates that one of the conformational states of TM5 could involve transient hydrogen-bonding interactions involving these residues, although the difference in activity observed with R227A and E216A in the FP-PEG-rhodamine assay suggests that these two residues have different functional roles in rhomboid activity.

This is the first study to identify a functional role for R227, a residue located at the bottom of TM5 that projects out into the lipid phase (Fig. 3.7c), suggesting that it interacts with the lipid bilayer. Some of the higher-resolution x-ray structures have shown electron density for protein-bound detergent, with one structure showing a nonyl glucoside molecule in the cleft between TM4 and TM5 such that its head group is within hydrogen bonding distance of R227 (65). Presumably this interaction would be lost in the mutant, potentially decreasing the affinity of this cleft for lipid. More recently, *ec*GlpG crystallized in

DMPC/CHAPS bicelles identified a DMPC molecule bound in the interface between TM2 and TM5. There is a hydrogen bond between the phosphodiester group and the amide side chain of Q226, placing the choline group in close proximity to the R227 side chain (Figure 3.7d) and providing further evidence that R227 interacts with lipid head groups. Molecular dynamics simulations of *ec*GlpG in POPC and POPE also suggest that this type of interaction occurs in a bilayer environment (76). In addition, these studies identified functional coupling between the active site, L1 and TM5, with perturbations in TM5 affecting local structure and the L1-mediated positioning of *ec*GlpG in the membrane. Mutation of R227 to alanine could lead to loss of a specific lipid-binding site, potentially altering the conformational dynamics of TM5 and coupled parts of the protein. Not only could this explain the loss of activity for the R227A mutant, but it may also provide a mechanism for previous observations that different lipid environments promote various levels of rhomboid activity (63).

3.4.3 Casein as a water-soluble rhomboid substrate.

We have found the casein assay to be a particularly useful probe of rhomboid function since it is commercially available, sensitive, and easy to handle due to its high solubility in water (49, 65). The β-casein isoform used in this assay is an amphiphilic instrinsically unstructured protein, having a hydrophobic C-terminal sequence and largely acidic N-terminal sequence. It has been shown that *ec*GlpG digestion of this fluorogenic casein probe gives rise to a small number of cleavage products, the most abundant of which is approximately half the molecular weight of uncleaved β-casein (65). Visual inspection of the β-casein sequence suggests that there are as many as six potential sites that this rhomboid could cleave according to the rhomboid substrate specificity determined by the

Freeman group (61). Of particular interest is the presence of two overlapping putative rhomboid cleavage sites (residues 108 – 118, Table 3.3) just prior to the C-terminal hydrophobic domain. Based on the size of the predominant cleavage product, it is likely that this is the primary site of casein cleavage in this assay. Based on these properties, casein could therefore be considered as a water-soluble rhomboid substrate. Most fortuitously, based on membrane-binding properties of other casein proteins (177), it is likely that this reaction is facilitated by interactions between the hydrophobic C-terminal domain and the micelle, helping to bring this pair of cleavage consensus sequences in close proximity to the rhomboid active site.

3.4.4 Activity-based profiling of the rhomboid protease – an insightful assay.

In contrast with the casein-based assay, reactivity of the rhomboid against FP-PEG-rhodamine allows the activation state of the active site serine to be probed with a relatively small electrophile. This allowed us to identify mutations that affect the ability of the rhomboid to act on a polypeptide substrate, while maintaining the active site Ser in a nucleophilic state (e.g. E216A). This fluorescent probe also provided an additional advantage in that it allowed rhomboid activity to be assessed in crude membrane extracts that are presumed to preserve specific lipid-protein interactions. This is particularly important in this study since most of the residues investigated are located at the membrane-protein interface. These experiments showed that the impact of mutations that were largely inactivating in purified samples could undergo some labeling in crude membranes, consistent with a stabilizing role for these lipids. One definitive example is TMD91 which shows the same level of labeling as S201A when purified in DDM, while retaining a higher level of activity in crude membrane. These types of differences have also been seen when

Table 3.3 Potential *ec*GlpG rhomboid cleavage sites of Bovine β-casein.

Sequence	Cleavage Site[a]
MKVLILACLVALALARELEELNVPGEIVESLSSSEESITRINKKIEKFQSE	ILAC-LV
EQQQTEDELQDKIHPFAQTQSLVYPFPGPIPNSLPQNIPPLTQTPVVVPPF	VALA-LA
LQPEVMGVSKVKEAMAPKHKEMPFPKYPVEPFTESQSLTLTDVENLHLPLP	MGVS-KV
LLQSWMHQPHQPLPPTVMFPPQSVLSLSQSKVLPVPQKAVPYPQRDMPIQA	VKEA-MA
FLLYQEPVLGPVRGPFPIIV	LLQS-WM
	LSQS-KV

[a] Residues involved in specificity are indicated in blue with cleavage after residues indicated in red, as determined by Strisovsky et al, 2009 (61).

comparing *in vitro* versus *in vivo* activities of other mutants (13, 62, 74, 75). For example, mutation of a highly conserved tryptophan in the L1 structure revealed that it was not required for activity *in vitro,* but was needed for in vivo activity (74, 75). In contrast, the conserved arginine adjacent to this residue was necessary in both systems. This finding is similar to our results for R217 and D218, with lower rhomboid activity in crude membranes but wild-type-like activity in detergent micelles.

In summary, this is the first study to identify a sequence N-terminal to residues visualized in the x-ray crystal structures that is required for full rhomboid activity. Future structural studies of *ec*GlpG that include this N-terminal sequence will be required to determine how these residues influence the active site in both micelle and membrane environments, and whether it functions through membrane positioning and/or interactions with the TMD cytoplasmic face. This study is also the first to identify R227 as a functionally important residue, suggesting a role for specific protein-lipid interactions. These findings, in combination with previously published data from molecular dynamics simulations, are starting to coalesce into a new picture of rhomboid activation where interactions at the membrane surface can be translated into changes in the active site that modulate function.

Chapter 4: Three-dimensional domain swapping in the E. coli rhomboid cytoplasmic domain.

Contributions of collaborators:

Dr. Houman Ghasriani (University of Ottawa) refined the CYANA-calculated structures using XPLOR-NIH. Dr. Tara Sprules (Québec/Eastern Canada High Field NMR Facility) and Dr. Yves Aubin (Health Canada) recorded NMR spectra of the *ec*GlpG CytD dimer. Backbone chemical shift assignments of the dimeric form of *ec*CytD, and the preliminary model of the domain-swapped dimer based on NOE analysis were provided by Dr. Natalie Goto.

4.1 Introduction

Structural studies of the isolated *ec*CytD were performed in parallel with mutational studies described in Chapter 3. As shown in Figure 2.1, this domain is homologous to the *pa*CytD that was proposed to bind to the lipid membrane (79), displaying 23% sequence identity in the structured part of the domain. Over the course of this study the *ec*CytD structure was solved using solution NMR spectroscopy, which led to a number of surprising discoveries regarding the oligomerization properties of this domain. First, purification *of ec*CytD revealed that it forms a weakly associated dimer that is in slow exchange with the monomeric form. This type of slow exchange is a common characteristic of three-dimensional domain swapping interactions, where non-covalent intramolecular bonds in each subunit must be broken and replaced by interactions with identical structural elements from its dimerization partner (178), as schematically shown in Figure 4.1. In this chapter, I present evidence that *pa*CytD does undergo a domain swapping interaction, and characterize the effect of detergent micelles on this interaction. Results from this study provide the first demonstration of detergent micelle-catalyzed interconversion of domain swapped states in a domain-swapping interaction. These findings have implications for potential interactions in the full-length rhomboid, and how this might give rise to domain-swapping events *in vivo*.

4.2 Materials & Methods

4.2.1 ecCytD expression and purification.

*ec*CytD (residues 1-61) was cloned into the pET30a vector between the NdeI and XhoI restriction sites to contain a C-terminal His_6-tag preceded by a Leu-Glu sequence. *ec*CytD was expressed and purified using a similar approach to that used for the *pa*CytD in

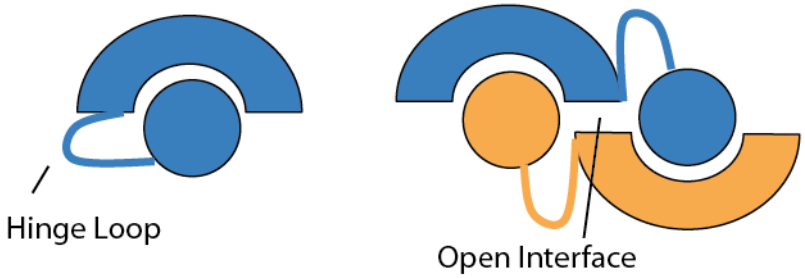


Figure 4.1 Schematic representation of domain swapping. Intramolecular interactions of exchanging structural elements are disrupted within the monomer (left) and form identical contacts with another subunit to form the dimer (right). Exchange is often facilitated by a flexible hinge loop that can change conformation between monomer and domain-swapped dimer. The closed interface is found in both monomeric and dimeric states, while the open interface is found only in the dimer and formed by new contacts in the oligomer.

Sections 2.2.2 and 2.2.5. This *ec*CytD expression plasmid was transformed into BL21 (DE3) cells, which were then grown in LB media or M9 containing 0.1% (w/v) ^{15}N-labeled ammonium chloride and 0.3% (w/v) glucose (U-^{13}C) as the sole nitrogen and carbon sources, respectively, as well as 50 μg/ml kanamycin. The bacterial pellet from a 1 L culture was resuspended in 25 mM phosphate buffer (pH 6.5) containing 150 mM NaCl, 1 X EDTA-free protease inhibitor cocktail (Roche) and 10 mM imidazole and lysed by sonication. Cell debris was centrifuged at 16,000 x g for 20 minutes and the supernatant was applied to a Ni-NTA column. The bound protein was washed in the above buffer before the *ec*CytD construct was eluted in buffer augmented with 250 mM imidazole. Concentration and size-exclusion chromatography steps were similar to those in Section 2.2.5, using phosphate buffer (pH 6.5) containing 150 mM NaCl and 100 μM EDTA. *ec*CytD eluted from the Superdex 75 size-exclusion column in two peaks at 12.5 and 15 ml. MALDI-MS performed on purified samples using a Microflex MALDI-TOF mass spectrometer (Bruker) using the same method described in Section 3.2.3 confirmed the identity of both peaks as the *ec*CytD.

4.2.2 Measurement of interconversion rates between monomeric and domain-swapped dimeric ecCytD.

Monomeric and dimeric species isolated by size exclusion chromatography on a Superdex-75 10/300 GL column (GE Healthcare) and stored for 24 hours at 4ºC showed no evidence for interconversion to monomer/dimer mixtures. To see if the two oligomeric states could reach equilibrium at higher temperatures, 0.25 mM samples of monomeric *ec*CytD were incubated at 45°C, 50°C, 55°C and 60°C. Aliquots of each sample were removed after a range of incubation times, and placed on ice for 5 minutes to "freeze" the

exchange mixture. Size exclusion chromatography of each sample on the Superdex-75 column showed no more than two peaks corresponding to monomeric and dimeric states of *ec*CytD. Peak volumes of the monomeric and dimeric species in each elution profile were determined through peak integration (Unicorn FPLC software) and used to calculate the amount of each species present. Beginning with a purely monomeric state of *ec*CytD the increase in dimer concentration over time, expressed as a percentage of the total protein concentration (D) could be fit to equation 2:

$$D = D_{eq}*(1 - e^{-kt}) - D_0 \quad [2]$$

using GraphPad Prism (Version 5.0), where D_{eq} is the percentage of dimer at equilibrium, D_0 is the initial percentage of dimer and k is the rate constant for dimerization. Apparent dissociation constants (K_d) were determined using

$$K_d = [M]^2 / [D] \quad [3]$$

where [M] and [D] represent the concentration of monomer and dimer at equilibrium. The effect of additives such as detergents on the rate of dimer formation was also investigated using this same method, with all experiments performed at 45°C to allow direct comparison to rate constants determined in the absence of additives. Detergents were obtained from Anatrace (DDM, Fos10, Fos12, Fos16) and Avanti Polar Lipids (LPPG).

4.2.3 Preparation of a ^{15}N, ^{13}C- / ^{14}N, ^{12}C-labeled heterodimeric ecCytD sample.

To identify NOEs arising from the dimer interface, ^{15}N, ^{13}C-labeled and unlabeled monomeric *ec*CytD was mixed to produce a heterodimer of labeled and unlabeled *ec*CytD. Equimolar concentrations of monomeric labeled and unlabeled *ec*CytD were mixed and incubated with 2.5 mM Fos16 at room temperature for 1 hour to catalyze subunit exchange and allow the system to reach equilibrium. The dimer was then isolated using size exclusion chromatography as described above.

4.2.4 NMR spectroscopy and monomeric ecCytD structure calculation.

All NMR spectra of the monomer were recorded at 25°C on a Varian Inova 500 spectrometer equipped with a triple resonance probe and pulsed field gradient probe at the University of Ottawa NMR Facility. Experiments used to determine the monomer structure were the ^{15}N-HSQC, ^{13}C-HSQC, HNCO, HNCA, HNCACB, HBHA(CO)NH, HN(CO)CA, HCCH-TOCSY, HCCH-COSY, ^{13}C-NOESY-HSQC and ^{15}N-NOESY-HSQC (157). Data were processed by NMRPipe (105) and analyzed with NMRView (156). NOE distance restraints and Phi/Psi dihedral angles derived from the program TALOS (99) were used in the torsion angle dynamics program CYANA 2.1 (112) which performed automatic NOE assignment to calculate 200 conformers. The 20 low-energy structures from the final round of CYANA were individually refined in explicit solvent (113) by Dr. Houman Ghasriani using XPLOR-NIH (179, 180) and scripts available from ftp://ftp.pasteur.fr/pub/BIS/linge/. Subsequently, the quality of the solvent-refined structures was validated by using AQUA, PROCHECK (115, 116) and WHAT IF (117). Structural statistics are shown in Table 4.1. Chemical shift assignments and structure coordinates were deposited in the Biological Magnetic Resonance Data Bank (17720) and PDB (accession ID 2LEP), respectively.

Table 4.1 Structural statistics for the final ensemble of 20 water-refined structures of monomeric *ec*CytD

NMR constraints	
Non-redundant angle and distance restraints	
Total distance restraints	1070
Short-range ($\| i\text{-}j \| \leq 1$)	530
Medium-range ($\| i\text{-}j \| < 5$)	205
Long-range ($\| i\text{-}j \| \geq 5$)	335
Dihedral angles	
ϕ (TALOS)	42
φ (TALOS)	42
Structure statistics	
Violations	
NOE violations > 0.5 Å	0
Dihedral angle violations > 5º	0
Deviations from idealized geometry	
Bond lengths (Å)	0.017 ± 0.001
Bond angles (º)	0.740 ± 0.020
Impropers (º)	0.770 ± 0.040
Agreement with experimental restraints	
NOE	0.065 ± 0.002
Dihedral angles	0.620 ± 0.140
Average RMSD (Å)[a]	
Backbone	0.34 ± 0.07
Heavy Atoms	0.95 ± 0.11
Ramachandran plot [a]	
Most favoured (%)	80.5
Additionally allowed (%)	19.3
Generously allowed (%)	0.1
Disallowed (%)	0

[a] Based on the well-defined regions of the protein, namely residues 3-31 and 35-57.

Backbone assignment experiments for the dimer were recorded on a Bruker AVANCE III 600 spectrometer equipped with a triple resonance cryoprobe at Health Canada by Dr. Yves Aubin. A filter-edited carbon NOESY (181) was also acquired for the dimer using a Varian Inova 800 spectrometer equipped with a triple resonance cryoprobe at the Québec/Eastern Canada High Field NMR Facility run by Dr. Tara Sprules. 3D HCCH-COSY and ^{13}C-NOESY-HSQC were also acquired on this sample to facilitate side chain chemical shift assignment of the dimer.

4.2.5 Purification of full-length ecGlpG and isolation of TMD core.

Full-length *ec*GlpG was purified as described earlier (section 3.2.2) in 50 mM HEPES pH 7.3, 0.1% DDM, 200 mM NaCl, 10% glycerol and 100 μM EDTA and digested with trypsin in a 1:20 ratio at 4°C overnight (65) to obtain the *ec*GlpG-TMD core. Full-length and digested *ec*GlpG were purified on a Superdex-200 size exclusion column.

4.3 Results

4.3.1 Identification of dimeric ecCytD.

During the purification of *ec*CytD, size exclusion chromatography profiles showed two distinct peaks, only one of which corresponded to the expected molecular weight of the isolated *ec*CytD (8 kDa). However, SDS-PAGE of fractions from these peaks showed that they both comprised species with the expected molecular weight of *ec*CytD (Fig. 4.2). This indicated that the higher molecular mass species observed in the size exclusion chromatography profile was an oligomer of *ec*CytD, with the elution volume consistent with that of a dimeric state. However, reapplication of pure monomeric or dimeric species on the size exclusion column did not give rise to appreciable amounts of the other species. This

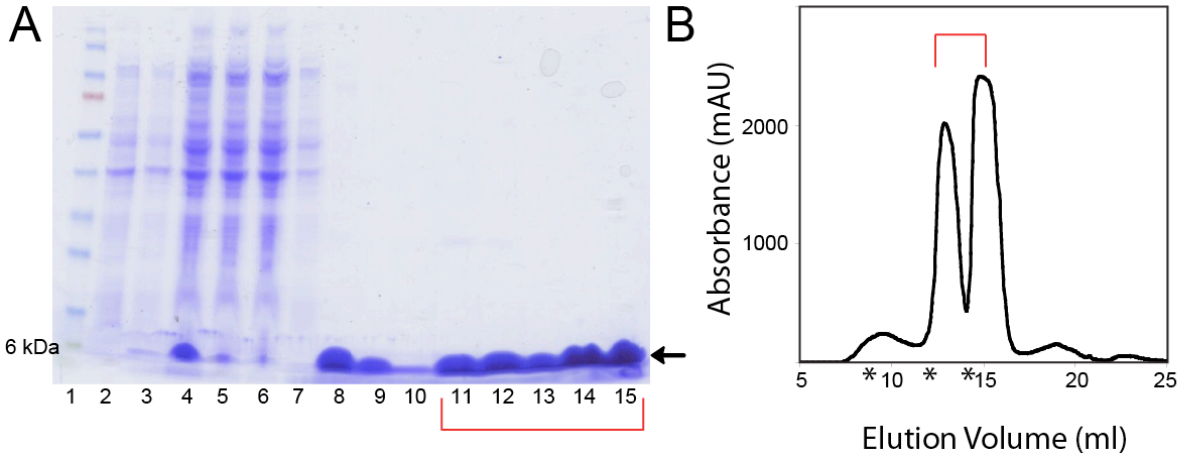


Figure 4.2 Evidence of *ec*CytD dimerization. (A) Coomassie-stained SDS-page gel of *ec*CytD expression and purification. Lane 1 – molecular weight ladder; lanes 2 & 3 – pre- and post-induction of *ec*CytD expression in *E. coli* BL21 (DE3) cells, normalized for OD_{600}; lane 4 – soluble fraction applied to Ni-affinity column; lanes 5 & 6 – column flow through after initial and re-application; lane 7 – wash of the column with 10 mM imidazole; lanes 8 to 10 – elution of *ec*CytD with 250 mM imidazole; and lanes 11-15 –fractions spanning 12.5-15 ml (dimer peak) of the size exclusion chromatography profile shown in (B). Protein in fractions from both peaks, indicated by the red brackets, had the same mobility shown in this gel. Asterisks mark the elution volumes obtained for (from left to right) BSA (66 kDa), chymotrypsinogen (25 kDa) and ribonuclease A (13 kDa). The void volume for this column is 7.6 mL, corresponding to globular proteins in excess of 70 kDa. The molecular weight of the *ec*CytD monomer is approximately 8 kDa.

demonstrated that interconversion between the two species was very slow under these conditions.

4.3.2 Energetics of dimerization.

Although the two species were found to be very stable under the room temperature chromatography conditions used to isolate them, formation of the dimeric state from pure monomer or dissociation of the dimer back into the monomer could be induced by exposure to elevated temperatures (Fig. 4.3). Using size exclusion chromatography to measure the amount of monomer and dimer as a function of time (described in Section 4.2.2), it was possible to monitor the rates of monomer or dimer formation to obtain rate constants and equilibrium populations at each temperature (Fig. 4.4). As expected, the rate of dimer formation increased with temperature, and followed Arrhenius behavior when the log of these rates was plotted as a function of inverse temperature (Fig. 4.5). The activation energy for dimer formation determined from this plot was 80 kcal/mol. Equilibrium populations of monomer and dimer forms at these various temperatures could be similarly used in a Van't Hoff analysis to determine that the enthalpy change that occurs upon dimer formation (ΔH) is -15 kcal mol^{-1} while the entropy change (ΔS) is -32 cal K^{-1} mol^{-1} (Fig. 4.5b).

4.3.3 Structural characterization of ecCytD by solution NMR

The slow rate of interconversion made it possible for each species to be isolated and analyzed by solution NMR spectroscopy. As shown in Figure 4.8a, the two spectra were largely similar, with only a subset of the peaks showing differences between the two states. This allowed the structure of each state to be investigated using methods described in Sections 1.3.3 and 1.3.4. For this purpose, a suite of triple resonance 3D experiments were performed to allow chemical shifts to be assigned for backbone atoms using procedures

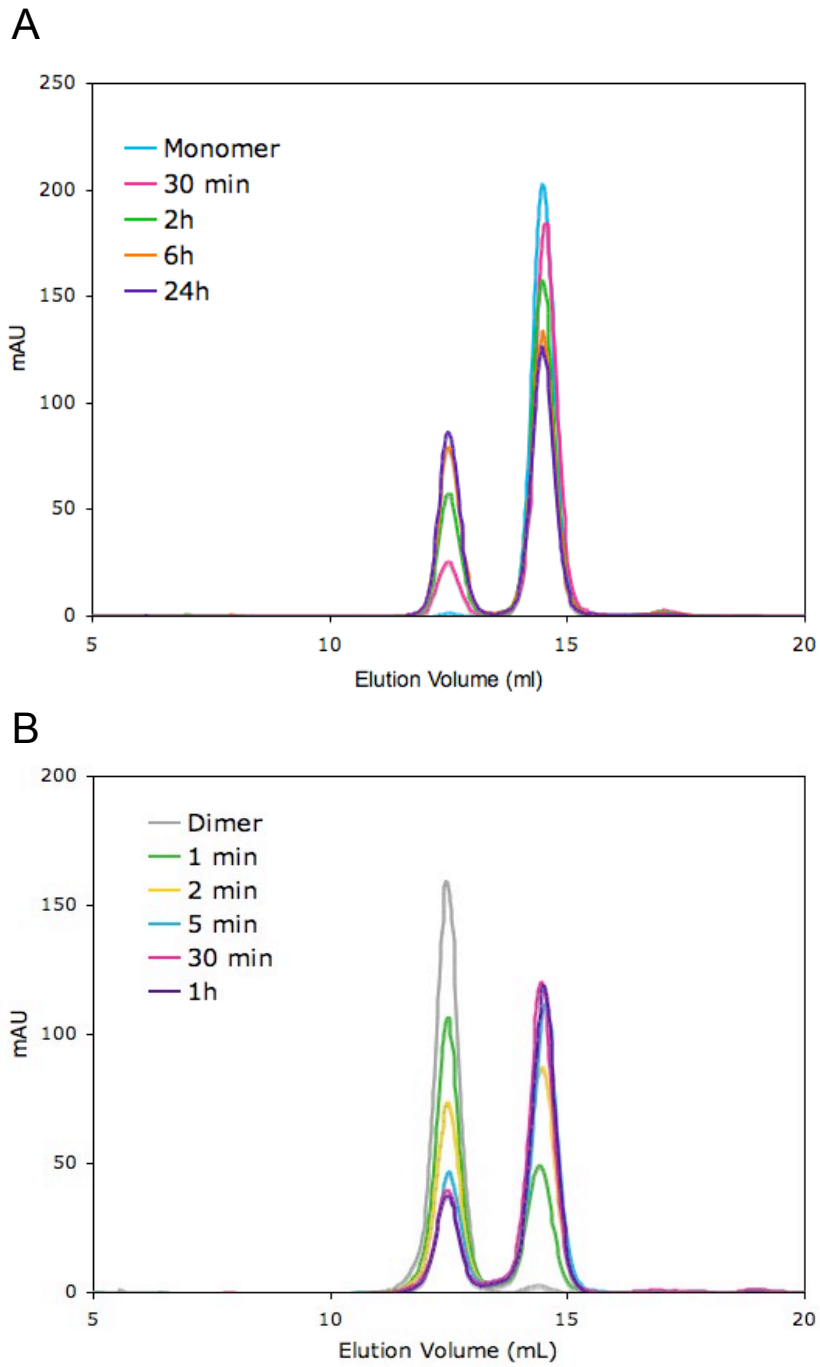


Figure 4.3 Interconversion between monomer and dimer at elevated temperatures. (A) Representative SEC profiles for 0.25 mM pure monomer incubated at 50°C in 25 mM phosphate buffer pH 6.5, 150 mM NaCl and 100 µM EDTA for indicated periods, showing conversion of dimer back to monomer. (B) SEC profiles for 0.125 mM dimer at 60ºC in the same buffer. Conversion of dimer to monomer confirms the reversibility of the reaction.

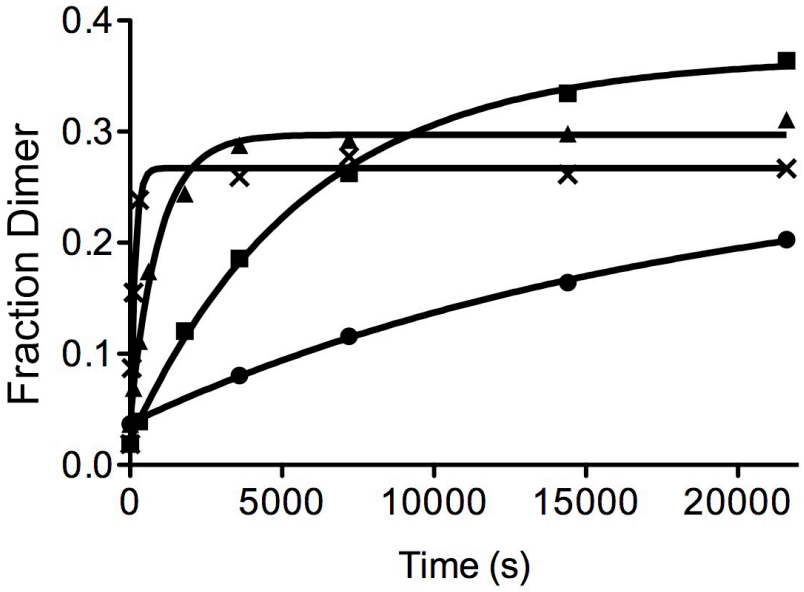


Figure 4.4 Kinetic profiles of *ec*CytD dimerization at 45°C (circles), 50°C (squares), 55°C (triangles) and 60°C (crosses). The fraction of *ec*CytD in the dimeric state was determined by integration of peaks in size exclusion chromatography profiles such as those shown in Figure 4.3a. The time dependent increase in dimer fraction (D) was fit to $D = D_{eq}*(1 - e^{-k_{obs}t}) - D_0$ where D_{eq} is the proportion of the population in a dimeric state at equilibrium, D_0 is the initial fraction of the population in a dimeric state and k_{obs} is the observed rate constant for dimerization. Observed rate constants for dimer formation are $2.2x10^{-5}$, $1.4x10^{-4}$, $1.1x10^{-3}$ and $6.6x10^{-3}$ s^{-1} at 45°C, 50°C, 55°C and 60°C, respectively. Points from longer incubation periods performed for the 45°C and 50°C profiles are not shown, but were used to generate the fits shown.

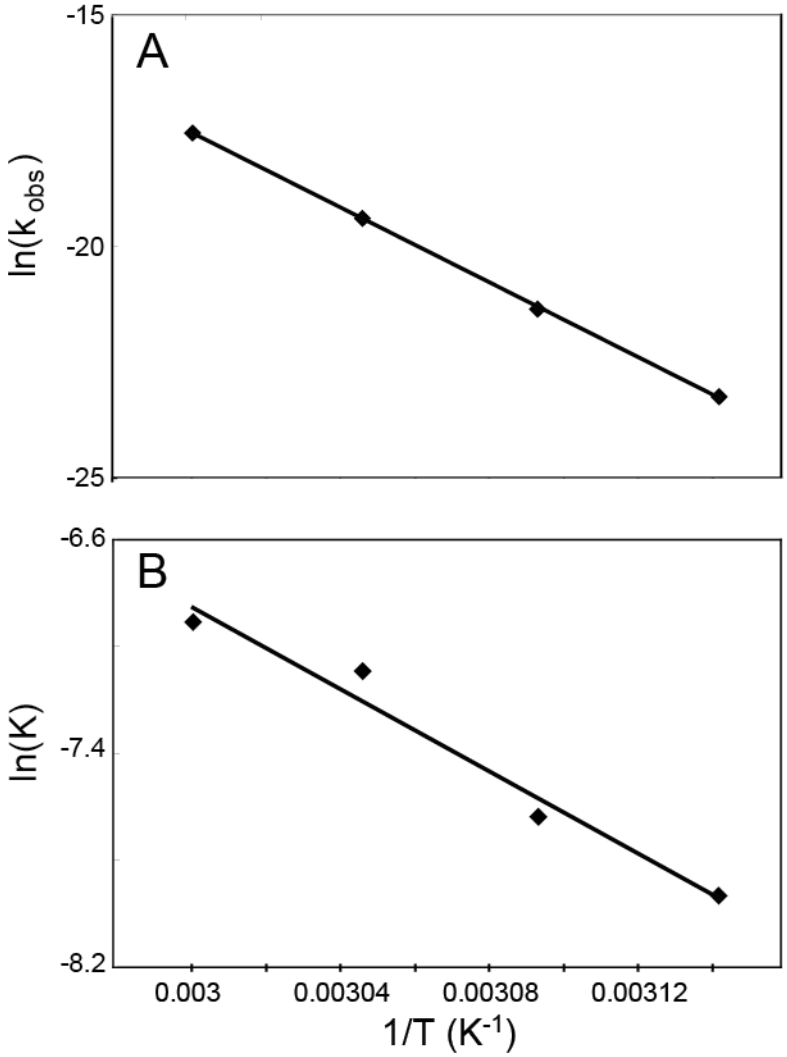


Figure 4.5 Energetics of *ec*CytD dimerization. (A) Arrhenius plot of ln(k_{obs}) versus 1/T to determine the activation energy of dimer formation. Rate constants were obtained from Figure 4.4. The activation energy for dimerization determined from this analysis is 80.2 kcal/mol. (B) Van't Hoff plot of ln(K) versus 1/T to determine changes in enthalpy and entropy for dimer dissociation. Equilibrium concentrations of monomer and dimer were used to determine equilibrium constants at 45°C, 50°C, 55°C and 60°C (found to be 0.36 mM, 0.48 mM, 0.83 mM and 1.0 mM, respectively). Change in enthalpy (ΔH) and entropy (ΔS) determined from this plot is -15 kcal mol^{-1} and -32 cal K^{-1} mol^{-1}, respectively. Errors are within the symbol dimensions for both plots.

outlined in Section 1.3.3. This analysis allowed 98 % of backbone atoms in the monomer to be assigned, with all peaks in the ^{1}H-^{15}N HSQC being assignable (Figure 4.6). A similar level of assignment was also obtained for the dimeric species. Side chain chemical shift assignments could be achieved to 89 % completion for the monomeric state, which were then used to determine the NMR structure. Some peaks had very low or no intensity due to conformational exchange. Therefore, it was not possible to assign these peaks.

As shown in Figure 4.7, the *ec*CytD structure is comprised of a mixed β-strand/α-helix fold, similar to the *pa*CytD structure, (backbone Cα rmsd of 6.45 Å). The *ec*CytD structure consists of a three-stranded β-sheet, (β1, residues 2-7, β2, 27-30, β3, 35-39) with two α-helices (α1, residues 10-22, α2, 42-57) packed against one face of the sheet. The α-helices interact extensively with each other and the β-sheet via a cluster of hydrophobic and aromatic residues that form the core. This is in contrast with the *pa*CytD structure which shows almost orthogonal relative positioning of the two helices with respect to each other, and a much smaller interaction surface between these helices.

To identify regions of *ec*CytD that are involved in dimerization, backbone amide chemical shift differences between the monomeric and dimeric states were measured (Fig. 4.8). These shift differences were plotted on the structure of the *ec*CytD monomer (Fig. 4.8c) and found to localize to residues in the loop connecting β2 to β3, with smaller changes observed for residues in close proximity to this structural element. Although Cα shifts were difficult to assign in the loop connecting β2 to β3, Cα secondary shift analysis (182) suggests that this part of the protein undergoes a structural change upon dimerization, while the structure of other parts of the protein are unaffected (Fig. 4.9a). Although this suggests that the loop region could be the primary interaction surface, the small surface area that

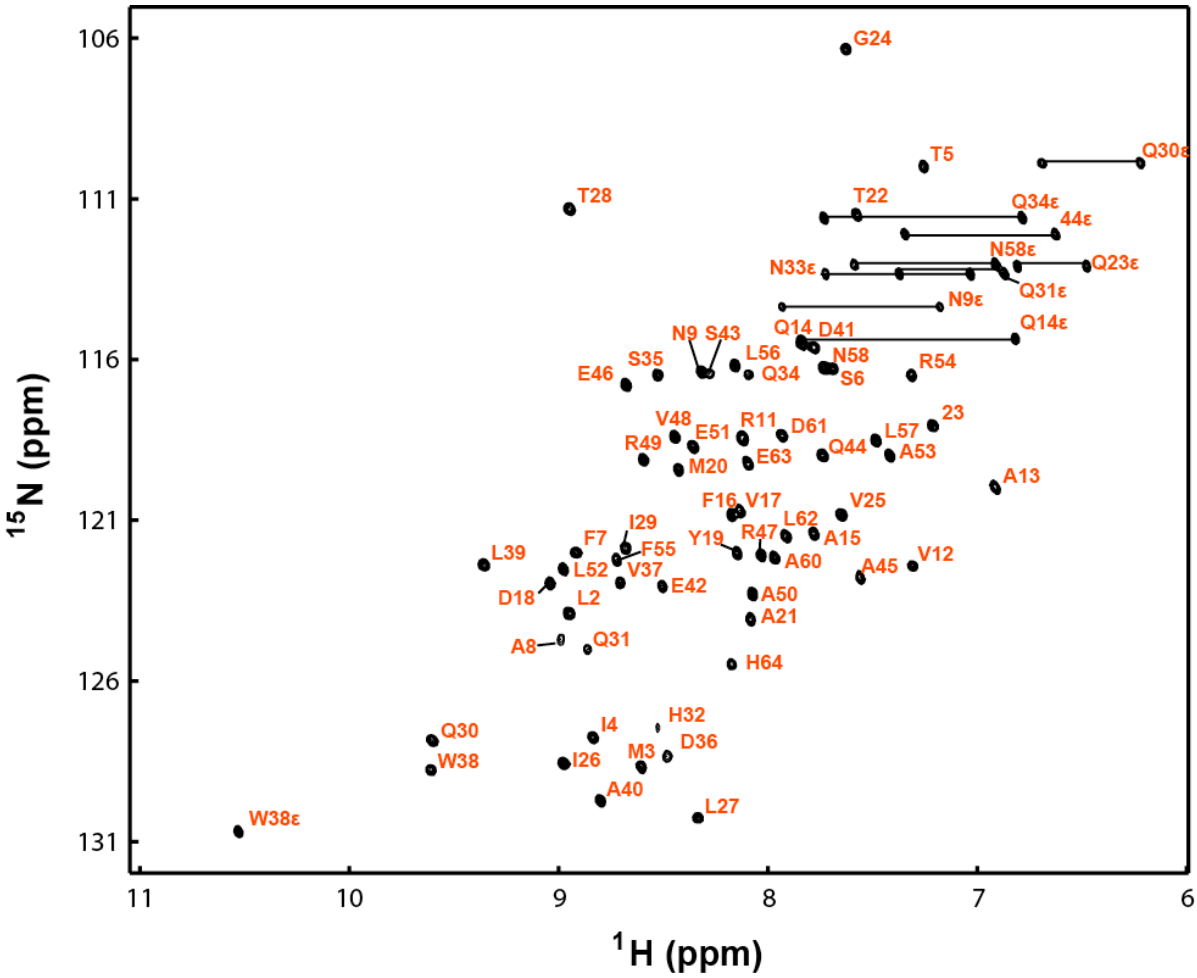


Figure 4.6 The assigned ^{1}H-^{15}N HSQC spectrum of monomeric *ec*CytD, recorded at pH 6.5, 25˚C. Side chain amide peaks are indicated by the horizontal lines.

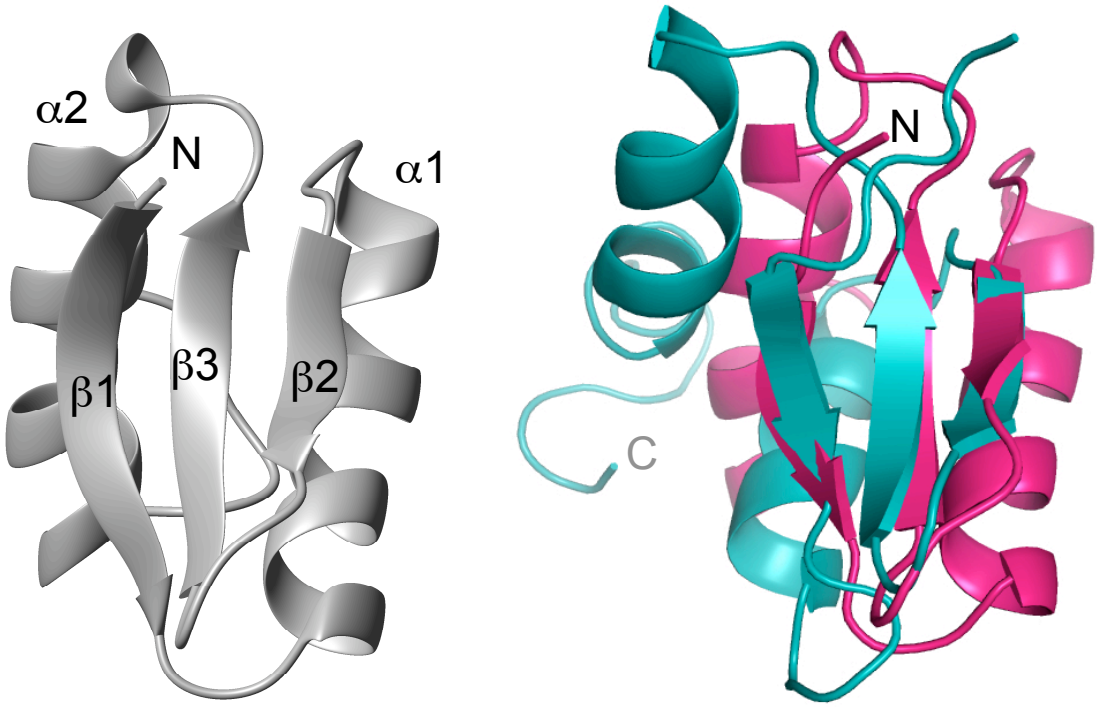


Figure 4.7 Solution NMR structure of the *E. coli* GlpG CytD (grey and pink) aligned with the *P. aeruginosa* CytD (cyan). The *ec*CytD contains 23% sequence identity (and 38% similarity) with the *pa*CytD. Cα RMSD is 6.45 Å with greatest similarity in the β-sheet region (2.21 Å).

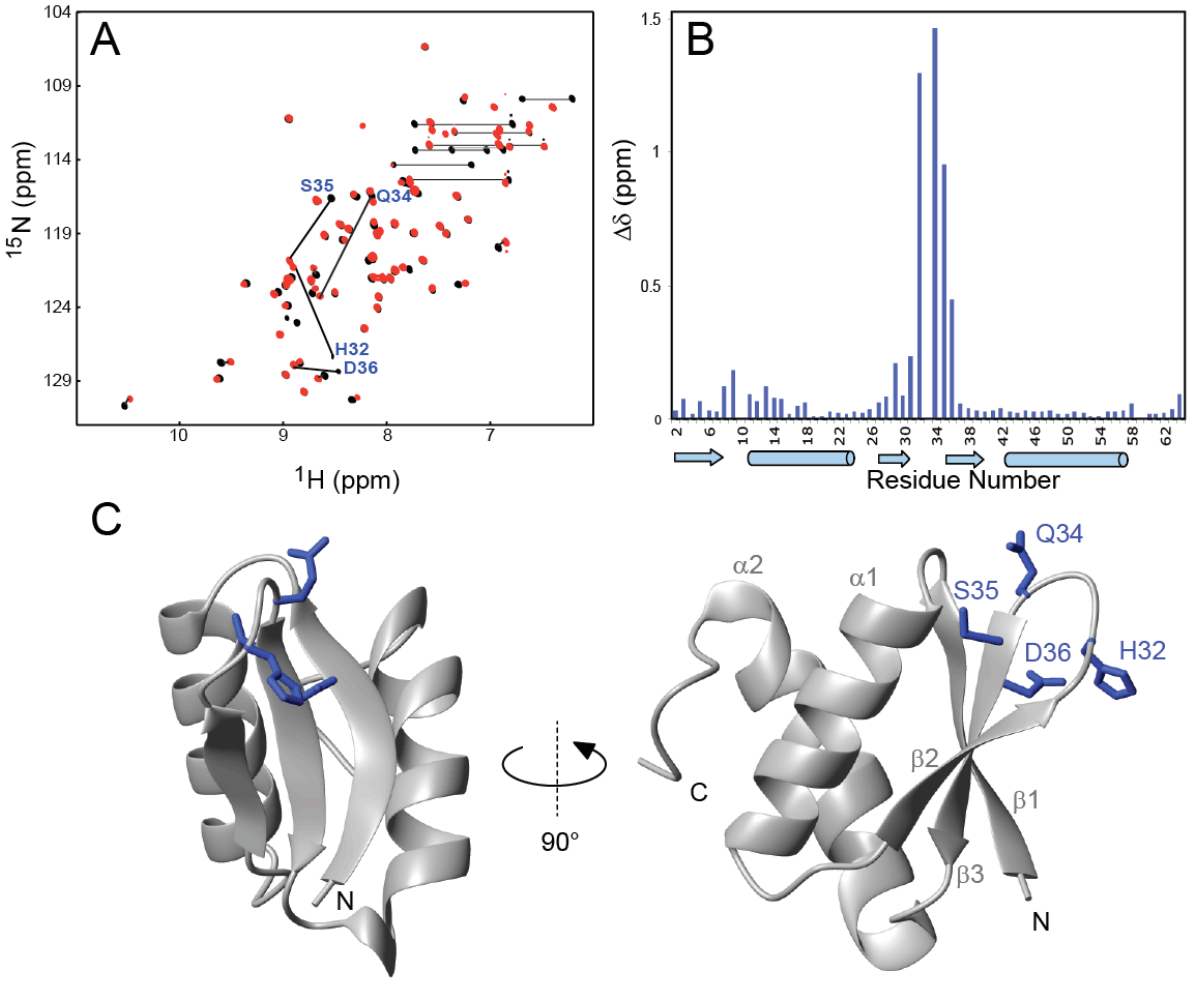


Figure 4.8 Chemical shift differences between monomeric and dimeric *ec*CytD. (A) ^{1}H-^{15}N HSQC spectrum of the *ec*CytD monomer (black) with that of the dimer (red). Assignments for peaks showing significant chemical shift changes are indicated in dark blue. (B) Average backbone amide chemical shift differences (Δδ) between monomeric and dimeric *ec*CytD displayed as a function of residue number, with corresponding secondary structure elements in the monomer structure shown below the sequence. (C) Residues experiencing the largest average amide shift changes in the monomer are mapped onto the *ec*CytD structure. Dark blue side chains are shown for residues with shift changes that are larger than the average plus one standard deviation.

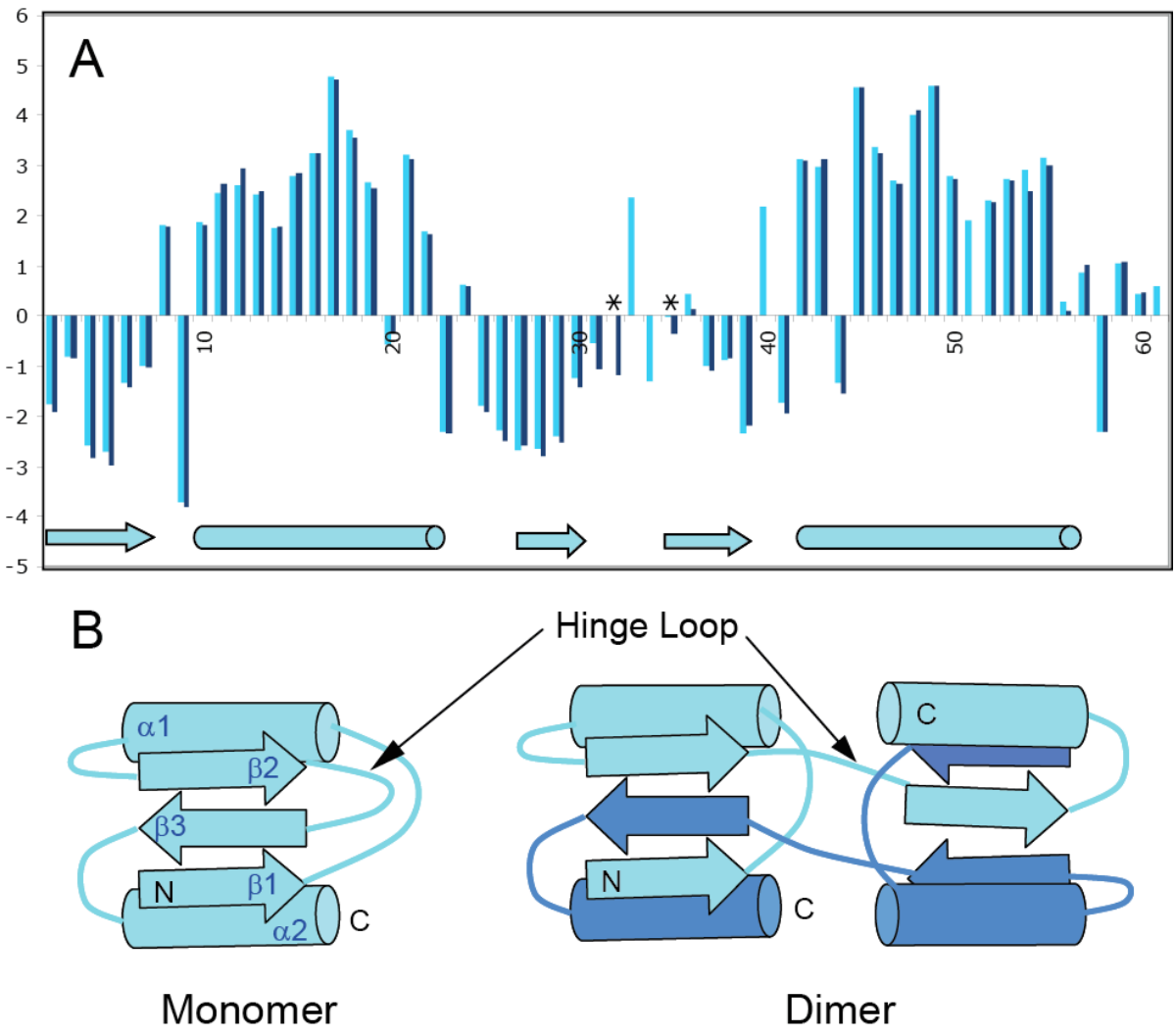


Figure 4.9 Cα secondary shift analysis and model for the domain swapping interaction of the *ec*CytD. (A) Cα secondary shift analysis of monomeric and dimeric *ec*CytD. Shown are differences between experimental Cα shifts and published random coil values (182) at pH 6.5 for the monomer (light blue) and the dimer (dark blue) with the secondary structure determined from the NMR structure of the monomer drawn schematically in blue at the bottom. The largest differences in Cα secondary shifts are for residues in the turn between β2 and β3, although the extent of assignment also differs between the two samples for these residues (31-36). Those that could not be assigned have Cα secondary shifts shown as zero, with the exception of residues 32 and 35 in the monomer, which were calculated to be close to zero (indicated by an asterisk). (B) Model of domain swapping exchange for *ec*CytD based on preliminary analysis of filter-edited ^{13}C-NOESY acquired on a ^{15}N, ^{13}C- / ^{14}N, ^{12}C-labeled heterodimer.

would be available for an interaction is not consistent with the large energy barrier measured for dimer formation. More consistent with this large activation energy would be a structural rearrangement caused by a domain swapping interaction, where these residues would act as the hinge region to accommodate the swap (Fig. 4.1).

To map the dimer interface in more detail a ^{13}C-edited filtered NOESY was acquired to allow the specific detection of inter-molecular NOEs. Intermolecular NOEs were found outside the regions showing large amide chemical shift differences between monomer and dimer, suggesting that a large part of the intermolecular interface has the same local structure as is found in the monomer, as would be expected in a domain-swapping interaction. Assignment of this intermolecular NOESY spectrum currently underway is consistent with an interaction involving a swap of the C-terminal β3-turn-α2 elements, with the loop connecting β2 to β3 acting as a hinge-loop, shown as a schematic model in Figure 4.9b. The extensive intermolecular interface in this dimer is consistent with the slow rate of exchange between dimer and monomeric states, since these interactions would need to be broken for dimer dissociation to occur.

4.3.4 Effect of detergent micelles on the ecCytD

It has been shown that the cytoplasmic domain from the homologous *pa*GlpG has a tendency to bind detergent micelles (79) through a denaturing interaction (Chapter 2). To determine whether the *ec*CytD undergoes similar interactions, the lipid-like detergent hexadecylphosphocholine (Fos16) was titrated into a 0.25 mM monomeric ^{15}N-labeled *ec*CytD sample (as described in section 2.2.6) and HSQC spectra acquired at 25ºC. In contrast to the denaturing interaction observed for the *pa*CytD in Chapter 2, addition of Fos16 gave rise to a mixture of monomeric and dimeric species in the HSQC spectrum (Fig.

4.10a) with the distinct set of peaks for each species demonstrating that they are in slow exchange on the NMR timescale. Size exclusion analysis of this sample confirmed that both monomer and dimer species were present after incubation with Fos16, and that there was no higher molecular weight aggregate present. These results showed that Fos16 could induce dimerization in a monomeric sample of *ec*CytD under conditions that would normally give rise to a very slow rate of dimerization.

The effect of Fos16 on the dimerization rate and equilibrium concentrations of *ec*CytD was analyzed at 45^{o}C as described in sections 4.2.2 and 4.3.2. As shown in Figure 4.10b, the rate of dimerization in the presence of 2.5 mM Fos16 was ~1300-fold faster than that measured under identical conditions in the absence of detergent. A similar effect was also found for the reverse reaction, with equilibrium reached in less than 10 minutes in the presence of Fos16 compared to the 24 hours required to equilibrate the monomer-dimer mixture at this temperature. This effect was observed over a range of *ec*CytD concentrations, with a plot of the equilibrium concentrations of monomer squared versus that of the dimer showing a linear correlation. As expected the slope of this line that corresponds to the K_d for this interaction is the same as that measured for in the absence of Fos16 (0.37 mM) (Fig. 4.10c). This ability of Fos16 to accelerate the interconversion of monomer and dimer without altering the affinity of the interaction demonstrates that Fos16 is acting as a catalyst in this system.

The 2.5 mM concentration of Fos16 used to accelerate interconversion is well above the critical micelle concentration for this detergent, suggesting that the detergent micelle is responsible for this effect. To confirm that the micelle was the species catalyzing this

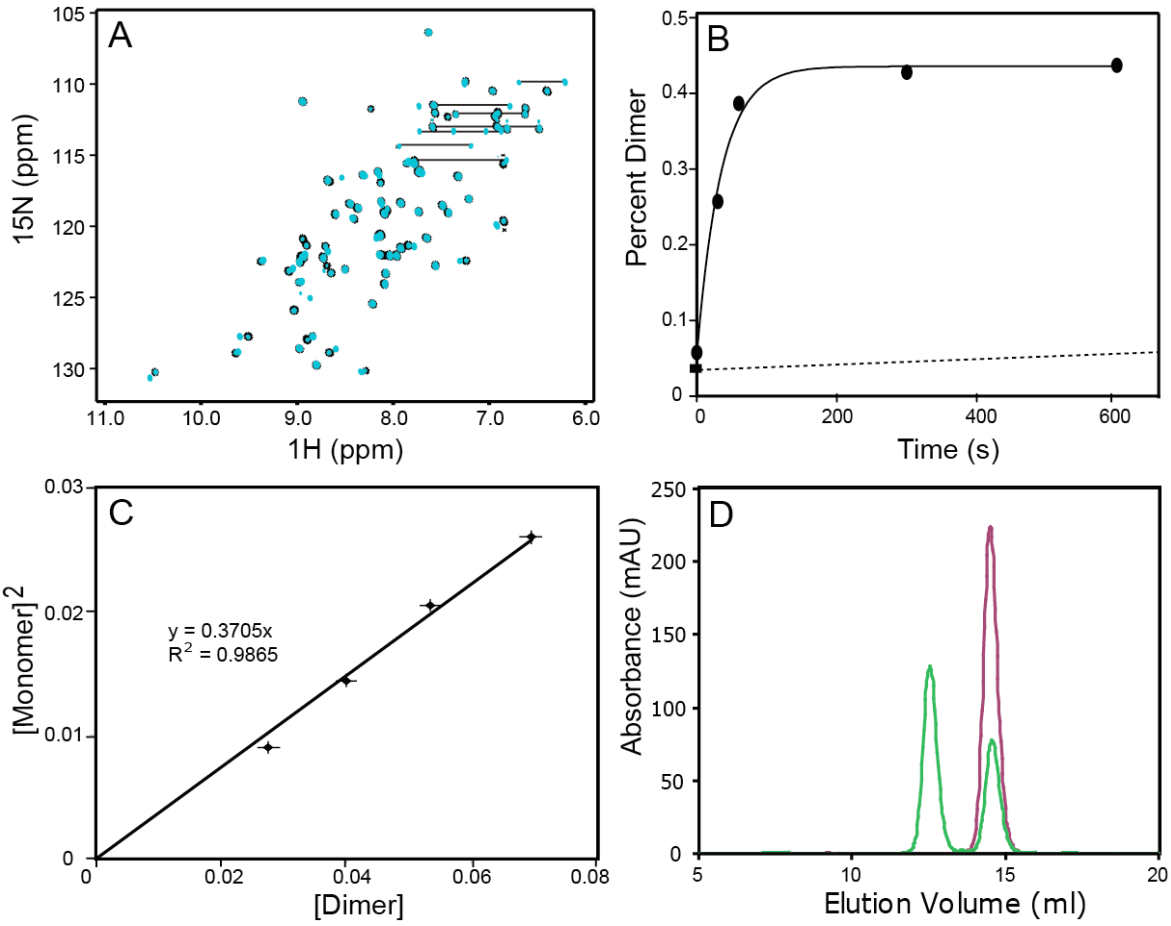


Figure 4.10 Catalysis of *ec*CytD dimerization by the detergent micelle. (A) ^{1}H-^{15}N HSQC spectrum of a 0.25 mM solution of dimeric *ec*CytD in the absence (black) and presence of 2.5 mM Fos16 detergent (blue) at 25 ^{o}C. Peaks arising from both monomeric and dimeric species are present in this sample, as confirmed by size exclusion chromatography (not shown). (B) Rate profile for dimer formation by 0.25 mM monomeric *ec*CytD in 25 mM phosphate buffer pH 6.5, 150 mM NaCl and 100 µM EDTA incubated in the presence (solid line) or absence (dashed line) of 2.5 mM Fos16 at 45^{o}C. The observed rate constant and K_d in the presence of Fos16 were determined to be 0.29 s^{-1} and 0.37 mM, respectively. (C) Confirmation that the Fos16 micelle does not alter the monomer-dimer equilibrium constant. The squared monomer concentration versus dimer concentration is shown for the a range of total *ec*CytD concentrations. All samples were incubated with 2.5 mM Fos16 for 10 minutes at 45°C, and resulting equilibrium concentrations of each species were measured at room temperature using size exclusion chromatography. The equilibrium constant for dimer dissociation determined from the graph is 0.37 mM +/- 0.01 mM. (D) The micellar form of Fos detergent is responsible for the catalytic effect of the detergent on dimer formation. SEC profiles are shown for 0.25 mM *ec*CytD in 25 mM phosphate buffer pH 6.5, 150 mM NaCl and 100 µM EDTA incubated at room temperature with sub-micellar (2.5 mM, purple) or micellar (15 mM, green) Fos10.

interconversion these experiments were repeated with an equivalent amount of decylphosphocholine (Fos10). While the headgroup is the same as that in Fos16, the shorter alkyl chain gives rise to a higher CMC (~11 mM) and hence at 2.5 mM the detergent does not form micelles. When monomeric *ec*CytD was incubated with 2.5 mM Fos10 at room temperature and analyzed by size exclusion chromatography, no peaks from the dimeric state were observed.. In contrast, if the Fos10 concentration exceeded its CMC, the *ec*CytD chromatography profile showed a mixture of monomer and dimer (4.10d). Hence this confirms that the micelle assembly is responsible for the catalytic effect on *ec*CytD domain swapping.

At a detergent monomer concentration of 2.5 mM, the concentration of Fos16 micelles would be ~14 μM, which under the conditions of the dimerization experiments would correspond to a micelle:*ec*CytD ratio of >1:15. This sub-stoichiometric concentration of micelles implies that the rate of dimerization could be further augmented by using higher concentrations of detergent. This was confirmed in Figure 4.11a, which shows that the rate of dimer formation increases with micelle concentration in a linear manner up to ~0.08 mM (micelle:*ec*CytD ratio of ~1:3). Beyond this concentration, dimerization rates were too rapid to measure by the method used in our study.

4.3.5 Effect of detergent size and charge on dimerization.

To gain insight into how the Fos16 micelle catalyzes dimerization, monomeric *ec*CytD was incubated with detergents that differ in their head group chemistry and alkyl chain length. First, micelle charge was investigated through comparison of Fos16 to anionic lysopalmitoylphosphoglycerol (LPPG) and non-ionic dodecylmaltoside (DDM). When 2.5 mM LPPG was incubated with monomeric *ec*CytD at 45°C, accelerated conversion of

monomer to dimer was observed. When compared to dimerization rates with a comparable concentration of Fos16 micelles (Fig. 4.11a), the LPPG-catalyzed rate was found to be approximately the same as that for Fos16. These results suggest that catalysis is not sensitive to the type of charge in the head group. In contrast, when *ec*CytD was incubated with 2.5 mM of non-ionic DDM, the rate of dimerization measured was only ~7-fold faster than the uncatalyzed rate (Fig. 4.11b), which was much slower than that observed for comparable amounts of Fos16.

The effect of micelle size on dimerization rate was also investigated to determine if catalysis levels would be sensitive to the dimensions of the micelle. This was tested by comparison of dimerization rates in Fos16 to rates with the same concentration of micelles comprised of shorter alkyl chain analogues Fos10 or Fos12 (Fig. 4.11a). In general, the smaller micelles gave rise to smaller enhancements in dimerization rates compared to that promoted by Fos16. When dimerization rates were plotted as a function of micelle aggregation number, which is proportional to micelle surface area, a linear relationship was observed (Fig. 4.11c) showing that micelles with larger surface areas are more effective catalysts for domain-swapping interconversion.

Although charged micelles showed a propensity to increase the rate of dimer formation, in some cases the affinity of the interaction could be altered. For example, while the use of 2.5 to 15 mM Fos16 gave rise to a highly similar dissociation constant (0.37 ± 0.05 mM), concentrations beyond 15 mM Fos16 shifted the equilibrium more to the monomeric form. This may reflect a dilution effect, where the effective concentration of monomer is reduced by interactions between the micelle and a single subunit of *ec*CytD.

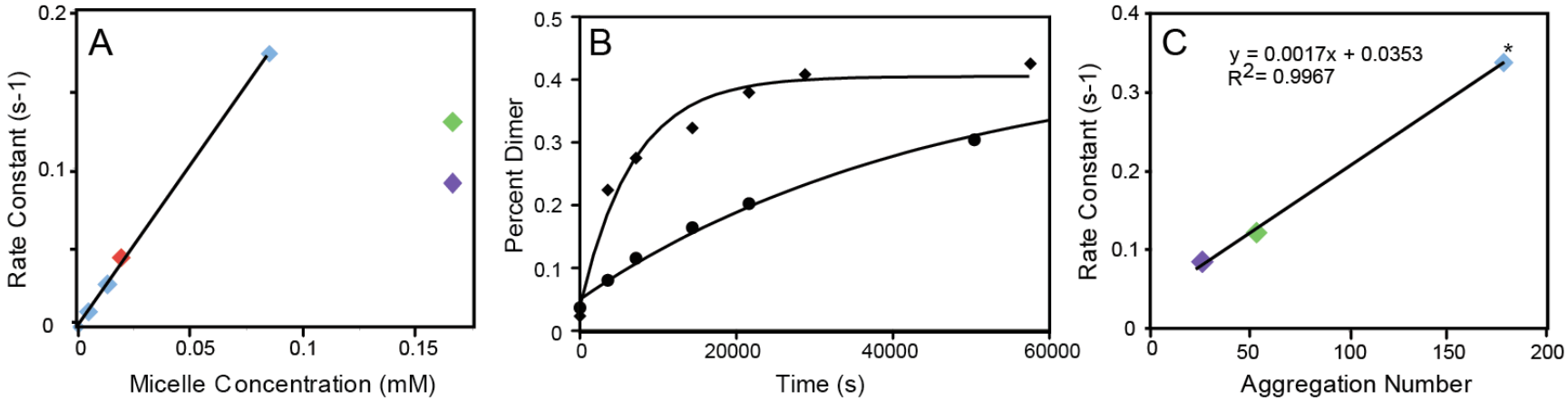


Figure 4.11 Effect of micelle size and charge on *ec*CytD dimerization rates. (A) The effect of micelle concentration on the rate of dimerization in 25 mM phosphate buffer pH 6.5, 150 mM NaCl and 100 µM EDTA at 45ºC. Observed rate constants for *ec*CytD dimerization with various concentrations of Fos16 (cyan) plotted as a function of micelle concentration, suggesting a linear relationship between the speed of dimer formation and micelle concentration ($y = 2.033\ x$ with $R^2 = 0.9999$). Observed rate constants obtained in the presence of 2.5 mM LPPG (red), 15 mM Fos10 (purple) and 10.5 mM Fos12 (green) are also indicated. (B) The effect of DDM on the *ec*CytD dimerization rate. Rate profile of 0.25 mM monomeric *ec*CytD in 25 mM phosphate buffer pH 6.5, 150 mM NaCl and 100 µM EDTA incubated with 2.5 mM DDM (squares) is compared to the same conditions without DDM (circles) at 45°C. The observed rate constant and K_d in the presence of DDM were determined to be 1.5×10^{-4} s^{-1} and 0.44 mM, respectively. (C) The effect of micelle size on the rate of dimerization. k_{obs} for Fos-detergents of increasing alkyl chain length (10, 12 and 16 carbons) appear to follow a linear relationship with detergent aggregation numbers. Since the rate of dimerization for 0.166 mM Fos16 micelles is too fast to measure, the rate was extrapolated from the linear relationship in (A) and is indicated by the asterisk. (Symbol color follows that shown in A.)

This could reduce the likelihood that the dimerization-competent conformation will find its binding partner. In contrast Fos10 and Fos12 micelles appeared to favor the dimeric state, with dimer dissociation constants of 0.27 ± 0.002 mM and 0.30 ± 0.001 mM for Fos10 and Fos12. Since a true catalyst should only alter reaction rates while keeping the equilibrium energetics the same, only the 2.5 – 15 mM Fos16 conditions can be considered to be truly catalytic.

4.3.6 Dimerization induced by incubation with a denaturant.

Although the mechanism for micelle-catalyzed domain swapping is not yet known, it is likely that some unfolding is involved in this process. To evaluate the importance of promoting the unfolded state, the rate of dimerization induced by the chemical denaturant urea was evaluated. These studies showed that 1M urea had a modest effect on the rate with a 3-fold increase (Fig. 4.12). However, urea also had an impact on monomer and dimer concentrations at equilibrium, leading to a two-fold reduction in the affinity of the dimer (K_d = 0.70 mM). These results suggest that simple exposure of *ec*CytD to denaturing conditions is not sufficient to significantly accelerate the rate of monomer-dimer interconversion.

4.3.7 ecCytD-mediated dimerization in the full-length GlpG protein.

Although the isolated *ec*CytD dimerizes through a domain swapping interaction, it is not clear if a similar interaction would occur in the full-length *ec*GlpG. During the size exclusion chromatography purification step of *ec*GlpG, a broad elution profile was obtained likely due to the large amounts being loaded for the purpose of purification. (Fig. 4.13). However, direct application of fractions taken from early- or late-eluting regions of the peak back onto the same SEC column gave rise to two distinct elution profiles corresponding to a higher and lower molecular weight species, respectively. It is possible that this could arise

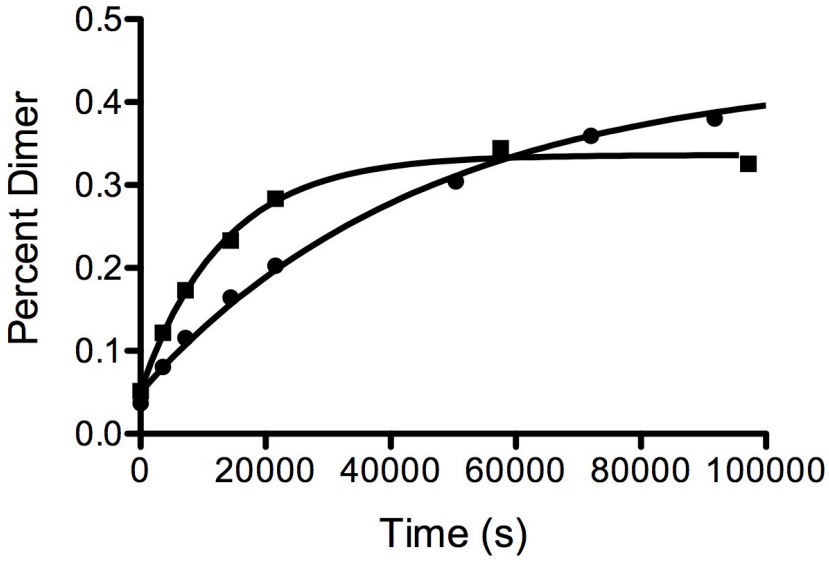


Figure 4.12 The effect of 1M urea on the *ec*CytD dimerization rate. 0.25 mM monomeric *ec*CytD in 25 mM phosphate buffer pH 6.5, 150 mM NaCl and 100 μM EDTA was incubated with 1M urea (squares) and without urea (circles) at 45°C yielding observed rate constants of 7.5×10^{-5} and 2.2×10^{-5} s^{-1}, respectively. The K_d's were determined to be 0.70 and 0.36 mM with and without 1M urea, respectively.

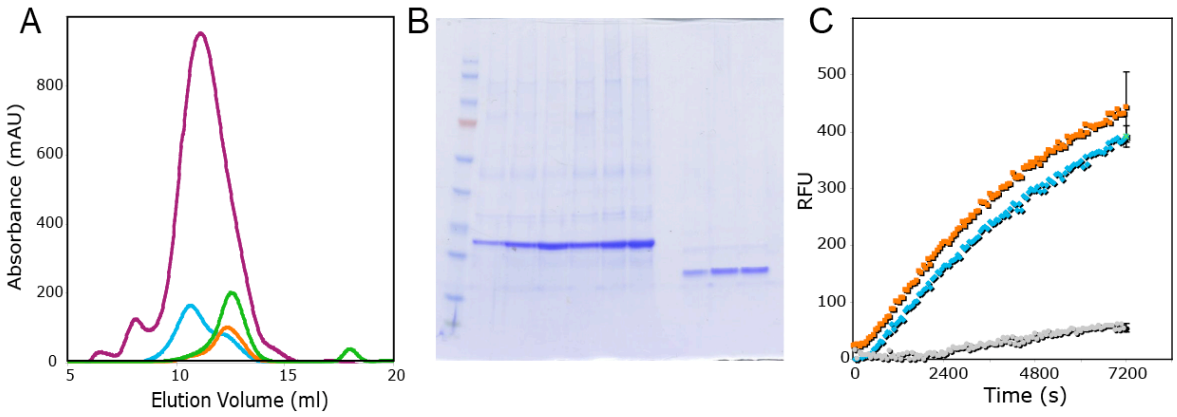


Figure 4.13 Size exclusion chromatography profiles suggest full-length, but not TMD *ec*GlpG also undergoes dimerization. (A) Early and late size exclusion fractions from the initial *ec*GlpG purification (purple) were directly reinjected onto the same SEC column (Superdex-75 10/300, GE Healthcare). Fractions taken from the first part of the peak eluted as 2 species when reapplied (blue profile), with the larger molecular weight species being dominant. Fractions from a later part of the peak eluted primarily as the lower molecular weight species (orange). This lower molecular weight species was also obtained when a limited trypsin digest of the purified sample was performed to remove the *ec*CytD prior to SEC (green). Approximate molecular weights of the two peaks as determined from comparison of protein molecule weight standards are ~260 and ~140 kDa. All samples were purified in 50 mM HEPES pH 7.3, 0.1% DDM, 200 mM NaCl, 10% glycerol and 100 µM EDTA. (B) Coomassie-stained SDS-page gel of fractions between 10-13 ml (lanes 1-6) of the *ec*GlpG SEC purification (purple profile in a) and fractions between 11-12.5 ml (lanes 8-10) of the pooled full-length fractions digested with trypsin overnight at 4ºC in a 1:20 ratio. (C) Monomeric *ec*GlpG (orange) and dimeric *ec*GlpG (blue) are compared to the inactive S201A mutant (grey) where the specific activities were determined to be 0.10 ± 0.02, 0.08 ± 0.01 and 0.013 ± 0.007 RFU/s/µM, respectively. Error bars are shown for the last time point and represent standard deviation for samples assayed in duplicate.

due to differences in the amount of detergent that is associated with each complex. However, it is unlikely that complexes with different amounts of detergent would undergo slow exchange to allow observation of the discrete species observed in the two elution profiles. Instead we suggest that the two species arise due to the slowly-exchanging domain swapping interaction in the cytoplasmic domain. This idea is substantiated by a similar experiment performed with the TMD generated using limited trypsin digestion. In this case the size exclusion profile of the catalytic TMD core revealed the presence of only one species with a molecular weight corresponding to that of the smaller species observed with the full-length *ec*GlpG. This suggests that the TMD on its own is unable to undergo the interactions that give rise to the higher molecular weight species, with the *ec*CytD being required for this slowly-exchanging dimerization.

To determine if oligomerization state influenced rhomboid activity, fractions from purified monomeric and dimeric *ec*GlpG was incubated with the fluorogenic casein substrate described in Chapter 3. The specific activities were not significantly different in this *in vitro* assay (Figure 4.13c), although this does not rule out potential differences that might occur in a lipid bilayer environment.

4.4 Discussion

4.4.1 ecCytD dimerizes through 3D domain swapping.

In this chapter we have identified a domain swapping dimer interaction in the GlpG cytoplasmic domain. This interaction occurs in the isolated *ec*CytD with evidence that the full-length rhomboid protease can dimerize through this domain as well (Fig. 4.13). The domain-swapping capabilities of this extramembranous domain of a membrane protein

places *ec*GlpG in a small subset of proteins having the same type of interaction, with only one other known example. Specifically, the *Pseudomonas fluorescens* inner membrane protein, LapD, has a domain swapped periplasmic domain (183), although it is unclear if this domain can exist in the monomeric form, or if there are conditions that can induce interconversion. *ec*CytD is therefore unique since both monomeric and dimeric states are stable, and interconversion can be stimulated by exposure to elevated temperatures or detergent micelles.

Interconversion of the two oligomeric states in *ec*CytD was slow at room temperature, reflecting the large energy barrier of 80 kcal/mol separating the two states. The size of this energy barrier is comparable to that for other domain-swapped dimers: e.g. 99 kcal/mol for stefin A, (184), and 80 kcal/mol for BCL-$_{XL}$, 80 kcal/mol (185). Furthermore, the enthalpy of dimer formation was found to be favorable (-15 kcal mol^{-1}), compensating for the decrease in entropy (-32 cal K^{-1} mol^{-1}). The magnitude of the enthalpic term suggests that additional interactions occur in the dimer that do not exist in the monomeric state, as has been seen for other domain-swapped dimers (186-189).

The NOEs detected in the filter-edited ^{13}C-NOESY experiment on the ^{15}N, ^{13}C-/^{14}N, ^{12}C-labeled heterodimer revealed a pattern of intermolecular NOEs involving the C-terminal helix and β3, and other parts of the structure that they interact with in the monomer. While structure determination of the domain-swapped dimer is currently underway, this data indicates that the interaction involves domain-swapping of these structural elements, with the loop connecting β2 to β3 acting as the hinge that undergoes the largest structural change between the two states (Fig. 4.9b).

4.4.2 The detergent micelle as a catalyst for domain swapping.

To the best of our knowledge, we believe that the work presented in this chapter provides the first example of micelle-catalyzed domain swapping. Although it has been shown that submicellar concentrations of SDS can facilitate dimerization in the pollen allergen, Bet v 4 (190), a small population was also found to aggregate under these conditions, with the aggregate being the dominant species at micellar concentrations of SDS. Detergent micelles have also been implicated in dimerization of the anti-apoptotic protein, BCL-$_{XL}$, which can bind to membranes and oligomerize to form a pore (191). Although this protein has been shown to form a domain-swapped dimer in solution under various conditions (e.g. acidic or basic solutions at elevated temperatures (192, 193)) in non-ionic detergents a very small fraction of the population exists as a disulfide-bonded dimer with an interaction surface that differs significantly from that of the aqueous domain-swapped dimer (192, 194). It is not known if the membrane-bound dimer interacts through a domain-swapping interaction, with no evidence yet available that could distinguish between domain-swapping versus non-swapping interactions. However, it is interesting that the detergents that give rise to the alternate dimer do *not* favor formation of the domain-swapped dimer, with the monomer being the only other observed state under these conditions (192).

4.4.3 How does the detergent micelle catalyze domain-swapping?

Given the central role played by β3 in the structure of the *ec*CytD monomer, it is likely that a significant amount of unfolding is required for domain swapping to occur. In previously characterized examples of high barrier domain swapping interactions, the transition state has been proposed to involve either partial unfolding (178, 195) or nearly complete denaturation (188, 196). For proteins that exchange secondary structure elements,

like chicken cystatin, the transition state tends to be largely unfolded (196). This may also be the case for *ec*CytD domain swapping, since multiple hydrogen-bonding and van der Waals interactions involving the central β3 strand would have to be disrupted for domain-swapping to occur.

Although unfolding is likely to be a significant part of *ec*CytD domain-swapping, it is not possible to know if this reaction can be approximated as a two-state system with a single transition state, or if one or more intermediates are involved in the process. For cyclin-dependent kinase 1 (p13suc1), several pieces of evidence were obtained supporting the hypothesis that exchange of the central β-strand must occur via a folding intermediate (197). Specifically, tryptophan fluorescence changes were monitored during p13suc1 unfolding and refolding, allowing the identification of the exchanging β-strand as a key component of the folding nucleus for both monomeric and dimeric states. Moreover, dimer formation occurs early in the folding coordinate from the monomeric state, a result that was supported in a molecular dynamics simulation of p13suc1 unfolding (198). These refolding experiments were done under conditions that would maintain a state of slow-exchange between monomeric and dimeric p13suc1. The fact that both forms were isolated at the end of the refolding reaction shows that the dimer folds independently from the unfolded state, without any requirement to first form a folded monomeric species (188). Although not directly evaluated in our work, it is possible that the mechanism of *ec*CytD domain swapping occurs via a similar pathway since, like p13suc1, the domain-swapped elements in *ec*CytD interact extensively with all other parts of the protein. If this is the case then interactions with the Fos16 micelle could promote unfolding to allow two largely unfolded *ec*CytD subunits to interact via their C-termini.

4.4.4 Micelle properties important for catalysis

One of the trends that was observed in the micelle-catalyzed interconversion was that changes in the size of the Fos micelle could alter the relative stability of the monomer and domain-swapped dimer. In general, smaller micelles tended to slightly shift the equilibrium toward the dimeric state. Although the mechanism for this change in equilibrium is not known, it is possible that smaller micelles may be less likely to interact with the dimeric state of *ec*CytD due to the smaller surface area available for binding. While these smaller micelles may still be able to bind the monomer to facilitate domain-swapping, interactions of the dimer with the micelle may be less favorable, reducing the ability of the micelle to liberate the domain swapped elements from the dimer. In contrast, larger micelles should be able to accommodate interactions with both monomer and dimer, helping both forms overcome the large energy barrier to similar extents (Fig. 4.14). In the case of Fos16, the micelle dimensions should significantly exceed those of the *ec*CytD dimer.

While this trend helps explain the effect of the phosphocholine detergents on domain-swapping rates, it should be noted that other factors are also important, particularly when anionic micelles such as LPPG are used. Although the micelle size for LPPG is predicted to be similar to Fos16, the equilibrium was shifted toward the monomeric state by this detergent. Since the rate of dimerization was not affected, the stability of the dimer relative to the monomer must be reduced by the LPPG micelle, although the reasons for this difference are not clear.

In summary, we have uncovered a unique role for Fos16 micelles as a catalyst for domain swapping in the N-terminal cytoplasmic domain of *ec*GlpG. This effect appears to be mediated by the zwitterionic detergent head group and the large binding surface of the

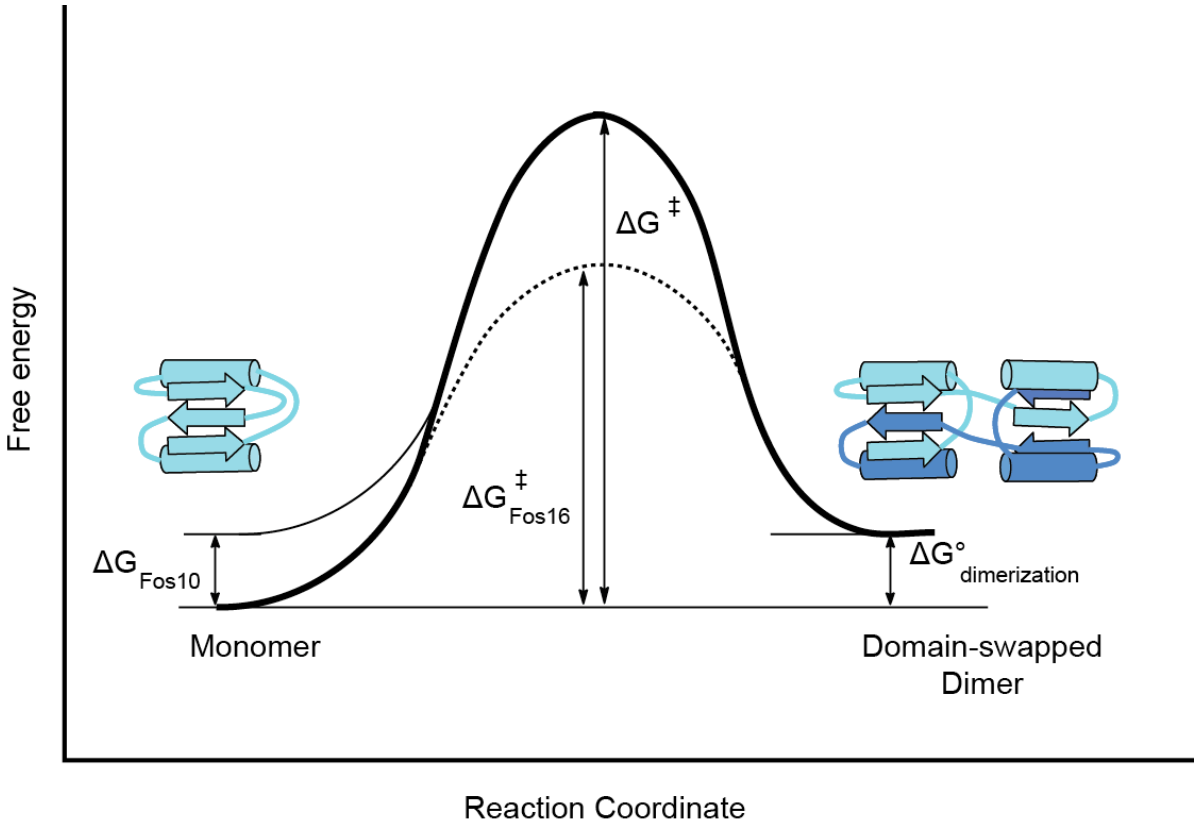


Figure 4.14 Energy diagram for the interconversion between monomer and dimer, assuming that it can be as a two-state system with a single transition state. In the presence of Fos10, monomeric *ec*CytD may have increased energy due to destabilizing interactions with the micelle (ΔG_{Fos10}). In contrast, Fos16 micelle dimensions should exceed those of the dimer and therefore accommodate interactions with both species to decrease the activation barrier ($\Delta G^{\ddagger}_{Fos16}$). Since equilibrium concentrations of monomer and dimer are not affected by the presence of Fos16, the larger micelles either selectively decrease the transition state energy or increase the energy of the monomer and dimer relative to the transition state by an equal extent.

Fos16 micelle, and may act to selectively stabilize unfolded intermediates separating the monomeric and dimeric states. An interesting question that remains to be explored is whether domain swapping could be catalyzed by micelles for other proteins, provided that the oligomer can fit within the dimensions of the micelle. Also, these findings raise the possibility that micelle-like structures that arise in biological systems (e.g. phase boundaries of lipid domains or regions of membrane curvature (199-203)) could act to catalyze domain swapping interactions for other proteins that are known to undergo domain-swapping interactions in vivo. This will be a particularly important question to investigate in the future, given that domain-swapping interactions form the basis of some pathological forms of protein aggregation (204).

Chapter 5: Discussion and Future Directions

5.1 Effect of membrane mimetics on the full-length rhomboid protease.

In this thesis I have characterized the role of interactions at the membrane involving different parts of the cytoplasmic side of the bacterial rhomboid. However, one of the original reasons that we began this work with the *P. aeruginosa* rhomboid was to investigate the feasibility of pursuing solution NMR studies on a full-length rhomboid protein. At the time, x-ray crystal structures were not available for any intramembrane protease, and crystals of *pa*GlpG obtained at the Structual Genomics Consortium in Toronto displayed poor diffraction (C. Koth, personal communication). Since that time it has become apparent that elimination of the cytoplasmic domain was paramount in obtaining well-diffracting protein crystals, with all rhomboid structures (*ec*GlpG and *hi*GlpG) obtained with the catalytic transmembrane domain alone. Although these structures gave rise to a wealth of information regarding how a hydrolysis reaction can occur in the hydrophobic membrane environment, we were interested in studying the full-length rhomboid since evidence from the Ha group (65) as well as our own work (Chapters 2 and 3) suggested a functional role for the N-terminal CytD.

One of the first detergents that was tested in our studies was DDM, since it retains an active state for the rhomboids from *B. subtilis (*YqgP) (62, 63), *E. coli* (*ec*GlpG) (49, 62, 63), *P. stuartii* (AarA) (63) and *A. aeolicus* (63). Although DDM-solubilized *pa*GlpG did remain functional in our hands, virtually no part of the protein could be detected in even the most basic NMR experiments. One possible explanation for this was that the estimated molecular mass of the protein-detergent complex (PDC) was ~100 kDa (*pa*GlpG alone is ~32 kDa), which is extremely large with respect to the limitations of a 500 MHz NMR spectrometer. More precisely, at this molecular mass transverse (T_2) relaxation times tend to

be very short, leading to broad linewidths and inefficient coherence transfer in multidimensional experiments (142, 205). It was therefore not surprising that peaks from the catalytic TMD were not detected for DDM-solubilized *pa*GlpG in the ^{1}H-^{15}N HSQC spectrum (Fig. 2.4a).

To obtain a solution NMR spectrum of a large, slowly tumbling protein-detergent complex there are some approaches that can be used to increase signal-to-noise (reviewed in (143, 145, 206)). For example, substitution of all non-exchangeable protons with deuterium reduces the number of relaxation pathways (92), leading to significant increases in transverse relaxation times and therefore sensitivity. In addition, transverse relaxation optimized spectroscopy (TROSY) experiments can be done on large, slowly tumbling molecules to isolate the slowly-relaxing component of the magnetization (142, 207-209). The effectiveness of the TROSY experiment for acquisition of spectra on larger proteins increases with magnetic field strength (up to ~1 GHz for ^{1}H-^{15}N) (142). However, in the case of the *pa*GlpG, we suspect that the large PDC molecular weight was not the only problem hindering acquisition of a NMR data since a ^{2}H, ^{15}N-labeled sample in DDM was analyzed on a 800 MHz spectrometer at the Québec/Eastern Canada High Field NMR Facility, but did not allow observation of any peaks from the TMD region (data not shown).

The fact that the signals from the TMD could not be obtained, even when higher field strengths, TROSY experiments and deuteration were used, suggests that this region of the protein could be undergoing conformational exchange on a timescale that is unfavorable for solution NMR (i.e. μs – ms) (91). One possible source of this conformational exchange may come from the large size of the DDM micelle, and how it compares to the size of the *pa*GlpG. In a study on a model 2-TM membrane protein by Columbus and coworkers,

spectral quality appeared to depend on micelle thickness (210); they found that highest quality NMR spectra were obtained in detergents that could match the size of the hydrophobic regions of the protein. One detergent that seemed to promote conformational exchange in this protein was DDM. It was proposed that this detergent destabilized the interaction between TM helices in this case, since the polar head group separation in the DDM micelle is longer than the hydrophobic regions of the TM-helices (210) and created a thicker hydrophobic region in the micelle. Since the rhomboid is believed to have a relatively narrow hydrophobic region (76), conformational exchange arising from hydrophobic mismatch may be more significant for the rhomboid sample.

Conformational exchange is a recurring theme in the study of integral membrane proteins through solution NMR (127, 211, 212). In a study that screened eight *Thermotoga maritima* membrane proteins in eleven different detergents, the longer alkyl chain detergents were found to be better at solubilizing these proteins but had a greater tendency to promote conformational exchange that degraded the quality of NMR spectra (213). In this work detergents with an aggregation number of ~70 appeared to give rise to the best conditions for solution NMR of this subset of proteins, which is smaller than the aggregation number for DDM micelles (Table 2.1).

These discoveries regarding the relationship between detergent micelle dimensions and quality of solution NMR spectra of membrane proteins were released after we had completed our solution NMR feasibility studies on the full-length protein. However, these results, in addition to our discoveries regarding the effect of detergents on the CytD and the role of residues 81-91 on the N-terminal side of TM1, have the potential to be applied to future efforts to study the rhomboid TMD by solution NMR. Options worth investigating

include the use of *ec*TMD81, which will be more stable than the full-length protein, and the use of detergent mixtures designed to generate micelles with hydrophobic regions that match the narrow band of hydrophobic surface area that has been identified in the x-ray structures.

An additional conclusion that we can draw based on our study on the isolated cytoplasmic domain is that it may not be feasible to obtain an NMR spectrum of the full-length rhomboid. In particular, the higher temperatures that are typically required to increase the rate of molecular reorientation of the large PDC complex, and therefore decrease the loss of signal from rapid relaxation processes (127), would promote dimerization of a monomeric sample. Since the dimer affinity is weak, an additional source of heterogeneity in a full-length rhomboid protein sample would arise from the mixture of monomeric and dimeric rhomboid that would be present under these conditions. Solution NMR studies of the full-length *ec*GlpG will therefore likely require that *ec*CytD mutants be engineered that prevent domain swapping-dimerization interactions and stabilize the monomer structure, thereby allowing a homogeneous population of stable full-length rhomboid to be maintained under higher temperature conditions.

Regardless of whether the low quality *pa*GlpG spectrum is a result of conformational exchange and/or complex size, it remains interesting that signals from the soluble cytoplasmic domain were not detected in DDM. This domain is connected to the TMD by a ~30 residue sequence that is expected to be unstructured based on secondary structure prediction using JPred (150) (Fig. 2.1). The lack of structure in this region was confirmed in my own work with the *pa*CytD, which had degenerate shifts characteristic of random coil values for these residues. The first constructs with the *ec*CytD also contained this region of the sequence, and gave rise to ^{1}H-^{15}N HSQC spectra containing some strong peaks in the

central region of the spectrum characteristic of unfolded polypeptide (data not shown). These peaks were not present in the final construct used to determine the NMR structure.

One potential explanation for the absence of NMR signal from the CytD in full-length *pa*GlpG in DDM was that this domain may interact with the detergent micelle or with other parts of the protein. Since we could rule out the possibility that the CytD had an intrinsic affinity for DDM (Section 2.3.4) this pointed to a structural interaction with the TMD. However, while not appreciated at the time, my results showing *ec*CytD dimerization indicates that the CytD was likely engaged in intermolecular homodimeric interactions. This would significantly reduce the mobility of the domain, giving rise to CytD molecular tumbling rates that would be much closer to those of the TMD-micelle complex. Although CytD dimerization was characterized in *ec*GlpG while the solution NMR spectra were acquired on *pa*GlpG, the two domains are very similar in overall structure (Fig 4.7). Hence, it is possible that the *pa*CytD also forms a domain-swapped dimer in DDM micelles and therefore prevents detection of signals from this domain in the HSQC spectrum. However, future experiments similar to those described in Chapter 4 with *ec*CytD would be required to confirm this hypothesis.

5.2 The effect of detergent monomers on aqueous regions of membrane proteins.

This study highlighted the impact of interactions between detergent monomers and aqueous solvent-exposed regions of detergent-solubilized proteins. These underappreciated interactions could help to explain lower activities observed in future membrane proteins solubilized in bicelle systems. For example, diacyl glycerol kinase (DAGK), a 3-TM helix trimer, may have been influenced by these types of interactions when reconstituted in

DMPC bicelles, since its activity was lower than that in DMPC vesicles, or in bicelles composed of phospholipids with longer acyl chains (162). Detergent monomers have also been found to bind inappropriately to membrane proteins, as was seen in the x-ray crystal structures for PagP and the M2 channel. In this case PagP was found to have a lauryldimethylamine-oxide (LDAO) detergent molecule bound in the active site of this β-barrel protein (214). Similarly, an x-ray structure of the M2 channel from the influenza A virus was found to have a β-octylglucoside (βOG) molecule within the channel (215). These undesired interactions with monomeric detergent are not limited to micellar solutions, but can also occur in bicelles as was revealed in Chapter 2 for *pa*GlpG. In particular, lower activity against the Spitz substrate was obtained in bicelles, possibly caused by interactions involving CHAPS or DHPC monomers that we showed to interact with the *pa*CytD.

One way to avoid these issues involving undesirable interactions with monomeric detergent is to use alternate membrane mimetics that could solubilize the rhomboid in an environment without detergent monomers in solution. One possibility is the nanodisc, which is a bilayer composed of less than 200 lipids wrapped in an apolipoprotein membrane scaffold protein (216, 217). This scaffold protein replaces the short chain detergent that shields the hydrophobic rim of the bilayer from solvent water. However, incorporation of the rhomboid into nanodiscs would create a large complex (>200 KDa) that would make it difficult to observe by solution NMR, although relaxation-optimized sequences tailored to these very large complexes could potentially allow some insights to be obtained (reviewed in (218)). Another detergent-free alternative that could be considered is the amphipathic polymer – or amphipol. These molecules essentially consist of a hydrophilic backbone with hydrophobic side chains to create an amphipathic polymer that can maintain membrane

proteins in a soluble state (219-221). While the feasibility of amphipols for the study of multi-helical membrane proteins through solution NMR has yet to be determined, promising results have been obtained for the β-barrel OmpA with only a small increase in line-broadening compared to the spectrum of OmpA solubilized in DHPC (222). Use of these alternate membrane mimetics to stabilize the rhomboid may one day comprise a reasonable approach to avoid undesirable interactions with monomeric detergent in solution.

5.3 Interactions at the membrane in the *E. coli* rhomboid

The functional importance of the *ec*GlpG CytD was initially suggested from the activity of a TMD fragment produced using chymotrypsin (65), results we were able to reproduce using trypsin in place of chymotrypsin (Fig. 3.1). However, functional analysis of TMD truncation constructs revealed that the structured region of CytD was not responsible for the observed decrease in activity but instead a ~10-residue sequence N-terminal to TM1 was the only element required to restore full activity. This sequence (^{80}YRRYPFF**A**AL**R**ERAG^{94}) contains four potential trypsin cleavage sites, while N-terminal sequencing performed on the chymotrypsin-isolated TMD indicated that this sample began at either A87 or R90 (65). We suggest that fully active rhomboid must include Y83, which is absolutely conserved in the *Enterobacteriaceae* species. Future studies should target Y83 for mutagenesis, as well as perform alanine scanning of this N-terminal sequence to specifically identify residues involved in this activating effect.

Our results with the fluorophosphonate probe provided confirmation of the functional importance of residues 81 to 90, and suggest that it helps maintain the active site serine in a nucleophilic state. However, another possibility that we cannot fully exclude is

that the truncation does not perturb the active site, but instead causes a structural change that restricts access of the probe to the active site. While it has been shown that the membrane-impermeable, bulky thiol alkylating agent 4-acetamido-4′-maleimidylstilbene-2,2′-disulfonic acid (AMS) can label cysteine mutants in the TMD aqueous cavity (223), it is possible that this accessibility is altered by the truncations. Nonetheless, significant differences in accessibility do not seem likely since all x-ray crystal structures available for *ec*GlpG were made with truncation mutants starting at A87 or later in the sequence. These structures show variable accessibility of the active site Ser residue that depends on the conformation of L5. The range of conformations observed for this loop implies that it is a dynamic element, with both open and closed states existing in solution. This suggests that the active site Ser remains water accessible in these truncated TMD constructs. Further support for the high solvent accessibility of active-site residues in truncated TMD constructs comes from structures cocrystalized with either a lipid head group (64) or isocoumarin (72) bound in the active site. The electrophilic fluorophosphonate moiety is smaller than both of these molecules, suggesting that it should be freely accessible to the active site residues in the truncated rhomboid. Taken together this evidence suggests that the reduced labeling observed with the fluorophosphonate probe reflects a change environment surrounding the active-site Ser such that its nucleophilicity is significantly reduced. Future experiments probing the ability of active-site cysteine mutations to be modified by AMS could be done as described by Maegawa and coworkers (223) to confirm the absence of an effect on accessibility by TMD truncation.

Future studies incorporating both *in vivo* western analysis of substrate cleavage and FP-PEG-rhodamine labeling of crude extracts could provide a more complete picture of how

different mutations affect substrate gating and/or the active site directly. Activity-based profiling may also be applied to labeling in live cells with some optimization of the labeling procedure and/or the probe itself. In a preliminary set of experiments, C43 (DE3) *E. coli* cells expressing either *ec*GlpG or the inactive S201A mutant were incubated with the FP-PEG-rhodamine in the attempt to label live cells; however in-gel fluorescence analysis of whole cell lysates did not reveal significant labeling of the rhomboid. One possibility for this absence of labeling is that the probe may not be able to cross the outer bacterial membrane. Outer membrane porins typically have a molecular weight cut-off of 600 Da (224) and the larger size of FP-PEG-rhodamine (838 Da) may prevent it from crossing the outer membrane. If this is the case, it may be possible to modify the probe such that it could cross this barrier, for example, by replacing the rhodamine fluorophore with an alkyne or azide functional group. Once labeling was complete, the cells could then be lysed and reacted with a fluorophore through an azide-alkyne Huisgen cycloaddition reaction, commonly referred to as a "click" reaction (225) where the azide/alkyne coupling partners of the probe and tag form a stable triazole product. This process could allow for the active site of the rhomboid to be probed directly *in vivo* for a better understanding of how this protein behaves in its native membrane environment. For example, activity-based profiling could be used in the characterization of activity states in parasitic life cycles and other pathological states attributed to the rhomboid detailed in Section 1.2.1.

There are only a handful of inhibitors known to bind to the rhomboid active site: isocoumarin, tosyl phenylalanyl chloromethyl ketone (TPCK) (13, 63) and, more recently, β-lactams (226). Diisopropylfluorophonate has also been described as a rhomboid inhibitor, however concentrations between 0.1-0.5 mM were required for inhibition. Here we present

the less bulky ethyl fluorophosphonate group of FP-PEG-rhodamine (Fig. 3.6a) as a more effective fluorophosphonate inhibitor since *ec*GlpG was labeled with 10 nM of probe. The FP-PEG-rhodamine probe can also be used to screen for more potent and/or selective inhibitors using a competitive assay. Identification of new and more selective rhomboid inhibitors not only has a medical relevance in drug development against pathogens, but could also be used to speed progress in uncovering the biological roles of these intramembrane proteases.

5.4 Dimerization of the *ec*GlpG CytD through 3D domain-swapping.

As described in Chapter 4, during the course of our structural studies on the GlpG CytD, we discovered a domain swapping interaction, with monomeric and dimeric states separated by a large energy barrier. In fact, domain swapping in membrane proteins is extremely rare, with the CytD of *ec*GlpG being the first example of this type of oligomerization occurring in an intramembrane protease. As described in Chapter 4, to the best of our knowledge, this constitutes just the second example of domain swapping in a membrane protein extramembranous domain.

Since the original description of 3D domain swapping, over 290 domain-swapped structures have been published as of the beginning of 2011 (227). A subset of these structures are likely to be an artifact of protein crystallization created by high protein concentrations and non-physiological pH; nonetheless there is evidence that other domain-swapping interactions could have physiological relevance. Bovine seminal RNase (BS-RNase), for example, is found as an N-terminally domain-swapped dimer *in vivo* where the two active sites are composed of histidines from two separate polypeptide chains (228).

Based on this structure, it was proposed that the dimer interface could be involved in allosteric control through communication between the two sites (228, 229). Furthermore, it was shown that dimeric BS-RNase is less active than its monomeric form (230), and displays a unique cytotoxicity in tumor cells, while monomeric BS-RNase does not (231). Another well-characterized example of physiologically relevant domain swapping is provided by glyoxalase I from *Pseudomonas putida.* This protein was isolated as both monomeric and dimeric species, with the active site being formed by residues from the two swapped domains (231). Particularly interesting was the observation that the active dimer could be reversibly converted into the less active monomer through modulation of glutathione concentrations *in vitro*, giving rise to the proposal that functional regulation could occur through changes the redox environment *in vivo*.

The numerous examples of functionally important domain-swapping interactions raise the possibility that the GlpG CytD domain swapping interaction may have functional relevance *in vivo*. This possibility is supported by the observation that both monomeric and dimeric states could be isolated for the full-length *ec*GlpG, while the TMD truncation was largely monomeric, as determined by SEC. Although we did not observe any difference in specific activity against casein for the purified monomer and dimer forms of *ec*GlpG, the possibility remains that activity differences could result when GlpG is constrained to a lipid bilayer environment, or when assayed against the actual, albeit currently unidentified, native *ec*GlpG substrate. For example, dimerization of GlpG could obstruct substrate entry between TM2 and TM5. Furthermore, interactions may occur between the CytD and the native *ec*GlpG rhomboid substrate, similar to what was observed for the human RHBDL2 rhomboid against thrombomodulin (77). In this case, domain swapping in the CytD could

modulate the affinity of the CytD towards the *ec*GlpG substrate, adding a level of regulation to the bacterial rhomboid protease.

5.5 Catalytic mechanism of 3D domain swapping by detergent micelles.

One of the most novel findings to come out of our study of the domain swapping for CytD was the fact that it could be catalyzed by zwitterionic micellar detergents. To better understand how the micelle could catalyze domain swapping of CytD, it will be important to explore the mechanism of interconversion between monomer and dimer for this protein. CytD interconversion can be considered as either a 2-state system with a single transition state, or as a multi-state system with one or several intermediate(s).

Among studies that have attempted to describe a mechanism for domain-swapping interactions, two of these propose a two-state model for the transition similar to the model forwarded for *ec*CytD in Section 4.4.2 (Fig. 4.13). This model is similar to that proposed for domain swapping in the diphtheria toxin (DT) (232) where a 148-residue globular structural domain is exchanged between two subunits. In this system the dimer could only be formed if DT was exposed to low pH conditions when it was present in high concentrations. The DT domain-swapped dimer is metastable at neutral pH, and will gradually revert back to the monomer with an activation barrier of 27 ± 1 kcal/mol at pH 7.5 (232, 233). In this system, the barrier is thought to represent the open form of the structure, with the globular domain being swapped dissociating from the main body of the structure.

In many cases domain swapping does not involve exchange of globular structural domains as was seen for DT, but instead swaps a small group of secondary structure elements as was described for well-characterized p13suc1 in Section 4.4. Although a strong

argument was provided that unfolding of p13suc1 is required for domain-swapping (188), the Wodak group has since proposed that domain swapping does not require the formation of an unfolded intermediate (234). They hypothesized that exchange of N- or C-terminal secondary structure elements could occur through the formation of a non-specific encounter complex involving a domain-swapping interaction between just a small number of residues at the domain-swapping end of the protein. Formation of this type of encounter complex could then allow the exchange of *intra-molecular* interactions for *inter-molecular* interactions for an increasing number of residues, gradually approaching the completely domain-swapped dimer configuration. This type of mechanism would minimize exposure of hydrophobic groups to the solvent and maximize the number of contacts maintained throughout the process, leading to a wide range of states with very similar energies over the course of the transition. They substantiated this hypothesis by computation of free energy profiles for their proposed domain swapping pathway in immunoglobulin G binding protein B1 domain from *Streptococcus* (GB1) (234). They found that the highest energy barrier was the formation of the initial slightly domain-swapped encounter complex. Once this barrier was reached, the intermediate structures of various degrees of domain swapping were very similar in energy, and hence should be freely sampled until the lower energy domain-swapped dimeric state is found. These same calculations were also performed for this type of swapping pathway for p13suc1, giving rise to a similarly isoenergetic profile along this part of the dimerization reaction coordinate and suggesting that unfolding is not required for domain-swapping interactions to occur in this protein either.

The GB1 domain-swapping interaction is generally only seen if a number of mutations are introduced that destabilize the monomeric state and concomitantly stabilize

the domain-swapped dimer. The ability of a destabilized monomer to undergo domain swapping at a relatively quick rate (~minutes) implied that a destabilized form of the monomer is the state that could form the encounter complexes with a small amount of swapping (234). This gave rise to a model of swapping for GB1 where an activated state of the monomer, potentially molten-globule-like in structure, is an intermediate along the domain-swapping pathway. A similar molten-globule-like dimer structure was also proposed in this swapping pathway (234). While the generality of the gradual domain swapping mechanism has not been established, it is interesting that this mechanism is actually the most consistent with the ability of the micelle to catalyze domain swapping for *ec*CytD. According to this model (Fig. 5.1), conversion of monomeric *ec*CytD into the dimer involves an activated state that is unfolded to some degree (M*), which can then collide with another activated monomer to form the initial encounter complex with one or two swapped residues. This complex could be considered to be an activated dimer (D*), which would progressively increase the number of domain swapping interactions by sampling the relatively smooth energy potential surface. If *ec*CytD undergoes this type of swapping mechanism, then it would be expected that interactions with the detergent micelle could act by decreasing the energy of the activated state, thereby increasing its population. Similarly, the micelle may stabilize a less folded, activated state of the dimer, allowing the reverse reaction to be accelerated to the same degree. This would account for the ability of the micelle to increase interconversion rates catalytically, that is without disturbing the equilibrium. This idea is substantiated by the increase in stability of the dimer relative to the monomer when smaller detergent micelles were used. This model is also consistent with the small effect that other types of mildly-denaturing conditions have on the interconversion

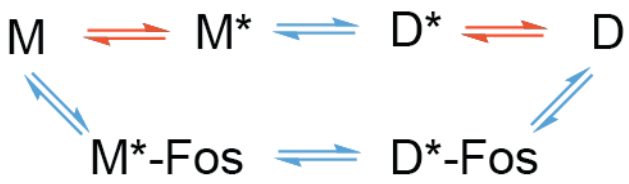


Figure 5.1 Model of domain swapping for *ec*CytD with the proposed mechanism for detergent micelle catalysis. M and D represent the folded state of monomeric and dimeric CytD, respectively, while M* and D* represent activated states having some degree of unfolding. Binding of CytD to the phosphocholine micelle is denoted by FOS with high energy barriers indicated by red arrows and low energy barriers indicated by blue arrows.

rates for *ec*CytD. These denaturing conditions would destabilize the folded state, but may also destabilize the activated state since it is not completely unfolded.

While interactions with the micelle may be acting to reduce the energy of the activated states, it is not yet clear how the micelle does so. It is possible that the increase in hydrophobic surface area exposed to solvent in the activated state would be reduced by interactions with a micelle, making the less folded state more stable. Since our results suggest that micelle charge also appears to be important for this effect, it is possible that electrostatic types of interactions between the swapping termini and micelle headgroup charges may compensate for the loss of these interactions when the protein adopts a less folded activated state. It is also not known if domain swapping occurs on the micelle itself, or if the activated state must be released from the micelle before the appropriate encounter complex can form. However, it is important to note that catalysis does not seem to be occurring via a simple proximity effect where binding to the micelle increases the local concentration of monomeric CytD. If this was the case, then the equilibrium should shift toward the dimeric state, since proximity effects would not be expected to accelerate dissociation of the dimer.

In the future, NMR spin relaxation experiments can be done to measure exchange rates for monomeric and dimeric species (reviewed in (235)). If high quality spin-relaxation data can be obtained for multiple nuclei types, then it should be possible to fit these relaxation dispersion profiles to the Bloch-McConnell equations to obtain information about the structure of the activated species involved in this domain-swapping interaction (236-238). Similar studies could also be performed in the presence of catalytic quantities of micelles, providing a window into the structure of the activated states under these

conditions. If successful, these studies would provide the first structural insights into the intermediates of a domain-swapping interaction and provide new fundamental information into domain-swapping mechanisms in general.

5.6 Biological implications from micellar catalysis of CytD domain swapping.

The ability of micelles to catalyze GlpG CytD domain swapping interactions may provide a clue for a role of the lipid membrane for domain swapping provided it occurs *in vivo*. The bacterial membrane is not a homogeneously-distributed mixture of phospholipids since cardiolipin, phosphatidylglycerol and phosphatidylethanolamine have been found to segregate within the membrane (200, 201, 203). More specifically, cardiolipin is found in the cell poles where there is negative membrane curvature (200, 202) while phosphatidylglycerol and phosphatidylethanolamine tend to segregate into distinct domains that differ in composition (201, 203).

The presence of local irregularities in bilayer structures in native lipid membranes raises the possibility that interactions with certain lipid domains could alter the orientation of the catalytic TMD with respect to the membrane surface and/or alter the dimerization state of the CytD. In addition, local structural features of these membrane domains may give rise to environments that could facilitate interactions with the CytD. Specifically, regions with a high degree of membrane curvature would make the hydrophobic phase of the bilayer more accessible for interactions with proteins with a weak tendency to bind to the membrane surface. This type of structure could promote a micelle-like interaction with the CytD that could catalyze domain swapping. Although we tested the effect of small (~100 nm) *E. coli* lipid vesicles to catalyze the domain-swapping interaction, no effect in dimerization rate was

found. Nonetheless, it is possible that the vesicle size did not match the curvature potentially required to induce dimerization. This would follow the example of other membrane-curvature sensing elements, such as the ALPS motif, an unstructured 20-40 residue sequence that folds into an α-helix upon binding of small vesicles with a radius of less than 50 nm (reviewed in (199)) and with lower affinity to larger vesicles (239). It is possible that the vesicles used in Chapter 4 were not appropriately sized to have an effect if the CytD senses a specific membrane curvature. Alternatively, the CytD might sense lipid packing defects in the bilayer that arise from membrane deformation (199), potentially caused by the rhomboid itself (76), or by specific localization of the rhomboid within the membrane.

5.7 Future directions

The biological function and substrate of the bacterial rhomboids, excluding *P. stuartii* AarA, are generally still unknown. Based on general rhomboid specificity features determined by the Freeman group (72), it is now possible to search transmembrane segments from single-pass type I bacterial membrane proteins for potential rhomboid cleavage sites. Identification of the substrate may help to reveal a specific function for the CytD and improve our understanding of how domains outside the membrane can alter the activity of an active site that is embedded within the membrane.

The approaches that were developed in this thesis for the study of the functional role of extramembraneous regions of the bacterial rhomboid have the potential to be applied to rhomboid proteases with clinical applications. For example, in addition to the α and β-cleavage events that modify function of the mammalian mitochondrial rhomboid, Parl

(described in Section 1.2.1.3), a third cleavage event has been described (γ-cleavage) to produce the Parl-Rhomboid-Domain (240) (PROD, Fig. 5.2). This γ-cleavage event essentially eliminates the additional N-terminal α-helix found in all mitochondrial rhomboids and results in the inactivation of Parl. Since the γ-cleavage site is very close to the N-terminal side of the first TM segment and the resulting PROD is reminiscent of inactive TMD91 described in Chapter 3 (Fig. 5.2), future structural comparisons of active TMD81 versus inactive TMD91 may provide insight into how γ-cleavage is inactivating for Parl. Furthermore, it has already been shown that the RHBDL2 cytoplasmic domain is required for cleavage of thrombomodulin (77) (described in Section 1.2.4.2); it is possible that removal of this domain through an event similar to γ-cleavage may be a way to modulate RHBDL2 activity *in vivo.* These and other clinically relevant rhomboids containing extramembraneous domains (e.g. TgROM, PfROM, RHBDL2) remain tantalizing prospects for future studies into the role of these domains in regulation of rhomboid function.

5.8 Conclusion

In this thesis, novel interactions at the membrane were identified that could modulate the function of the bacterial rhomboid protease. As detailed in Chapter 2, experiments probing *P. aeruginosa* rhomboid structure and function revealed a correlation between activity and CytD association with the protein-detergent complex. A model consistent with these results suggested that the CytD might interact with the catalytic TMD as a way to modulate rhomboid function. Chapter 3 sought to uncover this potential interaction in the homologous *E. coli* rhomboid, leading to the identification of other novel interactions

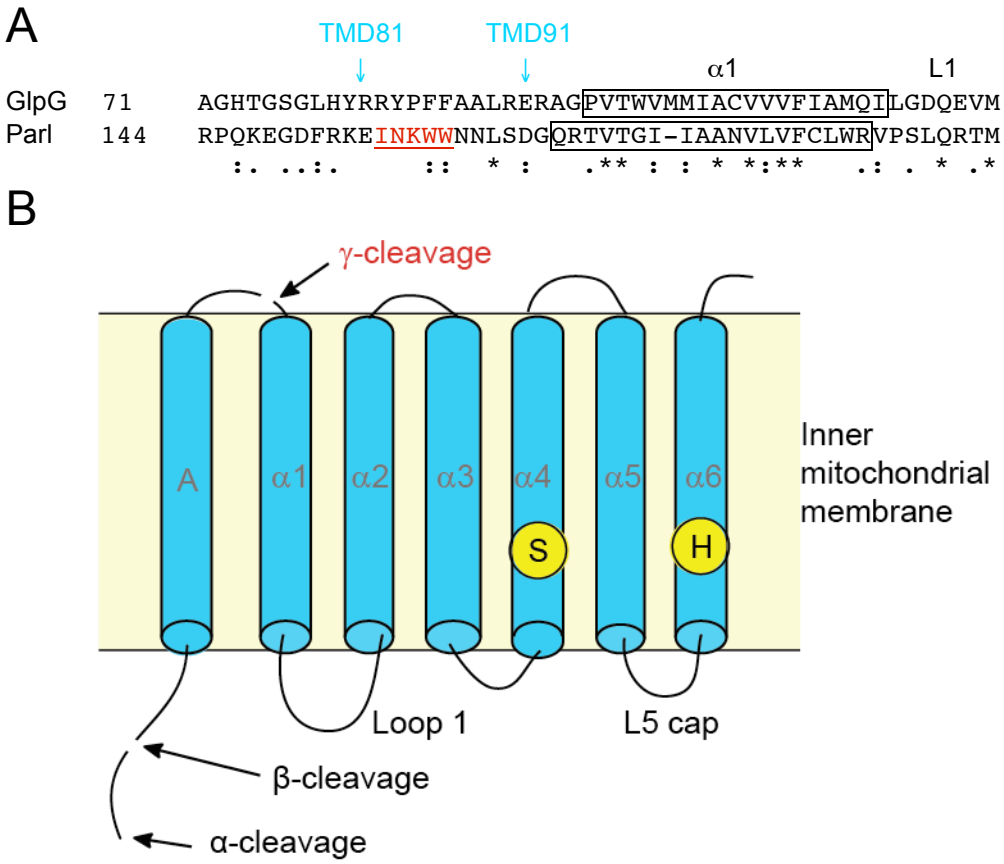


Figure 5.3 Comparison of inactivating γ-cleavage of mammalian Parl with the critical N-terminal sequence of *ec*GlpG. (A) Sequence alignment of the *ec*GlpG with Parl highlighting N-terminal residues of the active *ec*TMD81 and inactive *ec*TMD91, and the consensus sequence for the γ-cleavage site in Parl (red sequence) (240). Transmembrane helices are indicated by boxes. (B) Schematic representation of Parl cleavage sites. Core TM helices are indicated α1-α6 with the additional N-terminal TMH labeled A. α-cleavage generates mature Parl, β-cleavage is activating, and results in mitochondrial fragmentation while liberating the N-terminal peptide involved in mitochondrial retrograde signaling. γ-cleavage produces the Parl rhomboid domain (PROD) that is inactive.

occurring at the cytoplasmic face. Finally, structural analysis of the *ec*GlpG CytD revealed that the CytD dimerizes through domain swapping, an interaction with appears to be preserved in the full-length protein. This dimerization of the CytD provided an explanation for the absence of NMR signals for this domain in detergents supporting CytD function, since a CytD-mediated dimer would not have the same ability to dissociate from the detergent-protein complex. Although the biological function of this class of bacterial rhomboid is not yet known, this CytD dimerization could provide a mechanism for the modulation of rhomboid activity *in vivo*. Additional interactions at the TMD cytoplasmic face may also allow the rhomboid to sense membrane composition potentially adding another level of regulation. Overall, these findings open new avenues of investigation for the regulation of rhomboid protein function. Given the widespread conservation of rhomboid family members through all kingdoms of life, it is likely that regulatory mechanisms found in these bacterial rhomboids will be utilized by other members of the family, including those that are potential targets for drug development.

References

1. Turk, B. 2006. Targeting proteases: successes, failures and future prospects. *Nat Rev Drug Discov* 5:785-799.

2. Vermeulen, K., Z.N. Berneman, and D.R. Van Bockstaele. 2003. Cell cycle and apoptosis. *Cell Prolif* 36:165-175.

3. Vermeulen, K., D.R. Van Bockstaele, and Z.N. Berneman. 2003. The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer. *Cell Prolif* 36:131-149.

4. Patel, T.G., G.J. Kaufmann, S.H. 1996. The role of proteases during apoptosis. *FASEB J.* 10:587-597.

5. Riewald, M., and W. Ruf. 2003. Science review: role of coagulation protease cascades in sepsis. *Crit Care* 7:123-129.

6. Mittl, P.R., and M.G. Grutter. 2006. Opportunities for structure-based design of protease-directed drugs. *Curr Opin Struct Biol* 16:769-775.

7. Rawlings, N.D., and A.J. Barrett. 1993. Evolutionary families of peptidases. *Biochem J* 290 (Pt 1):205-218.

8. Hedstrom, L. 2002. Serine protease mechanism and specificity. *Chem Rev* 102:4501-4524.

9. Storer, A.C., and R. Menard. 1994. Catalytic mechanism in papain family of cysteine peptidases. *Methods Enzymol* 244:486-500.

10. Auld, D.S. 1997. Zinc catalysis in metalloproteases. *Struct Bond* 89:29-50.

11. Polgar, L. 1987. The mechanism of action of aspartic proteases involves 'push-pull' catalysis. *FEBS Lett* 219:1-4.

12. Rawson, R.B., N.G. Zelenski, D. Nijhawan, J. Ye, J. Sakai, M.T. Hasan, T.Y. Chang, M.S. Brown, and J.L. Goldstein. 1997. Complementation cloning of S2P, a gene encoding a putative metalloprotease required for intramembrane cleavage of SREBPs. *Mol Cell* 1:47-57.

13. Urban, S., J.R. Lee, and M. Freeman. 2001. Drosophila Rhomboid-1 defines a family of putative intramembrane serine proteases. *Cell* 107:173-182.

14. Wolfe, M.S., J. De Los Angeles, D.D. Miller, W. Xia, and D.J. Selkoe. 1999. Are presenilins intramembrane-cleaving proteases? Implications for the molecular mechanism of Alzheimer's disease. *Biochemistry* 38:11223-11230.

15. Duncan, E.A., U.P. Dave, J. Sakai, J.L. Goldstein, and M.S. Brown. 1998. Second-site cleavage in sterol regulatory element-binding protein occurs at transmembrane junction as determined by cysteine panning. *J Biol Chem* 273:17801-17809.

16. Erez, E., D. Fass, and E. Bibi. 2009. How intramembrane proteases bury hydrolytic reactions in the membrane. *Nature* 459:371-378.

17. Wolfe, M.S., and R. Kopan. 2004. Intramembrane proteolysis: theme and variations. *Science* 305:1119-1123.

18. Ye, J., U.P. Dave, N.V. Grishin, J.L. Goldstein, and M.S. Brown. 2000. Asparagine-proline sequence within membrane-spanning segment of SREBP triggers intramembrane cleavage by site-2 protease. *Proc Natl Acad Sci USA* 97:5123-5128.

19. De Strooper, B., P. Saftig, K. Craessaerts, H. Vanderstichele, G. Guhde, W. Annaert, K. Von Figura, and F. Van Leuven. 1998. Deficiency of presenilin-1 inhibits the normal cleavage of amyloid precursor protein. *Nature* 391:387-390.

20. Hardy, J., and D.J. Selkoe. 2002. The amyloid hypothesis of Alzheimer's disease: progress and problems on the road to therapeutics. *Science* 297:353-356.

21. Tanzi, R.E., and L. Bertram. 2005. Twenty years of the Alzheimer's disease amyloid hypothesis: a genetic perspective. *Cell* 120:545-555.

22. Koonin, E.V., K.S. Makarova, I.B. Rogozin, L. Davidovic, M.C. Letellier, and L. Pellegrini. 2003. The rhomboids: a nearly ubiquitous family of intramembrane serine proteases that probably evolved by multiple ancient horizontal gene transfers. *Genome Biol* 4:R19.

23. Urban, S., D. Schlieper, and M. Freeman. 2002. Conservation of intramembrane proteolytic activity and substrate specificity in prokaryotic and eukaryotic rhomboids. *Curr Biol* 12:1507-1512.

24. Mayer, U., and C. Nusslein-Volhard. 1988. A group of genes required for pattern formation in the ventral ectoderm of the Drosophila embryo. *Genes Dev* 2:1496-1511.

25. Urban, S. 2010. Taking the plunge: integrating structural, enzymatic and computational insights into a unified model for membrane-immersed rhomboid proteolysis. *Biochem J* 425:501-512.

26. Hill, R.B., and L. Pellegrini. 2010. The PARL family of mitochondrial rhomboid proteases. *Semin Cell Dev Biol* 21:582-592.

27. Freeman, M. 2009. Rhomboids: 7 years of a new protease family. *Semin Cell Dev Biol* 20:231-239.

28. Rawson, R.B. 2008. Intriguing parasites and intramembrane proteases. *Genes Dev* 22:1561-1566.

29. Freeman, M. 2008. Rhomboid proteases and their biological functions. *Annu Rev Genet* 42:191-210.

30. Adrain, C., K. Strisovsky, M. Zettl, L. Hu, M.K. Lemberg, and M. Freeman. 2011. Mammalian EGF receptor activation by the rhomboid protease RHBDL2. *EMBO Rep* 12:421-427.

31. Stevenson, L.G., K. Strisovsky, K.M. Clemmer, S. Bhatt, M. Freeman, and P.N. Rather. 2007. Rhomboid protease AarA mediates quorum-sensing in *Providencia stuartii* by activating TatA of the twin-arginine translocase. *Proc Natl Acad Sci USA* 104:1003-1008.

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

33. Brossier, F., T.J. Jewett, L.D. Sibley, and S. Urban. 2005. A spatially localized rhomboid protease cleaves cell surface adhesins essential for invasion by Toxoplasma. *Proc Natl Acad Sci USA* 102:4146-4151.

34. O'Donnell, R.A., F. Hackett, S.A. Howell, M. Treeck, N. Struck, Z. Krnajski, C. Withers-Martinez, T.W. Gilberger, and M.J. Blackman. 2006. Intramembrane proteolysis mediates shedding of a key adhesin during erythrocyte invasion by the malaria parasite. *J Cell Biol* 174:1023-1033.

35. Baxt, L.A., R.P. Baker, U. Singh, and S. Urban. 2008. An Entamoeba histolytica rhomboid protease with atypical specificity cleaves a surface lectin involved in phagocytosis and immune evasion. *Genes Dev* 22:1636-1646.

36. Herlan, M., F. Vogel, C. Bornhovd, W. Neupert, and A.S. Reichert. 2003. Processing of Mgm1 by the rhomboid-type protease Pcp1 is required for maintenance of mitochondrial morphology and of mitochondrial DNA. *J Biol Chem* 278:27781-27788.

37. Urban, S., J.R. Lee, and M. Freeman. 2002. A family of Rhomboid intramembrane proteases activates all Drosophila membrane-tethered EGF ligands. *Embo J* 21:4277-4286.

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

39. Freeman, M., B.E. Kimmel, and G.M. Rubin. 1992. Identifying targets of the rough homeobox gene of Drosophila: evidence that rhomboid functions in eye development. *Development* 116:335-346.

40. Gallio, M., C. Englund, P. Kylsten, and C. Samakovlis. 2004. Rhomboid 3 orchestrates Slit-independent repulsion of tracheal branches at the CNS midline. *Development* 131:3605-3614.

41. Urban, S., G. Brown, and M. Freeman. 2004. EGF receptor signalling protects smooth-cuticle cells from apoptosis during Drosophila ventral epidermis development. *Development* 131:1835-1845.

42. Guichard, A., M. Roark, M. Ronshaugen, and E. Bier. 2000. brother of rhomboid, a rhomboid-related gene expressed during early Drosophila oogenesis, promotes EGF-R/MAPK signaling. *Dev Biol* 226:255-266.

43. Schulz, C., C.G. Wood, D.L. Jones, S.I. Tazuke, and M.T. Fuller. 2002. Signaling from germ cells mediated by the rhomboid homolog stet organizes encapsulation by somatic support cells. *Development* 129:4523-4534.

44. Blobel, C.P. 2005. ADAMs: key components in EGFR signalling and development. *Nat Rev Mol Cell Biol* 6:32-43.

45. Peschon, J.J., J.L. Slack, P. Reddy, K.L. Stocking, S.W. Sunnarborg, D.C. Lee, W.E. Russell, B.J. Castner, R.S. Johnson, J.N. Fitzner, R.W. Boyce, N. Nelson, C.J. Kozlosky, M.F. Wolfson, C.T. Rauch, D.P. Cerretti, R.J. Paxton, C.J. March, and R.A. Black. 1998. An essential role for ectodomain shedding in mammalian development. *Science* 282:1281-1284.

46. Abba, M.C., E. Lacunza, M.I. Nunez, A. Colussi, M. Isla-Larrain, A. Segal-Eiras, M.V. Croce, and C.M. Aldaz. 2009. Rhomboid domain containing 2 (RHBDD2): a novel cancer-related gene over-expressed in breast cancer. *Biochim Biophys Acta* 1792:988-997.

47. Rather, P.N., X. Ding, R.R. Baca-DeLancey, and S. Siddiqui. 1999. Providencia stuartii genes activated by cell-to-cell signaling and identification of a gene required for production or activity of an extracellular factor. *J Bacteriol* 181:7185-7191.

48. Baba, T., T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K.A. Datsenko, M. Tomita, B.L. Wanner, and H. Mori. 2006. Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol Syst Biol* 2:2006 0008.

49. Maegawa, S., K. Ito, and Y. Akiyama. 2005. Proteolytic action of GlpG, a rhomboid protease in the Escherichia coli cytoplasmic membrane. *Biochemistry* 44:13543-13552.

50. Zeng, G., S. Ye, and T.J. Larson. 1996. Repressor for the sn-glycerol 3-phosphate regulon of Escherichia coli K-12: primary structure and identification of the DNA-binding domain. *J Bacteriol* 178:7080-7089.

51. Yang, B., and T.J. Larson. 1998. Multiple promoters are responsible for transcription of the glpEGR operon of Escherichia coli K-12. *Biochim Biophys Acta* 1396:114-126.

52. Clemmer, K.M., G.M. Sturgill, A. Veenstra, and P.N. Rather. 2006. Functional characterization of Escherichia coli GlpG and additional rhomboid proteins using an aarA mutant of Providencia stuartii. *J Bacteriol* 188:3415-3419.

53. Gallio, M., G. Sturgill, P. Rather, and P. Kylsten. 2002. A conserved mechanism for extracellular signaling in eukaryotes and prokaryotes. *Proc Natl Acad Sci USA* 99:12208-12213.

54. Conseil, V., M. Soete, and J.F. Dubremetz. 1999. Serine protease inhibitors block invasion of host cells by Toxoplasma gondii. *Antimicrob Agents Chemother* 43:1358-1361.

55. Pellegrini, L., B.J. Passer, M. Canelles, I. Lefterov, J.K. Ganjei, B.J. Fowlkes, E.V. Koonin, and L. D'Adamio. 2001. PAMP and PARL, two novel putative metalloproteases interacting with the COOH-terminus of Presenilin-1 and -2. *J Alzheimers Dis* 3:181-190.

56. Sesaki, H., S.M. Southard, A.E. Hobbs, and R.E. Jensen. 2003. Cells lacking Pcp1p/Ugo2p, a rhomboid-like protease required for Mgm1p processing, lose mtDNA and mitochondrial structure in a Dnm1p-dependent manner, but remain competent for mitochondrial fusion. *Biochem Biophys Res Commun* 308:276-283.

57. McQuibban, G.A., J.R. Lee, L. Zheng, M. Juusola, and M. Freeman. 2006. Normal mitochondrial dynamics requires rhomboid-7 and affects Drosophila lifespan and neuronal function. *Curr Biol* 16:982-989.

58. Cipolat, S., T. Rudka, D. Hartmann, V. Costa, L. Serneels, K. Craessaerts, K. Metzger, C. Frezza, W. Annaert, L. D'Adamio, C. Derks, T. Dejaegere, L. Pellegrini, R. D'Hooge, L. Scorrano, and B. De Strooper. 2006. Mitochondrial rhomboid PARL regulates cytochrome c release during apoptosis via OPA1-dependent cristae remodeling. *Cell* 126:163-175.

59. Urban, S., and M. Freeman. 2003. Substrate specificity of rhomboid intramembrane proteases is governed by helix-breaking residues in the substrate transmembrane domain. *Mol Cell* 11:1425-1434.

60. Akiyama, Y., and S. Maegawa. 2007. Sequence features of substrates required for cleavage by GlpG, an *Escherichia coli* rhomboid protease. *Mol Microbiol* 64:1028-1037.

61. Strisovsky, K., H.J. Sharpe, and M. Freeman. 2009. Sequence-specific intramembrane proteolysis: identification of a recognition motif in rhomboid substrates. *Mol Cell* 36:1048-1059.

62. Lemberg, M.K., J. Menendez, A. Misik, M. Garcia, C.M. Koth, and M. Freeman. 2005. Mechanism of intramembrane proteolysis investigated with purified rhomboid proteases. *Embo J* 24:464-472.

63. Urban, S., and M.S. Wolfe. 2005. Reconstitution of intramembrane proteolysis in vitro reveals that pure rhomboid is sufficient for catalysis and specificity. *Proc Natl Acad Sci USA* 102:1883-1888.

64. Ben-Shem, A., D. Fass, and E. Bibi. 2007. Structural basis for intramembrane proteolysis by rhomboid serine proteases. *Proc Natl Acad Sci USA* 104:462-466.

65. Wang, Y., Y. Zhang, and Y. Ha. 2006. Crystal structure of a rhomboid family intramembrane protease. *Nature* 444:179-180.

66. Wu, Z., N. Yan, L. Feng, A. Oberstein, H. Yan, R.P. Baker, L. Gu, P.D. Jeffrey, S. Urban, and Y. Shi. 2006. Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry. *Nat Struct Mol Biol* 13:1084-1091.

67. Lemieux, M.J., S.J. Fischer, M.M. Cherney, K.S. Bateman, and M.N. James. 2007. The crystal structure of the rhomboid peptidase from *Haemophilus influenzae* provides insight into intramembrane proteolysis. *Proc Natl Acad Sci USA* 104:750-754.

68. Brooks, C.L., C. Lazareno-Saez, J.S. Lamoureux, M.W. Mak, and M.J. Lemieux. 2011. Insights into substrate gating in H. influenzae rhomboid. *J Mol Biol* 407:687-697.

69. Russ, W.P., and D.M. Engelman. 2000. The GxxxG motif: a framework for transmembrane helix-helix association. *J Mol Biol* 296:911-919.

70. Wang, Y., and Y. Ha. 2007. Open-cap conformation of intramembrane protease GlpG. *Proc Natl Acad Sci USA* 104:2098-2102.

71. Wang, Y., S. Maegawa, Y. Akiyama, and Y. Ha. 2007. The role of L1 loop in the mechanism of rhomboid intramembrane protease GlpG. *J Mol Biol* 374:1104-1113.

72. Vinothkumar, K.R., K. Strisovsky, A. Andreeva, Y. Christova, S. Verhelst, and M. Freeman. 2010. The structural basis for catalysis and substrate specificity of a rhomboid protease. *EMBO J* 29:3797-3809.

73. Vinothkumar, K.R. 2011. Structure of rhomboid protease in a lipid environment. *J Mol Biol* 407:232-247.

74. Baker, R.P., K. Young, L. Feng, Y. Shi, and S. Urban. 2007. Enzymatic analysis of a rhomboid intramembrane protease implicates transmembrane helix 5 as the lateral substrate gate. *Proc Natl Acad Sci USA* 104:8257-8262.

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

76. Bondar, A.N., C. del Val, and S.H. White. 2009. Rhomboid protease dynamics and lipid interactions. *Structure* 17:395-405.

77. Lohi, O., S. Urban, and M. Freeman. 2004. Diverse substrate recognition mechanisms for rhomboids; thrombomodulin is cleaved by Mammalian rhomboids. *Curr Biol* 14:236-241.

78. Pascall, J.C., and K.D. Brown. 2004. Intramembrane cleavage of ephrinB3 by the human rhomboid family protease, RHBDL2. *Biochem Biophys Res Commun* 317:244-252.

79. Del Rio, A., K. Dutta, J. Chavez, I. Ubarretxena-Belandia, and R. Ghose. 2007. Solution structure and dynamics of the N-terminal cytosolic domain of rhomboid intramembrane protease from Pseudomonas aeruginosa: insights into a functional role in intramembrane proteolysis. *J Mol Biol* 365:109-122.

80. Mota, C., J. Ridenoure, J. Cheng, and F.L. de Los Reyes, 3rd. 2005. High levels of nitrifying bacteria in intermittently aerated reactors treating high ammonia wastewater. *FEMS Microbiol Ecol* 54:391-400.

81. Yip, C.K., T.G. Kimbrough, H.B. Felise, M. Vuckovic, N.A. Thomas, R.A. Pfuetzner, E.A. Frey, B.B. Finlay, S.I. Miller, and N.C. Strynadka. 2005. Structural characterization of the molecular platform for type III secretion system assembly. *Nature* 435:702-707.

82. Jeyaraju, D.V., L. Xu, M.C. Letellier, S. Bandaru, R. Zunino, E.A. Berg, H.M. McBride, and L. Pellegrini. 2006. Phosphorylation and cleavage of presenilin-associated rhomboid-like protein (PARL) promotes changes in mitochondrial morphology. *Proc Natl Acad Sci USA* 103:18562-18567.

83. Sik, A., B.J. Passer, E.V. Koonin, and L. Pellegrini. 2004. Self-regulated cleavage of the mitochondrial intramembrane-cleaving protease PARL yields Pbeta, a nuclear-targeted peptide. *J Biol Chem* 279:15323-15329.

84. Bier, E., L.Y. Jan, and Y.N. Jan. 1990. Rhomboid, a gene required for dorsoventral axis establishment and peripheral nervous system development in *Drosophila melanogaster*. *Genes Dev* 4:190-203.

85. Klambt, C. 2002. EGF receptor signalling: roles of star and rhomboid revealed. *Curr Biol* 12:R21-23.

86. Snyder, D.A., Y. Chen, N.G. Denissova, T. Acton, J.M. Aramini, M. Ciano, R. Karlin, J. Liu, P. Manor, P.A. Rajan, P. Rossi, G.V. Swapna, R. Xiao, B. Rost, J. Hunt, and G.T. Montelione. 2005. Comparisons of NMR spectral quality and success

in crystallization demonstrate that NMR and X-ray crystallography are complementary methods for small protein structure determination. *J Am Chem Soc* 127:16505-16511.

87. Yee, A.A., A. Savchenko, A. Ignachenko, J. Lukin, X. Xu, T. Skarina, E. Evdokimova, C.S. Liu, A. Semesi, V. Guido, A.M. Edwards, and C.H. Arrowsmith. 2005. NMR and X-ray crystallography, complementary tools in structural proteomics of small proteins. *J Am Chem Soc* 127:16512-16517.

88. Kalodimos, C.G. 2011. NMR reveals novel mechanisms of protein activity regulation. *Protein Sci* 20:773-782.

89. Abragam, A. 1961. Principles of Nuclear Magnetism. Oxford University Press, Oxford.

90. Cavanagh, J.F., W.J. Palmer, A.G. Skelton, N.J. 1996. Protein NMR spectroscopy: principles and practice. Academic Press, San Diego.

91. Claridge, T.D.W. 1999. High-resolution NMR techniques in organic chemistry. Pergamon, New York.

92. Gardner, K.H., and L.E. Kay. 1998. The use of 2H, 13C, 15N multidimensional NMR to study the structure and dynamics of proteins. *Annu Rev Biophys Biomol Struct* 27:357-406.

93. Aue, W.P., E. Bartholdi, and R.R. Ernst. 1976. Two-dimensional spectroscopy. Application to nuclear magnetic resonance. *J. Chem. Phys.* 64:2229.

94. Jeener, J. 1994. NMR and More, in Honour of Anatole Abragam (M. Goldman, M Portneuf, eds.). *Les Ulis, France: Les Editions de Physique* 1-388.

95. Bodenhausen, G., and D.J. Ruben. 1980. Natural abundance nitrogen-15 NMR by enhanced heteronuclear spectroscopy. *Chem Phys Lett* 69:185-189.

96. Zuiderweg, E.R. 2002. Mapping protein-protein interactions in solution by NMR spectroscopy. *Biochemistry* 41:1-7.

97. Rajagopal, P., E.B. Waygood, J. Reizer, M.H. Saier, Jr., and R.E. Klevit. 1997. Demonstration of protein-protein interaction specificity by NMR chemical shift mapping. *Protein Sci* 6:2624-2627.

98. Kay, L.E., M. Ikura, R. Tschudin, and A. Bax. 1990. Three-dimensional triple-resonance NMR spectroscopy of isotopically enriched proteins. *J Magn Reson* 89:496-514.

99. Cornilescu, G., F. Delaglio, and A. Bax. 1999. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* 13:289-302.

100. Grzesiek, S., and A. Bax. 1992. Correlating backbone amide and side chain resonances in larger proteins by multiple relayed triple resonance NMR. *J Am Chem Soc* 114:6291-6293.

101. Grzesiek, S., and A. Bax. 1992. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. *J Magn Reson* 99:201-207.

102. Seavey, B.R., E.A. Farr, W.M. Westler, and J.L. Markley. 1991. A relational database for sequence-specific protein NMR data. *J Biomol NMR* 1:217-236.

103. Johnson, B.A. 2004. Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol Biol* 278:313-352.

104. Goddard, T.D.a.K., D.G. 2007. SPARKY 3. In University of California, San Francisco.

105. Delaglio, F., S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, and A. Bax. 1995. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6:277-293.

106. Bartels, C.X., T.H. Billeter,M. Guntert,P. and Wuthrich,K. 1995. The program XEASY for computer-supported NMR spectral analysis of biological macromolecules. *J Biomol NMR* 5:1-10.

107. Bax, A., G.M. Clore, and A.M. Gronenborn. 1990. 1H-1H correlation via isotropic mixing of 13C magnetization, a new three-dimensional approach for assigning 1H and 13C spectra of 13C-enriched proteins. *J Magn Reson* 88:425-431.

108. Bax, A., G.M. Clore, P.C. Driscoll, A.M. Gronenborn, M. Ikura, and L.E. Kay. 1990. Practical aspects of proton-carbon-carbon-proton three-dimensional correlation spectroscopy of 13C-labeled proteins. *J Magn Reson* 87:620-627.

109. Overhauser, A.W. 1953. Polarization of nuclei in metals. *Phys Rev* 92:411-415.

110. Marion, D., P.C. Driscoll, L.E. Kay, P.T. Wingfield, A. Bax, A.M. Gronenborn, and G.M. Clore. 1989. Overcoming the overlap problem in the assignment of 1H NMR spectra of larger proteins by use of three-dimensional heteronuclear 1H-15N Hartmann-Hahn-multiple quantum coherence and nuclear Overhauser-multiple quantum coherence spectroscopy: application to interleukin 1 beta. *Biochemistry* 28:6150-6156.

111. Marion, D., L.E. Kay, S.W. Sparks, D.A. Torchia, and A. Bax. 1989. Three-dimensional heteronuclear NMR of nitrogen-15 labeled proteins. *J Am Chem Soc* 111:1515-1517.

112. Guntert, P. 2004. Automated NMR structure calculation with CYANA. *Methods Mol Biol* 278:353-378.

113. Linge, J.P., M.A. Williams, C.A. Spronk, A.M. Bonvin, and M. Nilges. 2003. Refinement of protein structures in explicit solvent. *Proteins* 50:496-506.

114. Montelione, G.T., D. Zheng, Y.J. Huang, K.C. Gunsalus, and T. Szyperski. 2000. Protein NMR spectroscopy in structural genomics. *Nat Struct Biol* 7 Suppl:982-985.

115. Laskowski, R.A., M.W. MacArthur, D.S. Moss, and J.M. Thornton. 1993. PROCHECK - a program to check the steriochemical quality of protein structures. *J App Crys* 26:283-291.

116. Laskowski, R.A., J.A. Rullmann, M.W. MacArthur, R. Kaptein, and J.M. Thornton. 1996. AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J Biomol NMR* 8:477-486.

117. Vriend, G. 1990. WHAT IF: a molecular modeling and drug design program. *J Mol Graph* 8:52-56, 29.

118. Bax, A., G. Kontaxis, and N. Tjandra. 2001. Dipolar couplings in macromolecular structure determination. *Methods Enzymol* 339:127-174.

119. Tjandra, N., J.G. Omichinski, A.M. Gronenborn, G.M. Clore, and A. Bax. 1997. Use of dipolar 1H-15N and 1H-13C couplings in the structure determination of magnetically oriented macromolecules in solution. *Nat Struct Biol* 4:732-738.

120. Battiste, J.L., and G. Wagner. 2000. Utilization of site-directed spin labeling and high-resolution heteronuclear nuclear magnetic resonance for global fold determination of large proteins with limited nuclear overhauser effect data. *Biochemistry* 39:5355-5365.

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

122. Garavito, R.M., and S. Ferguson-Miller. 2001. Detergents as tools in membrane biochemistry. *J Biol Chem* 276:32403-32406.

123. Krueger-Koplin, R.D., P.L. Sorgen, S.T. Krueger-Koplin, I.O. Rivera-Torres, S.M. Cahill, D.B. Hicks, L. Grinius, T.A. Krulwich, and M.E. Girvin. 2004. An evaluation of detergents for NMR structural studies of membrane proteins. *J Biomol NMR* 28:43-57.

124. Poget, S.F., and M.E. Girvin. 2007. Solution NMR of membrane proteins in bilayer mimics: small is beautiful, but sometimes bigger is better. *Biochim Biophys Acta* 1768:3098-3106.

125. Seddon, A.M., P. Curnow, and P.J. Booth. 2004. Membrane proteins, lipids and detergents: not just a soap opera. *Biochim Biophys Acta* 1666:105-117.

126. Tanford, C. 1973. The Hydrophobic Effect: Formation of Micelles and Biological Membranes. John Wiley & Sons Inc, New York.

127. Sanders, C.R., and F. Sonnichsen. 2006. Solution NMR of membrane proteins: practice and challenges. *Magn Reson Chem* 44 Spec No:S24-40.

128. le Maire, M., P. Champeil, and J.V. Moller. 2000. Interaction of membrane proteins and lipids with solubilizing detergents. *Biochim Biophys Acta* 1508:86-111.

129. Chou, J.J., J.D. Kaufman, S.J. Stahl, P.T. Wingfield, and A. Bax. 2002. Micelle-induced curvature in a water-insoluble HIV-1 Env peptide revealed by NMR dipolar coupling measurement in stretched polyacrylamide gel. *J Am Chem Soc* 124:2450-2451.

130. Vold, R.R., R.S. Prosser, and A.J. Deese. 1997. Isotropic solutions of phospholipid bicelles: a new membrane mimetic for high-resolution NMR studies of polypeptides. *J Biomol NMR* 9:329-335.

131. Bocharov, E.V., Y.E. Pustovalova, K.V. Pavlov, P.E. Volynsky, M.V. Goncharuk, Y.S. Ermolyuk, D.V. Karpunin, A.A. Schulga, M.P. Kirpichnikov, R.G. Efremov, I.V. Maslennikov, and A.S. Arseniev. 2007. Unique dimeric structure of BNip3 transmembrane domain suggests membrane permeabilization as a cell death trigger. *J Biol Chem* 282:16256-16266.

132. Sanders, C.R., 2nd, and J.P. Schwonek. 1992. Characterization of magnetically orientable bilayers in mixtures of dihexanoylphosphatidylcholine and dimyristoylphosphatidylcholine by solid-state NMR. *Biochemistry* 31:8898-8905.

133. Glover, K.J., J.A. Whiles, G. Wu, N. Yu, R. Deems, J.O. Struppe, R.E. Stark, E.A. Komives, and R.R. Vold. 2001. Structural evaluation of phospholipid bicelles for solution-state studies of membrane-associated biomolecules. *Biophys J* 81:2163-2171.

134. Marcotte, I., and M. Auger. 2005. Bicelles as model membranes for solid and solution-State NMR studies of membrane peptides and proteins. *Concept Magn Reson A* 24A:17-37.

135. Luchette, P.A., T.N. Vetman, R.S. Prosser, R.E. Hancock, M.P. Nieh, C.J. Glinka, S. Krueger, and J. Katsaras. 2001. Morphology of fast-tumbling bicelles: a small angle neutron scattering and NMR study. *Biochim Biophys Acta* 1513:83-94.

136. Vold, R.R., and R.S. Prosser. 1996. Magnetically oriented phospholipid bilayered micelles for structural studies of polypeptides: does the ideal bicelle exist? *J Magn Reson B* 113:267-271.

137. McKibbin, C., A.M. Toye, P.J. Reeves, H.G. Khorana, P.C. Edwards, C. Villa, and P.J. Booth. 2007. Opsin stability and folding: the role of Cys185 and abnormal disulfide bond formation in the intradiscal domain. *J Mol Biol* 374:1309-1318.

138. Poget, S.F., R. Harris, S.M. Cahill, and M.E. Girvin. 2010. 1H, 13C, 15N backbone NMR assignments of the Staphylococcus aureus small multidrug-resistance pump (Smr) in a functionally active conformation. *Biomol NMR Assign* 4:139-142.

139. Kovacs, H., D. Moskau, and M. Spraul. 2005. Crygenically cooled probes -- a leap in NMR technology. *Prog Nucl Mag Res Sp* 46:131-155.

140. Goto, N.K., K.H. Gardner, G.A. Mueller, R.C. Willis, and L.E. Kay. 1999. A robust and cost-effective method for the production of Val, Leu, Ile (delta 1) methyl-protonated 15N-, 13C-, 2H-labeled proteins. *J Biomol NMR* 13:369-374.

141. Goto, N.K., and L.E. Kay. 2000. New developments in isotope labeling strategies for protein solution NMR spectroscopy. *Curr Opin Struct Biol* 10:585-592.

142. Pervushin, K., R. Riek, G. Wider, and K. Wuthrich. 1997. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci USA* 94:12366-12371.

143. Tzakos, A.G., C.R. Grace, P.J. Lukavsky, and R. Riek. 2006. NMR techniques for very large proteins and rnas in solution. *Annu Rev Biophys Biomol Struct* 35:319-342.

144. Chen, J., D. Nietlispach, A.J. Shaka, and V.A. Mandelshtam. 2004. Ultra-high resolution 3D NMR spectra from limited-size data sets. *J Magn Reson* 169:215-224.

145. Patching, S.G. 2011. NMR structures of polytopic integral membrane proteins. *Mol Membr Biol* 28:370-397.

146. Zhou, Y., T. Cierpicki, R.H. Jimenez, S.M. Lukasik, J.F. Ellena, D.S. Cafiso, H. Kadokura, J. Beckwith, and J.H. Bushweller. 2008. NMR solution structure of the integral membrane enzyme DsbB: functional insights into DsbB-catalyzed disulfide bond formation. *Mol Cell* 31:896-908.

147. Maslennikov, I., C. Klammt, E. Hwang, G. Kefala, M. Okamura, L. Esquivies, K. Mors, C. Glaubitz, W. Kwiatkowski, Y.H. Jeon, and S. Choe. 2010. Membrane domain structures of three classes of histidine kinase receptors by cell-free expression and rapid NMR analysis. *Proc Natl Acad Sci USA* 107:10902-10907.

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

149. Larkin, M.A., G. Blackshields, N.P. Brown, R. Chenna, P.A. McGettigan, H. McWilliam, F. Valentin, I.M. Wallace, A. Wilm, R. Lopez, J.D. Thompson, T.J. Gibson, and D.G. Higgins. 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* 23:2947-2948.

150. Cole, C., J.D. Barber, and G.J. Barton. 2008. The Jpred 3 secondary structure prediction server. *Nucleic Acids Res* 36:W197-201.

151. Zer, H., M. Vink, S. Shochat, R.G. Herrmann, B. Andersson, and I. Ohad. 2003. Light affects the accessibility of the thylakoid light harvesting complex II (LHCII) phosphorylation site to the membrane protein kinase(s). *Biochemistry* 42:728-738.

152. Chou, J.J., J.L. Baber, and A. Bax. 2004. Characterization of phospholipid mixed micelles by translational diffusion. *J Biomol NMR* 29:299-308.

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

154. Stafford, R.E., T. Fanni, and E.A. Dennis. 1989. Interfacial properties and critical micelle concentration of lysophospholipids. *Biochemistry* 28:5113-5120.

155. McKibbin, C., N.A. Farmer, C. Jeans, P.J. Reeves, H.G. Khorana, B.A. Wallace, P.C. Edwards, C. Villa, and P.J. Booth. 2007. Opsin stability and folding: modulation by phospholipid bicelles. *J Mol Biol* 374:1319-1332.

156. Johnson, B.A. 1994. NMRView: A computer program for the visualization and analysis of NMR data. *J Biomol NMR* 4:603-614.

157. Sattler, M.S., J.; Griesinger, C. 1999. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog Nucl Mag Res Sp* 34:93-158.

158. Da Costa, G., L. Mouret, S. Chevance, E. Le Rumeur, and A. Bondon. 2007. NMR of molecules interacting with lipids in small unilamellar vesicles. *Eur Biophys J* 36:933-942.

159. Bocharov, E.V., K.S. Mineev, P.E. Volynsky, Y.S. Ermolyuk, E.N. Tkach, A.G. Sobol, V.V. Chupin, M.P. Kirpichnikov, R.G. Efremov, and A.S. Arseniev. 2008. Spatial structure of the dimeric transmembrane domain of the growth factor receptor ErbB2 presumably corresponding to the receptor active state. *J Biol Chem* 283:6950-6956.

160. Howell, S.C., M.F. Mesleh, and S.J. Opella. 2005. NMR structure determination of a membrane protein with two transmembrane helices in micelles: MerF of the bacterial mercury detoxification system. *Biochemistry* 44:5196-5206.

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

162. Czerski, L., and C.R. Sanders. 2000. Functionality of a membrane protein in bicelles. *Anal Biochem* 284:327-333.

163. Andersson, A., and L. Maler. 2005. Magnetic resonance investigations of lipid motion in isotropic bicelles. *Langmuir* 21:7702-7709.

164. Mazer, N.A., G.B. Benedek, and M.C. Carey. 1980. Quasielastic light-scattering studies of aqueous biliary lipid systems. Mixed micelle formation in bile salt-lecithin solutions. *Biochemistry* 19:601-615.

165. Sheetz, M.P., and S.I. Chan. 1972. Effect of sonication on the structure of lecithin bilayers. *Biochemistry* 11:4573-4581.

166. Lei, X., K. Ahn, L. Zhu, I. Ubarretxena-Belandia, and Y.M. Li. 2008. Soluble oligomers of the intramembrane serine protease YqgP are catalytically active in the absence of detergents. *Biochemistry* 47:11920-11929.

167. Fenyo, D., Q. Wang, J.A. DeGrasse, J.C. Padovan, M. Cadene, and B.T. Chait. 2007. MALDI sample preparation: the ultra thin layer method. *J Vis Exp* 192.

168. Crooks, G.E., G. Hon, J.M. Chandonia, and S.E. Brenner. 2004. WebLogo: a sequence logo generator. *Genome Res* 14:1188-1190.

169. Evans, M.J., and B.F. Cravatt. 2006. Mechanism-based profiling of enzyme families. *Chem Rev* 106:3279-3301.

170. Kidd, D., Y. Liu, and B.F. Cravatt. 2001. Profiling serine hydrolase activities in complex proteomes. *Biochemistry* 40:4005-4015.

171. Liu, Y., M.P. Patricelli, and B.F. Cravatt. 1999. Activity-based protein profiling: the serine hydrolases. *Proc Natl Acad Sci USA* 96:14694-14699.

172. Blais, D.R., R.K. Lyn, M.A. Joyce, Y. Rouleau, R. Steenbergen, N. Barsby, L.F. Zhu, A.F. Pegoraro, A. Stolow, D.L. Tyrrell, and J.P. Pezacki. 2010. Activity-based protein profiling identifies a host enzyme, carboxylesterase 1, which is differentially active during hepatitis C virus replication. *J Biol Chem* 285:25602-25612.

173. Banerjee, P., J.B. Joo, J.T. Buse, and G. Dawson. 1995. Differential solubilization of lipids along with membrane proteins by different classes of detergents. *Chem Phys Lipids* 77:65-78.

174. Drin, G., and B. Antonny. 2010. Amphipathic helices and membrane curvature. *FEBS Lett* 584:1840-1847.

175. Zhang, W., E. Crocker, S. McLaughlin, and S.O. Smith. 2003. Binding of peptides with basic and aromatic residues to bilayer membranes: phenylalanine in the myristoylated alanine-rich C kinase substrate effector domain penetrates into the hydrophobic core of the bilayer. *J Biol Chem* 278:21459-21466.

176. Kumar, S., and R. Nussinov. 2002. Close-range electrostatic interactions in proteins. *Chembiochem* 3:604-617.

177. Sokolovski, M., T. Sheynis, S. Kolusheva, and R. Jelinek. 2008. Membrane interactions and lipid binding of casein oligomers and early aggregates. *Biochim Biophys Acta* 1778:2341-2349.

178. Bennett, M.J., M.P. Schlunegger, and D. Eisenberg. 1995. 3D domain swapping: a mechanism for oligomer assembly. *Protein Sci* 4:2455-2468.

179. Schwieters, C.D., J.J. Kuszewski, and G.M. Clore. 2006. Using Xplor-NIH for NMR molecular structure determination. *Prog Nucl Mag Res Sp* 48:47-62.

180. Schwieters, C.D., J.J. Kuszewski, N. Tjandra, and G.M. Clore. 2003. The Xplor-NIH NMR molecular structure determination package. *J Magn Reson* 160:65-73.

181. Zwahlen, C., P. Legault, S.J.F. Vincent, J. Greenblatt, R. Konrat, and L.E. Kay. 1997. Methods for Measurement of Intermolecular NOEs by Multinuclear NMR Spectroscopy: Application to a Bacteriophage λ N-Peptide/boxB RNA Complex. *J Am Chem Soc* 119:6711-6721.

182. Wishart, D.S., B.D. Sykes, and F.M. Richards. 1992. The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* 31:1647-1651.

183. Navarro, M.V., P.D. Newell, P.V. Krasteva, D. Chatterjee, D.R. Madden, G.A. O'Toole, and H. Sondermann. 2011. Structural basis for c-di-GMP-mediated inside-out signaling controlling periplasmic proteolysis. *PLoS Biol* 9:e1000588.

184. Jerala, R., and E. Zerovnik. 1999. Accessing the global minimum conformation of stefin A dimer by annealing under partially denaturing conditions. *J Mol Biol* 291:1079-1089.

185. Denisov, A.Y., T. Sprules, J. Fraser, G. Kozlov, and K. Gehring. 2007. Heat-induced dimerization of BCL-xL through alpha-helix swapping. *Biochemistry* 46:734-740.

186. Kuhlman, B., J.W. O'Neill, D.E. Kim, K.Y. Zhang, and D. Baker. 2001. Conversion of monomeric protein L to an obligate dimer by computational protein design. *Proc Natl Acad Sci USA* 98:10687-10691.

187. Liu, Y., G. Gotte, M. Libonati, and D. Eisenberg. 2001. A domain-swapped RNase A dimer with implications for amyloid formation. *Nat Struct Biol* 8:211-214.

188. Rousseau, F., J.W. Schymkowitz, H.R. Wilkinson, and L.S. Itzhaki. 2001. Three-dimensional domain swapping in p13suc1 occurs in the unfolded state and is controlled by conserved proline residues. *Proc Natl Acad Sci USA* 98:5596-5601.

189. Schymkowitz, J.W., F. Rousseau, H.R. Wilkinson, A. Friedler, and L.S. Itzhaki. 2001. Observation of signal transduction in three-dimensional domain swapping. *Nat Struct Biol* 8:888-892.

190. Magler, I., D. Nuss, M. Hauser, F. Ferreira, and H. Brandstetter. 2010. Molecular metamorphosis in polcalcin allergens by EF-hand rearrangements and domain swapping. *Febs J* 277:2598-2610.

191. Losonczi, J.A., E.T. Olejniczak, S.F. Betz, J.E. Harlan, J. Mack, and S.W. Fesik. 2000. NMR studies of the anti-apoptotic protein Bcl-xL in micelles. *Biochemistry* 39:11024-11033.

192. Feng, Y., Z. Lin, X. Shen, K. Chen, H. Jiang, and D. Liu. 2008. Bcl-xL forms two distinct homodimers at non-ionic detergents: implications in the dimerization of Bcl-2 family proteins. *J Biochem* 143:243-252.

193. O'Neill, J.W., M.K. Manion, B. Maguire, and D.M. Hockenbery. 2006. BCL-XL dimerization by three-dimensional domain swapping. *J Mol Biol* 356:367-381.

194. Feng, Y., D. Liu, X. Shen, K. Chen, and H. Jiang. 2009. Structure assembly of Bcl-x(L) through alpha5-alpha5 and alpha6-alpha6 interhelix interactions in lipid membranes. *Biochim Biophys Acta* 1788:2389-2395.

195. Gotte, G., M. Libonati, and D.V. Laurents. 2003. Glycosylation and specific deamidation of ribonuclease B affect the formation of three-dimensional domain-swapped oligomers. *J Biol Chem* 278:46241-46251.

196. Sanders, A., C. Jeremy Craven, L.D. Higgins, S. Giannini, M.J. Conroy, A.M. Hounslow, J.P. Waltho, and R.A. Staniforth. 2004. Cystatin forms a tetramer through structural rearrangement of domain-swapped dimers prior to amyloidogenesis. *J Mol Biol* 336:165-178.

197. Schymkowitz, J.W., F. Rousseau, L.R. Irvine, and L.S. Itzhaki. 2000. The folding pathway of the cell-cycle regulatory protein p13suc1: clues for the mechanism of domain swapping. *Structure* 8:89-100.

198. Alonso, D.O., E. Alm, and V. Daggett. 2000. Characterization of the unfolding pathway of the cell-cycle protein p13suc1 by molecular dynamics simulations: implications for domain swapping. *Structure* 8:101-110.

199. Antonny, B. 2011. Mechanisms of membrane curvature sensing. *Annu Rev Biochem* 80:101-123.

200. Mileykovskaya, E., and W. Dowhan. 2000. Visualization of phospholipid domains in Escherichia coli by using the cardiolipin-specific fluorescent dye 10-N-nonyl acridine orange. *J Bacteriol* 182:1172-1175.

201. Nishibori, A., J. Kusaka, H. Hara, M. Umeda, and K. Matsumoto. 2005. Phosphatidylethanolamine domains and localization of phospholipid synthases in *Bacillus subtilis* membranes. *J Bacteriol* 187:2163-2174.

202. Renner, L.D., and D.B. Weibel. 2011. Cardiolipin microdomains localize to negatively curved regions of Escherichia coli membranes. *Proc Natl Acad Sci USA* 108:6264-6269.

203. Vanounou, S., A.H. Parola, and I. Fishov. 2003. Phosphatidylethanolamine and phosphatidylglycerol are segregated into different domains in bacterial membrane. A study with pyrene-labelled phospholipids. *Mol Microbiol* 49:1067-1079.

204. Zerovnik, E., V. Stoka, A. Mirtic, G. Guncar, J. Grdadolnik, R.A. Staniforth, D. Turk, and V. Turk. 2011. Mechanisms of amyloid fibril formation--focus on domain-swapping. *Febs J* 278:2263-2282.

205. Fernandez, C., and G. Wider. 2003. TROSY in NMR studies of the structure and function of large biological macromolecules. *Curr Opin Struct Biol* 13:570-580.

206. Wider, G. 2005. NMR techniques used with very large biological macromolecules in solution. *Methods Enzymol* 394:382-398.

207. Pervushin, K. 2000. Impact of transverse relaxation optimized spectroscopy (TROSY) on NMR as a technique in structural biology. *Q Rev Biophys* 33:161-197.

208. Riek, R., K. Pervushin, and K. Wuthrich. 2000. TROSY and CRINEPT: NMR with large molecular and supramolecular structures in solution. *Trends Biochem Sci* 25:462-468.

209. Salzmann, M., K. Pervushin, G. Wider, H. Senn, and K. Wuthrich. 1998. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc Natl Acad Sci USA* 95:13585-13590.

210. Columbus, L., J. Lipfert, K. Jambunathan, D.A. Fox, A.Y. Sim, S. Doniach, and S.A. Lesley. 2009. Mixing and matching detergents for membrane protein NMR structure determination. *J Am Chem Soc* 131:7320-7326.

211. Page, R.C., J.D. Moore, H.B. Nguyen, M. Sharma, R. Chase, F.P. Gao, C.K. Mobley, C.R. Sanders, L. Ma, F.D. Sonnichsen, S. Lee, S.C. Howell, S.J. Opella, and T.A. Cross. 2006. Comprehensive evaluation of solution nuclear magnetic resonance spectroscopy sample preparation for helical integral membrane proteins. *J Struct Funct Genomics* 7:51-64.

212. Tian, C., M.D. Karra, C.D. Ellis, J. Jacob, K. Oxenoid, F. Sonnichsen, and C.R. Sanders. 2005. Membrane protein preparation for TROSY NMR screening. *Methods Enzymol* 394:321-334.

213. Columbus, L., J. Lipfert, H. Klock, I. Millett, S. Doniach, and S.A. Lesley. 2006. Expression, purification, and characterization of *Thermotoga maritima* membrane proteins for structure determination. *Protein Sci* 15:961-975.

214. Ahn, V.E., E.I. Lo, C.K. Engel, L. Chen, P.M. Hwang, L.E. Kay, R.E. Bishop, and G.G. Prive. 2004. A hydrocarbon ruler measures palmitate in the enzymatic acylation of endotoxin. *Embo J* 23:2931-2941.

215. Stouffer, A.L., R. Acharya, D. Salom, A.S. Levine, L. Di Costanzo, C.S. Soto, V. Tereshko, V. Nanda, S. Stayrook, and W.F. DeGrado. 2008. Structural basis for the function and inhibition of an influenza virus proton channel. *Nature* 451:596-599.

216. Bayburt, T.H., and S.G. Sligar. 2010. Membrane protein assembly into Nanodiscs. *FEBS Lett* 584:1721-1727.

217. Denisov, I.G., M.A. McLean, A.W. Shaw, Y.V. Grinkova, and S.G. Sligar. 2005. Thermotropic phase transition in soluble nanoscale lipid bilayers. *J Phys Chem B* 109:15580-15588.

218. Sprangers, R., A. Velyvis, and L.E. Kay. 2007. Solution NMR of supramolecular complexes: providing new insights into function. *Nat Methods* 4:697-703.

219. Popot, J.L. Amphipols, nanodiscs, and fluorinated surfactants: three nonconventional approaches to studying membrane proteins in aqueous solutions. *Annu Rev Biochem* 79:737-775.

220. Sanders, C.R., A. Kuhn Hoffmann, D.N. Gray, M.H. Keyes, and C.D. Ellis. 2004. French swimwear for membrane proteins. *Chembiochem* 5:423-426.

221. Tribet, C., R. Audebert, and J.L. Popot. 1996. Amphipols: polymers that keep membrane proteins soluble in aqueous solutions. *Proc Natl Acad Sci USA* 93:15047-15050.

222. Zoonens, M., L.J. Catoire, F. Giusti, and J.L. Popot. 2005. NMR study of a membrane protein in detergent-free aqueous solution. *Proc Natl Acad Sci USA* 102:8893-8898.

223. Maegawa, S., K. Koide, K. Ito, and Y. Akiyama. 2007. The intramembrane active site of GlpG, an E. coli rhomboid protease, is accessible to water and hydrolyses an extramembrane peptide bond of substrates. *Mol Microbiol* 64:435-447.

224. Nikaido, H., and E.Y. Rosenberg. 1981. Effect on solute size on diffusion rates through the transmembrane pores of the outer membrane of Escherichia coli. *J Gen Physiol* 77:121-135.

225. Speers, A.E., and B.F. Cravatt. 2004. Profiling enzyme activities in vivo using click chemistry methods. *Chem Biol* 11:535-546.

226. Pierrat, O.A., K. Strisovsky, Y. Christova, J. Large, K. Ansell, N. Bouloc, E. Smiljanic, and M. Freeman. 2011. Monocyclic beta-lactams are selective, mechanism-based inhibitors of rhomboid intramembrane proteases. *ACS Chem Biol* 6:325-335.

227. Shameer, K., P.N. Shingate, S.C. Manjunath, M. Karthika, G. Pugalenthi, and R. Sowdhamini. 3DSwap: curated knowledgebase of proteins involved in 3D domain swapping. *Database (Oxford)* 2011:bar042.

228. Vitagliano, L., S. Adinolfi, F. Sica, A. Merlino, A. Zagari, and L. Mazzarella. 1999. A potential allosteric subsite generated by domain swapping in bovine seminal ribonuclease. *J Mol Biol* 293:569-577.

229. Mazzarella, L., S. Capasso, D. Demasi, G. Di Lorenzo, C.A. Mattia, and A. Zagari. 1993. Bovine seminal ribonuclease: structure at 1.9 A resolution. *Acta Crystallogr D Biol Crystallogr* 49:389-402.

230. Piccoli, R., A. Di Donato, and G. D'Alessio. 1988. Co-operativity in seminal ribonuclease function. Kinetic studies. *Biochem J* 253:329-336.

231. Cafaro, V., C. De Lorenzo, R. Piccoli, A. Bracale, M.R. Mastronicola, A. Di Donato, and G. D'Alessio. 1995. The antitumor action of seminal ribonuclease and its quaternary conformations. *FEBS Lett* 359:31-34.

232. Bennett, M.J., S. Choe, and D. Eisenberg. 1994. Domain swapping: entangling alliances between proteins. *Proc Natl Acad Sci USA* 91:3127-3131.

233. Carroll, S.F., J.T. Barbieri, and R.J. Collier. 1986. Dimeric form of diphtheria toxin: purification and characterization. *Biochemistry* 25:2425-2430.

234. Malevanets, A., F.L. Sirota, and S.J. Wodak. 2008. Mechanism and energy landscape of domain swapping in the B1 domain of protein G. *J Mol Biol* 382:223-235.

235. Kleckner, I.R., and M.P. Foster. 2011. An introduction to NMR-based approaches for measuring protein dynamics. *Biochim Biophys Acta* 1814:942-968.

236. Abergel, D., and A.G. Palmer. 2004. Approximate solutions of the Bloch-McConnell equations for two-site chemical exchange. *Chemphyschem* 5:787-793.

237. Hansen, D.F., and J.J. Led. 2003. Implications of using approximate Bloch-McConnell equations in NMR analyses of chemically exchanging systems: application to the electron self-exchange of plastocyanin. *J Magn Reson* 163:215-227.

238. Bouvignies, G., P. Vallurupalli, D.F. Hansen, B.E. Correia, O. Lange, A. Bah, R.M. Vernon, F.W. Dahlquist, D. Baker, and L.E. Kay. 2011. Solution structure of a minor and transiently formed state of a T4 lysozyme mutant. *Nature* 477:111-114.

239. Pranke, I.M., V. Morello, J. Bigay, K. Gibson, J.M. Verbavatz, B. Antonny, and C.L. Jackson. 2011. alpha-Synuclein and ALPS motifs are membrane curvature sensors whose contrasting chemistry mediates selective vesicle binding. *J Cell Biol* 194:89-103.

240. Jeyaraju, D.V., H.M. McBride, R.B. Hill, and L. Pellegrini. 2011. Structural and mechanistic basis of Parl activity and regulation. *Cell Death Differ* 18:1531-1539.

Appendix

Table 1a. Sequence alignment of *Enterobacteriaceae* species

```
                                                1                   20                  40                  60
E.coli_GlpG/1-276                               MLMITSFANPRVAQAFVDYMATQGVILTIQQH---NQSDVWLADESQAERVRAELARFLE
P.aeruginosa_gi|15598282/3-286                  AVQVLKFPLSVDLAGFVGLLRRLNVPHRVSEE--SGQQVLWVPDERLAEQVRELYRRYPE
A.nasoniae_gi|284008809/1-265                   -------------------MAMQNIVIEIRAEPQGKFFDLWLQDENKFEQAQQELKLFMQ
C.koseri_gi|157149004/1-276                     MLMITSFANPRVAQAFVDYMATQGVILTIQQH---HQSDVWLADESQAERVRAELARFLE
C.rodentium_gi|283787942/1-276                  MLMITSFANPRVAQAFVDYMATQGVILTIQQH---QQSDVWLADESQAERVRVELARFLE
C.youngae_gi|283835815/1-277                    MLMITSFANPRVAQAFVDYMATQGVILTIQQH---QQTDVWLADESQAERVRAELARFLD
D.dadantii_gi|271502340/1-274                   MIRIVALSNPRLAQAFVDYMHTQQIRLVARIN--GNEVDIWLPDESQQARVKQELERFLQ
D.zeae_gi|251787881/7-280                       MIRIVALSNPRLAQAFVDYMHTHQVRLVARIN--GDDVDIWLPDESQQERVKQELARFLQ
E.aerogenes_gi|336247568/1-276                  MLMITSFANPRVAQAFVDYMATQGVILTIQQH---TQSDVWLADESQAARVQAELARFLE
E.albertii_gi|170766588/1-276                   MLMITSFANPRVAQAFVDYMATQGVILTIQQH---NQSDVWLADESQAERVRAELARFHE
E.amylovora_gi|312174104/1-270                  -MRITHFTHLRPAQAFVDYMATRGIRLRIEQD---GGYTLWLDDESQIGVVENELNQFIR
E.billingiae_gi|300718791/2-274                 -MRITEFSNPRMAQAFVDYMATRGIQLRVEHD---NHYTVWLDDDSKVNLVENELSQFIR
E.cancerogenus_gi|261341920/1-276               MLMITSFTNPRVAQAFVDYMATQGVILTLQQH---TQTDVWLADESQAARVNEELARFLE
E.cloacae_gi|311277646/1-274                    --MITSFANPRVAQAFVDYMATQGIILTIQQH---TQSDIWLADESQAGRVQQELACFLE
E.fergusonii_gi|218550681/1-276                 MLMITSFANPRVAQAFVDYMATQGVILTIQQH---NQSDVWLADESLAERVRSELARFLE
E.hormaechei_gi|334126008/1-276                 MLMITSFSNPRVAQAFVDYMATQGIILTIQQH---TQTDVWLADETQASRVNAELARFLE
E.ictaluri_gi|238921529/1-273                   ---MITLPNPRLAQAFIDYMATQGVRLEAESG--SQGVALWLADDSRLAQVEQELQRFLQ
E.pyrifoliae_gi|259910066/1-276                 -MRITHFTHLRPAQAFVDYMATRGITLRIEQD---GGYTLWLDDESQIGVVENELNQFIG
E.tarda_gi|294638143/1-276                      MVRVITLPNPRLAQAFIDYMATQGVRLEAEAG--DHGIALWLADDSHLAQVEHELQRFLH
E.tasmaniensis_gi|188535345/1-271               IMRITHFTHLRPAQAFVDYMSTRGIRLRIERD---NGYTLWLDDESQSGVVENELDRFIR
K.pneumonia_gi|152972298/1-276                  MLMITSFANPRVAQAFVDYMATQGIILTIQQH---TQSDVWLADESQAGRVRAELARFLE
K.variicola_gi|288933195/1-274                  --MITSFANPRVAQAFVDYMATQGIILTIQQH---AQSDVWLADESQAGRVRAELARFLE
P.alcalifaciens_gi|212711656/1-281              MIHIISFENPRMAQAFVDYMAGQNIQLQVHPSNDQQHYELWLNDEQHTEQVRQELEAFLR
P.ananatis_gi|291619247/1-272                   -MRITQFNQPRMAQAFVDYMATRGVKLRIERE---SHYVIMLDDESQLSMVENELQQFVR
P.asymbiotica_gi|253987618/1-254                ----------------------------MHPGHDERHVELWLEDDTKLALVQQELEKFTR
P.atrosepticum_gi|50123058/2-277                MIRVIALSNPRLAQAFVDYMRTQQVHLDMRPQ--GHEAELWLEDETQLSKVQEALEIFLR
P.carotovorum_gi|227327619/1-276                MTRVIALSNPRLAQAFVDYMRTQQVHLEMRPQ--GHEAELWLEDETQLSKVQEALEIFLR
P.luminescens_gi|37524217/1-282                 MIRVIAISNPRLAQAFIDYMATHQVHLTMRPSHDGQHVELWLEDDTKLTLVQQELEQFTR
P.mirabilis_gi|197286757/1-279                  MIHIITLSNPQLADMFVNYMATKGVRIHAKTE--NQQVSLWLENEQQLNQVESELKTFLR
P.penneri_gi|226327545/1-264                    ----MSIIWQQKVCEFTQ-----------KNE--NQQISLWLEDQHQLDMVEKELNTFLR
P.rettgeri_gi|291326982/1-270                   -----------MAQAFVDYMASKNIHLKIKPSQEQAVFELWLEDEQLQPQVQKELDIFLQ
P.rustigianii_gi|261344537/1-281                MIHIISFANPRMAQAFVDYMASQHIHLQIQPSSEQQQFEVWLADDQQVELVRKELEIFLQ
P.stuartii_gi|188025399/1-270                   -----------MAQAFVDYMASQHIELKVHPVPEQQQIELWLQDEQYLAKVQQELNLFLQ
P.vagans_gi|308188432/1-273                     -MRITQFNQPRMAQAFVDYMATQGITLRIEHE---NHYVIMLDDESKISVVENELQQFLH
P.wasabiae_gi|261823352/2-277                   PIRVIALSNPRLAQAFVDYMRTQQVHLDMRPQ--GHEAELWLKDETQLSKVQEALEIFLR
S.bongori_gi|340001053/1-276                    MLMITSFANPRVAQAFVDYMATQGVILTIQQH---NQSDIWLADESQADRVQAELARFIE
S.boydii_gi|187733536/1-276                     MLMITSFANPRVVQAFVDYMATQGVILTIQQH---NQSDVWLADESQAERVRAELARFLE
S.dysenteriae_gi|309785831/1-276                MLMITSFANPRVAQAFVDYMATQGVILTIQQH---NQSDVWLADESQAERVRAELARFLE
S.flexneri_gi|110807258/1-257                   -------------------MATQGVILTIQQH---NQSDVWLADESQAERVRAELARFLE
S.glossinidius_gi|85060305/1-282                MIRVTALSNARLALAFIDYMKTQGVDMELRPQ--GRYAELWLAQDEKLAQVESALEAFLL
S.odorifera_gi|293393410/18-290                 MVRVIAVSNPRLAQAFVDYMATQGITLEVHNT--GEAAEIWLADDGRLEQVQHELQQFMI
S.proteamaculans_gi|157372871/1-278             MVRVIAVSNPRLAQAFVDYMTTQGIELRVHNT--GEAAEIWLADDSHLEQVQHELQQFLI
S.symbiotica_gi|320539837/1-278                 MVRVTAVSNPRLALAFVDYMATQGITLTLRNS--SNAAEIWLTDANHLEHVQHELQQFLV
X.bovienii_gi|290473227/1-277                   MIHVTSISNPRLAQAFIDYMVTQGIHLTMRPTHEPALVELWLEDENQLSFVEQELNQFSR
X.nematophila_gi|300725231/1-282                MIHITSISNPRLAQAFIDYMAMQSVHLIMRPTDEPSLVELWLKDDSQLNLVEQELRYFIR
Y.aldovae_gi|238757436/4-281                    MIRVIAISNLRLAQAFVDYMATRNVALEVRR--DSQGAEIWLADDEQLPLVQHELEQFLL
Y.bercovieri_gi|238786230/1-278                 MIRVIAISNLRLAQAFVDYMATRNVPLELRP--ETQGAEIWLADDEQLPLVQHELEQFLL
Y.enterocolitica_gi|123444164/1-278             MVRVIAISNLRLAQAFVDYMATRHVALEVKP--DSQGAEIWLTDDEQLPQVQHELEQFLL
Y.frederiksenii_gi|238788988/1-278              MIRVIAISNLRLAQAFVDYMATRHVALEVRP--DSQGAEIWLADDEQLSLVQHELEQFLL
Y.intermedia_gi|238792859/4-280                 MIRVIAISNPRLAQAFVDYMATRNVVLEVRP--DTQGAEIWLADDEQLPLVQQELEQFLL
Y.kristensenii_gi|238765189/1-278               MVRVIAISNLRLAQAFVDYMATRKVVLEVRP--DSQGAEIWLADDEQLPLVQHELEQFLL
Y.mollaretii_gi|238799344/1-278                 MIRVIAISNLRLAQAFVDYMATRNVALEVRP--DTQGAEIWLADDEQLSLVQQELEQFLL
Y.pestis_gi|162420860/260-537                   MTRVIVISNLRLAQAFVDYMATHHVALEIRP--DAQGVEIWLADDEQLSAVQHELEQFLL
Y.rohdei_gi|238751162/4-280                     MIRVIAISNLRLAQAFVDYMATRNVALEVKP--DTQGAEIWLADDEQLPLVQHELEQFLL
Y.ruckeri_gi|238754087/1-278                    MIRVIVISNLRLAQAFVDYMATQKVELSVRH--GTQDAEIWLAEDEKLPQVQHELEQFLL
```

Table 1b. Sequence alignment of *Enterobacteriaceae* species

```
                                          61                  80                  100                 120
E.coli_GlpG/1-276                         NPADPRYLAASWQAGHT--GSGLHYRRYP--FFAALRERAGPVTWVMMIACVVVFIAMQI
P.aeruginosa_gi|15598282/3-286            GDPQATLEAAPRPAGQG-FVRQLRMSPMT-AAVLLLTFVVAAVTYLGDNFATMRWLTFQD
A.nasoniae_gi|284008809/1-265             DPLNSRYQAASWQTGHI--DKKFGYRNYLT--FDYLKKQSGPLTIIVFFLAIFIYIWMLI
C.koseri_gi|157149004/1-276               NPADPRYLAASWQSGHT--GSGLHYRRFP--FIATLRERAGPVTWLIMIACILVFVVMSI
C.rodentium_gi|283787942/1-276            NPADPRYLAASWQSGHT--GSGLHYRRFP--FFATLRERAGPVTWFIMALCVLVFIVQSL
C.youngae_gi|283835815/1-277              NPGDPRYLAASWQSGHT--DSGLRYQRYP--FLATLRERAGPVTWVVMAACILVFIAMNV
D.dadantii_gi|271502340/1-274             EPWHERYQAASWQSGTT--NSGLRYPSYS--YLQTLRQRAGPLTLGMLVVAIAVYLLMQV
D.zeae_gi|251787881/7-280                 EPWHERYQAASWQSGTT--NSGLRYPSFS--YMQTLRQRAGPLTLSLLVVAIAVYLLMKV
E.aerogenes_gi|336247568/1-276            NPADPRYLAASWQSGHT--NSGLHYQRFP--FLATLRNNAGPFTWAILLACILVFILQNV
E.albertii_gi|170766588/1-276             NPADPRYLAASWQSGHT--GSGLHYRRFP--FFATLRERAGPVTWAVMIACVLVFIAMQI
E.amylovora_gi|312174104/1-270            DPNHARYQAASWHAGTT--KSGMHYQHTS--LLANLRERAGPLSLLVIAACVLVFILMQV
E.billingiae_gi|300718791/2-276           DPNHPRYQAASWHSGST--SSGIHYQGSS--LLANIRERAGPLTLSVMVMCIVVFVLMQI
E.cancerogenus_gi|261341920/1-276         NPADPRYLAASWQSGQT--GSGLQYRRFP--FIATLRERAGPFTLLLMAACILVFIIMNV
E.cloacae_gi|311277646/1-274              NPADPRYLAASWQSGHT--NSGLHYRRFP--FLATLKNNAGPFTWCIMMACILVFIVMNI
E.fergusonii_gi|218550681/1-276           NPADPRYLAASWQAGHT--GSGLHYRRYP--FFAALRERAGPVTWVMMIACVVVFIAMQI
E.hormaechei_gi|334126008/1-276           NPGDPRYLAASWQSGQT--DSGLHYQRFP--FLATLRERAGPFTLLLMAACIIVFIIMSV
E.ictaluri_gi|238921529/1-273             APQHPRYQAASWQSGST--HSGLRYARFG--YLRSFARQSGPLTIGVTLLCILVYLLQLA
E.pyrifoliae_gi|259910066/1-276           DPNHARYQAASWHAGTT--KSGIYYQHTS--LLANLRERAGPLTLLVIAACVLVFILMQV
E.tarda_gi|294638143/1-276                EPQHPRYQAASWQSGST--HSGLRYTRFA--YLRSFAGQSGPLTIGVTLLCILVYLCQLA
E.tasmaniensis_gi|188535345/1-271         DPNHARYQAASWHSGTT--HSGIHYQHTS--LLANLRERAGPLTLSVIAACVLVFILMQV
K.pneumonia_gi|152972298/1-276            NPADPRYLAASWQSGQT--NSGLRYQRFP--FFATLRHNAGPFTWAILLICIAVFILQNL
K.variicola_gi|288933195/1-274            NPADPRYLAASWQSGQT--NSGLRYQRFP--FFATLRHNAGPFTWAILLICIAVFILQNL
P.alcalifaciens_gi|212711656/1-281        NPNDPRYLEASWQTGRT--DAQFQYRNYLT--FGYLKQQSGPLTIAVILLSIAVYLWVTL
P.ananatis_gi|291619247/1-272             DPNHPRYQAASWHSGKT--DSGLRYERSN--IWANIRERAGPLTMVVMFACIAVFVLMQV
P.asymbiotica_gi|253987618/1-254          DPLNERYQAASWQSGKI--SHSLKYQNNL--NWHYLSRQAGPLTITVLILNIAVYLWMQF
P.atrosepticum_gi|50123058/2-277          DPTNPRYLAASWQTGSM--DTGIQYQRYS--YLQTLKQKAGPLTLSVMVVTIAVFILMQI
P.carotovorum_gi|227327619/1-276          DPTNPRYLAASWQTGSM--DTGIQYQRYS--YLQTLKQKAGPLTLSLMVVTIAVFILMQI
P.luminescens_gi|37524217/1-282           DPLNERYQAASWQSGDV--NHPLKYHNNL--NWQYLSRQAGPLTLTILLLNIVVYLWMQF
P.mirabilis_gi|197286757/1-279            EPFHPRYQAASWQTGTPQ-RTGLQYRPAF--SIKSMVQQSGPLTVLTIILCVLVFIWQQI
P.penneri_gi|226327545/1-264              EPFHPRYQAASWQTGNPQ-NTGIKYRPAF--SIKNMVQRSGPLTVLVIILCILVFIWQQV
P.rettgeri_gi|291326982/1-270             DPNHPRYLEASWHTGSA--DTQFQYRNYLT--WGYLKEQSGPLTIAVILLSIVVYLWVEM
P.rustigianii_gi|261344537/1-281          NPNDPRYLEASWQTGRT--DAQFQYRNYLT--FSYLKQQSGPLTIAVILLSIIVYLWVTM
P.stuartii_gi|188025399/1-270             DPSNPRYTEASWKIGNT--QTAFKYRNTLT--LANLKAQSGPLTIAVILICAGVYFWLAI
P.vagans_gi|308188432/1-273               DPNHPRYQAASWHSGKT--DSGLRYERSN--IWANIRERAGPLTMSLMVLCIAVFILMQV
P.wasabiae_gi|261823352/2-277             DPTNPRYLAASWQTGSM--DTGIQYQRYP--FLQTLQQKAGPLTLSVMVVTIAVFILMQI
S.bongori_gi|340001053/1-276              NPGDPRYLAASWQSGQT--NSGLHYRQFS--FLATLRERAGPVTWSVMAACVLVYVVMNL
S.boydii_gi|187733536/1-276               NPADPRYLAASWQAGHT--GSGLHYRRYP--FFAALRERAGPVTWVVMIACVVVFIAMQI
S.dysenteriae_gi|309785831/1-276          NPADPRYLAASWQAGHT--GSGLHYRRYP--FFAALRERAGPVTWVMMIACVVVFIAMQI
S.flexneri_gi|110807258/1-257             NPADPRYLAASWLAGHT--GSGLHYRRYP--FFAALRERAGPVTWVMMIACVVVFIAMQI
S.glossinidius_gi|85060305/1-282          DPQHPRYQAASWRTGSL-HYRGETGAGLRG-WLPALRQQAGPLTLGVTALCVLVFFLMAL
S.odorifera_gi|293393410/18-290           DPLNRRYQAASWQTGTT--GTDMPYEHVS--YLQIIRSKAGPLTLTVMVLCILVYLLMQV
S.proteamaculans_gi|157372871/1-278       DPLNSRYRAASWQAGNT--DADLHYQGFS--YLQTLRSKAGPLTLGVMALCIVVYILMQI
S.symbiotica_gi|320539837/1-277           DPLSRRYQAVSWQTGHT--DAGLHYQGYS--YLQALCSKAGPLTLSVMGLCIAVYILMQL
X.bovienii_gi|290473227/1-277             DPFNERYQTASWQAGKS--GYSFKYHNSLN--LSTLKSQSGPLTISVTALCILVYLWMQV
X.nematophila_gi|300725231/1-282          DPLHERYQAASWQTGKS--DILFKYRNNLN--WSTLRSQSGPLTISMTAICILVYFWIQI
Y.aldovae_gi|238757436/4-281              DPLNPRYQAASWQSGNL--HSDLNYQRFP--YLQTIRSQAGPLTLSVMVLCIAIYILMQF
Y.bercovieri_gi|238786230/1-278           DPLNPRYQAASWQSGNL--HSNLPYQRFS--YLQTLRSQAGPLTLSLMVLCIAVYILMQI
Y.enterocolitica_gi|123444164/1-278       DPLNPRYQAASWQSGNL--HSNLPYQRFS--YLQTLRSQAGPLTLSVMVLCIAIYILMQI
Y.frederiksenii_gi|238788988/1-278        DPLNPRYQAASWQSGNL--HSNLPYQRFS--YLQTIRSQAGPLTLSVMVLCIAIYILMQI
Y.intermedia_gi|238792859/4-280           EPLNPRYQAASWQSGNL--HSNLSYQRFS--YLQTLRSQAGPLTLSVMVLCIAIYILMQI
Y.kristensenii_gi|238765189/1-278         DPLNPRYQAASWQSGNL--HSNLPYQRFS--YLQTLRSQAGPLTLSIMVLCIAIYILMQI
Y.mollaretii_gi|238799344/1-278           DPLNPRYQAASWESGNL--HSHLSYQRFS--YLQTLRSQAGPLTLSVMMLCIAIYILMQI
Y.pestis_gi|162420860/260-537             DPLNPRYQAASWQAGNV--NSNLPYQRFS--YLQTLRSQAGPLTLSVMVLCIAIYILMLI
Y.rohdei_gi|238751162/4-280               DPLNPRYQAASWQSGNL--HSNLPYQRFS--YLQTIRSQAGPLTLGVMVLCIAVYILMQI
Y.ruckeri_gi|238754087/1-278              DPLNPRYQAASWQSGNL--HTEIPYQPVS--LLQNLRSQAGPLTLSMMVVCIAVYILMQI
```

Table 1c. Sequence alignment of *Enterobacteriaceae* species

```
                                                121                 140                 160                 180
E.coli_GlpG/1-276                               LGDQEVMLWLAWPFDPTLKF-EFWRYFTHALMHFSLMHILFNLLWWWYLGGAVEKRLGSG
P.aeruginosa_gi|15598282/3-286                  FRIQGQYALFTPLAESLAAG-QWWRLFTPMLIHFGWLHLAMNAMWFWELGRRIEFRQGRP
A.nasoniae_gi|284008809/1-265                   SGNDAVMAYLAWPSS-IQYT-EIWRIFTPALLHFSLMHILFNLMWWWYLGSQVERQLGTG
C.koseri_gi|157149004/1-276                     VGAQSVMVWLAWPFDPSLKF-EFWRYFTHAFMHFSLMHILFNLLWWWYIGGAVEKRLGSG
C.rodentium_gi|283787942/1-276                  IGYQTVMAWLAWPFDPSLKF-EVWRYFTHAFMHFSLMHILFNLLWWWYLGGAVEKRLGSG
C.youngae_gi|283835815/1-277                    VGDQAVMLWLAWPFDPSLKF-EFWRYFTHAFMHFSLMHILFNLLWWWYLGGAVEKRLGSG
D.dadantii_gi|271502340/1-274                   YGVNRMMTLLAFPDS-TQTL-QLWRWFSHALLHFSLLHLLFNLMWWWYLGGPVERVLGTR
D.zeae_gi|251787881/7-280                       YGVDRLMVLLAFPDS-TQDL-QLWRWFSHALLHFSLLHLLFNLMWWWYLGGPVERVLGSR
E.aerogenes_gi|336247568/1-276                  LGDQTMMIWLAWPYDPSLKF-EAWRYLTHAFMHFSLMHILFNLLWWWYLGGAVEKRIGSG
E.albertii_gi|170766588/1-276                   LGDQEVMLWLAWPFEPALKF-EFWRYFTHALMHFSLMHILFNLLWWWYLGGAVEKRLGSG
E.amylovora_gi|312174104/1-270                  LGDQTVMSLLSWPDE-TQHY-QLWRWFSHALLHFSLLHILFNLMWFWYLGGALEKRLGAG
E.billingiae_gi|300718791/2-274                 FGDQTVMSLLAWPDA--DQHFELWRWFTHALLHFSILHITFNLMWWWYLGGAVEKRLGSG
E.cancerogenus_gi|261341920/1-276               VGDQPVMLALAWPYDPSLQF-DVWRYFTHTLMHFSVMHILFNLLWWWYLGGIVEKRLGSG
E.cloacae_gi|311277646/1-274                    VGDRAVMGWLAWPYDPSLNF-ELWRYVTHAFMHFSLLHILFNLLWWWYLGGAVEKRLGSG
E.fergusonii_gi|218550681/1-276                 LGDQEVMLWLAWPFDPTLKF-EFWRYFTHALMHFSLMHILFNLLWWWYLGGAVEKRLGSG
E.hormaechei_gi|334126008/1-276                 VGDQSVMIALAWPYDPSLQF-DVWRYFTHALMHFSVMHILFNLLWWWYLGGIVEKRLGSG
E.ictaluri_gi|238921529/1-273                   LGDGAVMAYLAWPADAAQHY-QLWRYFTPALLHFSPLHIIFNLMWWWYLGGQVEKRLGAA
E.pyrifoliae_gi|259910066/1-276                 LGDQTVMSLLSWPDE-TQHY-QLWRWFSHALLHFSLLHILFNLMWFWYLGGALEKRLGAG
E.tarda_gi|294638143/1-276                      LGDGAVMTYLAWPADAAQHY-QLWRYFTPALLHFSPLHIIFNLMWWWYLGGQVEKRLGAS
E.tasmaniensis_gi|188535345/1-271               VGDQAVMSLLSWPDE-TQHY-QLWRWFSHALLHFSLLHILFNLMWFWYLGGALEKRLGSG
K.pneumonia_gi|152972298/1-276                  LGDQPVMIWLAWPYDPSLQF-EAWRYFSHAFMHFSLMHILFNLLWWWYLGGAVEKRIGSG
K.variicola_gi|288933195/1-274                  LGDQPVMIWLAWPYDPSLQF-EAWRYFSHAFMHFSLMHILFNLLWWWYLGGAVEKRIGSG
P.alcalifaciens_gi|212711656/1-281              TDPREVLYYLGWPIG-DQQS-ELWRWISPAFVHFSISHIGFNLALWWFLAGQVEKKMGTG
P.ananatis_gi|291619247/1-272                   LGDRTVMYWLAWPDA--DQHFEIWRWFSHVLLHFSLLHILFNLMWWWYLGGAVEKRLGSG
P.asymbiotica_gi|253987618/1-254                AGDYQVMYWLAWPDN-SQHM-ELWRWVTHGLLHFSLLHIIFNLMWWWYLGGQTEKRLGTG
P.atrosepticum_gi|50123058/2-277                SGYESVMTWLAFPAE-GQQL-QLWRWFSHALLHFSLLHILFNLMWWWYLGGPVEKVLGTG
P.carotovorum_gi|227327619/1-276                SGYESVMVWLAFPAE-GQQI-EVWRWFSHVLLHFSLLHILFNLMWWWYLGGPVEKVLGTG
P.luminescens_gi|37524217/1-282                 AGDYQVMSWLAWPND-SQHM-ELWRWVTHGLLHFSLLHIIFNLMWWWYLGGQTEKHLGTG
P.mirabilis_gi|197286757/1-279                  VGDYNVMLHLAWPYDASLDF-ELWRYITPAFVHFSLMHIAFNLAMWWYLASQTEKQLGTG
P.penneri_gi|226327545/1-264                    VGDYDVMLHLAWPYEESLNL-EVWRYITPAFIHFSLMHIAFNLAMWWYLASQTEQQLGTG
P.rettgeri_gi|291326982/1-270                   TDAREVLRYLAWPIG-DQQT-ELWRWFSPALVHFSLSHIGFNLALWWFLAGQVERKLGTG
P.rustigianii_gi|261344537/1-281                ADPREVLSYLGWPIG-DQQS-ELWRWFSPALVHFSISHIGFNLALWWFLAGQVEKKMGTG
P.stuartii_gi|188025399/1-270                   TQHLDVLRYLAWPRE-GQQF-ELWRWFSPAFVQFSLSQLIFNLVLWWYLASQVERRLGSG
P.vagans_gi|308188432/1-273                     MGDQTVMAWLSWPDA--DQHLQVWRWFSHALLHFSLLHILFNLMWWWYLGGAVEKRLGSG
P.wasabiae_gi|261823352/2-277                   SGYESVMTWLAFPAE-GQQV-QLWRWFSHALLHFSLLHILFNLMWWWYLGGPVEKVLGTG
S.bongori_gi|340001053/1-276                    VGDQTAMVWLAWPFDPALKF-EFWRYFTHIFMHFSLMHILFNLLWWWYLGGAVEKRLGSG
S.boydii_gi|187733536/1-276                     LGDQEVMLWLAWPFDPTLKF-EFWRYFTHALMHFSLMHILFNLLWWWYLGGAVEKRLGSG
S.dysenteriae_gi|309785831/1-276                LGDQEVMLWLAWPFDPALKF-EFWRYFTHALMHFSLMHILFNLLWWWYLGGAVEKRLGSG
S.flexneri_gi|110807258/1-257                   LGDQEVMLWLAWPFDPTLKF-EFWRYFTHALMHFSLMHILFNLLWWWYLGGAVEKRLGSG
S.glossinidius_gi|85060305/1-282                IGDDRVMAALLYPAGPAQYG-QIWRWVSHAFLHFSLLHLLFNVVWWWYLAGLTERYRGRA
S.odorifera_gi|293393410/18-290                 LGDDNVMYWLSWPQNSEQYS-QLWRWVSHALLHFSLLHIIFNLLWWWYLGGPMEKRLGAG
S.proteamaculans_gi|157372871/1-278             LGDDTLMYWLSWPQDSSQYL-QLWRWVSHAFLHFSLLHITFNLLWWWYLGGPLEKRLGSG
S.symbiotica_gi|320539837/1-278                 LGDSTLMYWLSWPQDRSQYL-QLWRGFSHAFLHFFLLHITFNLLWWWYLGGPVEKCLGTG
X.bovienii_gi|290473227/1-277                   AGDSDVMRWLAWPNG-EQYL-ELWRWVSPALLHFSLTHLLFNLALWWYLGSQVERHMGAG
X.nematophila_gi|300725231/1-282                AGTPDVMNWLAWPNN-DQYL-ELWRWVSHAVLHFSLTHLLFNLVLWWYLGSQVERKMSAG
Y.aldovae_gi|238757436/4-281                    VGDLAVMSWLAWPRDESQYL-QVWRWVSHALLHFSLLHILFNLMWWWYLAGQMEKRLGTG
Y.bercovieri_gi|238786230/1-278                 FGDYAVMSYLAWPRDSSQYL-QIWRWVSHAFLHFSLLHILFNLMWWWYLGGQMEKRLGTS
Y.enterocolitica_gi|123444164/1-278             VGDGAVMSWLAWPRDNSQYL-QIWRWVSHAFLHFSLLHILFNLMWWWYLAGQMEKRLGTG
Y.frederiksenii_gi|238788988/1-278              VGDAAVMSWLAWPRDSSQYL-QIWRWVSHALLHFSLLHILFNLMWWWYLAGQMEKRLGTG
Y.intermedia_gi|238792859/4-280                 VGDAAVMSWLAWPRDESQYL-QLWRWVSHGFLHFSLLHILFNLMWWWYLAGQMEKRLGTG
Y.kristensenii_gi|238765189/1-278               VGDATVMSWLAWPRDESQYL-QIWRWVSHALLHFSLLHILFNLMWWWYLAGQMEKRLGSG
Y.mollaretii_gi|238799344/1-278                 FGDNTVMSWLAWPRDESQYL-QIWRWISPAFLHFSLLHILFNLMWWWYLGGQMEKRLGTG
Y.pestis_gi|162420860/260-537                   TGDMAVMSWLAWPYNSSQYL-QIWRWVSHAFLHFSLLHILFNLMWWWYLGGQMEKRLGTS
Y.rohdei_gi|238751162/4-280                     AGDAAVMSWLAWPRDSSQYL-QVWRWGSHALLHFSLLHILFNLMWWWYLAGQMEKRLGTG
Y.ruckeri_gi|238754087/1-278                    VGDNTVMSWLAWPRNSSQYF-EFWRWVSHGLLHFSLMHILFNLMWWWYLGGQMEKRLGTG
```

Table 1d. Sequence alignment of *Enterobacteriaceae* species

```
                                                181                 200                 220                 240
E.coli_GlpG/1-276                               KLIVITLISALLSGYVQQKFSGP-WFGGLSGVVYALMGYVWLRGERDPQSGIYLQRGLII
P.aeruginosa_gi|15598282/3-286                  MLLGLTLLFGLVSNVVQYAVSGASLFGGLSGVLYGLLGHCWIFQYLAPNQAYRLPRGVVA
A.nasoniae_gi|284008809/1-265                   KLFVITIVSALFSNWGQSLFSGS-AFGGLSGVVFSLIGYVWLTGERHPAKGISIPRGLMI
C.koseri_gi|157149004/1-276                     KLIVITVISALLSGYVQQKFSGP-WFGGLSGVVYALMGYAWLRGERDPQSGIYLQRGLIA
C.rodentium_gi|283787942/1-276                  KLIVITVISALLSGYVQQKFSGP-WFGGLSGVVYALMGYVWLRGVRDPQSGIYLQGGLIV
C.youngae_gi|283835815/1-277                    KLIVITVISALLSGYVQQKFSGP-WFGGLSGVVYALMGYVWLRGERDPQSGIFLQRGLII
D.dadantii_gi|271502340/1-274                   RLCVILLLSSAISGWVQSWFSGV-YFGGLSGVVYALMGYVWLRGEREPGGPLHLQRGLMA
D.zeae_gi|251787881/7-280                       RLCVILLASSAFSGWVQSWFSGV-YFGGLSGVVYALMGYVWLRGEREPDGPLHFQRSLMA
E.aerogenes_gi|336247568/1-276                  KLVVITIISALLSGFVQHQVSGP-WFGGLSGVVYALMGYVWLRGERDPQSGIYLQRGLIL
E.albertii_gi|170766588/1-276                   KLIVITLISALLSGYVQQKFSGP-WFGGLSGVVYALMGYVWLRGERDPQSGIYLQRGLII
E.amylovora_gi|312174104/1-270                  KLFVIMLISALLSGWMQAKFSGL-LFGGLSGVVYALMGYSWLRGERDPDSGIFLPRGLMA
E.billingiae_gi|300718791/2-274                 KLFVILLISALLSGWMQSKYSGV-MFGGLSGAVYALMGYAWLRGERDPQSGIYLERGLMA
E.cancerogenus_gi|261341920/1-276               KLIVLTVISALLSGYVQHKFSGP-WFGGLSGVVYALMGYVWLRGERDPESGIYLQRGLIA
E.cloacae_gi|311277646/1-274                    KLIVITVISALLSGFIQQKFSGP-WFGGLSGVVYALMGYVWLRGERDPQSGVYLQRGLIL
E.fergusonii_gi|218550681/1-276                 KLIVITLISALLSGYVQQKFSGP-WFGGLSGVVYALMGYVWLRGERDPQSGIYLQRGLII
E.hormaechei_gi|334126008/1-276                 KLIVITLISALLSGYVQHKFSGP-WFGGLSGVVYALMGYAWLRGERDPESGIYLQRGLIT
E.ictaluri_gi|238921529/1-273                   KLAQIALVSALMSAWAQSLFSGI-FFGGLSGVVYALMGYVWWCGERAPQFGVAMPRGLML
E.pyrifoliae_gi|259910066/1-276                 KLFVIMLISALLSGWMQAKFSGV-LFGGLSGVVYALMGYSWLRGERDPDSGIFLPRGLMA
E.tarda_gi|294638143/1-276                      KLAQIMLVSALMSGWAQSLFSGI-YFGGLSGVVYALMGYVWWCGERAPQFGVSMPRGLMV
E.tasmaniensis_gi|188535345/1-271               KLFVIMLISALLSGWMQAKFSGV-LFGGLSGVVYALMAYSWLRGERDPDSGIFLQRGLMA
K.pneumonia_gi|152972298/1-276                  KLVVITVISALLSGFVQHQFSGP-WFGGLSGVVYALMGYVWLRGERDPQSGIYLQRGLIL
K.variicola_gi|288933195/1-274                  KLVVITVISALLSGFVQHQFSGP-WFGGLSGVVYALMGYVWLRGERDPQSGIYLQRGLIL
P.alcalifaciens_gi|212711656/1-281              KLFTILLVSALFSNWGQSLFSEN-NFGGLSGVVYALVSYVWVTGERRPEIGIGIPRGLMI
P.ananatis_gi|291619247/1-272                   KLFVIFIISALLSGWLQSKFSGI-WFGGLSGVVYALMGYCWLRGERDPDSGVYLERGLIG
P.asymbiotica_gi|253987618/1-254                KLFVITIVSAVFSGWGQSLFSGS-NFGGLSGVVYALIGYVWLTGERAPEHGIGVPRGLMA
P.atrosepticum_gi|50123058/2-277                KLLVIALVSALVSGWAQSWCSGT-YFGGLSGVVYALMGYVWLRGEREPDGYLSMPRSLMA
P.carotovorum_gi|227327619/1-276                KLLVITLVSALVSGWAQSWFSGT-HFGGLSGVVYALMGYVWLRGEREPDGYLSMPRSLMA
P.luminescens_gi|37524217/1-282                 KLFVIMIVSAVFSGWGQSLFSGS-HFGGLSGVVYALIGYVWLTGERAPERGIGVPRGLMA
P.mirabilis_gi|197286757/1-279                  KLFVIMLISALFSNWAQSLFAES-NFGGLSGVVYALISYVGLTGLRAPEKGIGVPTGLIV
P.penneri_gi|226327545/1-264                    KLFVIMLVSALFSNWAQSLFTDS-NFGGLSGVVYALIGYVGLTGIRNPEKGIGAPTGLIV
P.rettgeri_gi|291326982/1-270                   KLFTILLVSALFSNWGQSLFSEN-NFGGLSGVVYALVSYVWLTGERRPESGISVPRGLMI
P.rustigianii_gi|261344537/1-281                KLLTIFLVSALFSNWGQSLFSEN-NFGGLSGVVYALVSYVWITGERRPEIGIGIPRGLMV
P.stuartii_gi|188025399/1-270                   KLLTIFLVSALFCNWGQSLYSAS-NFGGLSGVVYALISYVWLMGERQPEKEVSAPRGLII
P.vagans_gi|308188432/1-273                     KLFVIMLISALLSGWLQAKFSGI-WFGGLSGVVYALMGYCWLRGERDPDSGVYLERGLIG
P.wasabiae_gi|261823352/2-277                   KLLVITLVSALVSGWAQSWFSGT-YFGGLSGVVYALMGYVWLRGEREPDGYLAMPRSLMV
S.bongori_gi|340001053/1-276                    KLIVITVISALLSGYVQQKFSGP-WFGGLSGVVYALMGYVWLRGERDPQSGIYLQRGLII
S.boydii_gi|187733536/1-276                     KLIVITLISALLSGYVQQKFSGP-WFGGLSGVVYALMGYVWLRGERDPQSGIYLQRGLII
S.dysenteriae_gi|309785831/1-276                KLIVITLISALLSGYVQQKFSGP-WFGGLSGVVYALMGYVWLHGERDPQSGIYLQRGLII
S.flexneri_gi|110807258/1-257                   KLIVITLISALLSGYVQQKFSGP-WFGGLSGVVYALMGYVWLRGERDPQSGIYLQRGLII
S.glossinidius_gi|85060305/1-282                SLAILFVLSAVVSGMVQSHFSGI-FFGGLSGVVYALMGYVWWHGEKNPGGPLFMPLGVIV
S.odorifera_gi|293393410/18-290                 KLFVLAVVSAFFSGWAQSLFSGA-LFGGLSGVVYALMGYSWLSGERAPERGLMLPRGLMV
S.proteamaculans_gi|157372871/1-278             KLFVLAVVSAFFSGWAQSLFSGA-LFGGLSGVVYALMGYCWLSGERAPERGLMLPRGLMV
S.symbiotica_gi|320539837/1-278                 KLLVLTVVSAFFSGWVQSLFSGA-LFGGLSGVVYALIGYVWLSGERAPACGLMLPRGLMV
X.bovienii_gi|290473227/1-277                   KLFEITIVSAVFTDWAQSLFSGS-HFGGLSGVVYALISYVWLTGEMSPKRGISVPRGLIA
X.nematophila_gi|300725231/1-282                KLFEITIVSAFFSGWAQSLFSGS-HFGGLSGVVYALIGYVWLTGELSPKRGISVPRGLIA
Y.aldovae_gi|238757436/4-281                    KLLVLTIVSALFSGWGQSLFSGA-NFGGLSGVVYALMGYVWLTGERAPERGISLPRGLMA
Y.bercovieri_gi|238786230/1-278                 KLLVLTIVSALFSGWGQSLFSGV-NFGGLSGVVYALMGYVWLTGERAPERGISLPRGLMA
Y.enterocolitica_gi|123444164/1-278             KLLVLTIVSALFSGWGQSLFSGV-NFGGLSGVVYALMGYVWLTGERAPERGISLPRGLMA
Y.frederiksenii_gi|238788988/1-278              KLLVLTIVSALFSGWGQSLFSGV-NFGGLSGVVYALMGYVWLTGERAPERGISLPRGLMV
Y.intermedia_gi|238792859/4-280                 KLLVLTVVSALFSGWGQSLFSGV-NFGGLSGVVYALMGYVWLTGERAPERGISLPRGLMA
Y.kristensenii_gi|238765189/1-278               KLLVLTIVSALFSGWGQSLFSGV-NFGGLSGVVYALMGYVWLTGERAPERGISLPRGLMA
Y.mollaretii_gi|238799344/1-278                 KLLVLTIVSALFSGWGQSLFSGV-NFGGLSGVVYALMGYVWLTGERAPERGISLPRGLMA
Y.pestis_gi|162420860/260-537                   KLLVLTIVSAVFSGWGQSLFSGA-NFGGLSGVVYALMGYVWLTGERAPEHGISLPRGLMA
Y.rohdei_gi|238751162/4-280                     KLLVLTLVSALFSGWGQSLFSGV-NFGGLSGVVYALMGYVWLTGERAPERGISLPRGLMA
Y.ruckeri_gi|238754087/1-278                    KLVVLTIVSAVFSGWGQSLFSGS-NFGGLSGVVYALMGYVWLTGQRAPERGISLPNGLMA
```

Table 1e. Sequence alignment of *Enterobacteriaceae* species

```
                                        241                 260                 280
E.coli_GlpG/1-276                       FALIWIVAGWFDLFGMS----MANGAHIAGLAVGLAMAFVDSLNAR--KRK----
P.aeruginosa_gi|15598282/3-286          MMLIWLLVCLSGVIDLLGFGSIANGAHVGGLLVGCLSGLLGGLLARRRR------
A.nasoniae_gi|284008809/1-265           FAVLWLFLGYFDIFGMK----IANTAHTAGLIIGLLMGLWDNRHKPDNTKHNQYI
C.koseri_gi|157149004/1-276             FALIWIVAGWFDVFGMS----MANGAHIAGLAVGLAMAFADTVNAR--KRT----
C.rodentium_gi|283787942/1-276          FALIWIVAGWFDLFGMS----MANGAHIAGLAVGLTMALADTINAR--KRT----
C.youngae_gi|283835815/1-277            FALIWIVAGWFDLFGMS----MANGAHIAGLAVGLAMAFADTMNVR--KRTL---
D.dadantii_gi|271502340/1-274           FALVWLVAGYFDILGMS----IANAAHASGLVIGVLMAWWDTRQH--RR------
D.zeae_gi|251787881/7-280               FALIWLVAGYFDVLGMS----IANAAHVAGLVMGVLMAWWDTRRR--SM------
E.aerogenes_gi|336247568/1-276          FSLVWLVAGWFDVFGMS----IANGAHVAGLAVGLAMAFVDTLNVR--KRT----
E.albertii_gi|170766588/1-276           FALIWIVAGWFDLFGMS----MANGAHIAGLAVGLAMAFVDSLNAR--KRK----
E.amylovora_gi|312174104/1-270          FAVLWLLVGWFGWFGLS----IANAAHVTGLLVGLAMAFVDTR-RR---------
E.billingiae_gi|300718791/2-274         FAVVWLVVGYMGWFGLS----IANAAHVTGLLVGLAMAFVDTR-HV--KRS----
E.cancerogenus_gi|261341920/1-276       FALIWLIAGWFDLFGMS----IANGAHVTGLAVGLAMAFADTLHAR--KRT----
E.cloacae_gi|311277646/1-274            FSLLWLVAGWFDLFGMS----IANGAHVAGLAVGLAMAFADTQNVR--KRT----
E.fergusonii_gi|218550681/1-276         FALIWIVAGWFDLFGMS----MANGAHIAGLAVGLAMAFVDSLNAR--KRK----
E.hormaechei_gi|334126008/1-276         FALIWLIAGWFDLFGMS----IANGAHVTGLAVGLAMALADTLHAR--KRT----
E.ictaluri_gi|238921529/1-273           FSLLWLVAGYFNWFGLA----IANAAHAGGLVIGLLMAFWDMRHFRRQ-------
E.pyrifoliae_gi|259910066/1-276         FAVIWLLVGWFGWFGLS----IANAAHVTGLLVGLAMAFVDTR-RR---------
E.tarda_gi|294638143/1-276              FSLLWLLAGYFNWFGLS----IANAAHIGGLVIGLLMAFWDMRHFRRQ-------
E.tasmaniensis_gi|188535345/1-271       FAVLWLLVGYFGWFGLS----IANAAHVTGLLVGLAMAFVDTR-RR---------
K.pneumonia_gi|152972298/1-276          FSLVWLIAGWFDVFGMA----IANGAHVAGLATGLAMAFVDTLHGR--KRA----
K.variicola_gi|288933195/1-274          FSLVWLIAGWFDVFGMA----IANGAHVAGLATGLAMAFVDTLHGR--KRA----
P.alcalifaciens_gi|212711656/1-281      FSIIWLFFGYFDLLGMD----IANAAHTSGLIIGLLMGIWDNRLSFKHQG-SK--
P.ananatis_gi|291619247/1-272           FALVWLVVGWFGAFGLA----IANAAHVGGLLVGLAMAFVDTR-KR--KR-----
P.asymbiotica_gi|253987618/1-254        FSILWLIVGYFDTFGLS----IANAAHFSGLTIGLLMALWDNRHTFKNNNRHF--
P.atrosepticum_gi|50123058/2-277        FALLWLVAGYFDILGMS----IANAAHVAGLIVGLLMAFWDTYNKTNPR------
P.carotovorum_gi|227327619/1-276        FALLWLIAGYFDILGMS----IANAAHVAGLVVGLLMAFWDTYNKTNPR------
P.luminescens_gi|37524217/1-282         FSLFWLIVGYFDAFGLS----IANAAHFSGLIIGLLMALWDNRHTFKNNNRHF--
P.mirabilis_gi|197286757/1-279          FSILWIIAGYMGVLGDS----IGNAAHFAGFLIGLLMALWDNRHQLSRQS-----
P.penneri_gi|226327545/1-264            FSILWIVAGYMGVLGDS----IGNAAHFAGFLIGLLMALWDNRHQLSRRS-----
P.rettgeri_gi|291326982/1-270           FSIIWLFFGYFDVLGMH----IANAAHTSGLIIGLLMGVWDNRHSFKHQS-SR--
P.rustigianii_gi|261344537/1-281        FSIVWLFFGYFDLLGMD----IANAAHTSGLIIGLLMGVWDNRLSFKHQS-SK--
P.stuartii_gi|188025399/1-270           LVVVWLVLGNLEILGPD----IANASHISGLIIGLLMGIWDNRHTFTSHR-SR--
P.vagans_gi|308188432/1-273             FALVWLVIGWFGVFGLA----IANAAHVAGLVVGLAMALVDTRHVT--RR-----
P.wasabiae_gi|261823352/2-277           FALLWLVAGYFDILGMS----IANAAHVAGLIVGLLMAFWDTYNKTNPR------
S.bongori_gi|340001053/1-276            FALLWIVAGWFDWFGMS----MANGAHIAGLVVGLAMAFADTLNAR--KRT----
S.boydii_gi|187733536/1-276             FALIWIVAGWFDLFGMS----MANGAHIAGLAVGLAMAFVDSLNAR--KRK----
S.dysenteriae_gi|309785831/1-276        FALIWIVAGWFDLFGMS----MANGAHIAGLAVGLAMAFVDSLNAR--KRK----
S.flexneri_gi|110807258/1-257           FALIWIVAGWFDLFGMS----MANGAHIAGLAVGLAMAFVDSLNAR--KRK----
S.glossinidius_gi|85060305/1-282        FALLWLIAGYFNILGIA----IANAAHVAGLVTGLLMAFWQTRRGQSHGYQR---
S.odorifera_gi|293393410/18-290         FSVLWLVAGYFDILGMS----IANAAHVAGLVLGLLMAFWDTRHRASHQQ-----
S.proteamaculans_gi|157372871/1-278     FSVLWLVAGYFDILGMS----IANAAHVAGLVLGLLMAFWDTRHRAHNEQ-----
S.symbiotica_gi|320539837/1-278         FSVLWLVVGYFDILGMS----IANAAHVAGLMIGLLMAFWDTHQRTRLEQ-----
X.bovienii_gi|290473227/1-277           ISVIWLLVGYFDMFSLN----IANAAHFSGLIIGLLMGLWDNLRKQKN-------
X.nematophila_gi|300725231/1-282        ITVIWLLVGYFDVFPLN----IANAAHVSGLIIGLFMGLWDNLRKQKNQEYSK--
Y.aldovae_gi|238757436/4-281            FSVLWLVAGYFDILGLS----IANAAHVSGLIIGLLMACWDTRNSARTAQ-----
Y.bercovieri_gi|238786230/1-278         FSVLWLVAGYFDILGLS----IANAAHVSGLIIGLLMAFWDTRSSVRTAQ-----
Y.enterocolitica_gi|123444164/1-278     FSVLWLVAGYFDILGLS----IANAAHVSGLIIGLLMAFWDTRNSAKTTQ-----
Y.frederiksenii_gi|238788988/1-278      FSVLWLVAGYFDILGLS----IANAAHISGLIIGLLMAFWDTRNSAKTTQ-----
Y.intermedia_gi|238792859/4-28          FSVLWLIAGYFDVLGLS----IANAAHVSGLIIGLLMAFWDTRSSVRTTQ-----
Y.kristensenii_gi|238765189/1-278       FSVLWLVAGYFDILGLS----IANAAHVSGLIIGLLMAFWDTRNSARTTQ-----
Y.mollaretii_gi|238799344/1-278         FSVLWLVAGYFDMLGLS----IANAAHVSGLIIGLLMAFWDTRSSVRTAQ-----
Y.pestis_gi|162420860/260-537           FSVLWLIAGYFDILGLS----IANAAHVSGLIIGLLMAFWDTRNSARTVQ-----
Y.rohdei_gi|238751162/4-280             FSVLWLVAGYFDILGLS----IANAAHISGLIIGLLMAFWDTRNSAKTA------
Y.ruckeri_gi|238754087/1-278            FSVLWLVAGYFDILGLS----IANAAHVSGLIIGLLMAFWDTRNSTRTTQ-----
```

Curriculum Vitae

Allison Rae Sherratt

EDUCATION

Doctorate (PhD – Biochemistry) Sept 2005 – 2012
University of Ottawa (Ottawa, ON, Canada)
Supervisor: Dr. Natalie K. Goto

Thesis: Beyond the active site of the bacterial rhomboid protease: Novel interactions at the membrane to modulate function.

Rhomboids are unique membrane proteins that use a serine protease hydrolysis mechanism to cleave their transmembrane substrate within the lipid bilayer. The proteolytic activity responsible for these functions is achieved by a core domain comprised of 6 transmembrane (TM) segments. In addition to the core domain, many rhomboids also possess aqueous domains of varying sizes at the N- and/or C-termini, the sequences of which tend to be rhomboid-type specific. The functional role of these domains is generally not well understood, although it is thought that they may be involved in regulation of rhomboid activity and specificity. To understand the role of the bacterial rhomboid cytoplasmic domain, full-length homologues from *Escherichia coli* and *Pseudomonas aeruginosa* were investigated using solution nuclear magnetic resonance (NMR) spectroscopy, protein engineering and activity analysis. Towards this goal, the *P. aeruginosa* was purified in a range of membrane-mimetic media, evaluated for its functional status *in vitro* and investigated for the NMR spectroscopic properties of these samples. Results from this study suggested an activity modulating interaction occurred between these two domains (**Publication #1**). However, investigation of this hypothesis with the *E. coli* rhomboid revealed that protease activity relied on a critical N-terminal sequence with a direct impact on the active site (**Publication #2, in preparation)**, and that the cytoplasmic domain homodimerized through a unique exchange of secondary structure elements called 3D domain swapping (**Publication #3, in preparation)**. Not only is domain swapping of membrane protein extramembranous domains extremely rare, this exchange was accelerated by large phosphocholine detergent micelles and appears to be the first example of domain swapping catalysis. If the membrane can induce the cytoplasmic domain to dimerize *in vivo*, this could provide a way for the bacterial rhomboid to sense its environment and provide a mechanism for regulation since this is still generally unknown.

Honours B. Sc. (Biochemistry, Co-operative Education Program) 2000 – 2005
University of Ottawa (University of Ottawa Heart Institute – Ottawa, ON, Canada)
Supervisor: Dr. Rosalind S. Labow

Thesis: The role of oxidation in U937-mediated biodegradation of polycarbonate-based polyurethanes.

Polycarbonate based polyurethanes (PCNUs) have been used in the development of medical devices intended for long-term implantation in the human body as well as drug delivery and tissue regeneration systems. Even though PCNUs are considered biocompatible, they still initiate the foreign body reaction (FBR) where neutrophils and monocytes are recruited to the polymer surface and monocytes differentiate to monocyte-derived macrophages. Part of the FBR involves the release of reactive oxygen species (such as H_2O_2 and HOCl) and hydrolytic enzymes (such as esterases) that have been shown to degrade PCNUs. Differentiated U937 cells were cultured on ^{14}C-labeled PCNUs differing in their surface chemistries in order to investigate the role of oxidation in hydrolytic biodegradation. The PCNUs (HDI431, HDI321 and MDI321) were treated with H_2O_2 prior to addition of cells, and radiolabel release (RR) and esterase activity was determined after 48 hours. There was an increase of RR on MDI321 following the H_2O_2 pretreatment, but not the aliphatic PCNUs, whereas esterase activity increased on HDI431. However, pretreatment with HOCl caused an increase in RR from the HDI-containing PCNU surfaces, as well as an increase in esterase activity. Understanding the link between oxidation and hydrolytic degradative processes is necessary in order to optimize materials for their intended use.

PUBLICATIONS

(1) Allison R. Sherratt, Michael V. Braganza, Elizabeth Nguyen, Thierry Ducat and Natalie K. Goto. *Insights into the effect of detergents on the full-length rhomboid protease from Pseudomonas aeruginosa and its cytosolic domain.* Biochim Biophys Acta. 2009 Nov;1788(11):2444-53.

(2) Allison R. Sherratt, David R. Blais, John Paul Pezacki and Natalie K. Goto. Activity-based protein profiling of *E. coli* GlpG rhomboid proteins identifies functional role for cytoplasmic region. (Manuscript in preparation).

(3) Allison R. Sherratt, Houman Ghasriani, Yves Aubin and Natalie K. Goto. Catalysis of domain swapping in the rhomboid protease cytoplasmic domain. (Manuscript in preparation).

PRESENTATIONS

ORAL

Allison R. Sherratt. Beyond the active site of the bacterial rhomboid protease: Novel interactions at the membrane to modulate function. Biochemistry PhD Research Seminar, Ottawa (June 15th 2011)

Allison R. Sherratt, Michael V. Braganza, Elizabeth Nguyen, Thierry Ducat and Natalie K. Goto. Gone, but not forgotten: Uncovering the role of the rhomboid protease cytoplasmic domain. BioMOLAR meeting, Ottawa (July 28th, 2010)

Allison R. Sherratt, Michael V. Braganza, Elizabeth Nguyen, Thierry Ducat and Natalie K. Goto. The effect of detergents on the full-length rhomboid protease using solution NMR. MOOT XXII Mini NMR Symposium, Ottawa (October 17th-18th, 2009)

Allison R. Sherratt and Natalie K. Goto. Solution NMR studies of the rhomboid intramembrane protease. BMI Seminar Symposium, University of Ottawa (February 18th, 2009)

Allison R. Melhuish and Natalie K. Goto. NMR analysis of a rhomboid protease. BMI Seminar Symposium, University of Ottawa (February 21st, 2007).

POSTER

Allison R. Sherratt, Houman Ghasriani and Natalie K. Goto. Structure And Function Of The *E. coli* Rhomboid Protease N-terminal Cytoplasmic Domain. Joint Euromar 2010 and 17th ISMAR Conference, Florence, Italy (July 4th-9th, 2010).

Allison R. Sherratt, Houman Ghasriani and Natalie K. Goto. Structure And Function Of The *E. coli* Rhomboid Protease N-terminal Cytoplasmic Domain. BMI Poster Day, University of Ottawa (May 20th, 2010).

Allison R. Sherratt, Houman Ghasriani and Natalie K. Goto. Structure And Function Of The E. coli Rhomboid Protease N-terminal Cytoplasmic Domain. McGill Biophysical Chemistry Symposium, Montreal, Quebec (May 13st, 2010).

Allison R. Sherratt, Thierry Ducat and Natalie K. Goto. Solution NMR studies of the rhomboid intramembrane protease. Groupe de Recherche Axé sur la Structure des Protéines (GRASP) Symposium, Montreal, Quebec (November 6th, 2008).

Allison R. Sherratt, Thierry Ducat and Natalie K. Goto. Solution NMR studies of the rhomboid intramembrane protease. International Conference on Magnetic Resonance in Biological Systems, San Diego, United States (August 24th-29th, 2008).

Allison R. Sherratt and Natalie K. Goto. Solution NMR study of the P. aeruginosa rhomboid intramembrane protease. McGill Biophysical Chemistry Symposium, Montreal, Quebec (May 1st, 2008).

Allison R. Melhuish and Natalie K. Goto. Functional study of an intramembrane protease by solution NMR. MOOT XX Mini NMR Symposium, Sainte-Adele, Quebec (September 29th-30th, 2007)

Allison R. Melhuish and Natalie K. Goto. Preliminary NMR analysis of a rhomboid protease. BMI Poster Day, University of Ottawa (May 18th, 2006).

Allison R. Melhuish and Natalie K. Goto. Preliminary NMR analysis of a rhomboid protease. McGill Biophysical Chemistry Symposium, Montreal, Quebec (May 16th, 2006).

AWARDS

2007-2010 PhD Admission Scholarship
2010 Faculty of Graduate and Postdoctoral Studies Travel Grant
2010 Travel Funding from the Biochemistry Graduate Studies Committee
2008 Faculty of Graduate and Postdoctoral Studies Travel Grant
2005 Dean's list, University of Ottawa

EMPLOYMENT

Post-doctoral Fellowship, National Research Council Canada (Ottawa)
PI: Dr. John P. Pezacki **Sept 2011—present**

- Development and characterization of chemical biology tools — including metabolic labelling and activity-based protein profiling of mammalian cells, pathogen-infected mammalian cells, and pathogenic bacterial cells — to gain a better understanding of unique metabolic pathways and host-pathogen interactions.

Co-op student, Defense Research & Development Canada (Ottawa) 2004

- Established a protocol to set up a bioreactor intended to co-culture human cell lines in a reduced-gravity environment.
- Identified problems with the bioreactor and worked with machinists to create better functioning parts.
- Evaluated metaphase chromosome spreads using the dicentric assay for signs of radiation damage to human white blood cells.

Co-op student, Health Canada (Ottawa) 2003

- Helped manage a process of compliance and enforcement for Canadian hospitals.
- Gained familiarity with Canadian health regulatory policies relating to the Food and Drugs Act and CSA Standards.
- Requested and processed data from Canadian hospitals regarding basic safety requirements in the handling of human cells, tissues and organs for transplantation.
- Prepared reports and presented findings at meetings and working groups on a weekly basis.

Co-op student, MDS Nordion (Ottawa) 2002

- Researched, developed and completed experiments to determine effective Y-90 radiation shielding properties of precious metals.
- Learned and practiced proper handling techniques for working with radioactive materials.

Laboratory Assistant, Dow Chemical: Research and Development (Sarnia) 1998-1999

- Prepared different types of solid and liquid samples for the Infrared spectrometer and eamined the acquired spectra.
- Prepared samples and examined the data gathered from the X-Ray Fluorescence.

- Created an Infrared spectra database of various substances.
- Completed independent projects testing polyethylene and epoxy.

Teaching Assistantships

2007, 2009 & 2010	Biochemistry Laboratory II (BCH3346), University of Ottawa
2005-2009	General Chemistry (CHM1311), University of Ottawa
2007	Enzymology & Biocatalysis (CHM4139), University of Ottawa