Structural and functional insights into prokaryotic rhomboids

by

Christelle Lazareno-Saez

A thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

Department of Biochemistry
University of Alberta

**Abstract**

Rhomboids are a family of intramembrane serine peptidases that cleaves a peptide bond of a substrate protein in the lipid bilayer. Various cellular processes such as cell differentiation and apoptosis have been linked to rhomboids proteolytic activty. More precisely, active and inactive rhomboids have been related to wound healing, type II diabetes, as well as respiratory defects and cancer. Currently, knowledge on rhomboids is limited with only two structures of prokaryotic rhomboids being elucidated: one from *Haemophilus influenzae* (hiGlpG) and a partial one (membrane domain from *Escherichia coli* (ecGlpG)). Initial studies have proposed a model for the reaction mechanism and the access of the substrate to the buried catalytic site of rhomboids. However, the structures have raised new questions and more structural details are needed. For example, different rhomboids contain a cytosolic domain, which has been hypothesized to be involved in regulation. In fact, the structure of the full-length ecGlpG (ecGlpG-FL) containing a large N-terminal cytoplasmic domain (ecGlpG-cyto) was never solved due to its arduous purification. To date, no structural information is available for any eukaryotic rhomboids, which often contain this cytosolic domain. This thesis aims to provide a detailed investigation of the catalytic mechanism, substrate access and regulation of prokaryotic rhomboids as they represent a valuable tool to study their eukaryotic counterparts. A new structural and functional study of hiGlpG proposed a model for the access of the substrate to the active site allowing us to gain insights into the catalytic mechanism. In addition structural and functional studies on ecGlpG-cyto revealed that this domain i) does not affect the apparent kinetic parameters of the enzyme and ii) oligomerize upon domain swapping.

**Preface**

Chapter 2 of this thesis has been published as Brooks C.L., Lazareno-Saez C., Lamoureux J.S., Mak M.W., Lemieux M.J.. "Insights into substrate gating in H. influenzae rhomboid." *Journal of Molecular Biology*, vol. 407, issue 5, 687-97. Protein characterization was performed by Lazareno-Saez C. and Mak M.W. Structural data analysis was mainly performed by Brooks CL, with the joint help of Lamoureux J.S., Lemieux MJ and myself. Finally, Brooks CL, Lemieux MJ and myself prepared the manuscript. Lemieux MJ was the supervisory author.

My original work (Chapter 4) has been published as Lazareno-Saez C., Arutyunova E., Coquelle N., Lemieux MJ. "Domain swapping in the cytoplasmic domain of the *Escherichia coli* rhomboid protease.." *Journal of Molecular Biology*, vol. 425, issue 7, 1127-42. I was responsible for experimental design, protein purification and characterization, structural data collection and analysis, manuscript preparation and edition. Arutyunova E. assisted with protein preparation and characterization, as well as manuscript preparation. Coquelle N. helped with structural data collection and analysis and manuscript preparation. Lemieux M.J. was the supervisory author involved with experimental design and manuscript preparation.

The literature review in Chapter 5 is part of my original work and has been published as Lazareno-Saez C, Brooks CL, Lemieux MJ. "Structural comparison of substrate entry gate for rhomboid intramembrane peptidases." *Biochemistry and Cell Biology.*, vol. 89, issue 2, 216-23. Brooks CL edited the manuscript. Lemieux MJ was the supervisory author.

# Acknowledgements

I would like to thank the department of Biochemistry at the University of Alberta and Dr. Lemieux for accepting me as a graduate student and giving me the opportunity of conducting this project. I really enjoy the world of research with its difficult and rewarding moments. I would also like to thank all the members of the Lemieux laboratory for their advice and support. Especially, I am really grateful to Elena Arutyunova for her support and long scientific (or not) discussions.

My committee members, Dr. Glover, Dr.Casey and Dr. Lemire have also been really supportive throughout my graduate studies and were always open for discussion.

Lastly, I would like to thank my husband, Nicolas Coquelle, for his constant support, love, faith and presence by my side in any situation. I also would like to thank my wonderful baby, Emma Coquelle, who complicated the writing phase of this thesis but is truly the sunshine of my life. She gives me the strength everyday to excel. She has made everything possible.

**Table of contents**

**Chapter 1. Introduction**

**Chapter 2. Insights into substrate gating and active site mechanism in *H. influenzae and E. coli* Rhomboid**

**Chapter 3. Towards structural determination of full length *E. coli* rhomboid (ecGlpG-FL)**

**Chapter 4. Domain swapping in the cytoplasmic domain of the *Escherichia coli* rhomboid protease**

**Chapter 5. General Discussion**

**Bibliography**

# List of Tables

# List of Figures

**Chapter 3**

**Chapter 4**

**Chapter 5**

List of Abbreviations

CAPF                    Cbz-Ala$^P$(O-iPr)F

$C_{12}E_8$                   octaethylene glycol monododecyl ether

$C_{12}E_9$                   polyoxyethylene(9)dodecyl ether

CFTR                    cystic fibrosis transmembrane conductance regulator

CHAPSO                  3-[(3-cholamidopropyl)dimethalammonio]-2-hydroxy-1-

                        propanesulfonate

DFP                     disopropylfluorophosphonate

DCI                     3,4-dichloroisocoumarin

DDM                     n-dodecyl-β-d-maltopyranoside

DM                      n-decyl-β-d-maltopyranoside

DMPC                    dimyristoyl phosphatidylcholine

DSP                     dithiobis[succinimidyl propionate]

DTSSP                   3,3′-dithiobis[sulfosuccinimidylpropionate]

DTT                     dithiothreitol

ecGlpG                  *E. coli* rhomboid

ecGlpG-cyto             N-terminal domain in ecGlpG

ecGlpG-FL               full-length ecGlpG

| | |
|---|---|
| ecGlpG-MD | membrane domain in  ecGlpG |
| EGFR | epidermal growth factor receptor |
| ERAD | ER-associated degradation pathway |
| GlpG | rhomboid |
| hiGlpG | *H. influenza* GlpG |
| HPLC | high performance liquid chromatography |
| I-CLiPs | intramembrane cleaving proteases |
| KC | potassium clavulanate |
| LDAO | lauryl dimethylamineoxide |
| MNG | maltose-neopentyl glycol |
| NG | nonyl-glucoside |
| OG | octylglucoside |
| paGlpG | *Pseudomonas aeruginosa* GlpG |
| PARL | presenilin-associated rhomboid like |
| PDB | protein data bank |
| PMSF | phenylmethanesylfanyl fluoride |
| S2Ps | site-2-proteases |

SEC                     size exclusion column

SPPs                    signal peptide peptidase

TatA                    translocase A

TEV                     tobacco etch virus

TFA                     tandem facial amphiphile

TM                      transmembrane

TMS                     transmembrane segment

TOC                     tylosis with esophageal cancer

TPCK                    N-*p*-tosyl-L-phenylalanine chloromethyl ketone

UIM                     ubiquitin-interaction motif

UM                      undecylmaltoside

## Standard Amino Acid Abbreviations

| | | |
|---|---|---|
| Glycine | Gly | G |
| Alanine | Ala | A |
| Valine | Val | V |
| Leucine | Leu | L |
| Isoleucine | Ile | I |
| Methionine | Met | M |
| Proline | Pro | P |
| Phenylalanine | Phe | F |
| Tryptophan | Trp | W |
| Serine | Ser | S |
| Threonine | Thr | T |
| Tyrosine | Tyr | Y |
| Asparagine | Asn | N |
| Glutamine | Gln | Q |
| Cysteine | Cys | C |
| Lysine | Lys | K |
| Arginine | Arg | R |
| Histidine | His | H |
| Aspartic Acid | Asp | D |
| Glutamic Acid | Glu | E |

**Chapter 1**

**Introduction**

## 1-1.    Thesis Rationale

Membrane proteins are a top research priority in both private and public research institutions worldwide. Located in the lipid bilayer of cells and organelles[1; 2], membrane proteins provide a pivotal gateway for the regulation of cell function and are therefore excellent drug targets. Rhomboids are a family of intramembrane serine peptidases that cleave a peptide bond of a substrate protein within the lipid bilayer[3]. Substrate cleavage occurs in various cellular processes[4] such as cell differentiation and apoptosis. More precisely, human rhomboids have been linked to wound healing[5], type II diabetes[6; 7], as well as respiratory defects[8]. Interestingly, inactive rhomboids are medically relevant as they were implied in esophageal and epithelial cancers[9; 10; 11].

Unfortunately, the exact mechanism by which rhomboids and especially human rhomboids cause diseases still remains unclear. This narrow link between rhomboid proteins and diseases highlights the importance of elucidating their structure and function for drug design. Details about the regulation of these proteins as well as substrate specificity and substrate gating will be critical. This thesis aims to provide insights into the molecular mechanism and regulation of rhomboid intramembrane peptidases. With the help of such a detailed investigation on prokaryotic rhomboids, which represent an ideal model to further study their eukaryotic counterparts, we are aiming to reach the first milestone to determine their role in diverse diseases.

**1-2.    Intramembrane Proteolysis**

*1-2.1.  Soluble proteases*

Soluble proteases have been extensively studied for decades. These proteins contain an active site located in the aqueous environment and process the hydrolysis of the amide bond that links amino acids together in peptides or proteins. Diverse functions of these proteins have been revealed through the years from proteolytic activity in the digestive system to cell growth and replication[12]. More, they play an important role in medicine as they represent important therapeutic targets. A specific protease terminology is used (**Figure 1-1**) to recognize the residues of the substrates on each side of the scissile amide bond (P and P' residues). Respectively, the amino acids of the proteases are located on binding pocket  (S and S' pockets) labeled depending on which residue of the substrate they interact with. Consequently, the catalytic residues of the protease are present between the S1 and S1' pockets, which are the key elements of protease specificity[12]. This protease terminology is used for rhomboid intramembrane peptidase, which are the focus of the thesis.

Four mechanistic classes of soluble proteases are distinguished (**Figure 1-2**): serine/threonine proteases, cysteine proteases, aspartyl proteases and metalloproteases. In serine/threonine proteases and cysteine proteases, the serine, threonine or cysteine catalytic residues are paired to a proton-withdrawing group in order to engage the nucleophilic attack to the carbonyl carbon of the amide bond. The serine (more rarely threonine) proteases generally contain a catalytic triad  (serine, histidine and aspartate), which allows a deprotonation of the serine residue for the nucleophilic attack of the
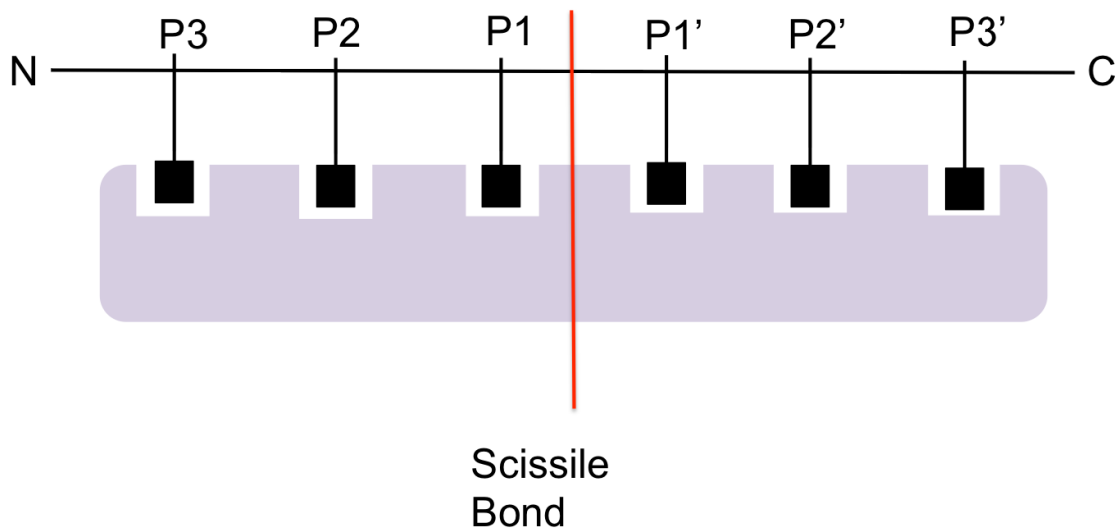
**Figure 1-1. General terminology of substrate and substrate binding pockets during proteolysis.**
The scissile peptide bond is represented in red. On the N-terminal side of the cleavage site, the residues are named P1, P2, P3, etc. successively. The related binding pockets of the protease are S1, S2, S3, etc. On the C-terminal side, a similar nomenclature, using P1', P2', P3', etc. and S1', S2', S3', etc., is indicated.
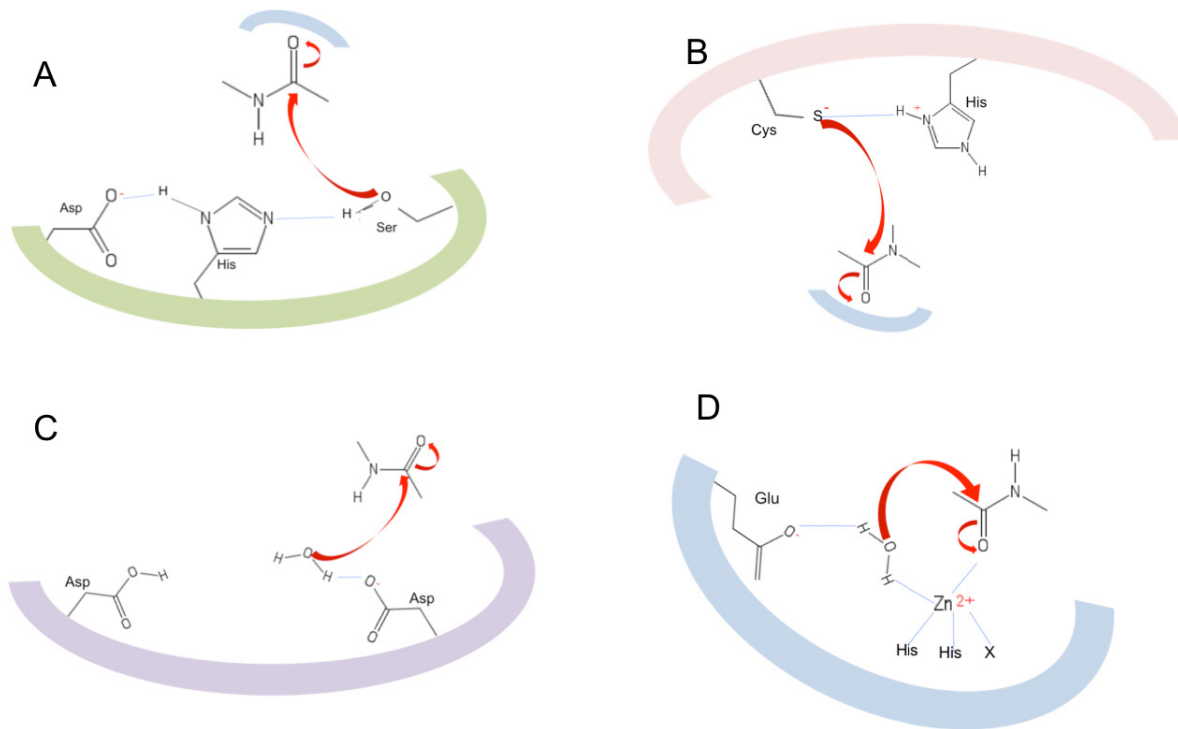
**Figure 1-2. Details of the four main mechanistic classes of soluble proteases.**

Serine (**A,** EC 3.4.21) and cysteine (**B,** EC 3.4.22) proteases mechanism combine a serine or cysteine residue, respectively, with a proton-withdrawing group during the nucleophilic attack of a peptide bond. Both mechanisms are really similar as they both use a strong nucleophile, a covalent enzyme-substrate complex and an oxyanion hole during the process. Aspartyl proteases (**C,** EC 3.4.23) and metalloproteases (**D,** EC 3.4.24) proceed to a nucleophilic attack on the polypeptide via a water molecule. A coordinated metal (often zinc) stabilizes the oxyanion in metalloproteases mechanism. The large colored curves and the narrow blue curves represent the enzyme and the oxyanion holes, respectively. For each mechanistic class, the catalytic residues are represented in black. For the metalloproteases (**D**), the letter X can represent a His, Asp or Glu. The grey lines, blue dots and red arrow are a schematic representation of the continuation of the substrate polypeptide, the electrostatic interactions and hydrogen bonds and the movement of electron pairs, respectively.

amide carbonyl (**Figure 1-2 A**). This is the first step of the catalytic mechanism (**Figure 1-3**) of such protease, during which a negatively charged tethraedral intermediate is generated (**Figure 1-3 B**). This intermediate is stabilized by a pocket in the enzyme delineated the oxyanion hole, which typically consists of amino acids able to hydrogen bond with the oxyanion. Depending on the protease, this oxyanion hole can vary. Cysteine proteases process with a similar mechanism using generally a catalytic dyad (cysteine and histidine) (**Figure 1-2 B**). Similarly to serine proteases, an enzyme-substrate complex is formed during the mechanism and stabilized by an oxyanion hole[12; 13]. Aspartyl proteases (**Figure 1-2 C**) and metalloproteases (**Figure 1-2 D**) proceed to an indirect nucleophilic attack of the scissile bond of the substrate, through a water molecule. Indeed, aspartyl proteases contain a catalytic dyad (two aspartate residues), which will deprotonate the water molecule prior to the nucleophilic attack. During the mechanism, no covalent enzyme-substrate complex is created and the tetrahedral intermediate is uncharged (**Figure 1-2 C**). Metalloproteases catalytic mechanism is often based on a zinc atom, which is coordinated to two histidines, an acidic side chain and the water molecule, which is also hydrogen bonded to a glutamate. The zinc ion stabilizes the oxyanion during the mechanism[12; 13] (**Figure 1-2 D**). I particularly detailed the catalytic mechanism of soluble serine protease (**Figure 1-3**), as it will be a good reference in the determination of molecular catalytic mechanism for rhomboids intramembrane serine peptidase (see details below).
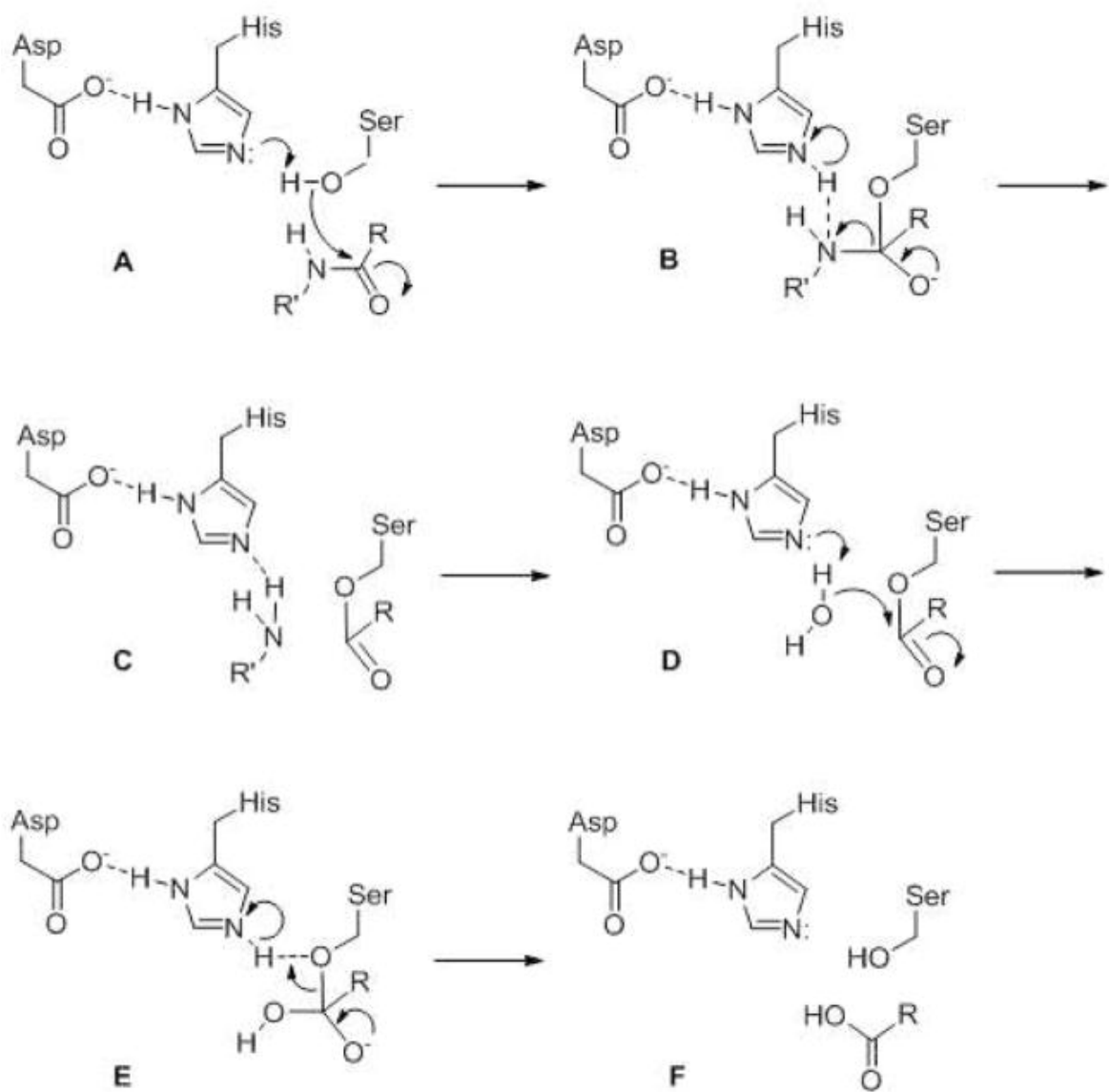
**Figure 1-3. Detailed general catalytic mechanism of serine protease.** The catalytic triad (serine, histidine and aspartate) works together to process the nucleopilic attack of the substrate carbonyl carbon. The aspartate represent a proton withdrawing group for histidine which deprotonates the serine hydroxyl group enabling it to attack the carbonyl carbon of the peptide bond (A). A tetrahedral intermediate is stabilized by the oxyanion hole of the protease (B). After release of the amino group (C), a water molecule is activated to hydrolyse the ester bond (D and E) and release the product (F).

*1-2.2.  Intramembrane Cleaving Proteases (I-CLiPs)*

Until a decade ago, all discovered proteases were soluble proteins, as described above, with the active site located in the aqueous environment either with the complete protein in this environment or the presence of a membrane anchor holding a soluble protease. In 1997, a new group of proteases was discovered: the Intramembrane Cleaving Proteases (I-CLiPs)[14]. These proteases process the regulated intramembrane proteolysis of the peptide bond within the lipid bilayer, a water limited environment. Four mechanistic classes of intramembrane proteases have been discovered so far (**Figure 1-4**): metalloproteases (Site-2-Proteases: S2Ps), aspartyl proteases (presenilin and signal peptide peptidase: SPPs), glutamic proteases and serine peptidases (Rhomboids named GlpG)[12; 13; 15; 16; 17]. They are polytopic membrane proteins present in all three domains of life and represent therapeutic targets, as they are involved in pathways such as Hepatitis C virus maturation, cholesterol and fatty acid synthesis and others[13].

Their catalytic residues, which are similar to their soluble counterparts, are located on transmembrane segments (**Figure 1-4**). Rhomboids, surprisingly, present a serine-histidine dyad, which is different to the common catalytic triad usually found in the soluble serine peptidases. These residues are connected to the aqueous environment by channel(s) allowing the presence of water in the active site. So far, most substrates have a single transmembrane segment; however polytopic proteins can also be cleaved by  rhomboids[15]. A lateral substrate gating was hypothesized for all intramembrane peptidases except SPPs. The Rhomboid was the last one discovered but it is now the best understood. Consequently, they represent a good candidate to understand and investigate some characteristics of I-CLiPs.
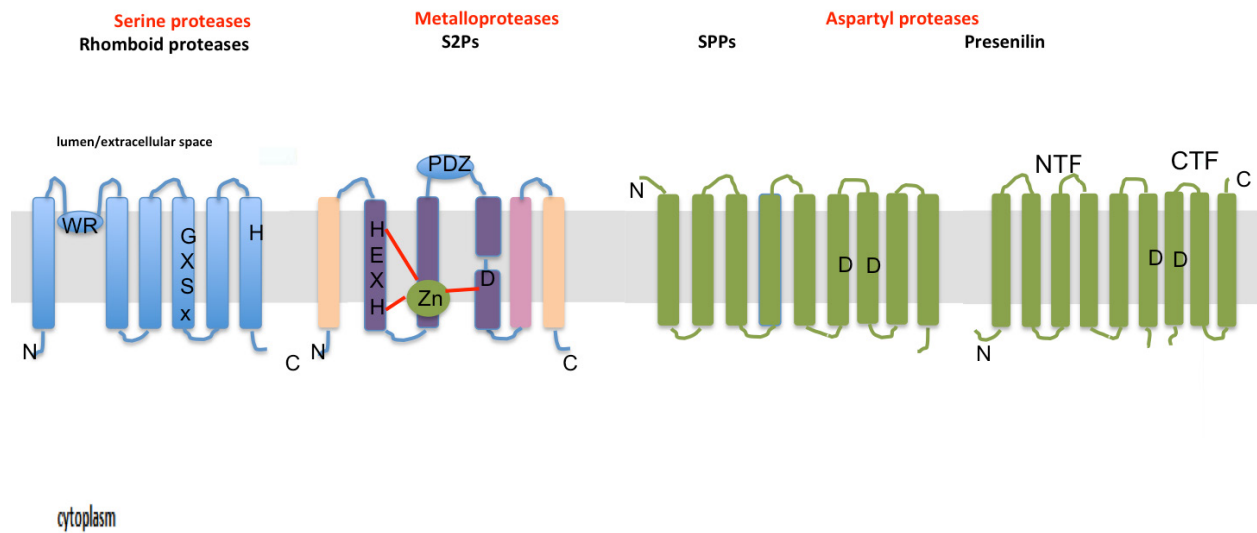
**Figure 1-4**. **Topology of the three mechanistic classes of intramembrane proteases.**
The catalytic residues forming the active site are indicated. For rhomboids and S2Ps, the minimal core is shown in blue and purple. Other transmembrane segments can be present. Presenilins undergo endoproteolysis creating active presenilin composed of NTF and CTF domains.

## 1-3. Rhomboid-family proteins

Transmembrane segment (TMS) predictions, multiple sequence alignment and phylogenetic analysis[18] helped to define classification of rhomboids. The term rhomboid-family is used for all the rhomboid homologous genes catalytically active (active proteases) or inactive (pseudoproteases)[19; 20] (**Figure 1-5**). The genes predicted to be catalytically active contain the sequence GxSx with the catalytic serine on one TMS and a histidine base on another TMS. A helix dimerization motif GxxG and a WR motif are noticeable on TM6 and the first loop, respectively[18]. Among the active rhomboids, three subfamilies can be distinguished: secretase A and B, and the PARL-type (presenilin-associated rhomboid like), which are until now found in endosymbiotic organelles such as mitochondria and plastids[20] (**Figure 1-5**). They all have different topologies with the minimal core of rhomboids represented by secretase B (6 TMS). It is noticeable that the 7 TMS topologies are mostly encountered in eukaryotes, with the exceptions of AarA (*Providencia stuartii*) and YqgP (*Bacillus subtilisis*) rhomboids for example. An additional N -or C-terminal cytoplasmic domain can be present in these three subfamilies.

Among the pseudoproteases, the catalytically inert iRhoms, Derlins and the other predicted inactive enzymes compose the rest of the rhomboid-family proteins (**Figure 1-5**). iRhoms are catalytically inert due to a proline in the GxSx motif (GPxx). One striking feature of iRhoms is the presence of an extended amino-terminal domain and, between the first two TMS, a large insertion loop (over 240 amino acids) predicted to have a globular fold. In addition, the WR motif is missing[18]. Derlins have been more recently added to rhomboids pseudoproteases[19; 21]. They are lacking active-site residues and interestingly, one Derlin (human Derlin-1) was shown to share the 6TMS secretase B topology with the presence of the WR motif[21].
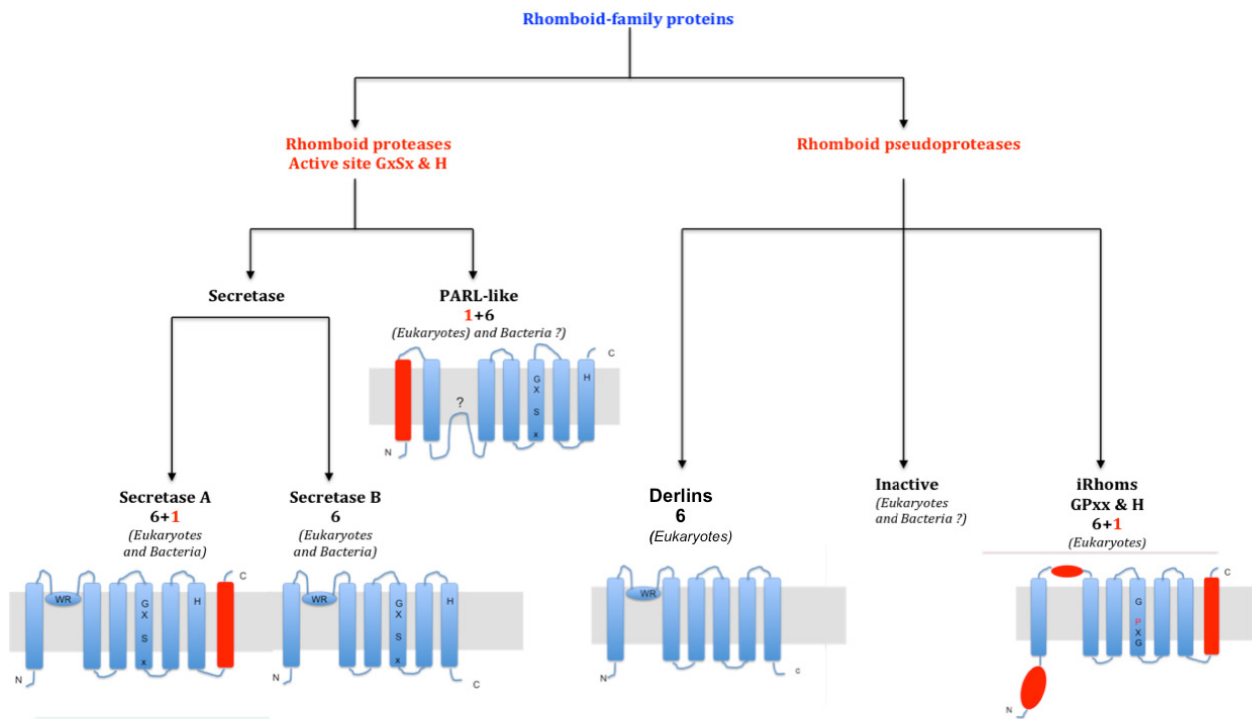
**Figure 1-5**. **Rhomboid-family proteins.**
The schematic outline between rhomboids proteases and psseudoproteases in the Rhomboid-family proteins. The presence of such members in eukaryotes and/ or prokaryotes is detailed.

Interestingly, mutations of this motif impaired its function[21]. Other inactive homologs, which cannot be classified as iRhoms or Derlins and miss key catalytic residues are thought to be mutants of active rhomboids[18]. Little is known about iRhom, Derlins and the inactive homologs like family. Among the rhomboid-family, the choice of *E. coli* rhomboid (ecGlpG) as model seems natural, as it is a rhomboid (active form) presenting the minimal common topology of 6 TMS. Investigating this model rhomboid will help to understand its counterparts.

## 1-4 Biological functions and medical implications of rhomboids

Rhomboids are well-conserved in all kingdom of life[18]. Although initial data strongly supported cell signaling regulation role[4], some diverse biological functions have been highlighted.

### 1-4.1. Biological function in eukaryotic rhomboids

Rhomboid-1 was the first discovered in *Drosophilia melanogaster* where it triggers the signal-generating component of epidermal growth factor receptor (EGFR)[22]. This protein was first found to be responsible for the release of a membrane-tethered EGFR ligand spitz. The spitz ligand is localized in the ER where it will interact with the membrane protein Star and be translocated to the Golgi asparatus **(Figure 1-6)**. Here spitz encounters rhomboid-1, is cleaved and released and then becomes an active EGFR ligand. This function is mainly accomplished by metalloproteases of ADAM family. However, other rhomboid-family protein are involved in EGFR signaling such as human rhomboid RHBDL2[23] and RHBDF2, an iRhom. In fact, this RHBDF2 rhomboid pseudoprotease, located in endoplasmic reticulum, promotes maturation of

an ADAM metalloprotease by trafficking from ER to the Golgi. This role as well as its N-terminal domain sequence are linked to tylosis with esophageal cancer (TOC)[24; 25].

In addition, various rhomboid-family proteins are part of the ER-associated degradation pathway (ERAD) such as eukaryotic RHBDL4 protease[26], Derlins and RHBDF2 pseudoproteases[27; 28; 29]. Especially, Derlin-1 acts in a complex network mediating ERAD for the disease-associated mutants of the polytopic membrane protein CFTR (cystic fibrosis transmembrane conductance regulator)[30; 31; 32].

Moreover, some evidences suggest that human rhomboids are linked to blood clotting and wound healing[5] (RHBDL2) and the growth of epithelial cancers (RHBDD2, an inactive rhomboid, and RHBDF1, an iRhom).

Studies in yeast revealed that PARL-type Rbd1 rhomboid deficiency results in mitochondrial fragmentation and growth arrest. Moreover, the mammalian PARL rhomboid is implicated in regulating progressive cachexia, type II diabetes and Parkinson's disease through apoptosis or mitophagy[33].

Finally, rhomboids were also identified in protozoan parasites, which are the most devastating pathogens. Rhomboids were identified in apicomplexan parasites which invasive pathogens are using transmembrane protein adhesins to create moving junction between the host and the parasite. In *Plasmodium falciparum* and *Toxoplasma gondii*, strong evidence suggest that rhomboids enzymes are part of the proteases involved in adhesins cleavage in the last stage of host invasion, in order to internalize the parasite and seal the host membrane[15; 34; 35]. Remarkably, in *T. gondii*, another rhomboid seems to affect intracellular parasite growth. These two organisms are responsible for toxoplasmosis infecting immunocompromised individuals,
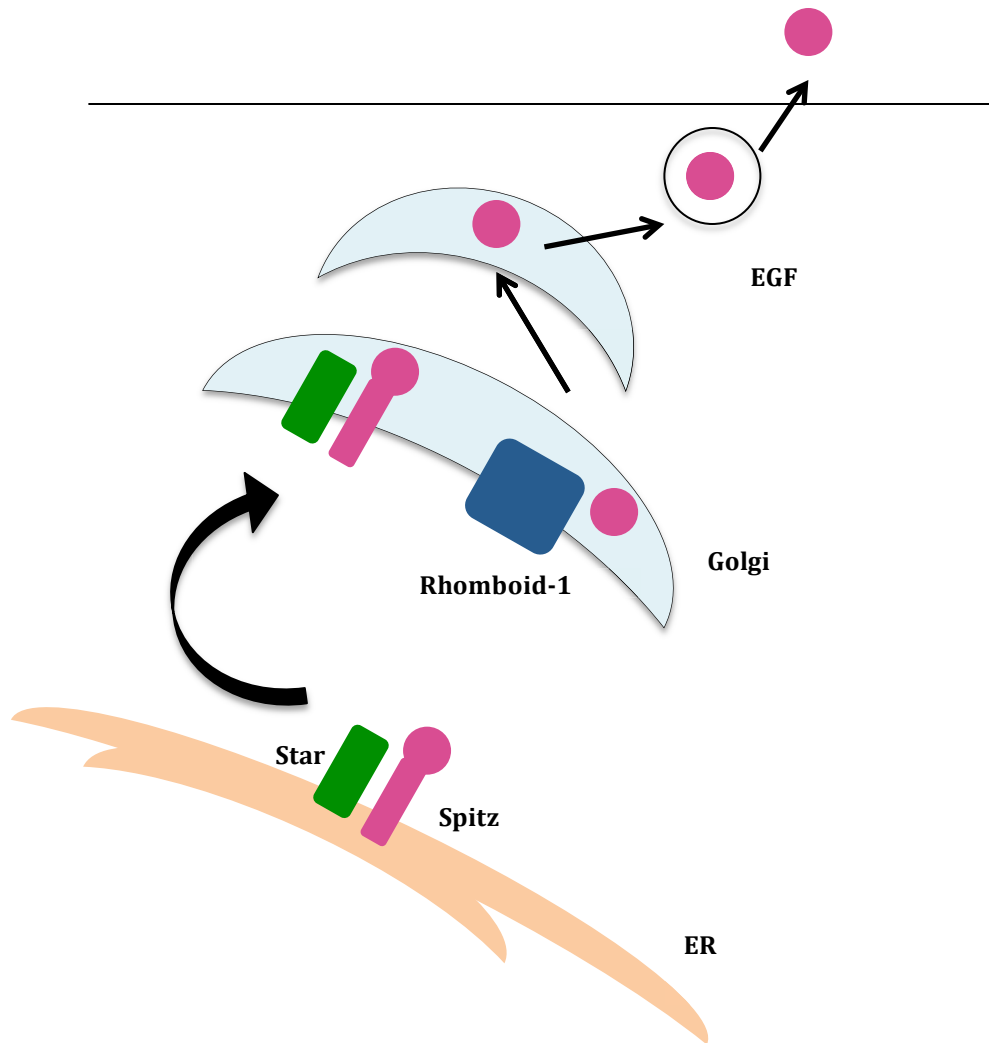
**Fig. 1-6. Rhomboid-1 mediated EGF signaling in *D. melanogaster***
In Golgi asparatus, Rhomboid-1 (blue) cleaves Spitz membrane protein (pink) for release of the extracellular domain to be exported out of the cell for epidermal growth factor (EGF) signaling.

pregnant women. They also cause neurologic birth defects, and are involved in the lethal form of malaria. Therefore, they represent important therapeutic targets. However, the presence of various rhomboid proteases in protozoan parasites suggests a wide range of potential functions for these proteases beyond the processing of cell surface adhesins[36].

*1-4.2. Biological function in prokaryotic rhomboids*

In the human pathogenic bacteria *Providencia stuartii*, AarA rhomboid cleaves an N–terminal extension of the twin arginine translocase A (TatA) thereby activating the TatA channel. The unknown signal responsible for quorum sensing is then translocated. The biological function of AarA is hardly generalizable as only a few bacteria encode an N-terminal extension form of TatA. Apart from *P. stuartii*, rhomboids, have been studied from the human pathogens *E. coli (ec*GlpG)*, H. influenzae (hi*GlpG)*, Mycobacterium tuberculosis* (rhomboid protease 1 and 2 encoded by Rv0110 and Rv1337), *Mycobacterium smegmatis* (rhomboid protease 1 and 2 encoded by MSMEG_5036 and MSMEG_4904)[37] and from the food contaminant *Bacillus subtilis* (YqgP). To date, only the roles of YqgP in glucose export and cell division as well as *M. smegmatis* rhomboid 2 in colony morphology, biofilm formation, antibiotic and DNA gyrase inhibitors resistance have been revealed[37]. More experiments are needed to determine if bacterial rhomboids play a pivotal role in the activity of a range of human pathogens and may serve as drug targets against infectious diseases.

A better understanding of rhomboid physiological functions is mandatory to clearly establish their link to these multiple diseases. Therefore, it is necessary to improve our knowledge about the function and regulation of rhomboids. Specifically, more insights about the

catalytic mechanism and substrate gating represent the first step to design specific drugs for various rhomboids.

**1-5.    Structural studies of the catalytic transmembrane core of rhomboids**

Although rhomboids were the last mechanistic class of intramembrane proteases discovered, the first two X-ray crystal structures of intramembrane proteases were those of two prokaryotic secretase B (6 TMS): hiGlpG[38; 39] and the membrane domain of ecGlpG[32; 40; 41; 42; 43; 44; 45].

*1-5.1. Prokaryotic rhomboid structure*

Aside from the presence of an N-terminal domain in ecGlpG (ecGlpG-cyto), hiGlpG and ecGlpG present a common fold with subtle differences. The membrane domain represents the minimal core of the rhomboid-family (**Figure 1-7**). For this domain, various structures in different environments (lipids[32],bicelles[44] or detergents[32; 40; 41; 42; 43; 45]) (**Table 1-1**) reveal different aspects of the structure, active site, mechanism and environment interaction. The different structures are remarkably similar except in one specific region (see below). The overall shape of the membrane domain is a compact helical bundle (**Figure 1-7**). The protein contains an internal hydrophilic cavity facing the periplasm. For ecGlpG, at the bottom of this cavity is located the catalytic serine (S201) which is approximately embedded 10 Å below the membrane surface. S201 is situated at topmost residue on helix 4, in the center of the protein, and surrounded by a ring of transmembrane helices **(Figure 1-8)**. Helix 4 is inclined relative to the plane of the membrane and does not completely cross the membrane. The general base histidine

**Table1-1. Prokaryotic rhomboids membrane core structure**
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 glycolmonododecyl ether. The PDB ID in bold highlights the structures solved during my thesis studies by other groups. Some structures (in particular those complexed with inhibitor) are detailed further in Chapter 5.

| Structure | PDB ID | Resolution (Å) | Space Group | Lipid bilayer mimetic | reference |
|---|---|---|---|---|---|
| **1. Membrane domain of ecGlpGFL** | | | | | |
| *Apo structures* | | | | | |
| | 2IC8 | 2.10 | R32 | NG | 40 |
| | 2IRV | 2.30 | P21 | LDAO/DDM | 42 |
| | **2XOV** | 1.65 | R32 | NG | 43 |
| | 3B45 | 1.90 | R32 | NG | 32 |
| | 2O7L | 2.50 | R32 | NG | 45 |
| *Helix 5 open structures* | | | | | |
| | 2NRF | 2.60 | P31 | NG | 41 |
| *Mutant structures* | | | | | |
| | 3B44 | 1.70 | R32 | NG | 32 |
| | **2XTV** | 1.70 | P212121 | DHPC/CHAPS (Bicelles) | 44 |
| | **2XTU** | 1.85 | R32 | | 44 |
| Inhibitor structures | | | | | |
| | **3TXT** | 2.30 | R32 | NG | 62 |
| | **2XOW** | 2.09 | R32 | NG | 43 |
| | **3UBB** | 2.60 | R32 | NG | 63 |
| **2. hiGlpG** | | | | | |
| *Apo structures* | | | | | |
| | 2NR9 | 2.20 | C2 | C12E8 | 38 |
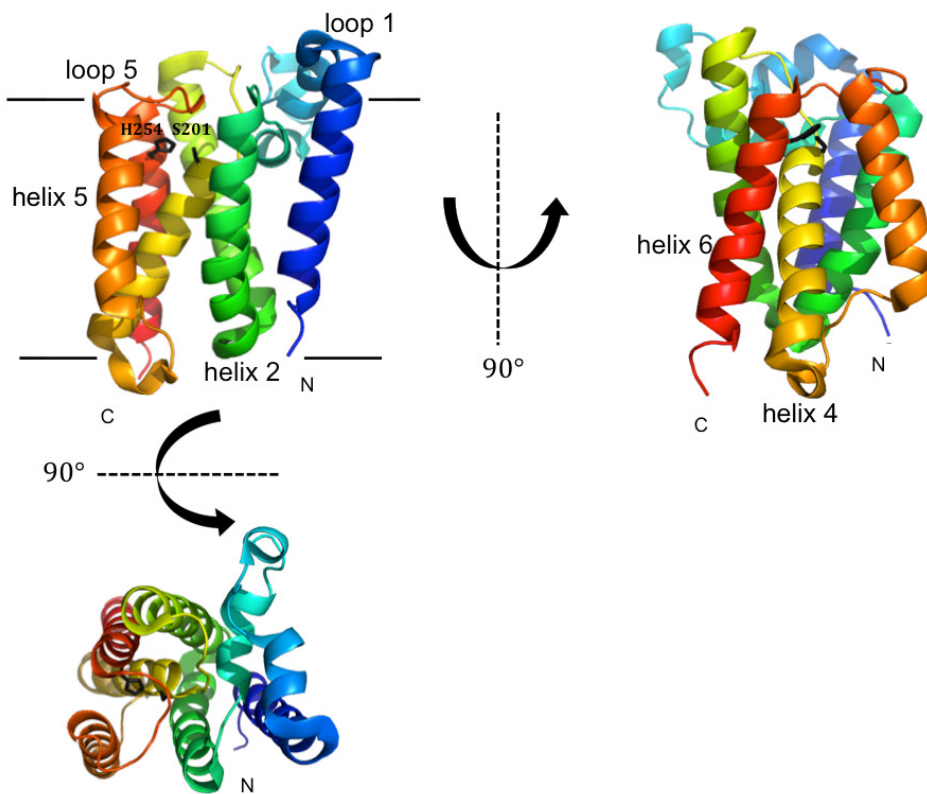
**Figure 1-7. X-ray crystal structure of the membrane domain of ecGlpG rhomboid protease**.
Rhomboid structure is shown in cartoon (PDB ID 2IC8). Active site dyad side chains are shown in black sticks. Black horizontal lines represent the lipid bilayer.

is situated on helix 6 forming the serine-histidine catalytic dyad in the center of the protein, providing the final proof of the widely accepted atypical catalytic dyad different from the common serine-histidine-aspartate catalytic triad found in their soluble counterparts. Structure-functions analysis highlighted different important parts of the structure. The amphiphilic loop 1 protrudes laterally forming a hairpin extension in the outer leaflet of the membrane (**Figure 1-7**). It contains the catalytically important WR motif, which is proposed to have a stabilizing structural role possibly orienting the enzyme in the lipid bilayer. Finally, loop 5 and helix 5 were highlighted as the most flexible part of the different structures and are proposed to take part to the substrate entry gate.

*1-5.2 Active site mechanism*

The active site mechanism of soluble serine proteases have been studied for decades and represent an instructive starting point for rhomboids. The available ecGlpG structures provided the proof of an active site embedded in and surrounded by the membrane: the uncommon catalytic dyad composed of S201 and H254 for ecGlpG (**Figure 1-8**). H254 is in good position to extract, *via* a hydrogen bond, a hydrogen atom from the serine residue during the nucleophile attack of the substrate. Water molecules were observed in the active site finally closing the controversy about their presence in the membrane environment during the hydrolysis process. The oxyanion hole of ecGlpG, which stabilizes the tetrahedral reaction intermediate, is mainly composed of the main chain amides of S201 and L200, the side chain amide of N154 and the side chain of H150[13] (**Figure 1-8**). Rhomboids provide an analogous mechanism to soluble serine peptidase (**Figure 1-3**) with two exceptions: the catalytic dyad active site (compared to the general triad) and the putative attack of the substrate on its si-face (compared to a re-face).
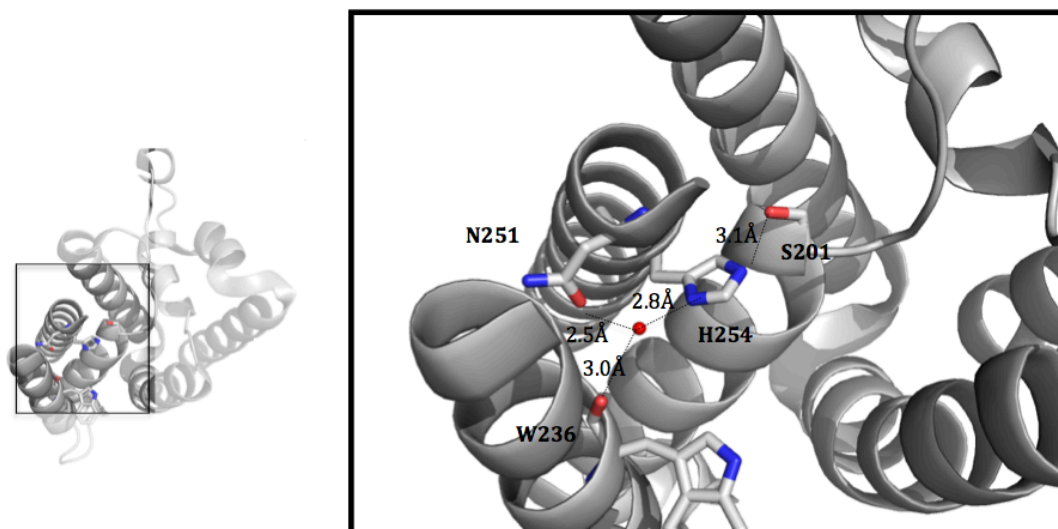
**Figure 1-8. Details of the active site of the x-ray crystal structure of the <u>membrane domain</u> of ecGlpG rhomboid protease**.
The rhomboid structure is shown (PDB 2IRV) in grey cartoon. Active site dyad (S201 and H254) side chains as well as N251 and W236 are shown in sticks. A water molecule is represented as a red sphere. The distance of the hydrogen bonds between residues and the water molecules are displayed. Here, H254 is in good position and through a hydrogen bond to extract an hydrogen from the serine residue during the nucleophile attack of the substrate.

Although the structures of the 6TMS minimal core of rhomboids represented an amazing input for these proteases, as well as for I-CLiPs, molecular mechanistic details were missing. These details can be generally provided by the structures of substrate-enzyme and/or inhibitor-enzyme complexes. Consequently, one part of my research thesis aimed to determine the inhibitor of hiGlpG and ecGlpG-FL and their optimal condition for co-crystallization or soaking technique (see chapter 2). However, during the period of my studies, three structures containing common serine peptidase inhibitors in the active site were solved revealing different details about rhomboid catalytic mechanism (see chapter 5 for discussion).

## 1-6      Structural features of the soluble domain(s) of rhomboids

Rhomboids are composed of a membrane domain and may possess a N- and/or C terminal domain(s). The structures of ecGlpG previously described were always restricted to its membrane domain. In fact, full-length ecGlpG (ecGlpG-FL) contains a large cytoplasmic domain and the structure was never solved due to its arduous purification. Furthermore, the removal of the N-terminal domain was presented by different groups as necessary to obtain diffracting crystals[41], as this domain was predicted to be unstructured[42] and subjected to proteolysis[41]. An NMR structure of a part of paGlpG (from *Pseudomonas aeruginosa*) N-terminal domain was obtained with a new a/b fold composed of a three-stranded anti-parallel b-sheet and two a-helices, shaped like a triangular wedge with a negatively charged face[46] (**Figure 1-9**). A systematic search of the protein databank (PDB) revealed a similar overall fold for the N-terminal domain of EscJ (from enteropathogenic *E. coli*), which is involved in type III secretion forming a key part of the oligomeric ring of the needle complex[47]. Also, the N-terminal domain of paGlpG was originally suggested to interact with the lipid bilayer[46], whereas an additional

study revealed the denaturing interaction between this domain and phosphocholine detergent micelles and suggested that the domain was interacting with its own membrane domain[48]. Consequently, the structural details of such extra-domains of rhomboids are of particular interest as they might give insights into their biological role and/or the one of the full-length protein. In addition, a subset of rhomboids possesses a N-terminal domain, which act on proteolytic activity: rhomboid-1[22], ecGlpG[40], paGlpG (from *Pseudomonas aeruginosa*)[48] and RHBDL2[49].

Some cytoplasmic domains of rhomboids have been annotated as zinc-fingers, EF hands, ubiquitin-binding domain and others. However, most of them are still uncharacterized, without any sequence homologies to provide any hints about their functions. Various studies focused then on the potential function and implication of cytoplasmic domains in rhomboid catalytic activity. Few functional roles of such cytoplasmic domains have been detailed. Thromobomodulin cleavage by RHBDL2, occurs after interacting with the N-terminal domain of the enzyme outside of the membrane[49]. Eukaryotic RHBDL4 is an ER-localized rhomboid acting in the ER-associated degradation pathway. Its C-terminal cytoplasmic domain contains a ubiquitin interaction motif that takes part in the recognition of ubiquitylated substrates with unstable transmembrane helices[26]. RHBDL2 and RHBDL4 studies suggested a substrate recognition function of their extra-membranous domains. Finally, N-terminal domains of two rhomboids from the protozoan parasite *Toxoplasma gondii* are essential for enzyme trafficking[1]. Regulation and oligomerization may also represent potential biological roles of such domains. A large number of both eukaryotic and prokaryotic rhomboids present a cytoplasmic domain at either the N- or C-terminal. Of particular interest, no structural information (including the structure of extra domains) is available for any human members of the rhomboid family, which
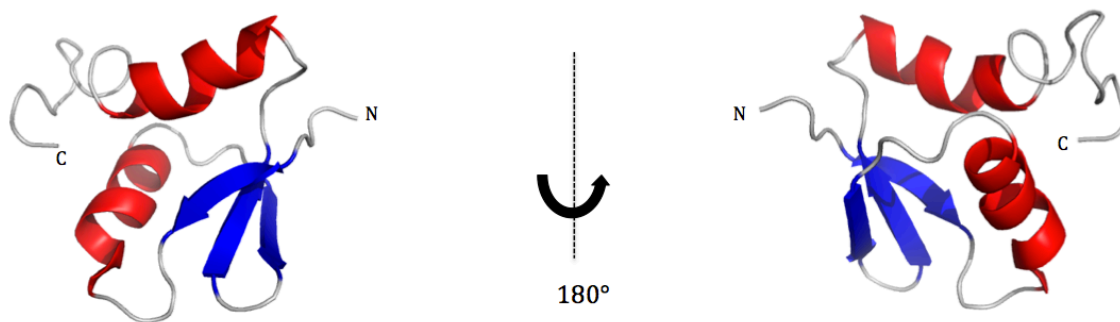
**Figure 1-9. NMR structure of a part of paGlpG N-terminal domain**
(PDB ID 2GQC [46]). The rhomboid structure is represented as cartoon. The two α-helices and the three-stranded antiparallel β-sheet are colored in red and blue, respectively.

have been directly linked to several diseases including cancer. However, contrary to prokaryotic rhomboids, these domains for eukaryotic counterparts are predicted to be unstructured[50]. Consequently, little is known about these rhomboids domains, and in particular about a high-resolution structure and its importance on rhomboid catalytic activity, which is one focus of my thesis. One important note is that an additional NMR structure of such domain was released during my research studies and is analyzed in Chapter 4.

## 1-7.    Models of substrate gating for rhomboid intramembrane peptidases

In the crystal structures of rhomboid membrane domains, the main structural differences were found in loop 5 and helix 5. A controversy remains regarding the substrate entry gate and the substrate access to the catalytic dyad. One hypothesis supports the movement of loop 5, needed to allow the substrate to enter the active site from the top of the protein, after binding an exosite potentially located between helix 5 and helix 2 (**Figure 1-7**). Other groups support the movement of the lateral helix 5, allowing a lateral entry of the substrate. A mutagenesis study enhancing displacement of helix 5 away from helix 2 in ecGlpG shows increased peptidase activity, while disulfide cross-linking of helix 5 with helix 2 results in a loss of substrate cleavage[40].

## 1-8.    Substrate specificity and inhibitors of rhomboids

One of the most important challenges in the field of intramembrane proteases remains the discovery of physiological substrates, elucidating their specificity and their mode of recognition. In fact, even if diverse cleavage activity assays have been performed, only few physiological substrates have been identified[15; 20] (only one in prokaryotes, TatA is cleaved by AarA *from*

24

*Providencia stuartii*). Contrary to the initial hypothesis that rhomboids only cleave monotopic type I membrane proteins, some single-spanning type II and/or polytopic membrane proteins were recently shown to be substrates or clients for rhomboid proteases and pseudoproteases, respectively[26; 27; 31; 51; 52; 53]. Various substrate recognition models have also been proposed. First, helix-destabilizing residues were suggested to be central in substrate proteolysis[54; 55]. Some consensus sequence of cleavage, at the top of the transmembrane segment, was discovered to be necessary and sufficient[3; 54; 55]. The specific requirements of substrate are controversial due to the lack of physiological substrates discovered. However, two complementary studies[3; 55] propose a negatively charged residue with a small side chain at position P1, a small side chain residue at position P1', a hydrophobic residue at position P2' and a large hydrophobic residue at position P4. An inhibitor bound structure confirmed the restriction on the size of P1[43]. For examples, ecGlpG and RHBDL2 were shown to cleave different consensus sequences of cleavage[55; 56; 57; 58] and more particularly, ecGlpG cleaves TatA[59] sequence with a high efficiency and RHBDL2 cleaves the anticoagulant cell-surface protein thrombomodulin[49]. These results seem to suggest that rhomboid may be exchanged to perform substrate cleavage, supporting the substrate sequence recognition model. A recent study challenges this hypothesis[60] suggesting that rhomboids identify substrates based on their intrinsic transmembrane dynamics. One can notice that even if a common sequence cleavage motif and the partially unfolding of the substrate region seem to be the most common hypotheses, recognition mode may vary between rhomboids. Differences could occur among rhomboid proteases, between rhomboid proteases and pseudoproteases and between rhomboids containing extra-membranous domains and others; these domains being implied in substrate recognition[26]. Of note, a rhomboid may also have multiple substrate requirements.

For soluble serine proteases, some inhibitors create acyl and tetrahedral intermediates. Among diverse results, two soluble serine peptidase inhibitors 3,4-dichloroisocoumarin (DCI) and N-p-tosyl-L-phenylalanine chloromethyl ketone (TPCK) were shown to inhibit some rhomboids and ecGlpG[22; 43; 57] in particular (even if some variability in the results were observed). Recently, a fluorescent peptide substrate based assay identified monocyclin b-lactams as a rhomboid inhibitor[61]. However, during the period of my studies, three structures containing 7-amino-4-chloro-3-methoxy-isocoumarin, Cbz-Ala$^P$(O-iPr)F (CAPF) and disopropylfluorophosphonate serine peptidase inhibitors in the active site were solved[43; 62; 63] (see chapter 5 for discussion).

Little is known about the substrates and inhibitors of rhomboids. For example, the physiological substrate of ecGlpG-FL, the most studied rhomboid protease, is still unknown which highly challenge functional assays of this protein. However, this information is relevant to design co-crystallization studies to reveal rhomboid molecular mechanism.

## 1-9.    Regulation of rhomboids

Uncontrolled protease activity is dangerous and could be linked to diseases. Until recently, rhomboids were the only I-CLiPs that do not require precleavage of the substrate[64]. However, some studies point out a possible precleavage regulation for PARL. PARL undergoes a, b and g-sites cleavages[65; 66; 67]. The a-sites cleavage corresponds to the N-terminal mitochondrial targeting sequence. b cleavage, between S77 and A78, is required for PARL activity[66; 67]. In contrast, g cleavage, between the two first TMS, was shown to eliminate it[65; 66]. Even if the understanding of rhomboid regulation is limited due to the lack of biological

examples, the most generally well supported hypotheses of rhomboid regulation are transcription or membrane trafficking, as for Rhomboid-1 (from *D. megalonaster).*

However, other hypotheses involved in enzymatic activity have emerged. For example, the lipidic environment influences ecGlpG-FL a lipid bilayer thinning is observed around the protein[57]. Oligomerization[64; 68; 69], iRhoms[64; 68; 69] and phosphorylation (phosporlation of PARL was shown to inhibit b cleavage and consequently the rhomboid activity[65]) could also play a role. Finally, the presence of cytoplasmic domain in a subset of rhomboids represent another valuable hypothesis for regulation[64; 68; 69]. Indeed, the cleavage of one of RHBDL2 substrate occurs after interacting with RHBDL2 N-terminal domain outside the membrane[49]. The characterization of the role played by the cytoplasmic domain of rhomboids can be critical for their regulation. Consequently, the soluble N-terminal domain, present in ecGlpG represents an interesting first step to study extra domains of rhomboids.

**1-10. Thesis Outline**

The primary goal of this thesis is to gain insights into molecular details of the catalytic mechanism, substrate gating and regulation of rhomboids, a family of proteins of great medical importance. More particularly, this work focuses on prokaryotic rhomboids, which represent an ideal basis to further study their eukaryotic counterparts.

Chapter 2 investigates two substrate entry gate models as well as the catalytic mechanism of rhomboids. Primarily based on a new partially disordered structure of hiGlpG, we proposed a model where both helix 5 and loop 5 are mobile and work jointly during substrate gating. Taking into account the access of the substrate to the active site, we discuss the reaction pathway. In

addition, a preliminary study of potential rhomboid inhibitors was performed as the first step to obtain enzyme:inhibitor complex structure on both ecGlpG-FL and hiGlpG.

In chapter 3 and 4, the focus shifts to the cytoplasmic domains of rhomboids. ecGlpG-FL was chosen as it is the best characterized rhomboid and contain two distinguished domains: the membrane domain (6 TMS) and a N-terminal cytoplasmic domain. This thesis attempts to gain structural details about this domain and investigate its putative regulatory role. Therefore, two strategies have been investigated: the direct one (crystallizing ecGlpG-FL) or the indirect one (crystallizing the cytoplasmic domain of ecGlpG-FL).

Chapter 3 examines the direct strategy focusing on obtaining the structure of the full-length protease. I managed to successfully over-express and purify the protein in order to obtain diffracting crystals. Diverse optimization methods including variations of detergents or other membrane-mimicking molecules as well as general methods such as removing flexible small part were tested. As no diffraction improvements were observed, more drastic strategies are envisioned.

Chapter 4 proposes to gain insights into the hypothetical role of this domain using an indirect strategy. We focused primarily on the structure of ecGlpG-cyto in order to gain information about its function and the one of ecGlpG-FL. The work showed that this domain undergoes an oligomerization upon domain swapping in solution. Also, steady-state kinetic assays showed no significant difference between ecGlpG membrane domain and ecGlpG-FL. We conclude that ecGlpG-FL cytoplasmic domain does not affect the structural integrity or the accessibility of the active site of the enzyme.

Chapter 5 will give a brief summary of the findings and discuss them regarding new inhibitor:enzyme complex structures. Future directions will be discussed.

## 1-11. References

1.    Sheiner, L., Dowse, T. J. & Soldati-Favre, D. (2008). Identification of trafficking determinants for polytopic rhomboid proteases in Toxoplasma gondii. *Traffic* **9**, 665-77.
2.    Cipolat, S., Rudka, T., Hartmann, D., Costa, V., Serneels, L., Craessaerts, K., Metzger, K., Frezza, C., Annaert, W., D'Adamio, L., Derks, C., Dejaegere, T., Pellegrini, L., D'Hooge, R., Scorrano, L. & De Strooper, B. (2006). Mitochondrial rhomboid PARL regulates cytochrome c release during apoptosis via OPA1-dependent cristae remodeling. *Cell* **126**, 163-75.
3.    Strisovsky, K., Sharpe, H. J. & Freeman, M. (2009). Sequence-specific intramembrane proteolysis: identification of a recognition motif in rhomboid substrates. *Mol Cell* **36**, 1048-59.
4.    Lee, J. R., Urban, S., Garvey, C. F. & Freeman, M. (2001). Regulated intracellular ligand transport and proteolysis control EGF signal activation in Drosophila. *Cell* **107**, 161-71.
5.    Cheng, T. L., Wu, Y. T., Lin, H. Y., Hsu, F. C., Liu, S. K., Chang, B. I., Chen, W. S., Lai, C. H., Shi, G. Y. & Wu, H. L. (2011). Functions of rhomboid family protease RHBDL2 and thrombomodulin in wound healing. *J Invest Dermatol* **131**, 2486-94.
6.    Hatunic, M., Stapleton, M., Hand, E., DeLong, C., Crowley, V. E. & Nolan, J. J. (2009). The Leu262Val polymorphism of presenilin associated rhomboid like protein (PARL) is associated with earlier onset of type 2 diabetes and increased urinary microalbumin creatinine ratio in an Irish case-control population. *Diabetes Res Clin Pract* **83**, 316-9.
7.    Walder, K., Kerr-Bayles, L., Civitarese, A., Jowett, J., Curran, J., Elliott, K., Trevaskis, J., Bishara, N., Zimmet, P., Mandarino, L., Ravussin, E., Blangero, J., Kissebah, A. & Collier, G. R. (2005). The mitochondrial rhomboid protease PSARL is a new candidate gene for type 2 diabetes. *Diabetologia* **48**, 459-68.
8.    Gottlieb, E. (2006). OPA1 and PARL keep a lid on apoptosis. *Cell* **126**, 27-9.
9.    Yan, Z., Zou, H., Tian, F., Grandis, J. R., Mixson, A. J., Lu, P. Y. & Li, L. Y. (2008). Human rhomboid family-1 gene silencing causes apoptosis or autophagy to epithelial cancer cells and inhibits xenograft tumor growth. *Mol Cancer Ther* **7**, 1355-64.
10.   Abba, M. C., Lacunza, E., Nunez, M. I., Colussi, A., Isla-Larrain, M., Segal-Eiras, A., Croce, M. V. & Aldaz, C. M. (2009). Rhomboid domain containing 2 (RHBDD2): a novel cancer-related gene over-expressed in breast cancer. *Biochim Biophys Acta* **1792**, 988-97.
11.   Zou, H., Thomas, S. M., Yan, Z. W., Grandis, J. R., Vogt, A. & Li, L. Y. (2009). Human rhomboid family-1 gene RHBDF1 participates in GPCR-mediated transactivation of EGFR growth signals in head and neck squamous cancer cells. *FASEB J* **23**, 425-32.
12.   Wolfe, M. S. (2009). Intramembrane proteolysis. *Chem Rev* **109**, 1599-612.
13.   Erez, E., Fass, D. & Bibi, E. (2009). How intramembrane proteases bury hydrolytic reactions in the membrane. *Nature* **459**, 371-8.
14.   Rawson, R. B., Zelenski, N. G., Nijhawan, D., Ye, J., Sakai, J., Hasan, M. T., Chang, T. Y., Brown, M. S. & Goldstein, J. L. (1997). Complementation cloning of S2P, a gene encoding a putative metalloprotease required for intramembrane cleavage of SREBPs. *Mol Cell* **1**, 47-57.
15.   Urban, S. (2009). Making the cut: central roles of intramembrane proteolysis in pathogenic microorganisms. *Nat Rev Microbiol* **7**, 411-23.

16. Ha, Y. (2007). Structural principles of intramembrane proteases. *Curr Opin Struct Biol* **17**, 405-11.

17. Urban, S. & Freeman, M. (2002). Intramembrane proteolysis controls diverse signalling pathways throughout evolution. *Curr Opin Genet Dev* **12**, 512-8.

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

19. Lemberg, M. K. (2013). Sampling the membrane: function of rhomboid-family proteins. *Trends Cell Biol* **23**, 210-7.

20. Urban, S. & Dickey, S. W. (2011). The rhomboid protease family: a decade of progress on function and mechanism. *Genome Biol* **12**, 231.

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

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

23. Adrain, C., Strisovsky, K., Zettl, M., Hu, L., Lemberg, M. K. & Freeman, M. (2011). Mammalian EGF receptor activation by the rhomboid protease RHBDL2. *EMBO Rep* **12**, 421-7.

24. Blaydon, D. C., Etheridge, S. L., Risk, J. M., Hennies, H. C., Gay, L. J., Carroll, R., Plagnol, V., McRonald, F. E., Stevens, H. P., Spurr, N. K., Bishop, D. T., Ellis, A., Jankowski, J., Field, J. K., Leigh, I. M., South, A. P. & Kelsell, D. P. (2012). RHBDF2 mutations are associated with tylosis, a familial esophageal cancer syndrome. *Am J Hum Genet* **90**, 340-6.

25. Wojnarowicz, P. M., Provencher, D. M., Mes-Masson, A. M. & Tonin, P. N. (2012). Chromosome 17q25 genes, RHBDF2 and CYGB, in ovarian cancer. *Int J Oncol* **40**, 1865-80.

26. Fleig, L., Bergbold, N., Sahasrabudhe, P., Geiger, B., Kaltak, L. & Lemberg, M. K. (2012). Ubiquitin-dependent intramembrane rhomboid protease promotes ERAD of membrane proteins. *Mol Cell* **47**, 558-69.

27. Ye, Y., Shibata, Y., Yun, C., Ron, D. & Rapoport, T. A. (2004). A membrane protein complex mediates retro-translocation from the ER lumen into the cytosol. *Nature* **429**, 841-7.

28. Christianson, J. C., Olzmann, J. A., Shaler, T. A., Sowa, M. E., Bennett, E. J., Richter, C. M., Tyler, R. E., Greenblatt, E. J., Harper, J. W. & Kopito, R. R. (2012). Defining human ERAD networks through an integrative mapping strategy. *Nat Cell Biol* **14**, 93-105.

29. Oda, Y., Okada, T., Yoshida, H., Kaufman, R. J., Nagata, K. & Mori, K. (2006). Derlin-2 and Derlin-3 are regulated by the mammalian unfolded protein response and are required for ER-associated degradation. *J Cell Biol* **172**, 383-93.

30. Carvalho, P., Goder, V. & Rapoport, T. A. (2006). Distinct ubiquitin-ligase complexes define convergent pathways for the degradation of ER proteins. *Cell* **126**, 361-73.

31. Younger, J. M., Chen, L., Ren, H. Y., Rosser, M. F., Turnbull, E. L., Fan, C. Y., Patterson, C. & Cyr, D. M. (2006). Sequential quality-control checkpoints triage misfolded cystic fibrosis transmembrane conductance regulator. *Cell* **126**, 571-82.

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

33. Chan, E. Y. & McQuibban, G. A. (2013). The mitochondrial rhomboid protease: its rise from obscurity to the pinnacle of disease-relevant genes. *Biochim Biophys Acta* **1828**, 2916-25.

34. Howell, S. A., Hackett, F., Jongco, A. M., Withers-Martinez, C., Kim, K., Carruthers, V. B. & Blackman, M. J. (2005). Distinct mechanisms govern proteolytic shedding of a key invasion protein in apicomplexan pathogens. *Mol Microbiol* **57**, 1342-56.

35. Brossier, F., Jewett, T. J., Sibley, L. D. & Urban, S. (2005). A spatially localized rhomboid protease cleaves cell surface adhesins essential for invasion by Toxoplasma. *Proc Natl Acad Sci U S A* **102**, 4146-51.

36. Sibley, L. D. (2013). The roles of intramembrane proteases in protozoan parasites. *Biochim Biophys Acta* **1828**, 2908-15.

37. Kateete, D. P., Katabazi, F. A., Okeng, A., Okee, M., Musinguzi, C., Asiimwe, B. B., Kyobe, S., Asiimwe, J., Boom, W. H. & Joloba, M. L. (2012). Rhomboids of Mycobacteria: characterization using an aarA mutant of Providencia stuartii and gene deletion in Mycobacterium smegmatis. *PLoS One* **7**, e45741.

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

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

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

41. Wu, Z., Yan, N., Feng, L., Oberstein, A., Yan, H., Baker, R. P., Gu, L., Jeffrey, P. D., Urban, S. & Shi, Y. (2006). Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry. *Nat Struct Mol Biol* **13**, 1084-91.

42. Ben-Shem, A., Fass, D. & Bibi, E. (2007). Structural basis for intramembrane proteolysis by rhomboid serine proteases. *Proc Natl Acad Sci U S A* **104**, 462-6.

43. Vinothkumar, K. R., Strisovsky, K., Andreeva, A., Christova, Y., Verhelst, S. & Freeman, M. (2010). The structural basis for catalysis and substrate specificity of a rhomboid protease. *EMBO J* **29**, 3797-809.

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

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

46. Del Rio, A., Dutta, K., Chavez, J., Ubarretxena-Belandia, I. & Ghose, R. (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**, 109-22.

47. Yip, C. K., Kimbrough, T. G., Felise, H. B., Vuckovic, M., Thomas, N. A., Pfuetzner, R. A., Frey, E. A., Finlay, B. B., Miller, S. I. & Strynadka, N. C. (2005). Structural characterization of the molecular platform for type III secretion system assembly. *Nature* **435**, 702-7.

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

49.    Lohi, O., Urban, S. & Freeman, M. (2004). Diverse substrate recognition mechanisms for rhomboids; thrombomodulin is cleaved by Mammalian rhomboids. *Curr Biol* **14**, 236-41.

50.    Ward, J. J., Sodhi, J. S., McGuffin, L. J., Buxton, B. F. & Jones, D. T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* **337**, 635-45.

51.    Wan, C., Fu, J., Wang, Y., Miao, S., Song, W. & Wang, L. (2012). Exosome-related multi-pass transmembrane protein TSAP6 is a target of rhomboid protease RHBDD1-induced proteolysis. *PLoS One* **7**, e37452.

52.    Tsruya, R., Wojtalla, A., Carmon, S., Yogev, S., Reich, A., Bibi, E., Merdes, G., Schejter, E. & Shilo, B. Z. (2007). Rhomboid cleaves Star to regulate the levels of secreted Spitz. *EMBO J* **26**, 1211-20.

53.    Erez, E. & Bibi, E. (2009). Cleavage of a multispanning membrane protein by an intramembrane serine protease. *Biochemistry* **48**, 12314-22.

54.    Urban, S. & Freeman, M. (2003). Substrate specificity of rhomboid intramembrane proteases is governed by helix-breaking residues in the substrate transmembrane domain. *Mol Cell* **11**, 1425-34.

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

56.    Pascall, J. C. & Brown, K. D. (2004). Intramembrane cleavage of ephrinB3 by the human rhomboid family protease, RHBDL2. *Biochem Biophys Res Commun* **317**, 244-52.

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

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

59.    Stevenson, L. G., Strisovsky, K., Clemmer, K. M., Bhatt, S., Freeman, M. & Rather, P. N. (2007). Rhomboid protease AarA mediates quorum-sensing in Providencia stuartii by activating TatA of the twin-arginine translocase. *Proc Natl Acad Sci U S A* **104**, 1003-8.

60.    Moin, S. M. & Urban, S. (2012). Membrane immersion allows rhomboid proteases to achieve specificity by reading transmembrane segment dynamics. *Elife (Cambridge)* **1**, e00173.

61.    Pierrat, O. A., Strisovsky, K., Christova, Y., Large, J., Ansell, K., Bouloc, N., Smiljanic, E. & Freeman, M. (2011). Monocyclic beta-lactams are selective, mechanism-based inhibitors of rhomboid intramembrane proteases. *ACS Chem Biol* **6**, 325-35.

62.    Xue, Y. & Ha, Y. (2012). Catalytic mechanism of rhomboid protease GlpG probed by 3,4-dichloroisocoumarin and diisopropyl fluorophosphonate. *J Biol Chem* **287**, 3099-107.

63.    Xue, Y., Chowdhury, S., Liu, X., Akiyama, Y., Ellman, J. & Ha, Y. (2012). Conformational change in rhomboid protease GlpG induced by inhibitor binding to its S' subsites. *Biochemistry* **51**, 3723-31.

64.    Urban, S. (2010). Taking the plunge: integrating structural, enzymatic and computational insights into a unified model for membrane-immersed rhomboid proteolysis. *Biochem J* **425**, 501-12.

65.    Jeyaraju, D. V., McBride, H. M., Hill, R. B. & Pellegrini, L. (2011). Structural and mechanistic basis of Parl activity and regulation. *Cell Death Differ* **18**, 1531-9.

66. Jeyaraju, D. V., Xu, L., Letellier, M. C., Bandaru, S., Zunino, R., Berg, E. A., McBride, H. M. & Pellegrini, L. (2006). Phosphorylation and cleavage of presenilin-associated rhomboid-like protein (PARL) promotes changes in mitochondrial morphology. *Proc Natl Acad Sci U S A* **103**, 18562-7.

67. Sik, A., Passer, B. J., Koonin, E. V. & Pellegrini, L. (2004). Self-regulated cleavage of the mitochondrial intramembrane-cleaving protease PARL yields Pbeta, a nuclear-targeted peptide. *J Biol Chem* **279**, 15323-9.

68. Freeman, M. (2009). Rhomboids: 7 years of a new protease family. *Semin Cell Dev Biol* **20**, 231-9.

69. Lemberg, M. K. & Freeman, M. (2007). Cutting proteins within lipid bilayers: rhomboid structure and mechanism. *Mol Cell* **28**, 930-40.

# Chapter 2

## Insights into substrate gating and active site mechanism in *H. influenzae* and *E. coli* Rhomboid

**Cory L. Brooks, Christelle Lazareno-Saez, Jason S. Lamoureux, Michelle W. Mak and M. Joanne Lemieux.**

**2-1. Introduction**

Rhomboids are intramembrane proteases (also termed peptidases[1]) that cleave a predicted transmembrane segment of peptide substrates.[2] Rhomboids were first discovered in Drosophila melanogaster, where they were shown to affect cell fate [3,4] and were later identified as serine peptidases.[2] As a new class of enzymes with critical roles in cell metabolism and signaling, the first identification of a membrane-embedded serine peptidase was of great importance.[2] Rhomboids have been identified in various membrane compartments throughout the cell. Their primary function is to cleave peptides, thereby releasing portions of the substrates from the membrane to participate in cellular signaling events. Diverse roles of the different members of the rhomboid peptidase family have been revealed through a combination of genetic screens, developmental biology, and cell biology studies.[5] Although rhomboid peptidases are ubiquitous membrane proteins that are conserved in all kingdoms of life, [6,7] the cellular function of only a few rhomboids has been elucidated. Rhomboid function has been linked to quorum sensing,[8] parasite invasion, [9-11] cell fate determination, [12-14] apoptosis, [15-16] type 2 diabetes, [17-19] Parkinson's disease,[20] blindness,[21] and epithelial cancers. [22,23]

The structure of the *Haemophilus influenzae* rhomboid hiGlpG,[24] as well as that of the *Escherichia coli* rhomboid ecGlpG, [25-28] revealed a common architecture for the basic secretase type of rhomboid having six transmembrane segments in a helical bundle. A striking feature revealed by these structures was a catalytic dyad in the active site, which was located approximately 10 Å below the periplasmic membrane surface. This active-site geometry was quite different from the triad observed with traditional soluble serine proteases such as trypsin.[29] From our initial hiGlpG structure, we presented a preliminary enzymatic reaction mechanism based on other soluble serine peptidase structures.[24] However, to comprehend the catalytic

mechanism, we need to answer the question: How do hydrophobic substrates gain access to the active site buried within the lipid bilayer? In the initial crystal structures, the substrate binding site was inaccessible from the lipid bilayer, yet an open cavity exposes the hydrated active site. A crystal structure showing disorder in loop 5 suggested that this was the mobile gate for substrate entry.[25] Mutagenesis experiments implied that helix 5 was the mobile region in substrate gating.[30] A recent inhibitor co-crystal structure of the *E. coli* rhomboid ecGlpG demonstrated that, indeed, both loop 5 and helix 5 are mobile.[31] Compared to the native *E. coli* rhomboid structure, in the presence of an inhibitor, large movements were observed in loop 5 (the cap), with smaller movements seen in helix 5 and loop 4. It is unclear whether the same degree of movement is required to permit substrate access to this buried active site. A close examination of the *E. coli* and *H. influenzae* rhomboid structures reveals important differences in the intramolecular connections between helix 5 and helix 2 of the central helical bundle (**Figure 2-1**). In *E. coli*, helix 5 is parallel with helix 2, while in *H. influenzae*, helix 5 is partially unwound and titled in comparison. In addition, there are key differences seen with loops 4 and 5. We seek to understand whether similar structural changes are required for the *H. influenzae* rhomboid hiGlpG to cleave substrates, as was observed with the *E. coli* rhomboid ecGlpG.

A comparison between the ecGlpG structure and the hiGlpG structure reveals key differences in the proposed gate region. In our first hiGlpG structure, the region proposed to be flexible in substrate gating (helix 5) is partially unwound; furthermore, B-factors for this region are in excess of 100 $\text{Å}^2$. Here we present a new structure of hiGlpG showing that loop 4, helix 5, and loop 5 are disordered, suggesting flexibility in both helix 5 and flanking regions. To test whether this flexibility is important for substrate binding and cleavage, we carried out amino acid

36
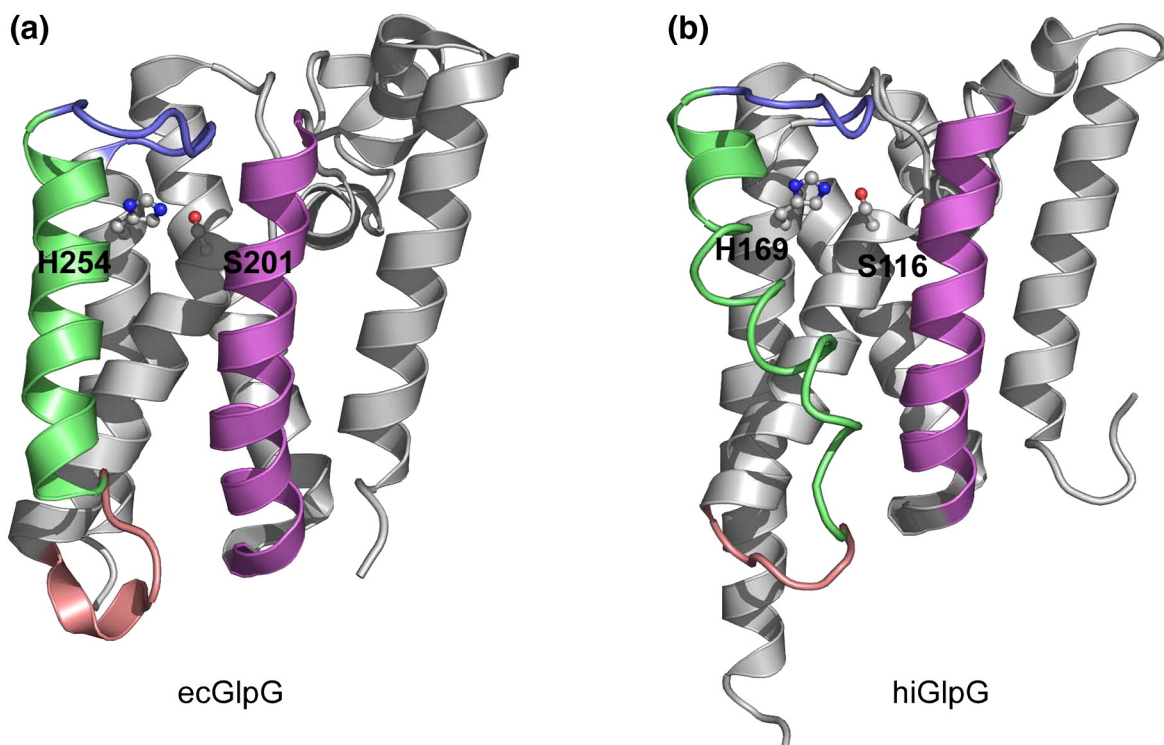
**Figure 2-1**. **A comparison of the overall structures of ecGlpG and hiGlpG**. Cartoon representation of differences in the proposed gate region of helices 5 and 2 between (a) hiGlpG (PDB code 2NR9) and (b) ecGlpG (PDB code 2IC8). Helix 2 (H2) and helix 5 (H5) are shown in purple and green, respectively. Catalytic serine and histidine residues are also labeled.

substitutions combined with functional assays. Substitutions in both helix 5 and loop 5 were tolerated, suggesting movements in these regions during substrate association and cleavage. In contrast, mutations in loop 4 are not tolerated, suggesting smaller movements in this region. These data imply that movements required for substrate access to the buried active site may be different from those observed with inhibitor binding.

To complement this analysis, a preliminary functional screening for GlpGs inhibtors was conducted to further obtain structural information on protein:inhibitor complexes. Valuable structural information would arise from such structure and would help the understanding of catalytic mechanism.


### 2-2. Results

*2-2.1. Helix 5 and flanking loops are conformationally flexible in H. influenzae rhomboid*

The overall architectures for both hiGlpG structure and ecGlpG structure are similar, consisting of six helical bundles (**Figure 2-1**). A major difference between the structures is observed in the regions proposed to act as a gate, allowing substrate access to the active site (**Figure 2-1**). In hiGlpG, helix 5 is partially unwound and tilted away from the helical bundle, while in ecGlpG, helix 5 is parallel with the helical bundle (**Figure 2-1**). Given the differences in helix 5 orientation and structure between the two rhomboids, a question arises: Does helix 5 in hiGlpG act as a lateral gate analogously to that observed in *E. coli* rhomboid? We identified hiGlpG crystals grown under identical conditions compared to our original structure that showed differences in the electron density surrounding helix 5—the region proposed to play a role in substrate gating (**Table 2-1**). Molecular replacement with our original hiGlpG model [Protein Data Bank (PDB) code 2NR9] revealed a $2F_o - F_c$ map with a weak electron density for a large

**Table 2-1. Data collection and refinement statistics for the hiGlpG structure (PDB code 3ODJ)**

| | |
|---|---|
| *Data collection* | |
| Space group | C2 |
| Unit cell dimensions | $a = 120.9$ Å, $b = 35.2$ Å, $c = 52.7$ Å, $\beta = 104.1°$ |
| Z | 1 |
| Wavelength (Å) | 1.1158 |
| Resolution range (Å)[a] | 25–2.85 (2.92–2.85) |
| Total number of reflections | 10,526 |
| Unique reflections | 4813 |
| $\langle I/\sigma I \rangle$[b] | 11.7 (2.6) |
| Completeness (%) | 96.8 (99.2) |
| Redundancy | 2.1 |
| $R_{\text{merge}}$[c] | 0.047 (0.29) |
| | |
| *Refinement* | |
| $R_{\text{work}}$[d] | 0.241 |
| $R_{\text{free}}$[e] | 0.294 |
| Number of atoms | 1279 |
| Solvent atoms | 18 |
| Deviations from ideal (RMSD) | |
|   Bond lengths (Å) | 0.015 |
|   Bond angles (°) | 1.593 |
| Ramachandran plot | |
|   Most favored (%) | 92.0 |
|   Allowed (%) | 7.2 |
|   Generously allowed (%) | 0 |
|   Disallowed (%) | 0.7 |

[a] Numbers in parentheses indicate values for the high-resolution shell.

[b] $\langle I/\sigma I \rangle$ is the ratio of mean intensity to the mean error of the intensity.

[c] $R_{\text{merge}} = \sum_{hkl}\sum_j |I_{hkl,\ j} - \langle I_{hkl}\rangle| / \sum_{hkl}\sum_j I_{hkl,\ j}$, where $\langle I_{hkl}\rangle$ is the average of the symmetry-related observations of a unique reflection.

[d] $R_{\text{work}} = \sum_{hkl}||F_{\text{obs}}(hkl)| - F_{\text{calc}}(hkl)||/\sum_{hkl}|F_{\text{obs}}(hkl)|$, where $F_{\text{obs}}$ and $F_{\text{calc}}$ are the observed and calculated structure factors, respectively.

[e] $R_{\text{free}}$ is calculated in the same manner as $R_{\text{work}}$ on 5% of structure factors that were not used in the model refinement.

region of the protein (**Figure 2-2A**). In order to remove bias from the model, we repeated the

molecular replacement with the original hiGlpG model (PDB code 2NR9) lacking residues 134–

164. The disordered residue range 134–164 corresponds to residues in loop 4, helix 5, and loop 5

(**Figure 2-3A and B**). Disorder was further evidenced by an examination of B-factor plots and

real-space correlation plots for the initial molecular replacement model  (**Figure 2-2B and C**).

We found six different data sets collected around 2.8 Å that lacked density in this region between

residues 134 and 164 compared to the original hiGlpG structure. This suggests that lack of

electron density was not due to lattice disorder but was rather a feature of the protein.


*2-2.2. Flexibility required for substrate cleavage in H. Influenzae rhomboid*

The new structure of hiGlpG clearly reveals disorder in the regions of loop 4, helix 5, and loop 5;

however, a question remains: Are they all involved in substrate gating? To address this question,

we generated point mutations of residues predicted to weaken hydrophobic interactions in loop 4,

in loop 5, or between helix 5 and helix 2 (**Figure 2-3C**). Disruption of hydrophobic interactions

between the protein core and the putative substrate gate would result in increased proteolytic

activity if that region is involved in substrate gating. The endogenous substrate for hiGlpG has

not been identified; therefore, a C100 functional assay[33] incorporating a rhomboid substrate

chimera was used. This substrate chimera consists of the C-terminal 100 residues of amyloid

precursor protein (APP) in which seven residues are mutated to introduce the rhomboid cleavage

motif of TatA from Providencia stuartii[34] (**Figure 2-4A**). This is similar to the C100Spitz Flag

assay developed by Baker et al.,[30]Li et al.,[33] and Urban and Wolfe.[35] Replacement of the

Drosophila Spitz substrate motif in the APP background with the cleavage motif of TatA from *P.*

*stuartii* results in a more efficient substrate cleavage by hiGlpG.

**(a)**

**(b)**

**(c)**
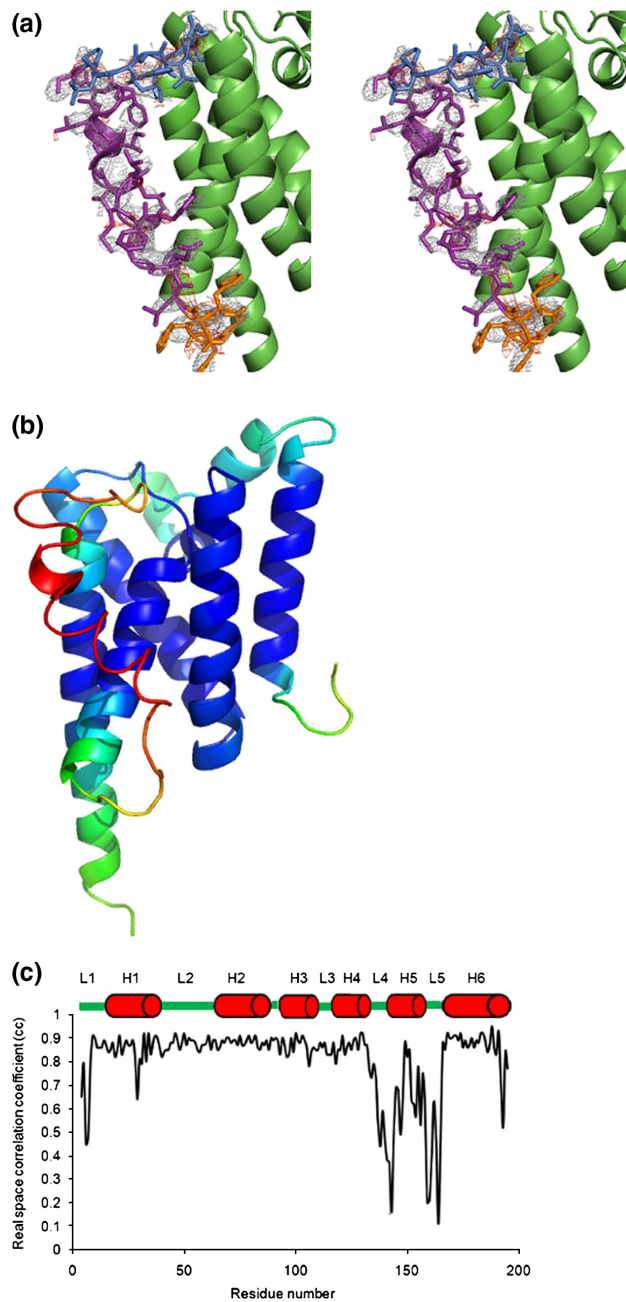
**Figure 2-2. Disorder in loop5 (L5), helix 5 (H5), and loop 4 (L4) for hiGlpG crystals.** (a) A stereo view of an electron density map ($2F_o - F_c$ contoured to $1\sigma$) surrounding regions L5 (blue), H5 (purple), and L4 (orange) after initial molecular replacement and refinement reveals a very weak electron density. (b) An initial model of hiGlpG following molecular replacement, colored according to B-factor (increasing relative B-factors; blue, teal, green, orange, and red), reveals a high degree of thermal motion in L5, H5, and L4. The B-factor range in PDB code 2NR9 is 26.2–115.6 $Å^2$. (c) Real-space correlation plot of the initial molecular replacement model reveals disorder in the regions of L5, H5, and L4.
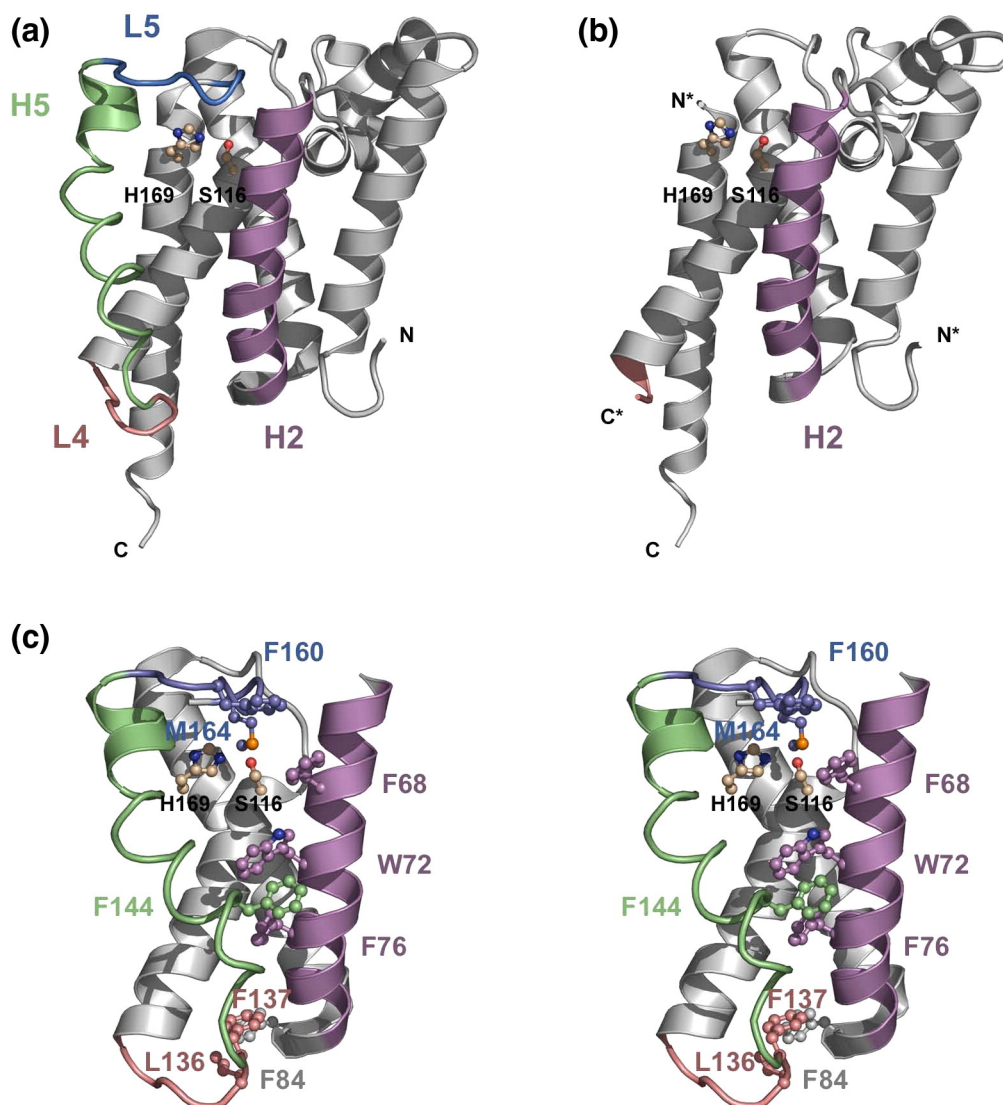
**Figure 2-3. A new hiGlpG structure shows disorder in loop 4, helix 5, and loop 5.** (a) Cartoon representation of the original hiGlpG, with loop 4 (L4) and loop 5 (L5) shown in pink and blue, respectively. This structure has been rotated slightly compared to **Figure 2-1a** in order to show the separation between helix 5 and helix 2. (b) Cartoon representation of a new structure for hiGlpG in which density is missing for L4, H5, and L5. Asterisks represent the new N-terminus and C-terminus. (c) Residues shown in segments L4, H5, and L5 are investigated for their involvement in substrate gating. A stereo cartoon representation of residues mutated in hiGlpG is presented to demonstrate potential connections in hiGlpG segments that may affect activity. Residues in three segments of the structure have been analyzed: loop 4 (L4; pink), helix 2 (H2; purple), and helix 5 (H5; green) interactions, and loop 5 (L5; blue). F84 is located in loop 2 at the base of helix 2. Catalytic residues are labeled in black. For clarity, helices 1 and 6 have been removed.

Single amino acid substitutions predicted to weaken interactions between the putative substrate gate and the protein helical bundle core in hiGlpG (**Figure 2-3C**) were tested for their ability to cleave substrates *in vitro* (**Figure 2-4B**). Quantification of cleavage was carried out by densitometry (**Table 2-2A**). In helix 2, an F68A substitution had no effect on activity, while W72A was not expressed and activity could not be assessed. F76A substitution, which participates in hydrophobic interactions with residues in helices 3 and 4, as well as with F144 from helix 5, resulted in a 95% decrease in activity compared with the wild-type cleavage by hiGlpG. F144A substitution in helix 5 had a less dramatic decrease in activity, with a 40% reduction in peptide cleavage. We observed a dramatic loss of activity with mutation F84A on helix 2, and with mutations L136A and F137A in loop 4, with only 5%, 22%, and 14% of wild-type activity remaining. In contrast, alanine substitution of residues F160 and M164 found in loop 5 retained only 54% and 40% of peptidase activity compared to wild type. In summary, the substitutions in loop 4 had the most dramatic decrease in hiGlpG activity.

In order to disrupt the specific residue partners predicted to interact between helix 2 and helix 5, we generated two groups of triple mutants that exchange paired hydrophobic groups for alanines (W72A, F76A, and F144A) and valines (W72V, F76V, and F144V) (**Figure 2-4C**). Quantification of cleavage was carried out by densitometry (**Table2-2B**). Our *in vitro* proteolysis assay shows an increase in substrate cleavage compared to wild type when these hydrophobic pairs are disrupted by the introduction of either an alanine or a valine. Alanine substitutions resulted in a 2.5-fold increase in activity, while valine substitutions resulted in a 2-fold increase in proteolytic activity relative to wild type. It is interesting to note that the single W72A could not be expressed in *E. coli*, while the triple mutant was easily expressed and purified. Lack of
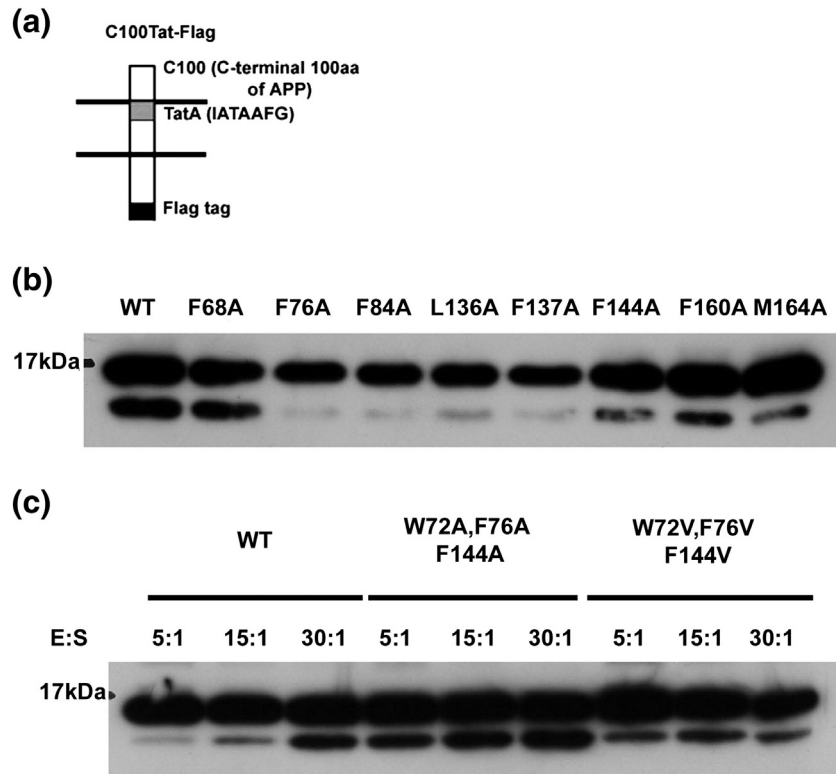
**Figure 2-4. Amino acid substitutions in the disordered regions of hiGlpG affect substrate cleavage efficiency.** (a) A symbolic representation of the 17-kDa substrate C100Tat-Flag, which is a chimera of the C-terminal 100 residues of APP, with seven residues of the P. stuartii TatA cleavage site substituted at the N-terminus. A Flag tag is used for Western blot visualization. (b) Single alanine substitutions were carried out on residues in the three disordered regions loop 4 (L4), helix 5 (H5), and loop 5 (L5), as well as in helix 2 (H2). An SDS-PAGE functional analysis of hiGlpG alanine mutations is analyzed for their ability to increase or decrease peptidase ability to cleave the C100Tat-Flag substrate. The Western blot was probed with anti-Flag antibody. (c) Triple amino acid substitutions breaking the hydrophobic linkage between H2 and H5 result in increased substrate cleavage. Triple alanine and valine substitutions in hiGlpG were generated and assayed for their ability to cleave the C100Tat-Flag substrate. Alanine and valine substitutions were carried out, with residues W72, F76, and F144 found on helices H2 and H5. Reactions with enzyme titrations and constant substrate concentrations were carried out to ensure conclusive results. The cleavage of the substrate C100TatA-Flag in the presence of different hiGlpG mutants was separated by SDS-PAGE. The Western blot was probed with anti-Flag antibody.

(a) Relative proteolytic activity of the single amino acid substitutions of residues in helix 2 (H2), loop 4 (L4), helix 5 (H5), and loop 5 (L5) of hiGlpG, calculated from Figure 2-4 b

| Mutation | Structural position in *hi*GlpG | Relative activity (%)[a] |
|---|---|---|
| Wild type | NA | 100 |
| F68A | H2 | 98 |
| F76A | H2 | 5 |
| F84A | L4 | 5 |
| L136A | L4 | 22 |
| F137A | L4 | 14 |
| F144A | H5 | 58 |
| F160A | L5 | 54 |
| M164A | L5 | 40 |

(b) Relative proteolytic activity of triple amino acid substitutions, calculated from Figure 2-4 c

| Mutation | Relative activity (%)[b] | | |
|---|---|---|---|
| | Enzyme/substrate ratio (wt/wt) | | |
| | 5:1 | 15:1 | 30:1 |
| Wild type | 43 | 85 | 100 |
| W72A+F76A+F144A | 166 | 247 | 263 |
| W72V+F76V+F144V | 134 | 193 | 223 |

NA, not applicable.
[a] Activity was normalized, with wild type being 100%.
[b] Activity was normalized to wild type using a 30:1 enzyme/substrate ratio.

**Table 2-2. Quantification of substrate cleavage by hiGlpG**

expression of the single substitution W72A suggests protein misfolding and possibly a loss of helix stabilization, since the triple mutant was expressed.

*2-2.3. Testing the different serine peptidase inhibitors in hi- and ec-GlpGs*

In order to gain insights into the catalytic mechanism of intra membrane proteases, the structural capture of intermediate states is a necessity. To reach this goal, a functional study was started using several already known and well-characterized inhibitors of soluble proteases. A total of four different compounds were tested: potassium clavulanate (KC - a β-lactam) and 3,4-dichloroisocoumarin (DCI – a coumarin) which are supposed to form acyl intermediates when bound to the protease, as well as phenylmethanesulfonyl fluoride (PMSF - a sulfonyl fluoride) and N-*p*-Tosyl-L-phenylalanine (TPCK – a chloromethyl ketone) which potentially form tetrahedral intermediates (**Figure 2-5a**). As these inhibitors tended to precipitate, a quick screen of solvent conditions and detergent concentrations was performed to ensure their complete solubilization at their usual working concentrations in the conditions used during the experiment.

The respective potency of these four compounds to inhibit rhomboids from *H. influenza* and *E. coli* was tested *in-vitro* using the same functional assay used for hiGlpG mutants. Regarding hiGlpG, the four compounds were first tested at a single concentration (1 mM – **Figure 2-5b**) and TPCK and DCI showed significant inhibitor effect. In a second experiment, three different concentrations of TPCK (100, 500 and 1000 µM – **Figure 2-5c**) were tested to further evaluate its potency. These results were used to validate the inhibitory effect of these compounds to further try to get a crystal structure of hiGlpG complexed to them, and also roughly estimate the working concentration of inhibitor to be used for the crystallization trials.
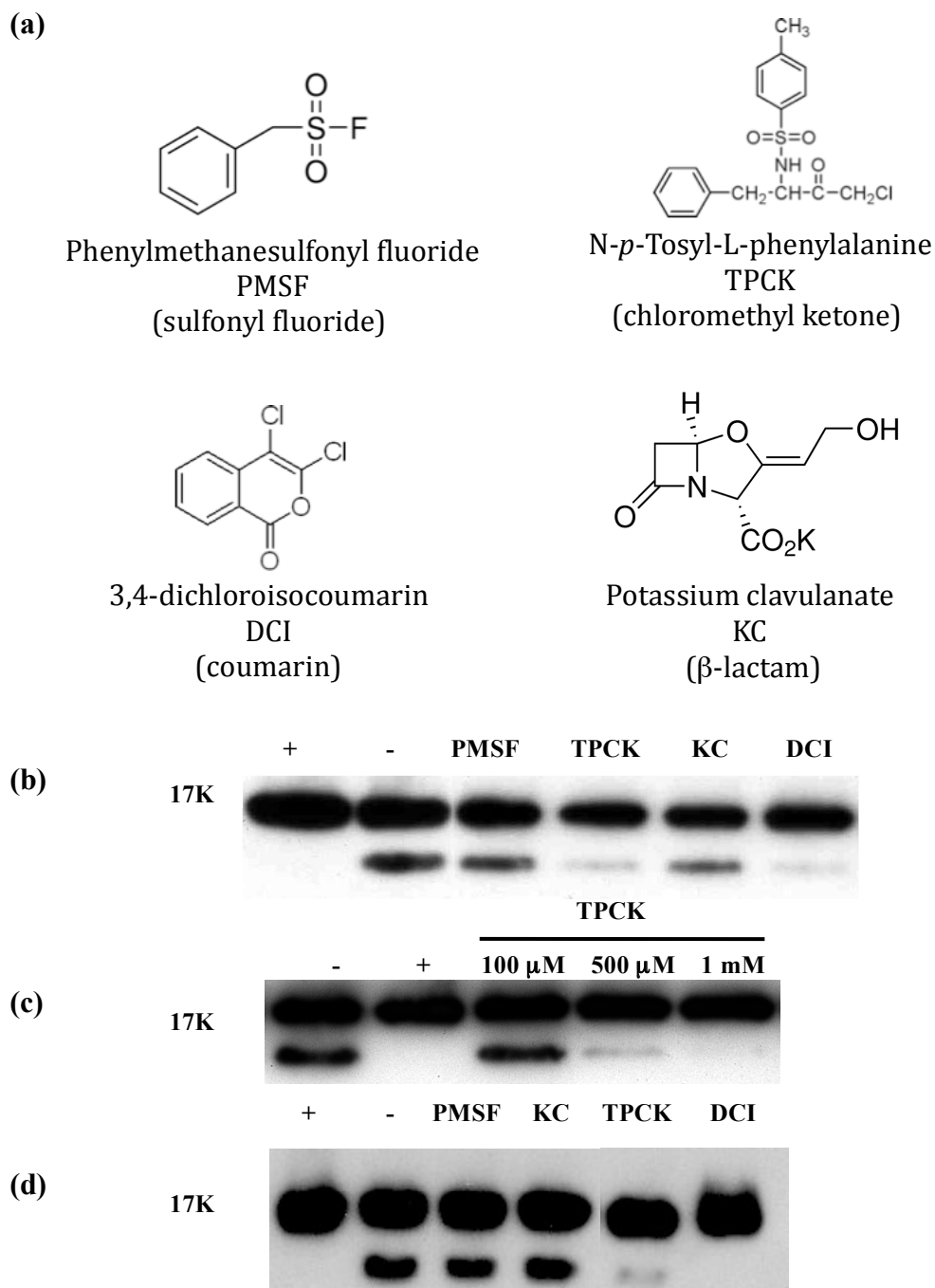
**Figure 2-5. Effect of serine protease inhibitors on hi- and ec-GlpGs proteolytic activity. (a)** Molecular structure of the different inhibitors tested. **(b)** Inhibitory effect of the four compounds on hiGlpG function at 1 mM final concentration. **(c)** Titration of TPCK inhibition of GlpG **(d)** Inhibitory effect of the four compounds on ecGlpG activity. PMSF, Concentrations of PMSF, KC, TPCK and DCI are 2.5, 0.133, 1 and 10 mM, respectively. In all experiments, (-): negative control (no inhibitor) and (+) positive control (no GlpG).

For *ec*GlpG, single concentrations of inhibitors were tested and only TPCK and DCI showed inhibitory effect (**Figure 2-5d**). Following these results, crystallization trials were started to obtain a crystal structure of hiGlpG in complex with TPCK or DCI. Two strategies were conducted in parallel: i) growing apo-crystal of hiGlpG to further soak them in a mother liquor solution complemented with TPCK or DCI or ii) co-crystallization of hiGlpG with the inhibitor (TPCK or DCI). Unfortunately, no protein crystal was obtained in any of the conditions tested.

## 2-3. Discussion

### 2-3.1. Substrate gating by helix 5 and loop 5 movements

Two prevalent hypotheses explain how substrates gain access to the active site of rhomboids. In the first hypothesis, Wang and Ha proposed that movements in the cap segment (termed loop 5 in this publication) result in exposure of the substrate to active-site residues.[25] Their *E. coli* rhomboid (ecGlpG) structure revealed that loop 5 (the cap) on top of helix 5 is flexible in the crystal structure. The movement of loop 5 would involve the substrate's recognition motif exiting the lipid bilayer. Indeed, molecular dynamics simulations have shown that *E. coli* rhomboid protease can induce thinning and deformations of lipid bilayers.[36] Additional studies revealed that the substrate recognition motif need not be embedded in the bilayer for cleavage to occur.[34,37] In an alternative model, Baker et al.[30] and Urban and Baker[38] postulated that a movement of helix 5 away from the rhomboid central helical bundle permits access of the substrate to the active-site residues. In agreement with this model, amino acid substitutions that displace helix 5 away from helix 2 in ecGlpG show increased peptidase activity, while disulfide cross-linking of helix 5 with helix 2 results in a loss of substrate cleavage.

Our new hiGlpG structure (**Figure 2-3B**) clearly reveals disorder in the region of the putative substrate gate of helix 5, loop 5, and loop 4, as evidenced by a lack of electron density in this region, B-factor analysis, and real-space correlation plots of this region in the original molecular replacement solution (**Figure 2-2**). In addition to the structure reported here, six other data sets from crystals that diffracted to less than 2.8 Å also showed disorder in this region (data not shown). Numerous data sets (approximately 150) from hiGlpG were collected while searching for covalent inhibitors and even heavy-atom derivatives. It is interesting to note that most crystals diffracted to a resolution similar to our initial data sets, with the average diffraction limit around 2.8–3.0 Å. In contrast to that reported for ecGlpG, crystals in different space groups that diffracted well were not readily found. Furthermore, hiGlpG crystals do not diffract to the same level as that observed with ecGlpG (1.9–2.1 Å). [25-28] This may be due to the flexibility in helix 5 observed with hiGlpG. In our original hiGlpG structure, the B-factors are indeed high in the helix 5 region, yet these crystals diffracted to 2.2 Å (PDB code 2NR9) (**Figure 2-2**).

In order to examine if indeed all three regions are involved in substrate gating, we used mutagenesis to disrupt the hydrophobic interactions anchoring the substrate gate to the core of the protein. A wild type or an enhanced substrate cleavage would indicate that these intramolecular interactions are not essential for substrate cleavage and furthermore would be suggestive that these elements are acting as a substrate gate.

In helices 2 and 5 of hiGlpG, triple mutations replacing key hydrophobic residues linking these helices clearly show a 2-fold increase in substrate cleavage (**Figure 2-4C**, **Table 2-2B**). The decrease in substrate cleavage observed with the single amino acid substitutions of residues F76 and F144 from helices 2 and 5 (**Figure 2-4B**, **Table 2-2A**) suggests that these regions may be intact during the closed state and not always flexible. As expected, F68A exhibited cleavage

similar to that of wild type, as this residue, found on helix 2, does not participate in hydrophobic interactions with residues on helix 5 (**Figure 2-3B**). Taken together, the observed increase in activity with the triple mutants, combined with the disorder in the crystal structure, strongly suggests that helix 5 is flexible and involved in substrate gating.

For these triple-mutant experiments, titrations of substrates were used to demonstrate cleavage efficiencies. Cleavage was monitored with *in vitro* cleavage using an artificial substrate, the C100TatA chimera (**Figure 2-4A**). It should be noted here that a 30-fold molar excess of enzyme is required to see a high degree of cleavage of the C100Tat-Flag. Even at this ratio, not all of the substrate is cleaved. The detergent-solubilized state of the membrane substrate, combined with the high degree of substrate oligomerization typically observed with APP, is most likely responsible for the elevated enzyme/substrate ratio required for cleavage. This phenomenon has been previously observed in other functional assays utilizing the C100Spitz Flag chimera substrate.[35] Therefore, any dramatic increase in the substrate cleavage observed is even more striking.

Single amino acid substitutions to the loop 5 "cap" region show a weak loss of activity (**Figure 2-4B**, **Table 2-2A**). This suggests that the intramolecular interactions of residues in this loop with those in the active site are not essential during substrate cleavage, supporting the premise that this area is most likely mobile during substrate gating in hiGlpG. The cap therefore does not hinder access to the catalytic residues.

Despite loop 4 being disordered in the crystal structure, alanine substitutions in this region exhibited a dramatic decrease in proteolytic activity (**Figure 2-4B**, **Table 2-2A**). This suggests that the intramolecular interactions in loop 4 remain intact, and that the large flexibility that we

observed in our new structure does not occur during substrate gating. Loop 4 may play an indirect role in enzyme catalysis. Our mutagenesis establishes its importance for substrate cleavage efficiency. Therefore, movements may occur, but intramolecular contacts may be intact during substrate cleavage. We propose that, in hiGlpG, both helix 5 and loop 5 work jointly in substrate gating, indicating that both the predicted model of substrate gating and the predicted model of domain movement, either by loop 5 (cap)[25] or by helix 5, [30, 38] are partially correct. Although similar, the crystal structure of the coumarin-bound ecGlpG shows slight movements in loop 4 and helix 5, with large movements only observed in loop 5.[31] Taken together, these data suggest that hiGlpG and ecGlpG may have comparable gating mechanisms with subtle differences. Since no crystal structure of an enzyme/substrate complex has been determined and, for that matter, the native substrates of either ecGlpG or hiGlpG have yet to be identified, docking simulations based on "lock-and-key" substrate recognition are problematic. Furthermore, both regions may not contact the substrate simultaneously. Strisovsky et al. described a model whereby the binding of the substrate's transmembrane segment to an exosite on rhomboid in the lipid bilayer precedes the binding to the active site.[34] Indeed, the hydrophobic residues exposed during the opening of helix 5 would facilitate this interaction and provide an exosite. This is further supported by the observation that mutations on loop 4 disrupt function. Disruption of this region may affect the angle at which the substrate is presented to the active site of the enzyme. In addition, various substrates may be identified for single rhomboid proteases, which may require different kinetics for cleavage. In fact, cleavage of the multispanning membrane proteins LacY[39] and MdfA[40] has been demonstrated with rhomboids, although they have not been shown to occur physiologically. Features in the substrates may indeed facilitate the method by which substrate gating and cleavage ensue. The fact that these movements are

observed with hiGlpG, a homolog of ecGlpG, indicates that this is a common feature of the rhomboid family.

*2-3.2. New rhomboid reaction pathway*

The structural evidence for substrate gating via helix 5 adds a new twist to the prediction for the catalytic mechanism carried out by rhomboid peptidases. With our initial hiGlpG rhomboid structure, we proposed a third mechanism for gating involving the movement of loop 1,[24] as this would preserve the stereochemistry for the standard peptide bond hydrolysis (namely cleavage on the re-face of the peptide) observed with the majority of soluble serine peptidases. In light of the convincing mutagenesis and crystallographic data presented here, this hypothesis appears to be highly improbable. It is more likely that loop 1 plays a role in "fine-tuning" the function of the receptor[41] possibly via orientation in the lipid bilayer.[36]

With the substrate approaching the active site from the opposite side of the peptidase compared with our earlier prediction,[24] and given the architecture for the active-site residues, we predict that the mechanism for substrate cleavage will be similar, with the exception that the scissile bond of the substrate will be attacked on its si-face[25] (**Figure 2-6**). The si-face attacked is, in fact, observed in the recent cocrystal structure of ecGlpG with the coumarin inhibitor.[31] The oxyanion hole, being composed of the main-chain NH from Ser116 and the protonated $N^\varepsilon$ from the neighboring His65, remains unchanged compared to our previous prediction. Cleavage of the scissile bond by serine peptidases is traditionally carried out on the re-face. Other examples of serine peptidases with a Ser-Lys catalytic dyad, which are predicted to cleave their substrates on the si-face, include the signal peptidase from *E. coli* (SPase),[42] Lon protease,[43] Lex A peptidase,[44] and VP4 protease.[45] A cocrystal of SPase with a β-lactam inhibitor confirmed this stereochemistry for attack.[42] A comparison of residues in the catalytic triad of trypsin with

52
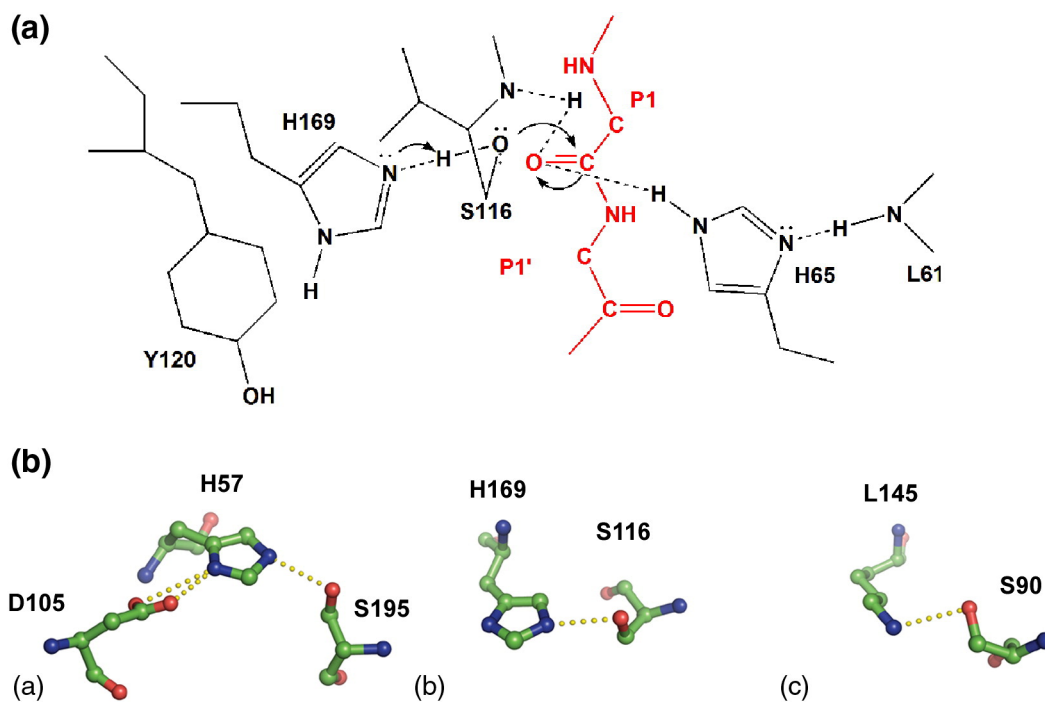
**Figure 2-6. Mechanism of si-face substrate cleavage by *H. influenzae* GlpG.** (A) Based on crystallographic and functional data, a new catalytic mechanism for substrate cleavage by hiGlpG is proposed (ChemDraw). Cleavage occurs on the si-face of the scissile bond. The oxyanion is proposed to be stabilized by the main-chain NH from Ser116 and by the protonated $N^{\varepsilon 2}$ from the neighboring His65. (B) Stereo chemistry of three different serine proteases— (a) trypsin (PDB code 1TPO), (b) hiGlpG (PDB code 2NR9), and (c) signal peptidase (PDB code 1KN9)—reveals a similar orientation of the hydroxyl group of serines found in peptidases with catalytic dyads.

residues in the catalytic dyads found in rhomboid and SPase demonstrates that the hydroxyl group on the catalytic serine from the dyad points to an opposite direction in relation to the central carbon atom. It has been observed that serine proteases with catalytic dyads have the hydroxyl groups oriented to the main-chain strand in the opposite direction.[43] This is important, since the stereochemistry for cleavage has been well established.[46] Known as the Bürgi–Dunitz angle, nucleophilic attack on the scissile bond must occur at 107° to the carbonyl. The hydroxyl group on the catalytic serine in the hiGlpG dyad has a similar nucleophilic attack on the signal peptidase (SPase). Rhomboid peptidases thus may utilize the Ser-Lys dyad arrangement found in the above mentioned peptidases but maintain the traditional histidine base as seen with most serine peptidases. Therefore, it appears that rhomboids carry out their reaction mechanism with similarities to serine peptidases, consisting of a catalytic triad in their active site, with the exception that they cleave the scissile bond on the si-face similarly to those with a catalytic dyad in the active site. A cocrystal structure with a substrate will be useful in determining further details on the mechanism of this interesting enzymatic reaction pathway.

## 2-3.3. Inhibition of hi and ecGlpGs

The complete understanding of rhomboid reaction pathway requires atomic resolution structures of catalytic intermediates with the help of rhomboid:inhibitor  or rhomboid:substrate complexes. To do so, we tested the effect of four know inhibitors of soluble protease on rhomboids protease function. Among the compounds tested, TPCK and DCI had inhibitory effect on both hi- and ec-GlpGs. Crystallization of hiGlpG was therefore undertaken to obtain complex structures but this work was stopped as a complex structure of ecGlpG with isocoumarin was published[47]. Since then, two other structures of ecGlpG bound to inhibitors (DFP and CAPF) were obtained [48,49].

54

The impact of these structure on our understanding of rhomboids reaction pathways will be further discussed in Chapter 5.


**2-4. Materials and Methods**

*2-4.1. Mutagenesis in H. influenzae rhomboid*

Our hiGlpG construct in pBAD-MycHisA[24] was mutated using the QuikChange lightening mutagenesis kit (Stratagene, USA). Nucleotide changes were confirmed with DNA sequencing. The pET21-C100Spi-Flag construct was mutated using the same kit to exchange the seven amino acids from Spitz (ASIASGA) with the cleavage site for TatA (IATAAFG), resulting in a pET21-C100Tat-Flag construct.

*2-4.2. Protein preparation (Expression and purification of hiGlpG)*

Cloning, expression, and purification were carried out as previously described,[24] with minor changes. Briefly, hiGlpG in the pBAD-Myc-HisA expression vector was transformed into Top10 cells (Invitrogen, USA). Cells were grown in LB media supplemented with ampicillin and induced with 0.002% arabinose at 24 °C for 5 h. Crude membrane fraction was isolated by high-speed ultracentrifugation and solubilized with a buffer containing 1% dodecylmaltoside (Anatrace, USA). Ni-NTA purification of hiGlpG was followed by thrombin removal of the tag using 30 U/mg GlpG for 1 h at room temperature. For crystallization, protein was purified further using gel filtration, as follows: after thrombin digestion, the protein was dialyzed to remove glycerol and concentrated using Millipore Amicon ultracentrifugal concentrators (30,000 Da molecular weight cutoff) to a concentration of 10–15 mg/ml. Further purification and detergent exchange (Anatrace) were carried out using a Superdex 200 gel-filtration column in 50 mM Tris, 200 mM NaCl, 0.005% AnaPOE

$C_{12}E_8$, 20% glycerol, and 0.5 mM ethylenediaminetetraacetic acid (pH 8.0). The pET21-C100Tat-Flag construct was purified as described by Urban and Wolfe.[35]

*2-4.3. Activity assay with mutants of hiGlpG*

Gel-based SDS-PAGE activity assays were conducted similarly to the detergent-solubilized *E. coli* GlpG protein.[35]hiGlpG (7.8 μg) in 0.1% dodecylmaltoside was incubated with 0.26 μg of pET21-C100Spi-Flag or 0.26 μg of pET21-C100Tat-Flag in a volume of 20–30 μl. The reaction was stopped with 2× SDS-PAGE buffer. Protein samples were resolved on SDS-PAGE, transferred to PVDF membranes, and probed using mouse anti-Flag antibodies (Sigma-Aldrich, USA), followed by rabbit anti-mouse antibodies conjugated to horseradish peroxidase (Sigma-Aldrich). To ensure equal protease amounts in each experiment, we ran a second gel in an identical manner but stained with Coomassie dye. Western blot analyses were repeated to ensure reproducibility. For all tables, digitization was carried out with ImageQuant LAS 4000 (GE Healthcare, USA). Quantification was carried out with ImageQuant software.

*2-4.4. Activity assays in presence of inhibitors*

PMSF, TPCK and DCI were dissolved in DMSO to reach final concentrations of 2.5, 1000 and 300 mM, respectively, whereas KC was dissolved in water at a concentration of 4 mM. Experiments were conducted as described above with hiGlpG, with the addition of the inhibitor at the first step of the experiment, at the desired concentration. For ecGlpG, detergent concentration was set to 0.075% to ensure ecGlpG proper function.

56

## 2-4.5. Crystallographic analysis

Crystals were prepared as previously described.[24] Data were collected on beamline 8.3.1 of the Advanced Light Source (ALS). Molecular replacement was carried out with MOLREP[50] using the hiGlpG coordinates (PDB code 2NR9) in their entirety and minus residues 134–164. Refinement was carried out with Refmac5.[51] Images were prepared with PyMOL. The coordinates and structure factors for H. influenzae GlpG have been deposited under PDB code 3ODJ.

## 2-4.6. Crystallization of hiGlpG:inhibitor complexes

Apo crystals of hiGlpG were grown as previously described[24], and wider screening of crystallization conditions was also attempted but no crystals were obtained. Co-crystallization was also tried, hiGlpG was mixed with TPCK at room temperature for 30 min (4,5 mg/ml and 4 mM, respectively). Same conditions were tested, and robot trays of Nextal PEG and PACT screens were also set up. Crystals appeared in manually set-up trays but these were inhibitors only crystals.

## 2-5. References

1.      Barrett, A.J., Rawlings, N.D. & O'Brien E.A. (2001) The MEROPS database as a protease information system. *J. Struct. Biol*., **134**, 95–102.

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

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

4       Brown, K.E., Kerr, M., Freeman, M. (2007). The egfr ligands spitz and keren act cooperatively in the drosophila eye. *Dev. Biol.*, **307**, 105–13.

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

6       Wasserman, J.D., Urban, S., Freeman, M. (2000). A family of rhomboid-like genes: drosophila rhomboid-1 and roughoid/rhomboid-3 cooperate to activate egf receptor signaling. *Genes Dev.*, **14**, 1651–63.

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

8       Stevenson, L.G.,Strisovsky, K., Clemmer, K.M., Bhatt, S., Freeman, M., Rather., P.N. (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

9       Brossier, F., Jewett, T.J., Sibley, L.D., Urban S. A spatially localized rhomboid protease cleaves cell surface adhesins essential for invasion by Toxoplasma. (2005). *Proc. Natl Acad. Sci. USA*, **102**, 4146–51

10      Urban, S., Freeman., M. (2003). Substrate specificity of rhomboid intramembrane proteases is governed by helix-breaking residues in the substrate transmembrane domain. *Mol. Cell*, **11**, 1425–34

11      Howell, S.A., Hackett, F., Jongco, A.M., Withers-Martinez, C., Kim, K., Carruthers, V.B., Blackman. M.J. (2005). Distinct mechanisms govern proteolytic shedding of a key invasion protein in apicomplexan pathogens. *Mol. Microbiol.*, **57**, 1342–56.

12      Freeman, M., Kimmel, B.E. (1992). Rubin Identifying targets of the rough homeobox gene of Drosophila: evidence that rhomboid functions in eye development. *Development*, **116**, 335–46.

13      Dutt, A., Canevascini, S., Froehli-Hoier, E., Hajnal., A. (2004). EGF signal propagation during C. elegans vulval development mediated by ROM-1 rhomboid. *PLoS Biol.*, **2**, e334

14      Dominguez, M., Wasserman, J.D., Freeman., M. Multiple functions of the EGF receptor in Drosophila eye development (1998) *Curr. Biol.*, **8**, 1039–48

15      Cipolat, S., Rudka, T., Hartmann, D., Costa, V., Serneels, L., Craessaerts, K. et al. (2006). Mitochondrial rhomboid PARL regulates cytochrome c release during apoptosis via OPA1-dependent cristae remodeling. *Cell*, **126**, 163–175

16      Frezza, C., Cipolat, S., Martins de Brito, O., Micaroni, M., Beznoussenko, G.V., Rudka, T. et al. (2006)**.** OPA1 controls apoptotic cristae remodeling independently from

mitochondrial fusion. *Cell*, **126**, 177–89.

17     Walder, K., Kerr-Bayles, L., Civitarese, A., Jowett, J., Curran, J., Elliott, K. et al. (2005). The mitochondrial rhomboid protease PSARL is a new candidate gene for type 2 diabetes *Diabetologia*, **48**, 459–68

18     Maassen, J.A., Janssen, G.M., Hart, L.M. (2005). Molecular mechanisms of mitochondrial diabetes (MIDD). *Ann. Med.*, **37**, 213–221

19     Hatunic, M., Stapleton, M., Hand, E., DeLong, C., Crowley, V.E., Nolan., J.J. (2009). The Leu262Val polymorphism of presenilin associated rhomboid like protein (PARL) is associated with earlier onset of type 2 diabetes and increased urinary microalbumin creatinine ratio in an Irish case–control population. *Diabetes Res. Clin. Pract.*, **83**, 316–319.

20     Whitworth, A.J., Lee, J.R., Ho, V.M., Flick, R. Chowdhury, R., McQuibban, G.A. (2008). Rhomboid-7 and HtrA2/Omi act in a common pathway with the Parkinson's disease factors Pink1 and Parkin. *Dis. Models Mech.*, **1** , 168–74

21     Alexander, C., Votruba, M., Pesch, U.E., Thiselton, D.L., Mayer, S., Moore, A. et al. (2000). OPA1, encoding a dynamin-related GTPase, is mutated in autosomal dominant optic atrophy linked to chromosome 3q28. *Nat. Genet.*, **26**, 211–15.

22     Yan, Z., Zou, H., Tian, F., Grandis, J.R., Mixson, A.J. Lu, P.Y., Li., L.Y. (2008). Human rhomboid family-1 gene silencing causes apoptosis or autophagy to epithelial cancer cells and inhibits xenograft tumor growth. *Mol. Cancer Ther.*, **7**, 1355–64.

23     Zou, H., Thomas, S.M., Yan, Z.W., Grandis, J.R., Vogt, A., Li, L.Y. (2008) Human rhomboid family-1 gene RHBDF1 participates in GPCR-mediated transactivation of EGFR growth signals in head and neck squamous cancer cells. *FASEB J.*

24     M.J. Lemieux, S.J. Fischer, M.M. Cherney, K.S. Bateman, M.N. James.The crystal structure of the rhomboid peptidase from Haemophilus influenzae provides insight into intramembrane proteolysis. (2007). *Proc. Natl Acad. Sci. USA*, **104**. 750–54.

25     Wang, Y., Ha. Y. (2007). Open-cap conformation of intramembrane protease GlpG.*Proc. Natl Acad. Sci. USA*, **104.** 2098–102.

26     Ben-Shem, A., Fass, D., Bibi., E. (2007). Structural basis for intramembrane proteolysis by rhomboid serine proteases. *Proc. Natl Acad. Sci. USA*, **104**. 462–66.

27     Wu, Z., Yan, N., Feng, L., Oberstein, A., Yan, H., Baker, R.P. et al. (2006). Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry. *Nat. Struct. Mol. Biol.*, **13**, 1084–91.

28     Wang, Y., Zhang, Y., Ha, Y. (2006). Crystal structure of a rhomboid family intramembrane protease. *Nature*, 1–5.

29     Radisky, E.S., Lee, J.M., Lu, C.J., Koshland Jr., D.E. (2006). Insights into the serine protease mechanism from atomic resolution structures of trypsin reaction intermediates. *Proc. Natl Acad. Sci. USA*, **103**, 6835–40.

30     Baker, R.P., Young, K., Feng, L., Shi, Y., Urban, S. (2007). Enzymatic analysis of a rhomboid intramembrane protease implicates transmembrane helix 5 as the lateral substrate gate. *Proc. Natl Acad. Sci. USA*, **104**, 8257–62.

31     Vinothkumar, K.R., Strisovsky, K., Andreeva, A., Christova, Y., Verhelst, S., Freeman, M. (2010). The structural basis for catalysis and substrate specificity of a rhomboid protease. *EMBO J.,* **29**, 3797-809.

32     Lazareno-Saez, C., Brooks, C. & Lemieux, M. (2010). Structural comparison of substrate entry gate for rhomboid intramembrane peptidases. *Biochem. Cell Biol*, **89**, 216-23.

33    Li, Y.M., Lai, M.T., Xu, M., Huang, Q., DiMuzio-Mower, J., Sardana, M.K. et al. (2000). Presenilin 1 is linked with gamma-secretase activity in the detergent solubilized state. *Proc. Natl Acad. Sci. USA*, **97**, 6138–43.

34    Strisovsky, K., Sharpe, H.J., Freeman. M. (2009). Sequence-specific intramembrane proteolysis: identification of a recognition motif in rhomboid substrates. *Mol. Cell*, **36**, 1048–59.

35    Urban, S., Wolfe. M.S. (2005). Reconstitution of intramembrane proteolysis in vitro reveals that pure rhomboid is sufficient for catalysis and specificity. *Proc. Natl Acad. Sci. USA*, **102**, 1883–88.

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

37    Maegawa, S., Koide, K., Ito, K., Akiyama, Y. (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**, 435-47.

38    Urban, S., Baker, R.P. (2008). In vivo analysis reveals substrate-gating mutants of a rhomboid intramembrane protease display increased activity in living cells. *Biol. Chem.*, **389**, 1107–15.

39    Maegawa, S., Ito, K., Akiyama, Y. (2005). Proteolytic action of GlpG, a rhomboid protease in the Escherichia coli cytoplasmic membrane. *Biochemistry*, **44**, 13543–52.

40    Erez, E., Bibi, E. (2009). Cleavage of a multispanning membrane protein by an intramembrane serine protease. *Biochemistry*, **48**, 12314–22.

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

42    Paetzel, M., Dalbey, R.E., Strynadka, N.C. (1998) Crystal structure of a bacterial signal peptidase in complex with a beta-lactam inhibitor. *Nature*, **396**, 186–90.

43    Botos, I., Melnikov, E.E., Cherry, S., Tropea, J.E., Khalatova, A.G., Rasulova, F. et al. (2004). The catalytic domain of Escherichia coli Lon protease has a unique fold and a Ser-Lys dyad in the active site. *J. Biol. Chem.*, **279**, 8140–8.

44    Luo, Y., Pfuetzner, R.A., Mosimann, S., Paetzel, M., Frey, E.A., Cherney M. et al. (2001). Crystal structure of LexA: a conformational switch for regulation of self-cleavage *Cell*, **106**, 585–94.

45    Feldman, A.R., Lee, J., Delmas, B., Paetzel, M. (2006). Crystal structure of a novel viral protease with a serine/lysine catalytic dyad mechanism.*J. Mol. Biol.*, **358**, 1378–89.

46    Burgi, H.B., Dunitz, J.D. (1974). Stereochemistry of reaction paths at carbonyl centers. *Tetrahedron*, **30**, 1563–1572

47    Vinothkumar, K.R., Strisovsky, K., Andreeva, A., Christova, Y., Verhelst, S., Freeman, M. (2010). The structural basis for catalysis and substrate specificity of a rhomboid protease. *EMBO J.*, **29**,3797–809.

48    Xue, Y., Hq, Y. (2012) Catalytic mechanism of rhomboid protease GlpG probed by 3,4-dichloroisocoumarin and diisopropyl fluorophosphonate. *J. Biol. Chem.*, **287**, 3099–107.

49    Xue, Y., Chowdhury, S., Liu, X., Akiyama, Y., Ellman, J., Ha, Y (2012). Conformational change in rhomboid protease GlpG induced by inhibitor binding to its S' subsites. *Biochemistry.*, **51**,3723–731.

50    Vagin, A., Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.*, **30**,1022–25.

60

51    Sheiner, L., Dowse, T.J., Soldati-Favre, D. (2008). Identification of trafficking determinants for polytopic rhomboid proteases in Toxoplasma gondii. *Traffic*, **9**,665–77.

# Chapter 3

## Towards structural determination of full length *E. coli* rhomboid (ecGlpG-FL)

## 3-1. Introduction

As mentioned earlier, a subset of rhomboids have an N-terminal cytosolic extension[1] which is thought to have a regulatory function. No structures of the full-length rhomboid or high-resolution structure of the N-terminal domain were reported. To investigate the regulatory function of this domain, ecGlpG was chosen as it is the best characterized rhomboid and it contains two domains: the membrane domain and the N-terminal domain (~90 residues). More precisely, the structure of ecGlpG-FL will provide insights into the function of ecGlpG-cyto. Information about the function of this N-terminal domain could be obtained by comparing this structure with structures already deposited in the database using DALI server (http://ekhidna.biocenter.helsinki.fi/dali_server/)[2]. Taken together, these details on ecGlpG-cyto will help us investigate its regulatory function and they may reveal the unknown physiological function and regulation of the *E. coli* rhomboid or other rhomboids containing a similar domain. To study the structure of ecGlpG-FL, two strategies can be used to fulfill this aim: the direct one (crystallizing ecGlpG-FL) or the indirect one (crystallizing ecGlpG-cyto). In this chapter, we will go down the road of the direct strategy. I will be describing the purification of ecGlpG-FL, its crystallization and strategies of optimization. The crystallization of such membrane protein often needs the exploration of various methods of crystallization to obtain good diffracting crystals.

## 3-2. Results

### 3-2.1. Optimized over-expression and purification of ecGlpG-FL

ecGlpG-FL is a 276 amino acid membrane protein that is highly susceptible to proteolysis. The presence of the N-terminal domain is thought to be detrimental for obtaining well diffracting crystals[3]. It also has caused difficulty with purification as it is proposed to be unstructured[4].

ecGlpG-FL was successfully cloned, over-expressed, and purified with a 19 mg yield of pure ecGlpG-FL from twelve liters of bacterial culture. At the time of writing, high quantities of this protein can be expressed, either for functional or structural studies. The DNA for ecGlpG-FL is inserted into the pBad Myc His A vector, resulting in, a Myc epitope and 6 histidine residues at the C-terminal. In a second step, a tobacco etch virus (TEV) protease cleavage site is inserted, prior to the Myc epitope. Following the expression and lysis of ecGlpG-FL in Top10 *E. coli* cells, two centrifugations steps are performed to remove the inclusion bodies and to obtain the membrane containing the protein of interest. To solubilize ecGlpG-FL, the membrane fragments are homogenized in a detergent solution. After another centrifugation step to remove the unsolubilized membrane fragments, two steps of purification are performed: a nickel affinity column (Ni-NTA resin) and a size-exclusion column, after cleavage of the Myc-His6 tag by TEV protease (**Figure 3-1**). Different detergents have been used for the homogenization and/or size-exclusion purification in order to improve the diffraction of the crystals (see below). No sign of proteolysis appeared during the purification, which allows me to provide a high yield of pure and reproducible ecGlpG-FL. The basic purification protocol involves n-dodecyl-b-d-maltopyranoside (DDM) for the homogenization, nickel affinity and size-exclusion columns.
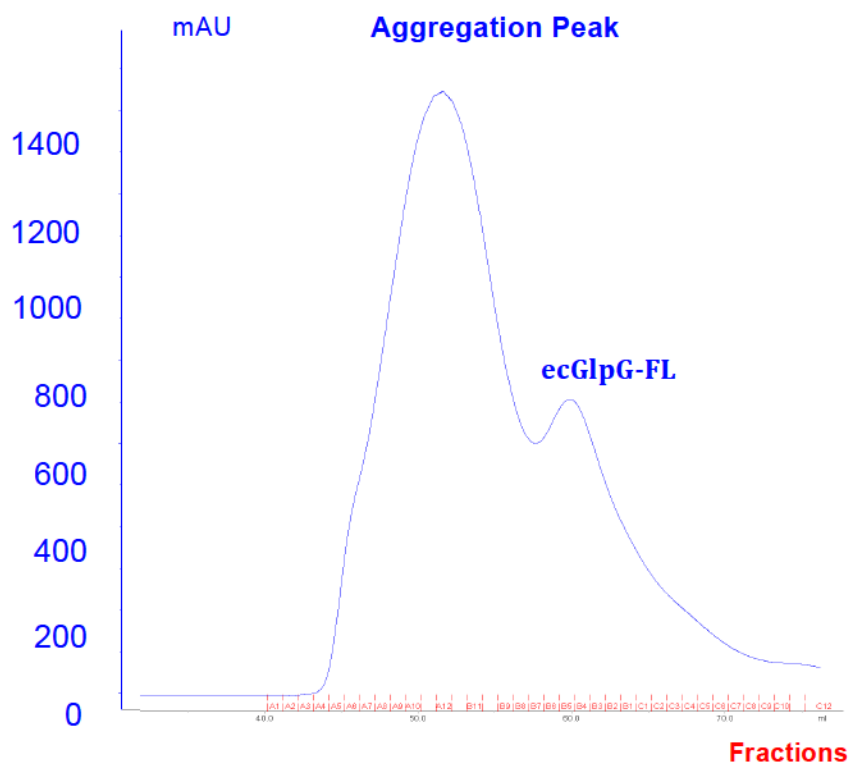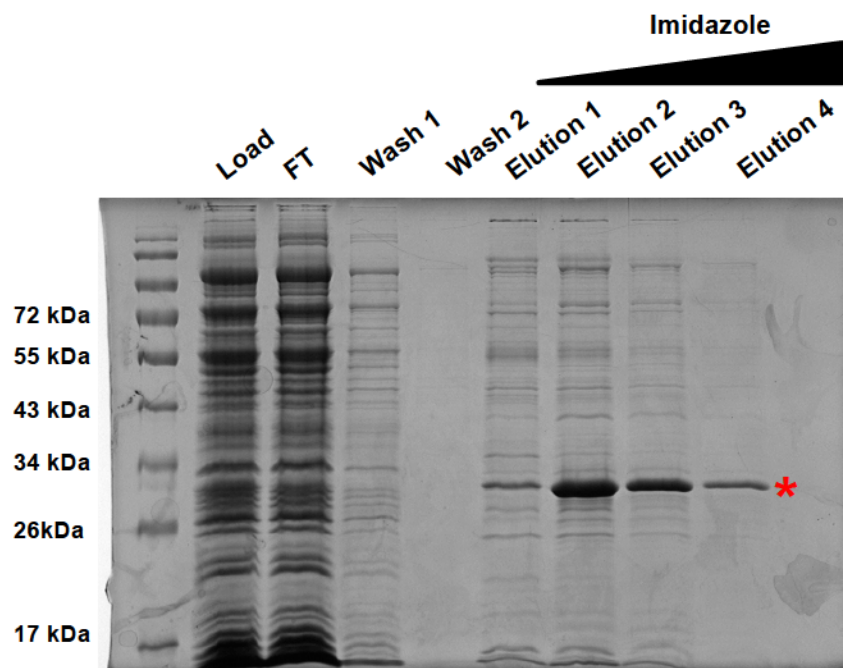
**Figure 3-1. SDS-PAGE and gel filtration analysis of ecGlpG-FL.**
**A**. SDS PAGE after Nickel column. The red star refers to ecGlpG. **B**. Size
exclusion column trace of absorbance at 280 nm.

## 3-2.2. *Obtaining and Optimizing diffracting crystals of full-length ecGlpGFL*

Expressing, purifying and obtaining well-diffracting crystals of any membrane protein is more challenging than for soluble proteins. Indeed, in July 2014, only 491 membrane protein structures have been solved in comparison to more than 101,000 soluble proteins (http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html and http://www.rcsb.org/pdb/home/home.do ). In fact, detergents represent an additional and crucial parameter and they can vary during the purification and/or the crystallization allowing different combinations. To solve the structure of ecGlpG-FL, different crystallization trials have been performed varying the purification protocol and/or the crystallization methods. For each crystallization trial, three steps were conducted: screening different conditions using a crystallization robot, reproducing the positive hits, and optimizing them manually (**Figure 3-2**).

### 3-2.2.1. *Initial crystallization trials and optimization strategies*

First hits were obtained with robot trays and crystals were reproduced manually. They diffracted to a resolution of 9 Å and further optimization was performed (additives screening, temperature variations or cryoprotectants) to improve resolution. As these different approaches were not successful, we developed various strategies to optimize the diffraction quality of ecGlpGFL crystals. These strategies and their rationale will be briefly described below and their results are summarized in **Table 3-1**.

66
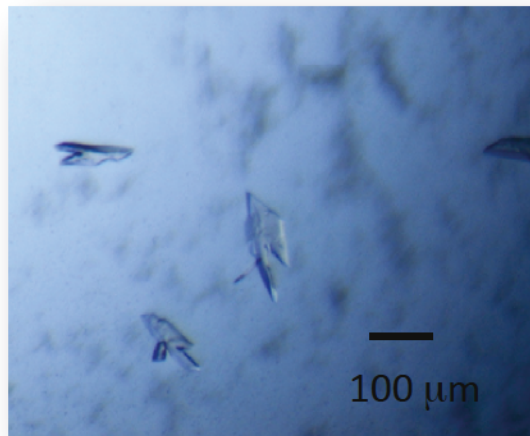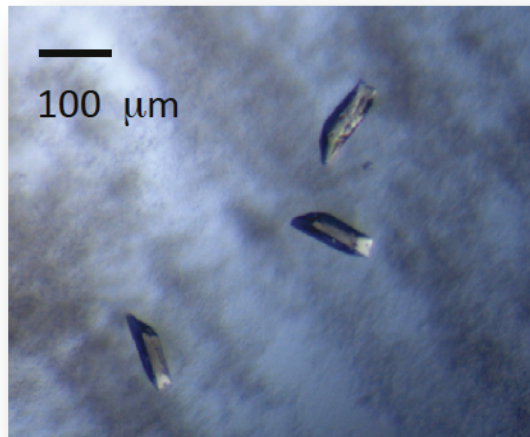
**Figure 3-2. Typical shapes and sizes observed for 9 Å resolution ecGlpG-FL crystals**

**Table 3-1. Strategies and results for ecGlpG-FL crystallization**

\* I have added detergents just prior to crystallization either as additives or by mixing the new detergent with the protein-DDM mixture.

\*\* others correspond to LDAO, NG and $C_{12}E_8$.

\*\*\* Reproducing means that hits from the robot could be reproduced in standard crystal trays.

| Optimization method | | Detergent homogenization/ purification | Steps | | Crystals | Best Diffraction |
|---|---|---|---|---|---|---|
| Basic purification | | DDM/DDM | Robot trays | | Yes | N/A |
| | | DDM/DDM | Reproducing\*\*\* | | Yes | 9 Å |
| | | DDM/DDM | Optimizing | Temperature factors | No | - |
| | | DDM/DDM | | Additives | Yes | 9.4 Å |
| | | | | Cryoprotectant | Yes | 11 Å |
| Detergent exchange | Mixing at RT\* | DDM/DDM | Robot trays | | Yes | N/A |
| | | | Reproducing | | Yes | 14.7 Å |
| | During purification | DM/DM | Robot trays | | No | - |
| | | DDM/others\*\* | Robot trays | | No | - |
| Lipid type detergent | | DDM/DDM | Robot trays | | Yes | N/A |
| Bicelles | | DDM/DDM | Robot trays and manual screening | | No | - |
| Removal of flexible regions | | - | - | | No | - |

As detergents are a critical parameter for membrane protein solubility, their role in the crystallization process is essential. A recent study revealed that 8 detergents (n-dodecyl-β-D-maltopyranoside (DDM), n-decyl-β-D-maltopyranoside (DM), undecylmaltoside (UM), nonyl-glucoside (NG), octyl glucoside (OG), lauryl dimethylamineoxide (LDAO), octaethylene glycol monododecyl ether ($C_{12}E_8$) and polyoxyethylene(9)dodecyl ether ($C_{12}E_9$)) are used in 80% of the crystallization conditions for α-helical polytopic membrane proteins[5]. Based on this study, some of these detergents were tested to replace the DDM used in the standard purification protocol. This exchange was performed at different stages of the purification protocol: i) during the homogenization, ii) during the size-exclusion column (**Figure 3-3**) or iii) just prior to crystallization (see **Table 3-1** for further details).

Three new classes of synthetic amphiphile molecules were tested: maltose-neopentyl glycol (MNG), tandem facial amphiphiles (TFA) and glycotripod amphiphiles (**Figure 3-4**). These amphiphiles were shown to be comparable or superior to a regular detergent for the solubilization and long-term stabilization of membrane proteins[6-8]. They represent an important tool for membrane protein crystallization as they are more rigid than detergent molecules and can improve tight packing in a crystal lattice.

Moreover, a recent structure of the membrane domain of ecGlpG has been obtained by using a lipid-detergent mixture which tends to create bicelles[9]. Bicelles are monodisperse and un-compartmentalized patches of lipid bilayers with detergent molecules lining the apolar edges[10]. There are two main advantages of such technique: membrane proteins are maintained in a native-like bilayer environment and the formation of bicelles at low temperature is not viscous[11].
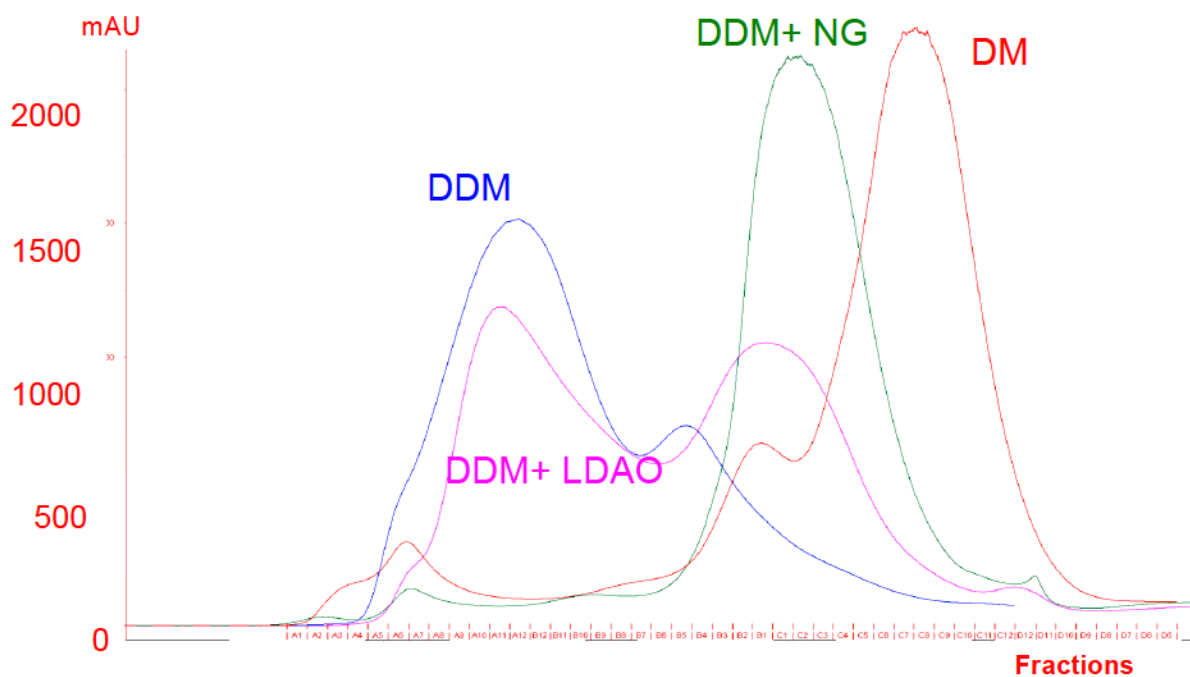
**Figure 3-3. Size exclusion column trace of 280 nm absorbance of ecGlpG-FL in different detergents.**

When one detergent is listed, the complete purification was done using only this one. However, when two detergents are listed, the first one was used for homogenization and the nickel affinity column purification step whereas it was partially exchanged by the second one during the SEC. Of note, all tested detergents are not presented here.
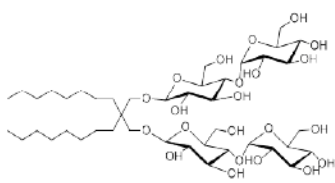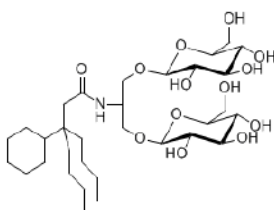
**Maltose Neopentyl Glycol**  **Glycotripod**  **Tandem Facial Amphiphiles**
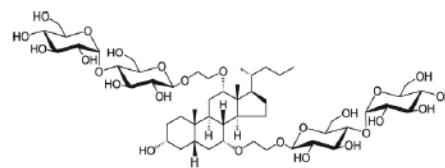
Figure 3-4. Molecular structures of the various synthetic amphilphile molecules tested in EcGlpG crystallization trials.

Lastly, in comparison to most of the groups[12-18] which crystallized the membrane domain of ecGlpG, a total of 12 residues, predicted to be flexible, are present at the C-terminus of ecGlpGFL after purification. As removal of an unstructured domain of a protein is a well-known process to improve crystallization, three different constructs were designed to reduce this C-terminal flexibility (**Figure 3-5**) and further purification and crystallization of these constructs are pursued in the laboratory.

Unfortunately, all these optimizations were not successful to substantially improve diffraction quality of ecGlpG-FL crystals.

## 3-3. Discussion

This chapter summarizes the methodology used to clone and purify the membrane protein on one part and the methodology used to crystallize and optimize the crystal resolution on the other part.

The cloning, over-expression and purification process were clearly optimized as a high yield of the pure and intact protein was obtained. Contrary to previous studies highlighting the high susceptibility of proteolysis of the cytoplasmic domain avoiding the purification of ecGlpG-FL, the intact protein was purified without any sign of proteolysis.

The full-length membrane protein was then successfully and reproducibly crystallized in presence of the cytoplasmic domain in contradiction to the hypothesis that its presence might be detrimental to obtain crystals. Even if a multitude of diverse and broad crystallization and optimization techniques were used, the diffracting resolution (9 Å) was not good enough to solve ecGlpG-FL X-ray structure. These results seem to indicate that there might be at least one part of

the protein, which is too flexible and would prevent good packing during crystallization. Removal of some amino acids at the C-terminus of the protein was performed but without any diffracting resolution improvement. Structure prediction software also reveals a structured N-terminus of ecGlpG-FL. Consequently, the linker between the two domains of ecGlpG-FL and its inherent flexibility could be the reason of such limiting resolution. To overcome this problem, two methods might be developed. The first one is the sequential removal of amino acids located at the center of the linker between the two domains. Different constructs have already been designed for different linkers of various shorter lengths. A second and more drastic method would be to co-crystallize the protein with an antibody[19] to fix the structure of the unfolded part and therefore allow a better packing of the protein during crystallization to obtain a better resolution. Even though these two methods will change to some extent the native structure of the protein: structural information on ecGlpG-FL would still be useful to hopefully highlighting the respective positions of the two domains and possible interactions.

## 3-4. Material and methods

### 3-4.1. Expression and purification of ecGlpG-FL

The gene of ecGlpG was cloned into pBAD-Myc/HisA plasmid (Invitrogen, Canada), having C-terminal tobacco etch virus (TEV) protease cleavage site, Myc-epitope and His$_6$-tag. TOP10 chemically competent cells were transformed. The protein was induced with 0.002% arabinose and expressed at 24°C for 6 hours in LB media. The cells were harvested and resuspended in 50 mM Tris-HCl pH 8.0, 150 mM NaCl and lysed using an EmulsiFlex (Avestin Inc, Ottawa,
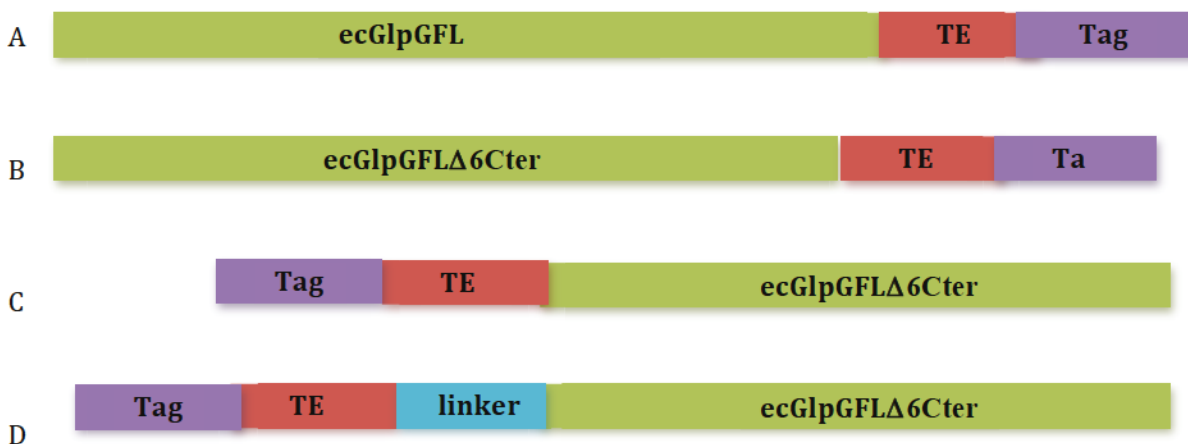
**Figure 3-5. Constructs used for flexible regions removal**. **A**. Original construct .**B**. Construct 1: deletion of 6 C-terminal residues of ecGlpG-FL normally cut by previous group during the cleavage of the N-terminal domain **C**. Construct 2: same deletion as construct 1 and the displacement of the Tag resulting in one additional residue at the N-terminus after TEV cleavage and a deletion of 12 residues in total at the C-terminus .**D** Construct 3: same as construct 2 with an additional linker to insure TEV cleavage. Construct A yielded the best diffracting crystals.

Canada) apparatus at 15,000 psi. Unbroken cells were removed with a centrifugation step at 16,500 rpm (22,320 $g$), and the membranes were isolated by ultracentrifugation at 35000rpm (95,800 $g$) for 2 h. The membranes were solubilized in 50 mM Tris-HCl pH 8.0, 300 mM NaCl, 10 mM imidazole, 20% glycerol, 1% (w/v) DDM and applied onto a Ni-NTA column (Qiagen, Ontario, Canada). Protein was eluted with 250-500 mM of imidazole, 50 mM Tris-HCl pH 8.0, 300 mM NaCl, 20% glycerol, 0.1% DDM. The His-tag was removed by TEV protease (1 mg per 100 mg of protein, overnight, 16°C). Eluted fractions were further dialyzed for 1 h at 4°C in 25 mM Tris-HCl pH 8.0, 200 mM NaCl, 5% glycerol. Dialysis membrane with a 12-14 kDa cutoff were used and were therefore impermeable to detergent micelles, so no detergent was needed in the dialysis buffer. The protein was concentrated up to 1 ml and loaded on a size exclusion column (Sephadex 200 16/60) in 25 mM Tris pH 8.0, 200 mM NaCl, 0.5 mM EDTA, 20% glycerol and 0.05% DDM. The eluted fractions were collected and dialyzed against a pre-crystallization buffer (5 mM Tris pH 8.0, 5 mM NaCl, 5% Glycerol) for 1 h at 4°C. The protein was finally concentrated using 30,000 Da MWCO concentrators (Millipore, USA) up to 4.4 mg/ml. The protein samples were kept at -80°C.

*3-4.2 Cloning and purification of different ecGlpG-FL constructs*

Primers used to generate the insertions/deletions were designed with the help of the QuickChange Primer Design Program (Agilent Technologies, Santa Clara, CA). The ecGlpG gene in the pBAD-Myc/HisA plasmid was used as a template to introduce the desired insertion/deletion with the QuickChange Lightning Multi Site-Directed Mutagenesis Kit (Agilent Technologies, Santa Clara, CA) and each construct was verified by DNA sequencing. Expression

and purifications of these ecGlpG constructs was identical to the one previously described, with the addition of a second Ni-NTA column, performed after His-Tag removal with TEV protease. Such step was needed as only 50% of ecGlpG was cleaved, even after optimization of the cleavage conditions.

### 3-4.3 Crystallization conditions of best diffracting crystals in DDM

Initial crystallization screening was performed using a Gryphon robot (Art Robbins Instruments, USA) at room temperature and conditions yielding crystals were reproduced with the sitting drop vapor diffusion technique. Best diffracting crystals were obtained with construct A, in a drop composed of a mixture of protein, mother liquor (0.1 M PCB pH 5.6-6.4, 25-27% PEG 1500) and water in a 1:1:2 ratios. Prior to data collection, crystals were cryo-cooled in liquid nitrogen using mother liquor complemented with 20% glycerol as cryo-protectant. Diffraction spots were observed at about 9 Å resolution.

### 3-4.4 Detergent Exchange

EcGlpGFL was subjected to detergent exchange at different stages before its crystallization.

DM was used throughout the purification protocol: from the homogenization (5%), to the size exclusion column (0.24%), through the NiNTA column (0.24%).

DDM was also exchanged during the size exclusion column with various detergents: C12E8, NG and LDAO. Based on CMC of these detergents, final detergent concentrations in the buffer were of 0.05, 0.6 and 0.5%, respectively. Further protein dialysis and concentration was performed as described above.

Detergents were also added just prior to crystallization. 96 detergent additives screen from Hampton Research was used at a final concentration of 1 CMC. NG, LDAO, OG and C12E8 detergents were mixed with the protein-DDM sample for 30 minutes at room temperature. Final concentrations of NG and LDAO ranged from 0.6 to 2.0% and from 0.4 to 1.2%, respectively, whereas unique final concentrations of OG and C12E8 were tested (1 and 0.2%, respectively).

*3-4.5 Amphiphiles*

These compounds were directly added to the drop with MNG3 concentrations ranging from 1 (0.001%) to 20 CMC while cs-tripgly was tested at 1 CMC (0.12%)

*3-4.6 Bicelles*

The protocol used for the structure determination of ecGlpG membrane domain obtained in bicelles of DMPC and CHAPSO was reproduced[9]. No crystals were observed after robot and manual tray screening.

## 3-5. References

1.  M. K. Lemberg, & M. Freeman. Functional and evolutionary implications of enhanced genomic analysis of rhomboid intramembrane proteases. *Genome Res* (2007) **17**, pp. 1634-46.
2.  Del Rio, A., Dutta, K., Chavez, J., Ubarretxena-Belandia, I. & Ghose, R. (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. *Journal of molecular biology* **365**, 109-22.
3.  Wu, Z., Yan, N., Feng, L., Oberstein, A., Yan, H., Baker, R. P., Gu, L., Jeffrey, P. D., Urban, S. & Shi, Y. (2006). Structural analysis of a rhomboid family intramembrane protease reveals a gating
4.  Ben-Shem, A., Fass, D. & Bibi, E. (2007). Structural basis for intramembrane proteolysis by rhomboid serine proteases. *Proc Natl Acad Sci U S A* **104**, 462-6.
5.  Newstead, S., Ferrandon, S. & Iwata, S. (2008). Rationalizing alpha-helical membrane protein crystallization. *Protein science : a publication of the Protein Society* **17**, 466-72.
6.  Chae, P. S., Rasmussen, S. G., Rana, R. R., Gotfryd, K., Chandra, R., Goren, M. A., Kruse, A. C., Nurva, S., Loland, C. J., Pierre, Y., Drew, D., Popot, J. L., Picot, D., Fox, B. G., Guan, L., Gether, U., Byrne, B., Kobilka, B. & Gellman, S. H. (2010). Maltose-neopentyl glycol (MNG) amphiphiles for solubilization, stabilization and crystallization of membrane proteins. *Nature methods* **7**, 1003-8.
7.  Chae, P. S., Gotfryd, K., Pacyna, J., Miercke, L. J., Rasmussen, S. G., Robbins, R. A., Rana, R. R., Loland, C. J., Kobilka, B., Stroud, R., Byrne, B., Gether, U. & Gellman, S. H. (2010). Tandem facial amphiphiles for membrane protein stabilization. *Journal of the American Chemical Society* **132**, 16750-2.
8.  Chae, P. S., Wander, M. J., Bowling, A. P., Laible, P. D. & Gellman, S. H. (2008). Glycotripod amphiphiles for solubilization and stabilization of a membrane-protein superassembly: importance of branching in the hydrophilic portion. *Chembiochem : a European journal of chemical biology* **9**, 1706-9.
9.  Vinothkumar, K. R. (2011). Structure of rhomboid protease in a lipid environment. *Journal of molecular biology* **407**, 232-47.
10. Sanders, C. R. & Prosser, R. S. (1998). Bicelles: a model membrane system for all seasons? *Structure* **6**, 1227-34.
11. Faham, S. & Bowie, J. U. (2002). Bicelle crystallization: a new method for crystallizing membrane proteins yields a monomeric bacteriorhodopsin structure. *Journal of molecular biology* **316**, 1-6.

12. Wang, Y., Zhang, Y. & Ha, Y. (2006). Crystal structure of a rhomboid family intramembrane protease. *Nature*, 1-5.
13. Wu, Z., Yan, N., Feng, L., Oberstein, A., Yan, H., Baker, R. P., Gu, L., Jeffrey, P. D., Urban, S. & Shi, Y. (2006). Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry. *Nat Struct Mol Biol* **13**, 1084-91.
14. Ben-Shem, A., Fass, D. & Bibi, E. (2007). Structural basis for intramembrane proteolysis by rhomboid serine proteases. *Proc Natl Acad Sci U S A* **104**, 462-6.
15. Wang, Y. & Ha, Y. (2007). Open-cap conformation of intramembrane protease GlpG. *Proc Natl Acad Sci U S A* **104**, 2098-102.

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

17.  Vinothkumar, K. R. (2011). Structure of rhomboid protease in a lipid environment. *Journal of molecular biology* **407**, 232-47.

18.  Vinothkumar, K. R., Strisovsky, K., Andreeva, A., Christova, Y., Verhelst, S. & Freeman, M. (2010). The structural basis for catalysis and substrate specificity of a rhomboid protease. *The EMBO journal* **29**, 3797-809.

19.  Hunte, C., and Michel, H. (2002) Crystallisation of membrane proteins mediated by antibofy fragments. *Curr Opin Struct Biol,* **12,** 503-8

## Chapter 4


## Domain swapping in the cytoplasmic domain of the *Escherichia coli* rhomboid protease

**4-1. Introduction**

Rhomboids are a family of intramembrane serine proteases (peptidases) that cleave membrane proteins substrates within or close to the hydrophobic environment of the lipid bilayer. Substrate cleavage results in a domain being released which acts in a variety of cellular processes [1; 2]. Rhomboids have been shown to be involved in cell signaling [3; 4], apoptosis [5; 6], Parkinson's disease [7], hereditary blindness [8], type 2 diabetes [9; 10] and epithelial cancer [11; 12]. They are also linked to the last stage of parasitic invasion with *Toxoplasma gondii* [13] and *Plasmodium falciparum* [14]. In prokaryotic systems, roles in glucose utilization [15] and quorum sensing [16] have been identified. The exact mechanism by which rhomboids are related to disease is still uncertain but this link highlights the importance of their study. As is the case for many rhomboids, the cellular role of *E. coli* GlpG (ecGlpG), a subject of various genetic and biochemical studies, is not known, and furthermore its native substrate has not been identified. Further investigation of the structural, functional and regulatory features of ecGlpG is important to reveal its cellular role.

Rhomboid proteases are found in all branches of life, excluding viruses [17; 18]. Topologically, the family contains a catalytic core of six (predominates in bacteria) or seven (mostly found in eukaryotes and in endosymbiotic organelles) transmembrane (TM) segments. Remarkably, the enzyme's water dependent active site, a catalytic Ser-His dyad, is located approximately 10 Å below the lipid bilayer while maintaining exposure to the aqueous solvent [19; 20; 21; 22]. In addition to the membrane domain, most of rhomboids also contain soluble domains at the N- and /or C-termini. These regions can vary in size and sequence depending on the rhomboid family [17; 18]. For the majority of rhomboids the role of these soluble domains remains

to be determined. In prokaryotic rhomboids, including the extensively studied *E. coli* rhomboid (ecGlpG), the role of the conserved N-terminal cytoplasmic domain also remains undetermined.

To date, crystal structures for two prokaryotic rhomboid proteases have been determined: *Haemophilus influenzae* (hiGlpG) [19] which does not contain any soluble domain and the membrane domain of *Escherichia coli* (ecGlpG) [20; 21; 22; 23]. These structures have provided valuable insight into the catalytic mechanism of the enzyme. No crystal structure of the cytoplasmic domain of any rhomboid has been determined. An NMR solution for the cytoplasmic domain from both *Pseudomonas aeruginosa* (paGlpG) and ecGlpG was solved revealing a compact monomeric helix-sheet bundle [24; 25]. In their structure the C-terminal region was disordered suggesting an unstructured linker between the membrane domain and soluble domains of ecGlpG.

Another subject of interest is the regulation of these proteases. Rhomboids do not require either cofactors or other protein partners and are solely responsible for the substrate recognition [26]. Remarkably, they are able to cleave rhomboid substrates not only from different species but also from different kingdoms of life [27; 28]. This suggests that their active site is not strictly specific in substrate recognition and another mechanism could exist allowing for tight regulation of unspecific proteolysis within the membrane. Several hypotheses exist about how rhomboids avoid unintentional cleavage: cellular localizations of enzyme and substrate [29], direct interaction of regulators with the protease or substrate [30] and special requirements for the substrate [31]. Different models for the regulation of rhomboid function have been proposed involving their cytoplasmic domain. For example, the cytoplasmic domain of rhomboid-like proteases in *T. gondii* is responsible for targeting them to different cell compartments[32]. The cytoplasmic domain of mammalian rhomboid RHBDL2 recognizes its substrate thrombomodulin [33; 34].

Another example of more complex substrate recognition was demonstrated for mammalian RHBDL4 through its cytoplasmic domain, which identifies ubiquitinated substrate via its ubiquitin-interaction motif (UIM) [35]. The cleavage fragments are dislocated in AAA+-ATPase p97-dependent manner, making the protease a member of ER-associated degradation pathway. Although UIM was not found in other rhomboid protease cytoplasmic domains, other extramembrane substrate recognition mechanisms are not to be excluded. Another possible means for regulation of rhomboid protease function by the cytoplasmic domain might be an interaction between these domains. This interaction could modulate the tilt and the position of the membrane domain within the lipid bilayer and help positioning the active site relative to the substrate TMD and the membrane plane [36]. This interaction might alter the structure and stability of, or access to the active site [24; 25].

For ecGlpG, the deletion of the cytoplasmic domain has previously been linked to a decrease of activity [20; 24; 37]. More recently, it has been proposed that the cytoplasmic domain and specifically the ten residues in a linker region (residues 80-90) preceding the membrane domain (GlpG$^{90\text{-}270}$) regulates rhomboid activity [25]. Removal of the cytoplasmic domain from the membrane domain resulted in a loss of enzymatic activity.

The focus of this study is the N-terminal cytoplasmic domain of ecGlpG, (ecGlpG-cyto). We present the high resolution structure of ecGlpG-cyto and assess whether this domain affects the proteolytic efficiency of the enzyme. The 1.35 Å resolution structure reveals domain swapping between two monomers. We provide evidence that the domain-swapped dimers exist outside of crystallization conditions. Higher amounts of the dimeric form of the protein are obtained with a mutation in the hinge region. In addition we could shift the equilibrium towards dimers by increasing temperature. We also demonstrate that the cytoplasmic domain isolated

from full-length enzyme form dimers. Overall this suggests the domain sharing occurs for the full-length protein and is not an artifact of crystallization. Furthermore, we have developed steady-state kinetic assays to compare the activity of full length GlpG (ecGlpG-FL) and its membrane domain using two model substrates, a water-soluble fluorescently-labeled casein and the membrane protein TatA from *Providencia stuartii*, (psTatA). Combined with a mutagenesis study, we conclude that with these model substrates, the cytoplasmic domain does not alter the conformation of the active site, its stability or accessibility.

**4-2. Results**

*4-2.1. Domain swapping in the crystal structure of the cytoplasmic domain of ecGlpG.*

The structure of ecGlpG-cyto, first crystal structure of a rhomboid cytoplasmic domain, was determined at 1.35 Å resolution using Se-Met substituted protein and MAD phasing (**Table 4-1**). The asymmetric unit contains a 74-residue monomer with an unusual elongated fold (**Figure 4-1A**), which is significantly different from the globular structure recently determined by NMR for the cytoplasmic domain of ecGlpG [25]. The dimension of the final model is 54 Å x 54 Å x 28 Å, with 67 of the 74 residues built into the electron density. Given the high resolution, automatic building was attempted, with almost the entire chain being traced. Although no hydrophobic core is observed, careful inspection of the electron density validated this main-chain trace. In this conformation, the protomer presents an α-helix/ß-strand fold, with 3 α-helices and 2 ß-strands, one of which spans over 11 residues (28-39, **Figure 4-1A**).

84

**Table 4-1. Data collection and refinement statistics.**

| Parameter | SeMet-ecGlpG- cyto | | Native ecGlpG-cyto |
|---|---|---|---|
| **Data collection** | Peak (E1) | Remote (E2) | |
| Space group | I422 | | I422 |
| Unit cell dimensions | | | |
| $a$ (Å) | 51.89 | 51.89 | 51.76 |
| $b$ (Å) | 51.89 | 51.89 | 51.76 |
| $c$ (Å) | 91.33 | 91.33 | 91.56 |
| $\alpha$ (°) | 90.0 | 90.0 | 90.0 |
| $\beta$ (°) | 90.0 | 90.0 | 90.0 |
| $\gamma$ (°) | 90.0 | 90.0 | 90.0 |
| | | | |
| mol/ASU | 1 | 1 | 1 |
| Wavelength (Å) | 0.9793 | 1.116 | 0.9795 |
| Resolution range (Å)[*] | 36.7-2.3 (2.38-2.3) | 36.7-2.3 (2.38-2.3) | 45.06-1.35 (1.39-1.35) |
| Unique reflections | 5309 | 5311 | 13985 |
| $<I/\sigma I>$[†] | 15.3 (4.2) | 24.5 (4.8) | 14.83 (2.07) |
| Completeness (%) | 99.9 (99.9) | 99.9 (99.9) | 98.6 (98.5) |
| Redundancy | 9.4 (9.5) | 9.3 (9.4) | 2.1 (2.0) |
| $R_{merge}$[‡](%) | 7.2 (23.4) | 7.4 (20.7) | 3.2 (48.1) |
| **Refinement** | | | |
| $R_{work}$ (%)[±] | | | 16.18 |
| $R_{free}$ (%)[§] | | | 18.20 |
| Number of atoms | | | 1068 |
| Solvent atoms | | | 66 |
| Acetate ion | | | 7 |
| Deviations from ideal (r.m.s.d.) | | | |
| Bond lengths (Å) | | | 0.014 |
| Bond angles (°) | | | 1.437 |
| Dihedral angles (°) | | | 16.056 |
| Ramachandran Plot | | | |
| Preferred (%) | | | 96.8 |
| Allowed (%) | | | 3.2 |
| Outliers (%) | | | 0 |
| **PDB code** | **4HDD** | | |

[*] Brackets indicate high resolution shell

[†] $<I/\sigma I>$ = ratio between mean intensity and the mean error of the intensity

[‡] $R_{merge} = \Sigma_{hkl} \Sigma_j | I_{hkl,j} - <I_{hkl}> | / \Sigma_{hkl} \Sigma_j I_{hkl,j}$ Where $<I_{hkl}>$ is the average of symmetry related observations of a unique reflection

[±] $R_{work} = \Sigma_{hkl} || F_{obs}(hkl)| - F_{calc}(hkl)|| / \Sigma_{hkl} |F_{obs}(hkl)|$. Where $F_{obs}$ and $F_{calc}$ are the observed and calculated structure factors

[§] $R_{free}$ is calculated in the same manner as $R_{work}$ on 5% of structure factors that were not used in the model refinement.
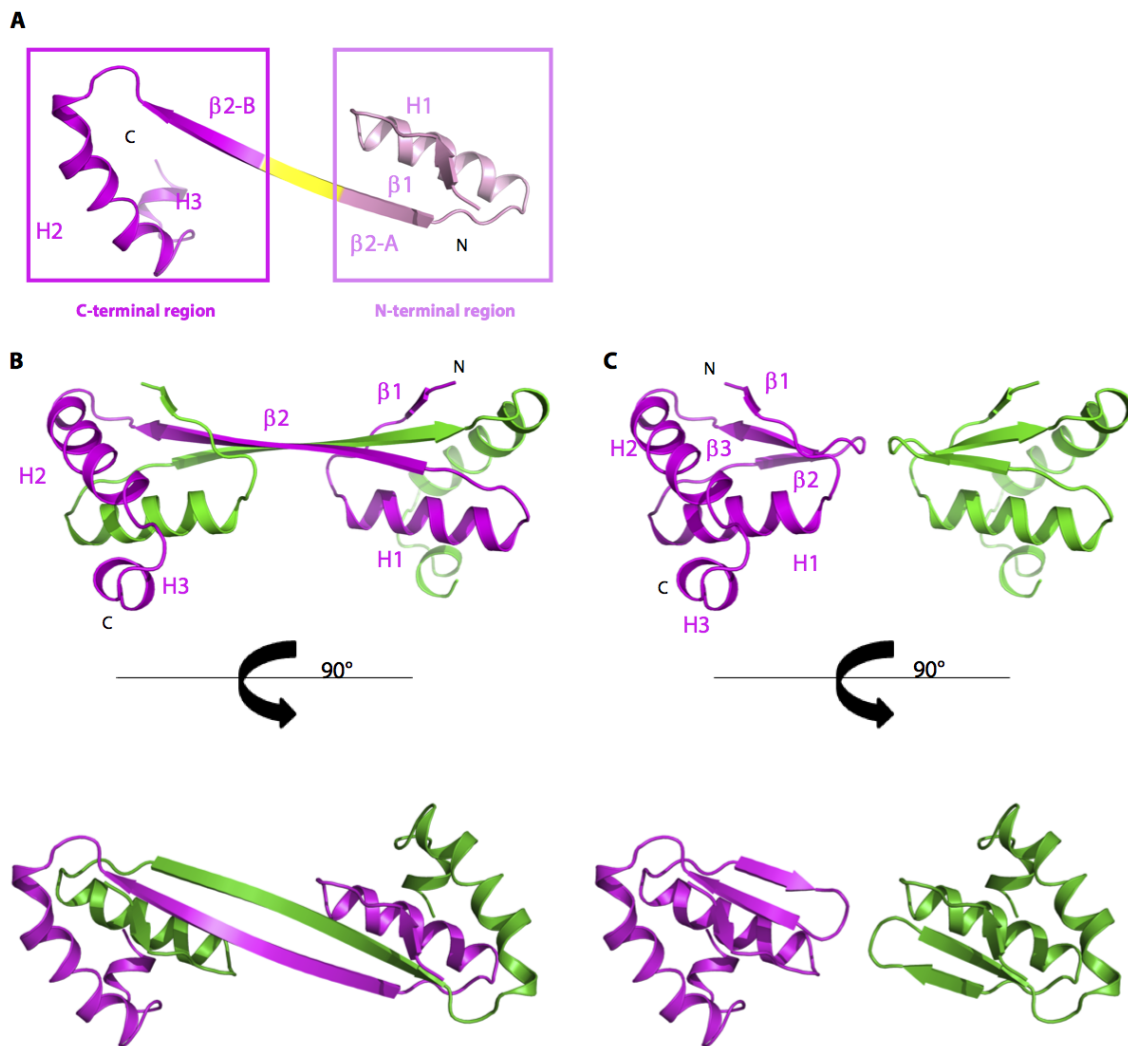
**Figure 4-1. Structure and representation of the 3D domain swapping of ecGlpG-cyto.**
**A.** Cartoon representation of asymmetric unit content of ecGlpG-cyto. The N-terminus region (light pink) and the C-terminus region (dark pink) are spanned around the hinge region 32-34, in yellow. The extended β strand is split into two parts β2-A and β2-B, which correspond to two distinct β strands in the globular monomeric form (β2 and β3, respectively) **B.** Cartoon representation of the dimer observed in the crystal structure with one protomer in green and the other in dark pink. **C.** Cartoon representation of a model for the globular monomeric form. The two protomers from the dimeric structure are shown in a globular monomeric form, with the secondary structure elements labeled for the dark pink monomer. These monomers were created assuming that the N-terminal region is exchanged during the 3D domain swap mechanism.

Generating the symmetry-related molecules helped to determine that, in the crystal, the cytoplasmic domain of ecGlpG is a dimer, in which the domains are swapped (**Figure 4-1B**). The main striking feature of the dimeric organization is the intertwined four-stranded anti-parallel ß-sheets, formed around the extended ß2 strands. Each extremity of the ß-sheet packs with the N- terminal helix H1 of one chain, as well as with helices H2 and H3 of the other chain. The hydrophobic core of each domain is mainly composed of helices H1 and H2 and the last 3 residues of ß2 strand (residues 37-39). Upon examination of the structure, the hinge region was identified being composed of residues 32-34, and a globular monomeric model was generated (**Figure 4-1C**). Using the NMR solution structure of ecGlpG-cyto recently deposited in the PDB (PDB code 2LEP, globular monomeric form), we performed a backbone dihedral analysis, which confirmed that residues 32-34 form the hinge loop (**Figure 4-2**). In the globular monomeric form, these residues connect ß2 and ß3 strands and, upon dimerization, they undergo structural rearrangement to join these strands together and form part of the extended ß-strand (ß2-A and ß2-B, respectively, **Figure 4-1A**).

As half of the protomer is swapped upon dimerization, either N-terminal or C-terminal regions (**Figure 4-1A**) could be exchanged. Moreover, domain swapping requires the disruption of a large number of interactions. The ß-sheet has to be completely remodeled; so numerous main-chain interactions have to be broken. Swapping of the helices (either H1 or H2-H3) would disrupt a large network of van der Waals interactions, exposing the hydrophobic core to the solvent. H-bonds network between side chains are observed at the dimer interface. For example, in the region surrounding the hinge, a large network of H-bonds is observed between residues S6, Q30 and H32 of one protomer, and D36 and Q34 of the other one (**Figure 4-3A**). In this dimeric conformation the extension of the ß-strand creates new main chain interactions as three
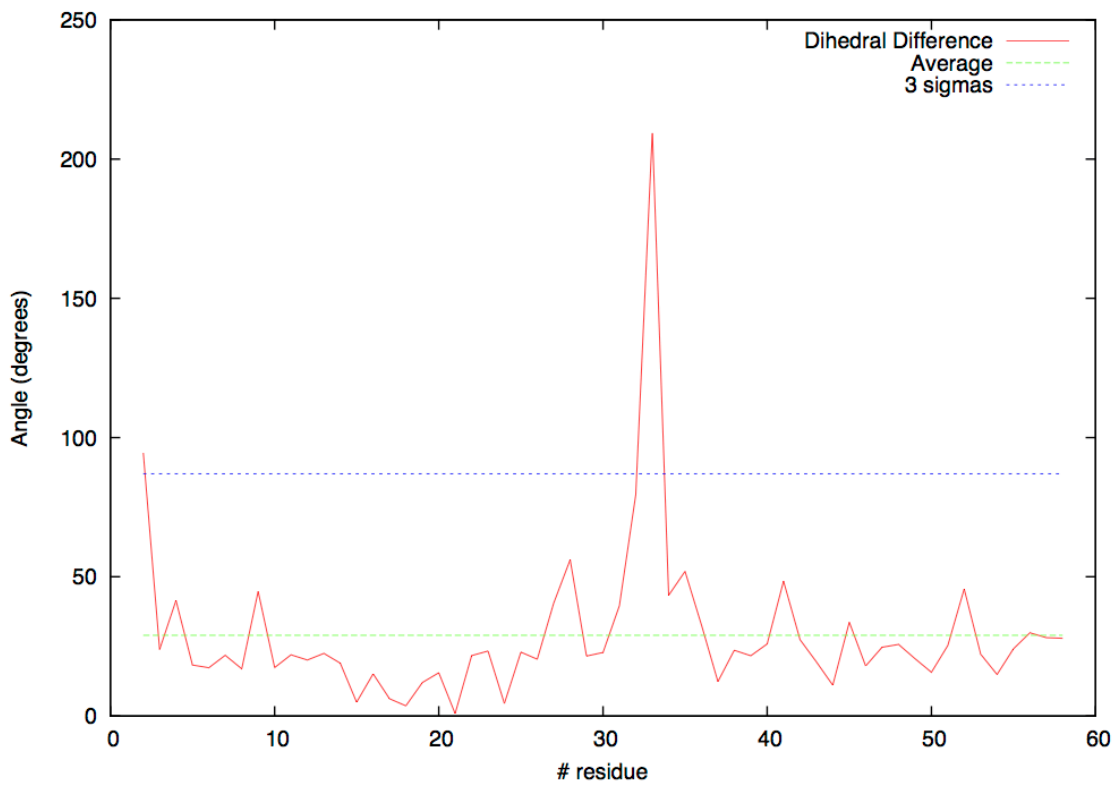
**Figure 4-2. Backbone dihedral analysis.** ecGlpG-cyto structure was subjected to backbone dihedral angle analysis.

more residues are in a β-strand conformation (residues 32-34). Furthermore, the crystal structure shows a bidendate interaction between side chains of the N33 in the hinge region and Q31, which is found in both protomers (**Figure 4-3B**), which is unlikely to be conserved in the monomeric form of ecGlpG-cyto. In our model of the monomeric form of ecGlpG-cyto, N33 is hydrogen-bonded to the side chain of S35 as well as the main chain of A8, but no longer interacts with Q31 (**Figure4-3C**). In the NMR models of the monomeric form of ecGlpG-cyto, these two residues do not interact.

Although the majority of prokaryotic rhomboids have an N-terminal cytoplasmic domain, there are only two NMR structures available to date. Structural alignment of the monomeric model of ecGlpG-cyto, generated from dimeric crystal structure, with the cytoplasmic domain of *P. aeruginosa* (paGlpG-cyto) [36] reveals an overall similar fold with the greatest variation in the N and C-terminal regions showing an r.m.s.d. of 4.6 Å (**Figure 4-4A**). Our globular model also superimposes well with the NMR solution structure of ecGlpG-cyto (1.2 Å for 46 Cα atoms superimposed) [25], with the exception of the C-terminal α-helix (**Figure 4-4B**). Dali search for structurally related proteins did not return any domain similar to our crystal structure of domain swapped ecGlpG-cyto. *E.coli* protein EscJ (PDB code 1YJ7) is the closest structure related to the monomeric form of ecGlpG-cyto, but exhibits a larger structural unit (186 residues for EscJ *vs.* 68 for ecGlpG-cyto).

*4-2.2. Domain swapping of E. coli rhomboid cytoplasmic domain exists in vitro.*

To determine if this structure is physiologically relevant purified ecGlpG-cyto (1 to 74 amino acids) was subjected to gel filtration. A monomer peak representing 95% of the protein injected was observed as was a dimer peak at retention time 12.2 ml reflecting the remaining 5%
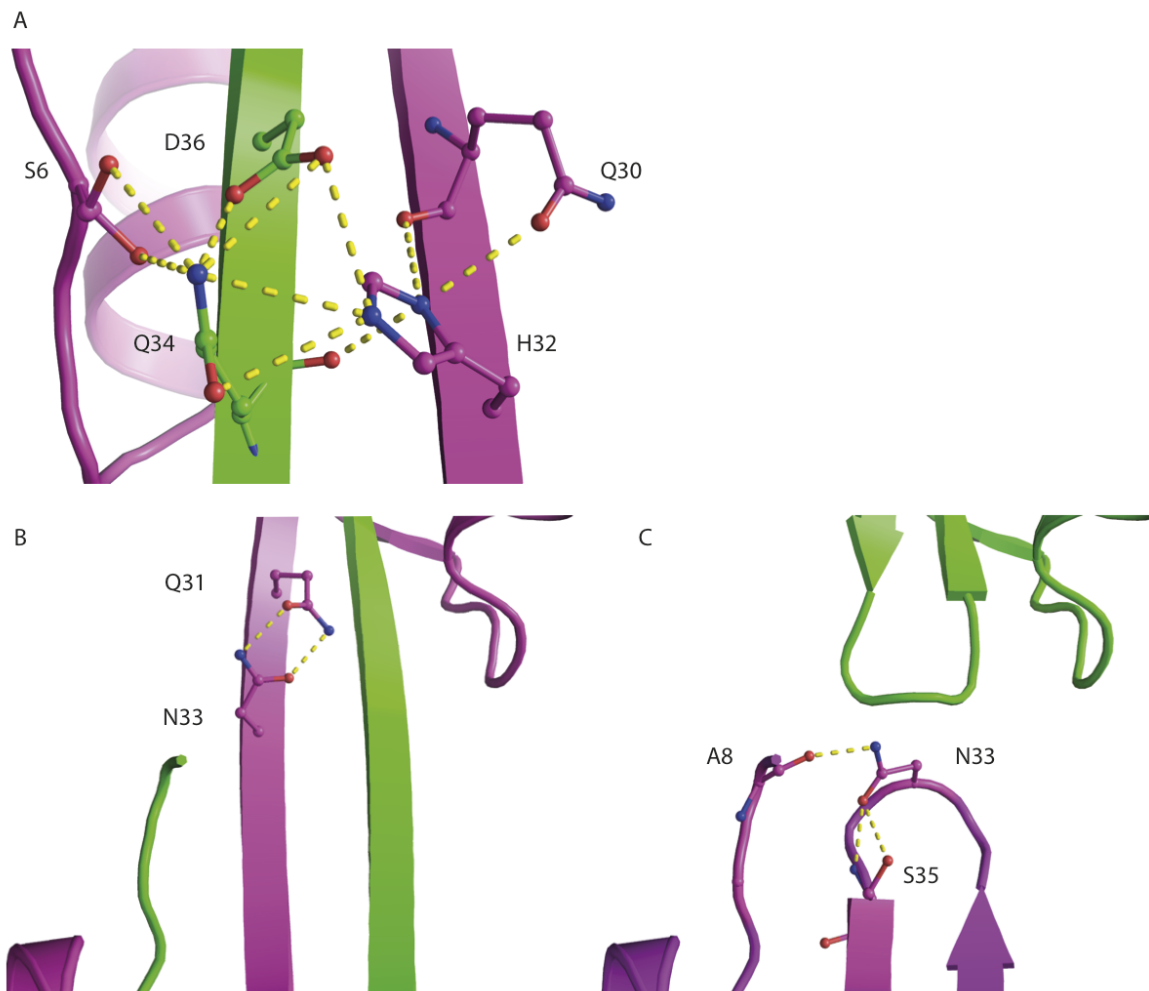
**Figure 4-3. Hydrogen bonds networks around the hinge region (32-34) for the dimer domain-swapped structure and the globular monomer model of ecGlpG-cyto. A, B**. Two views of a cartoon representation of the domain-swapped dimer with the same color code as Fig. 1. The hydrogen bonds networks are represented around Q34 and H32 (A) and N33 (B). The residues involved in hydrogen interactions are represented in sticks, with nitrogen in blue and oxygen in red. The hydrogen bonds are shown in yellow. The second bi-dentate interaction is under the second strand (green), and not shown. **C**. Cartoon representation of the two identical models of each globular monomer derived from the domain-swapped dimer structure. The hydrogen bonds network is represented around the hinge region. The position of these residues is derived from the model. For the clarity of the figures, helix 1 from the monomer in green and dark pink was removed in (B) and (C), respectively. As for Fig. 1, the figure was created assuming that domain N-terminal is exchanged during the 3D domain swap mechanism.

90
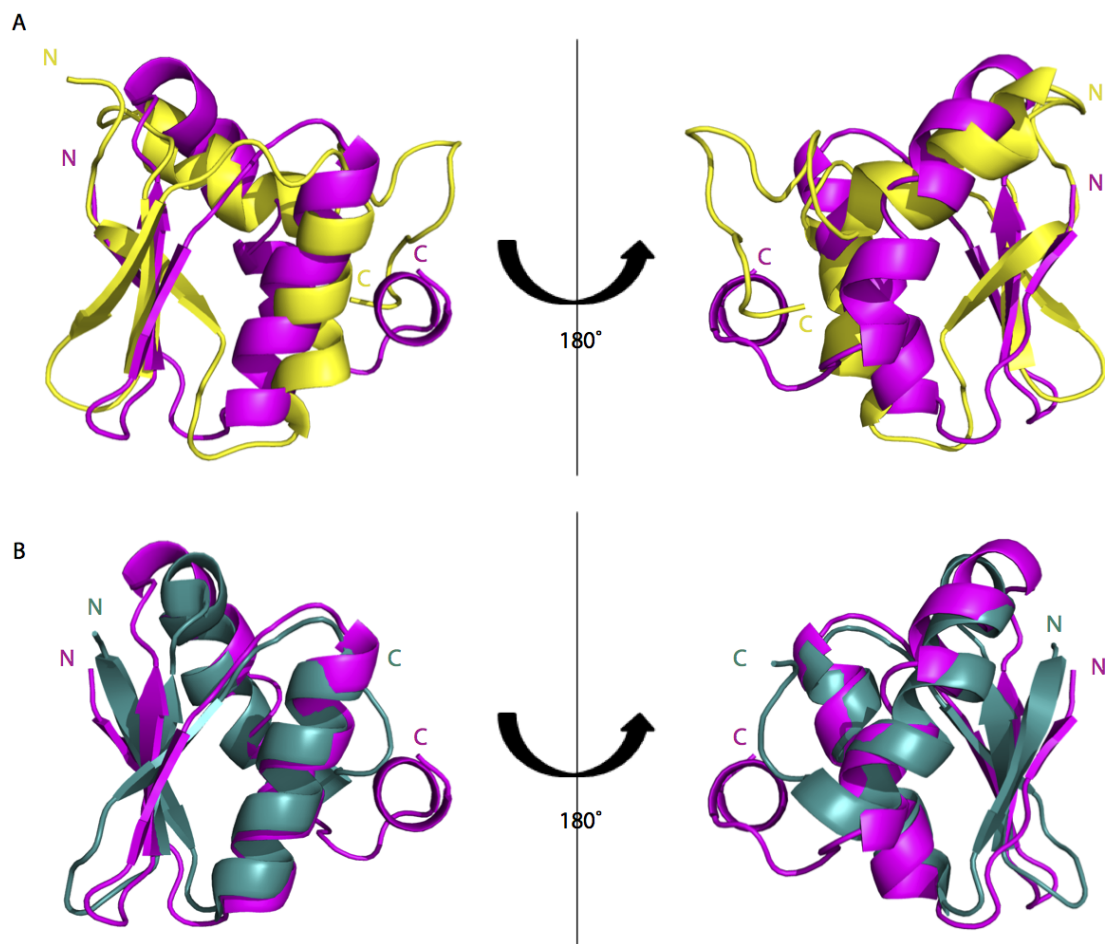
**Figure 4-4. Comparison of the globular structures of the cytoplasmic domains of rhomboids. A.** Superimposition of the globular monomeric model of ecGlpG-cyto in dark pink with the NMR model of minimal energy of the cytoplasmic domain of the rhomboid from *P. aeruginosa* in yellow (PDB code 2GQC). The r.m.s.d. value is 4.3 Å (45 Cα atoms superimposed). **B.** Superimposition of the globular monomeric model of ecGlpG-cyto in dark pink with the NMR model of minimal energy of the cytoplasmic domain of *E. coli* (PDB code 2LEP) in teal. The r.m.s.d. value is 1.2 Å (46 Cα atoms superimposed). Two views of the superimposed structures are presented, rotated trough 180°.

of the injected protein. SDS-PAGE analysis of both fractions confirmed the presence of the same protein in different oligomeric states (**Figure 4-5A, inset i**). Molecular weight estimates based on the gel filtration standard profile revealed mass 9 and 18 kDa suggesting a mixture of both monomers and dimers in solution. To confirm the gel-filtration observations, cross-linking experiments were undertaken. Purified cytoplasmic domain of ecGlpG was treated with crosslinkers – DSP and DTSSP **(Figure 4-5A, inset ii).** Cross-linked samples results in a second band at 18 kDa that was consistent with ecGlpG-cyto dimer.

The distinct feature of domain swapping is the large energetic barrier separating two oligomeric states or, more precisely, there is a high energy difference between the closed and open monomer[38]. The formation of domain-swapped oligomer from open monomers is not energetically demanding process, since open monomer and domain-swapped dimer share the same structure except for the hinge region. Therefore the conversion of monomers to dimers is a very slow process under purification conditions (4°C) and at room temperature. This energy barrier can be reduced under certain conditions, which favor a weakening of interdomain interactions and allow an open monomer to be formed. Elevating the temperature is one of the several approaches researchers use to shift this slow equilibrium and promote the dimers formation [39]. The protein sample with exactly the same concentration as for the previous SEC **(Figure 4-5A)** was incubated at 55°C for 2 and 14 hours and SEC was conducted. As represented in **Figure 4-5B** and **C**, the amount of dimers formed increased with temperature and reached 27% after 2 hours, and 52 % after 14 hours of incubation.

The oligomeric state of the cytoplasmic domain of ecGlpG isolated from the full length protein was also assessed. Full length ecGlpG was expressed and purified as described. The purified protein was incubated with immobilized α-chymotrypsin overnight to cleave between

the cytoplasmic and membrane domain. α-chymotrypsin resin was removed by centrifugation and the mixture of the membrane and the cytoplasmic domains of ecGlpG was subjected to gel-filtration. The cytoplasmic domain is partially degraded during the overnight incubation with immobilized α-chymotrypsin however the SDS-PAGE indicates there is sufficient protein to assess its oligomeric state. The SEC profile shows the presence of the membrane domain eluting at a retention time indicative of a dimeric form and two small peaks of the cytoplasmic domain that corresponded exactly to the retention times observed for dimeric and monomeric forms of the protein. The amount of cytoplasmic domain dimers in full-length protein was 20%, which is more than we observed after SEC of purified cytoplasmic domain (5%, **Figure 4-5D**).

Our previous work has demonstrated that full length ecGlpG forms dimers [40]. However we also showed that hiGlpG, a rhomboid that does not have a cytoplasmic domain, is also dimeric. The membrane domain of ecGlpG exists in dimeric state as well, as it was demonstrated by our gel-filtration results **(Figure 4-5D)**. To confirm this observation, we performed the crosslinking of purified ecGlpG-MD. We used both DSP and DTSSP, and the experiment resulted in a new band around 40 kDa corresponding to membrane domain dimers **(Figure 4-6)**. The crosslinking of ecGlpG-MD and ecGlpG-cyto were conducted at two protein concentrations – 1 and 0.1 mg/ml to eliminate the possibility of non-specific interactions of the molecules**.** This demonstrates that the membrane domain is dimeric on its own however it does not rule out the possibility that the cytoplasmic domain plays a role in ecGlpG full-length dimerization.

Another way to lower the energy barrier between monomers and dimers and shift the equilibrium towards dimeric fraction is to change the conformation of the hinge loop[39; 41]. Based on backbone dihedral analysis **(Figure 4-2)**, the residue Asn33 in the hinge loop was mutated to
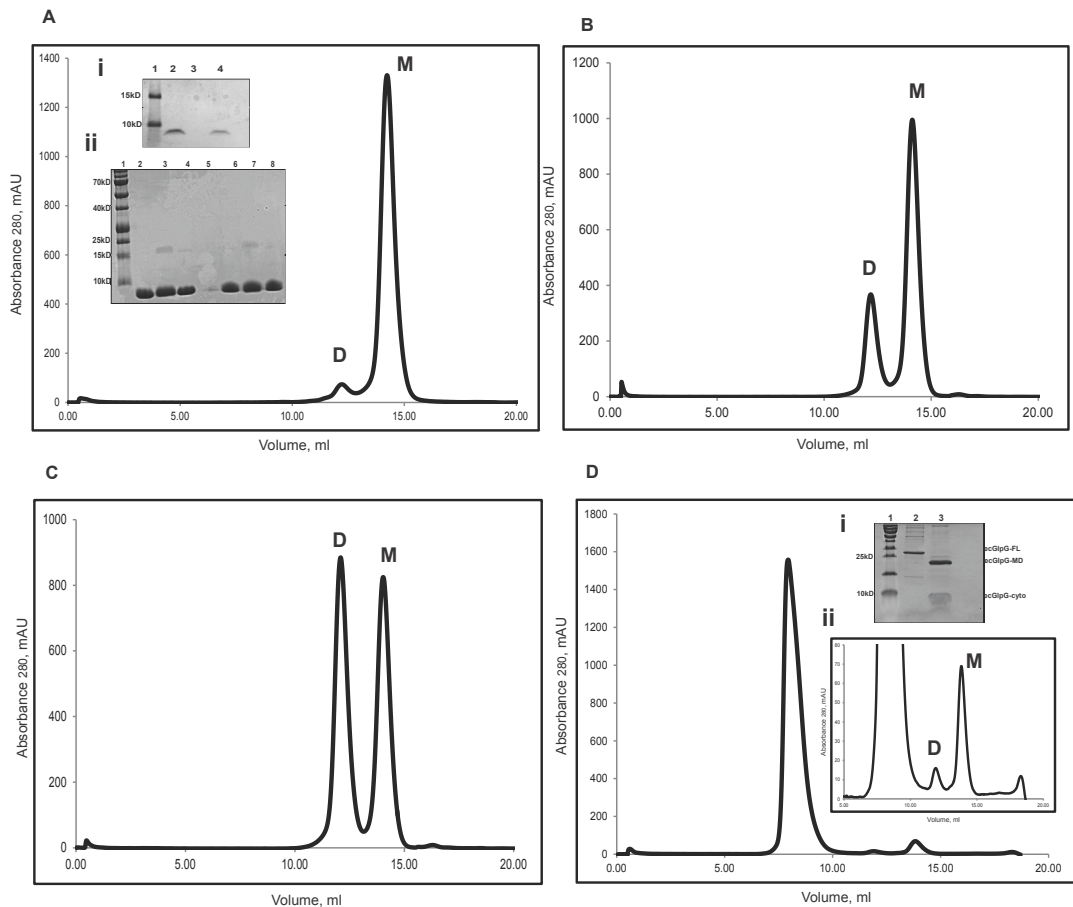
**Figure 4-5. Gel-filtration of cytoplasmic and membrane domains of ecGlpG.**
**A.** SEC of cytoplasmic domain of ecGlpG stored at 4°C. Inset i. SDS-PAGE of gel-filtration fractions. Lane 1, molecular mass markers; lane 2, the second peak fraction; lane 3, blank; lane 4, the first peak fraction. Inset ii. SDS-PAGE analysis of ecGlpG-cyto after DSP and DTSSP cross-linking. Lane 1, molecular mass markers; lane 2, ecGlpG-cyto not treated with DSP; lane 3, ecGlpG-cyto after incubation with DSP; lane 4, ecGlpG-cyto incubated with DSP followed DTT; lane 5, blank; lane 6, ecGlpG-cyto, no DTSSP; lane 7, ecGlpG-cyto treated with DTSSP; lane 8, ecGlpG-cyto incubated with DTSSP and DTT. M, monomer, D, dimer. **B.** SEC of ecGlpG-cyto after incubation at 55°C for 2h. **C.** SEC of ecGlpG-cyto after incubation at 55 °C for 14h. **D.** SEC of ecGlpG after cleavage of the membrane domain from the cytoplasmic domain with immobilized α-chymotrypsin. Inset i. SDS-PAGE analysis of the cleavage. Lane 1, molecular mass markers; lane 2, ecGlpG-FL; lane 3, ecGlpG after cleavage with α-chymotrypsin. Inset ii. The enlarged portion of the SEC curve. All SECs were performed on Superdex 75 (10/300) column (GE). Standards: thyroglobin, 8.3 ml, 670 kDa; r-globulin, 8.89 ml, 158 kDa; ovalbumin, 10.98 ml, 44 Da; myoglobin, 13.08 ml, 17 kDa; vitamin B12, 18.57 ml, 1.35 kDa.

proline. Proline residues play significant role in protein backbone conformation due to their unique properties [42] and substitution of amino acids to proline enhances the secondary structure stability by decreasing the configurational entropy of unfolding [44]. The SEC of N33P mutant of ecGlpG-cyto demonstrated that the amount of dimers increased up to 44%, showing that this mutation promoted domain-swapped dimer formation, due to probably a lower entropic barrier of protein-protein association **(Figure 4-7A)**. Both dimeric and monomeric forms were stable, as after 10 days of incubation at fairly low protein concentration (0.5 mg/ml) neither dissociation of dimers to monomers nor association of monomers to dimers was observed **(Figure 4-7B and C)**.

Thus we demonstrated that purified cytoplasmic domain exists in solution predominantly as monomers, with a very slow conversion to dimeric form. If the monomer is subjected to certain conditions, such as high temperature or mutation of the protein, it establishes faster equilibrium between monomers and dimers and stable dimers are produced. Our experiments strongly suggest that the observed dimers are domain-swapped.

*4-2.3.Optimization of protein purification and kinetic assay parameters*

We developed a kinetic assay to measure $K_m$, $k_{cat}$ and catalytic efficiency ($k_{cat}/K_m$) for ecGlpG to determine if the cytoplasmic domain plays a role in activity. This assay was based on commercially available BODIPY-FL-casein substrate (casein$^{Fluor}$). The idea of using casein was adapted from the literature [20; 37] with a more extensive analysis being conducted compared to previous studies. The ability of rhomboids to cleave water-soluble substrates as casein is a known paradox and can be probably explained by its partially unfolded structure and its
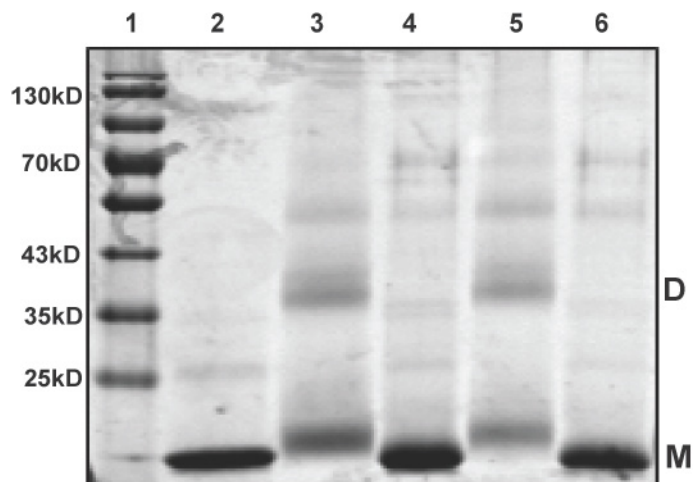
**Figure 4-6. SDS-PAGE analysis of ecGlpG-MD after DSP and DTSSP cross-linking.** Lane 1, molecular mass markers; lane 2, ecGlpG-MD no crosslinkers; lane 3, ecGlpG-MD after incubation with DSP; lane 4, ecGlpG-MD incubated with DSP and then with DTT; lane 5, ecGlpG-MD treated with DTSSP; lane 6, ecGlpG-MD incubated with DTSSP followed by DTT. M, monomer, D, dimer.
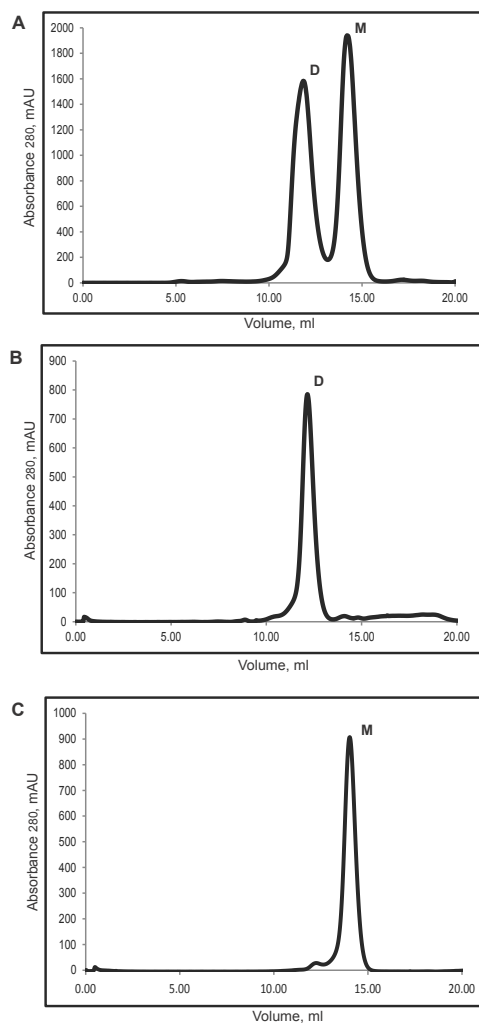
**Figure 4-7. Gel-filtration analysis of ecGlpG-cyto N33P**. A. The SEC of ecGlpG-cyto N33P mutant. The dimers and monomers fractions were pooled and stored at 4°C for 10 days and SEC was conducted separately for dimers (B) and monomers (C).

hydrophobicity. In addition rhomboids have been shown to have broad specificity in their substrate recognition motif [43]. While casein is not a native substrate, it provides a tool to analyze apparent kinetic parameters for the enzyme. The parameters for our assay were optimized prior to kinetic calculations including the minimal amount of enzyme that gave linearity between the amount of generated product and time, the time of reaction, the reaction mixture volume and the detergent concentration. Furthermore, to make sure that the cleavage of casein by rhomboid obeyed simple Michaelis-Menten model with zero or first order reaction, the cleavage products were revealed by SDS-PAGE gel showing only one major product around 12 kDa (**Figure 4-8A**). Similar parameters were optimized for a second substrate, psTatA. psTatA protein is a single spanning membrane protein and a component of twin arginine transport (Tat) system [44; 45]. It is the native substrate of AarA rhomboid from *Providencia stuartii.* It has been demonstrated that ecGlpG can process psTatA *in vitro* [43]. Using the psTatA substrate, kinetic parameters were assessed with a previously described gel-based analysis [26; 43; 46] however with further optimization of all parameters to ensure that the enzyme acts catalytically. Under these conditions, the time dependence of product formation was linear over a 3 hour interval.

In addition, we have optimized the purification protocol of ecGlpG to preserve its activity. It is known that endogenous phospholipids co-purifying with ecGlpG are important for protein stability and activity. It has also been shown that each chromatography step can remove up to 50% endogenous phospholipids from some membrane proteins causing precipitation out of solution [47]. Therefore all proteins used for the kinetic assays were not subjected to a gel filtration column. For the membrane domain preparation, immobilized α-chymotrypsin was used to separate it easily from the protein sample thus avoiding the gel filtration in purification protocol. Catalytic activity of both ecGlpG-FL and ecGlpG-MD are compared
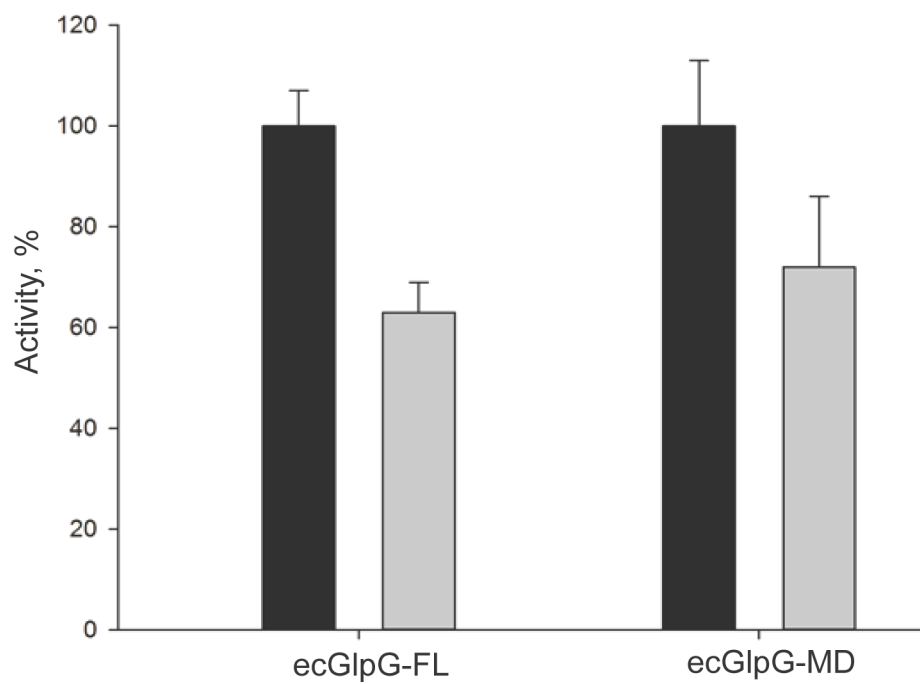
**Figure 4-8. Activity of ecGlpG-FL and ecGlpG-MD before and after gel filtration.** The $V_{max}$ of ecGlpG-FL and ecGlpG-MD is reflected as 100% activity before gel filtration (black bars). The percent of activity is also represented for both ecGlpG-FL and ecGlpG-MD after gel filtration (grey bars).

both before and after size-exclusion chromatography. Using Superdex-200 column, only 63% and 72% of initial activity remained for ecGlpG-FL and ecGlpG-MD respectively (**Figure 4-9).** This confirms that indeed gel filtration affects the properties of ecGlpG, most likely due to delipidation.

*4-2.4. Kinetic analysis using two model substrates reveals that the cytoplasmic domain does not affect the activity of full-length ecGlpG.*

To examine if ecGlpG-cyto plays a role in catalytic activity of protease, catalytic parameters including $K_M$, $V_{max}$ and $K_{cat}$ were determined. ecGlpG-catalyzed cleavage of casein$^{Fluor}$ follows Michaelis-Menten kinetics. A typical plot for these results is represented in **Figure 4-8B.** Analysis of residuals confirmed data obeyed Michaelis-Menten equation. The conversion of relative fluorescent units (RFUs) into μM of product was conducted as described in Materials and Methods. For ecGlpG-FL, apparent $K_M$ was $0.870\pm0.200$ μM, $V_{max}$ was determined to be $0.055\pm0.010$ μM/min. $k_{cat}$ value was $0.300\pm0.027$ min$^{-1}$, $k_{cat}/K_M$ value was $0.344\pm0.038$ min$^{-1}$μM$^{-1}$ (**Table 4-2A**).

This kinetic assay was also used to reveal if ecGlpG-cyto plays a role in catalytic event and explore the probability of its interaction with the membrane domain using mutagenesis. According to NSBI database, among 32 sequenced prokaryotic rhomboids, 29 have a similar N-terminal cytoplasmic domain consisting of approximately 90 residues. Alignment of these domains of 32 sequenced prokaryotic rhomboids **(Figure 4-10),** revealed their high similarity and 10 conserved residues were chosen to test such interaction (**Figure 4-11A**). All conserved

**A**



**B**
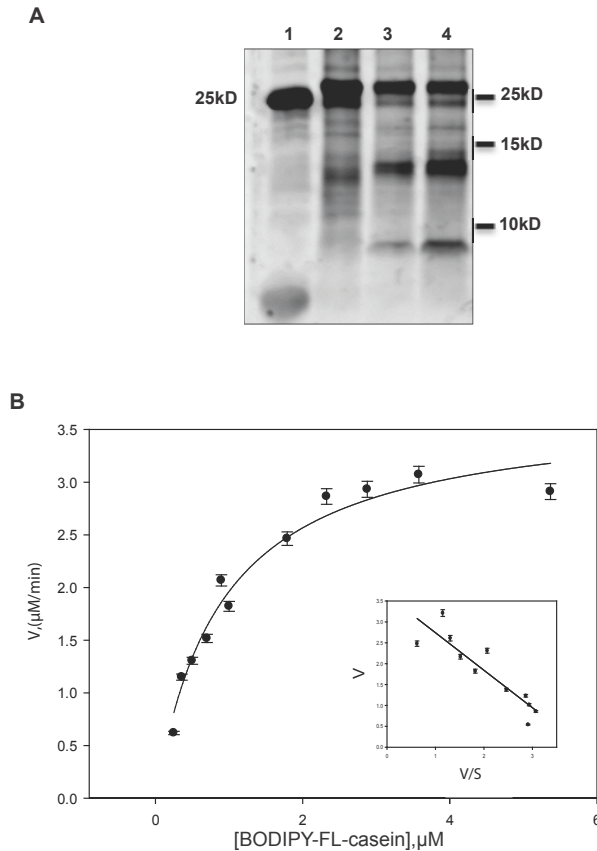


**Figure 4-9.  The cleavage assays of casein<sup>Fluor</sup> processed by ecGlpG. A.** The SDS-PAGE of casein<sup>Fluor</sup> cleavage by ecGlpG-FL Lane 1, fluorescent molecular mass marker; lane 2, casein<sup>Fluor</sup>, no enzyme; lanes 3 and 4, 0.179 μM of ecGlpG-FL was incubated with 0.179 μM (lane 3) or 6 μM (lane 4) of casein<sup>Fluor</sup> for 2 hours at 37°C. Non-fluorescent molecular mass markers are represented on the right. **B.** casein<sup>Fluor</sup> at the concentration range from 0.179 to 6 μM was reacted with 0.179 μM of ecGlpG-FL at 37°C for 2 hours. Velocity was determined as described in Experimental Procedures (n=4, mean±SD). Inset is Woolf-Augustinsson-Hofstee plot of the same data, which is linear ($r^2$=0.90).

residues were divided into two groups based on their location on two different surfaces of the protein molecule (**Figure 4-11B**). Alanine mutations of ecGlpG-FL were generated for all ten conserved residues. The kinetic parameters including $K_M$ and $k_{cat}$ and catalytic efficiency ($k_{cat}/K_M$) were assessed for all mutants using the BODIPY-FL-casein-based assay. The kinetic parameters determined for mutants of ecGlpG are represented in **Table 4-2B**. For all mutants, no significant change in activity was observed.

Concurrently, kinetic analysis of the ecGlpG-MD alone was performed. Surprisingly the catalytic parameters for the ecGlpG-MD are not significantly different in comparison to the full length protein and stay in the same range, with $K_M$ slightly increased to 2.9 µM (**Table 4-2A**). This finding is contradictory with previously published observations showing that the activity of membrane domain is lower than that of the full length protein and suggesting that the cytoplasmic domain enhances ecGlpG function [20; 25; 33; 48].

The kinetic analysis was also repeated for ecGlpG-FL and ecGlpG-MD using psTatA protein as a substrate. For this single pass transmembrane substrate, the apparent $K_M$ of ecGlpG-FL was found to be 2.65 µM±0.32 whereas $K_M$ of ecGlpG-MD was 3.85±0.46 µM ($p$=0.005). $V_{max}$ values for both proteins were determined to be roughly the same (2375±119 MGV/min for ecGlpG-FL and 3633.5±236.7 MGV/min for ecGlpG-MD) (**Figure 4-12**). These data correlate with our fluorescence-based assay demonstrating almost no change in $K_M$ and $V_{max}$ values for both the ecGlpG-FL and ecGlpG-MD.

Since ecGlpG is a dimer possessing two active sites, there is a possibility of cooperative substrate binding. To examine this possibility, the Hill coefficient was calculated. For ecGlpG-FL and ecGlpG-MD with the psTatA substrate, the Hill coefficient was 1.1 suggesting no

**Table 4-2. The catalytic parameters of ecGlpG-FL, ecGlpG-MD and ecGlpG-FL mutants.**
**A**. Comparison of catalytic parameters of ecGlpG-FL and ecGlpG-MD. The asterisks indicate the significant difference of ecGlpG-MD as compared to ecGlpG-FL (n=4, mean±SD; *, $p$=0.012; **, $p$=0.024; ***, $p$=0.004). **B.** Catalytic parameters of ecGlpG-FL mutants. n=2; mean±SD; for each mutant t-test was performed and showed no significant difference.

A

| Protein | $K_M$ ($\mu$M) | $k_{cat}$ (min$^{-1}$) | $k_{cat}/K_M$ (min$^{-1}$ . $\mu$M$^{-1}$) |
|---|---|---|---|
| ecGlpG-FL | 0.87 ± 0.20 | 0.307 ± 0.027 | 0.344 ±0.038 |
| ecGlpG-MD | 2.9 ± 1.13* | 0.418 ± 0.069** | 0.144 ± 0.078*** |

B

| | | | |
|---|---|---|---|
| ecGlpG M3A | 1.20 ± 0.16 | 0.253 ± 0.018 | 0.210 ± 0.043 |
| ecGlpG R11A | 0.93 ± 0.31 | 0.250± 0.039 | 0.268 ± 0.130 |
| ecGlpG Q14A | 1.26 ± 0.13 | 0.316±0.027 | 0.250 ± 0.046 |
| ecGlpG D18A | 0.95 ± 0.24 | 0.438 ± 0.048 | 0.461 ± 0.165 |
| ecGlpG T22A | 1.26±0.21 | 0.351 ± 0.031 | 0.278 ± 0.069 |
| ecGlpG Q30A | 1.04 ± 0.38 | 0.230 ± 0.030 | 0.220 ± 0.110 |
| ecGlpG W38A | 1.68 ± 0.27 | 0.416 ± 0.035 | 0.247 ± 0.060 |
| ecGlpG E42A | 1.27 ± 0.24 | 0.435 ± 0.038 | 0.342 ± 0.090 |
| ecGlpG R49A | 1.04 ± 0.39 | 0.230 ± 0.031 | 0.220 ± 0.111 |
| ecGlpG S68A | 0.74 ± 0.15 | 0.346 ± 0.029 | 0.467 ± 0.132 |

β1   H1   β2   H2   H3   *   TM1

Species labels (column 1):

E.coli
P.aeruginosa
P.multocida
V.cholerae
Shigella
Citrobacter
S.enterica
Klebsiella_sp.
P.stali
Pantoea_sp
Erwinia_sp.
Enterobacteriaceae
Serratia_sp.
P.atrosepticum
Y.pestis
Rahnella_sp.
P.asymbiotica
Vibrionales_sp.
Photobacterium_sp.
Shewanella_sp
Moritella_sp.
Psychromonas_sp.
Alteromonadales_sp.
Pseudoalteromonas_sp.
Gammaproteobacterium_sp.
G.agarilytica
Idiomarina_sp.
Rheinheimera_sp.
H.influenzae
P.Stuartii
B.subtilis_Ydca

TM1   TM2   TM3

TM4   TM5   TM6

**Figure 4-10. Sequence alignment of prokaryotic rhomboids.** The sequence alignment was performed using CLUSTAL-Ω server. The NCBI protein database accession numbers for aligned rhomboids; *E. coli* NP_756065; *P. aeruginosa* BAK91039; *P.multocida* NP_246377; *V.cholerae* NP_229758; *Shigella* ZP_08394383; *Citrobacter* ZP_04559220; *S. enterica* Q57IV1; *Klebsiella_sp.* ZP_08303075; *P. stali* ZP_08254696; *Pantoea_sp* ZP_07379711; *Erwinia_sp.* ADP10616; *Enterobacteriaceae* ZP_07952980; *Serratia_sp.* AEG30485; *P. atrosepticum* Q6CZL3; *Y. pestis* Q7CFX8; *Rahnella_sp.* YP_004210966; *P. asymbiotica* CAQ82228; *Vibrionales_sp.* ZP_01816437; *Photobacterium_sp.* ZP_01162446; *Shewanella_sp* YP_736023; *Moritella_sp.* ZP_01897275; *Psychromonas_sp.* ZP_01216866; *Alteromonadales_sp.* ZP_01611626; *Pseudoalteromonas_sp.* YP_004067872; *Gammaproteobacterium_sp.* YP_003810370; *G. agarilytica* YP_004432306; *Idiomarina_sp.* ZP_08622331; *Rheinheimera_sp.* ZP_08571587; *H. influenzae* NP_438776*; P. stuartii* P46116; *B. subtilis_Ydca* P96617; *B.subtilis_YqgP* BAA12519. Three transmembrane helices/regions prediction servers (TMHMM, HMMTOP, and TMPRED) were used to predict the membrane and cytoplasmic domains (present at the N and/or C termini) of the prokaryotic rhomboids. N and C termini cytoplasmic domains predicted by the three servers (and of maximal length) are highlighted in pink and orange respectively, and were used as the frontier with the membrane domains, highlighted in blue. For *E. coli* and *H. influenza*e, the known structures of the rhomboids were used to determine the two domains (2IC8 and 2NR9, respectively). For *P. aeruginosa*, *P. stuartii* and *B. subtilis_YqgP*, the various domains are the one previously predicted [18]. For the clarity of the sequence alignment, *B. subtilis_YqgP*, which contains N and C termini cytoplasmic domains, is not represented, because of its larger N terminal cytoplasmic domain (175 residues). For both domains, the secondary structure elements of ecGlpG are represented at the top of the sequence alignment. All prokaryotic rhomboids are predicted to contain six or seven transmembrane segments. A black star at the top of the sequence alignment represents the last amino acid modeled in the crystal structure of this study. For ecGlpG-cyto, the N-terminal and C-terminal regions, involved in 3D domain swapping, are represented on top of the sequence alignment with the same color code as Fig1. The hinge region (residues 32-34) is highlighted in yellow and yellow stars are positioned on top of the sequence alignment. In the membrane domain (blue), the highly conserved residues are highlighted in red with the WR motif located in loop1 and the catalytic dyad (S201 and H254). The catalytic residues are marked with a red star at the top of the sequence alignment.

# (a)

β1      H1      β2      H2      H3

```
E.coli                    ---MLKITSFANPRVAQAFVDYMATQ-GVILTI----QHN-QSDVLADISQ-AERVAELARFLENPAD-PRYLAASWQAGHTGSGL----HYRRYPP-FFAALR---------  91
Shigella                  ---MLKITSFANPRVAQAFVDYMATQ-GVILTI----QHN-QSDVLADISQ-AERVAELARFLENPAD-PRYLAASWQAGHTGSGL----HYRRYP--FFAALRERAGP----  95
Citrobacter               ---MLKITSFANPRVAQAFVDYMATQ-GVILTI----QHQ-QTDVLADISQ-AERVAELARFLENPGD-PRYLAASWQSGQTDSGL----HYQRYP--FLATLRERAGPVT--  97
S.enterica                ---MLKITSFANPRVAQAFVDYMATQ-GVILTI----QHN-QSDILADISQ-AERVVELARFIENPGD-PRYLAASWQSGQTNSGL----RYRRFP--FLATLRERAGPVT--  97
Klebsiella_sp             ---MLKITSFANPRVAQAFVDYMATQ-GIILTI----QHT-QSDVLADISQ-AGRVAELARFLENPAD-PRYLAASWQSGQTNSGL----RYQRFP--FFATLRHN-------  92
P.stali                   ---MKITQFNQPRMAQAFVDYMATR-GVTLRI----RES-HCVILDDISK-LEMVNELAQFVRDPNH-PRYQAASWHSGTTDSGL----RYERGN--IWANIRERAGP----  94
Pantoea_sp.               ---MMKITQFNQPRMAQAFVDYMATQ-GITLRI----HEN-HYVILDDISK-IAVVNELQQFLHDPNH-PRYQAASWHSGKTDSGL----RYERSN--IWANIRERAGP----  95
Erwinia_sp.               ----MKITHFTHLRPAQAFVDYMATR-GITLRI----QDG-GYTLLDDISQ-IGVVNELNQFIGDPNH-ARYQAASWHAGTTKSGI----HYQHTS--LLANLRERAGP----  94
Enterobacteriaceae        --MVKIISLSNPRLAQAFIDYMATQ-GVRLEL----VHNDEVVLLADISR-QAQVQELTRFLQEPLH-PRYQAASWHSGSTHSGL----KYSNFS--YLKSFTRQTGPL---  97
Serratia_sp.              --MVKIISLSNPRLAQAFVDYMATQ-GVRLEL----VHNDEVVLLADISR-QAQVQELTRFLQEPLH-PRYQAASWHSGSTHSGL----KYSNFS--YLKSFTRQTGPLT--  98
P.atrosepticum            ---MIRVIALSNPRLAQAFVDYMRTQ-QVHLDM----PQGHEAELLEDISTQ-LSKVEALEIFLRDPTN-PRYLAASWQTGSMDTGI----QYQRYS--YLQTLKQK-------  93
Y.pestis                  ---MTRVIVISNLRLAQAFVDYMATH-HVALEI----PDAQGVEILADIEQ-LSAVHELEQFLLDPLN-PRYQAASWQAGNVNSNL----PYQRFS--YLQTLRSQAGPLTLS 100
P.asymbiotica             -----------------------------MHP----HDERHVELLEDITK-LALVQELEKFTRDPLN-ERYQAASWQSGKISHSL----KYQNNL--NWHYLSRQAGP----  70
Rahnella_sp.              --MVRILNLNNPRLAQAFVDYMATQ-GFTLTM----HESQEVQLLEDISR-LEEAKQVALFVRDPLH-PRYQEASWKTGKAQPHL----KMSHTP--FLKTLRSR-------  93
Vibrionales_sp.           ---MIRLMVLDNPRLAQAFIDYMASR-QIPIQMS----EGEGRFALLLDQF-QVETAELNRFLSEPNH-KRYQAASWDMAETRKSH----FHYHT-PS-FMGMIKAKAGPVT--  100
Photobacterium_sp.        --MYKLAGFYNPRQAQAFIDYMAYR-GIEISMA----EPEGKFALLADQY-LVEATELNAFLLDPNN-KKYQAASWQVAESRTAK----FNYSN-PS-LINMVKAQAGPF---  99
V.cholerae                --MHKLTTFNNPRAAQAFIDYMAAH-HIEIQMM----DAGGQFTLVIQQH-IETAAELALFLENPYA-EKYQAASWEVADQKRPQ----FHYAS-PN-LLSLIKAK-------  95
Moritella                 --MIKIGVINNPRMAQAFVDYVLAVQ-KIDAKVA----QTDDG-FAICLPANTH-VEVAQELDLFLADPYQ-DKYQAASWDVADTRTAN----FSYGS-GN-MLSNFLMHAGK----  97
Psychromonas_sp.          --MFKLYYFQNLRIAQAFSDHLNTL-DIDNKIE----DEFN-YVLLLLKDQ-QEQAQHELDAFLENPNA-DKYLSASWSTGSTKQVRN----SAYDTHEN-LLNNFIAHA----  97
Alteromonadales_sp.       --MIKLGSLNNPRAAQGFIDYLKSQ-GLDGHLH----TDGSHVIIVSEIH--FHQVQPLWAEFTQNPGH-SRYQQASWSVGGTQSPL----KYQGQSLNLKARFKALSWLNQ--  100
Pseudoalteromonas_sp.     --MIKLGSLTNPRAAQGFIDYLKSQ-GFSGQIR----EDGHTLVIVAPID--FHAVQPLWNEFAKNPNH-ERYQQASWQVGSTQSPL----KYQGQSLNLVARFKALSWLNQS-  101
Shewanella_sp             --MIIGRLPNPRAAQAFVDYLKGE-QIDCQIQ----LPQGVAILVIRDCD--AEQAKEFAHFIANPYD-NKYLQASWEHGDTHTKL----DYGAPSLQLFAQFITGAGPVT--  100
Gammaproteobacterium_sp.  ------MSFAHPRAAHAFGDYLTSL-GIRNRIV----TDVG-VQVLLDIHY-RAAFEEFQRFWDNPND-ERYLAASWSVGGSTEPVDY----ALLGRSAGRRSGGFALRRM---  95
Rheinheimera_sp           ---MIKIGDFASSAALSLADYCKSQ-QLDVSVV----HSAEQAELVCAPIH--AELVAELERFLLNPYA-DRYQAAAWDRAEVVASG-----TGSAG---LSGYWQQ-------  91
G.agarilytica             MSMPFKLVSFNQERAARLLTNYLNSK-GIVAQY----FHSGEHAVQLLNISD-KDVADIASEFIQFPQD-KKYQQAAWDSSQTVTLSD-RNSGVSLNGMWQQLRDIPVTTS--  105
P.multocida               ---MRYLFGSEIPTLAFRFRDYIPAQFNVEVELREIQENGQQIQLLPQIAPHFEAIAEVPLFLADPFH-TKYEQASWQRGDTQSTTTSLKGMLQQSGLTLSALRSRSRTGP--  110
P.aeruginosa              -MSAVKVLKFPLSVDLAGFVGLLRRL-NVPHRV----EESGQQVLVPDIRL-AEQVELYRRYPEGDPQ-ATLEAAFRPAGQG---------------FVRQLRMSPMT----  88
Idiomarina_sp.            ---MKKLFSSTDPQQIAMLQQIMQRS-GVPVTV----VERGSALELLIQIGY-EPVAEIIREFKANPPNLDDFIHQQNAQPKQPQGP-----------ALWKILARQAG-----  91
```

        ←→ Surface 1

    ←- - - - - - - - - - - - → Surface 2

106

**Figure 4-11. Mutants design of ecGlpG-FL. A.** Sequence alignment of the N-terminal cytoplasmic domains of prokaryotic rhomboids. For the clarity of the sequence alignment, *B. subtilis_YqgP* is not represented as the N-terminal cytoplasmic domain is longer than the other ones (175 residues). Secondary structure elements of ecGlpG are represented at the top of the sequence alignment. **B.** Cartoon representation of nine of the ten conserved residues on the two different surfaces of ecGlpG-cyto homodimer. Surface 1, blue and surface 2, yellow.

cooperativity between the two active sites. With the soluble substrate casein^Fluor the Hill coefficient ranged from 1.3 to 1.4 for ecGlpG-FL, ecGlpG-MD and all mutants.

**4-3. Discussion**

In the past ten years, we have gained a great deal of knowledge on the rhomboid protease family, stemming from developmental genetics, biochemistry, biophysics and protein crystallography studies. Despite these advances, most studies have focused on the action of the membrane domain. As described earlier, the majority of prokaryotic rhomboids possess a conserved N-terminal soluble domain with unknown function.

In this paper, we have presented a novel example of domain swapping in the crystal structure of ecGlpG-cyto. This is a case of bona-fide domain swapping, as structures of both the globular monomer [25] and domain-swapped dimer (this study) have been determined. Approximately half of the domain is swapped; the newly created interface is rather small, only comprising residues from the hinge loop (32-34). However, this structural rearrangement results in a striking new structural feature: an intertwined four-stranded β-sheet, which bridges the two domains and creates a completely redesigned molecular surface. Given the conservation of the N-terminal cytoplasmic domain in prokaryotic rhomboids, domain swapping may be a common structural feature.

The gel filtration confirmed that ecGlpG-cyto exists in both monomeric and dimeric forms. The slow equilibrium and large activation energy separating the two oligomeric states, as we have observed for ecGlpG-cyto, are common features of domain swapping[49]. Multiple factors can affect the activation energy between monomers and dimers - the greater entropy makes monomers more thermodynamically stable, the conformational changes observed upon domain
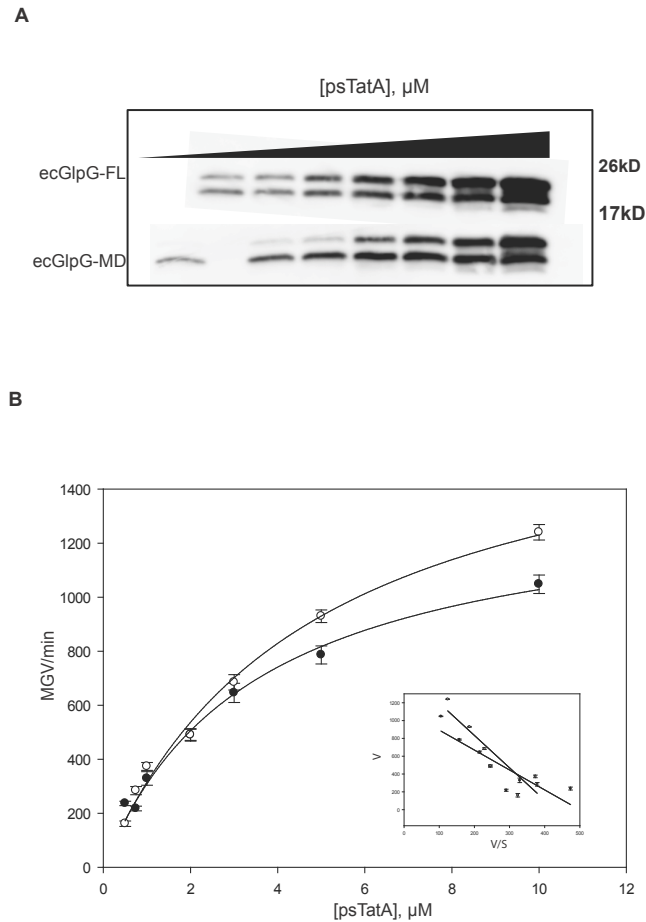
**Figure 4-12. Characterization of ecGlpG-FL and ecGlpG-MD activity with psTatA. A.** Western blot analysis of rhomboid cleavage activity on psTatA. psTatA at the concentration range of 0.5 to 10µM was reacted with 0.3µM of ecGlpG-FL or ecGlpG-MD at 37°C for 2 hours and the cleavage product was separated from the uncleaved substrate with SDS-PAGE. Molecular markers are represented for the top gel only. **B.** Typical formation of cleaved psTatA as a function of psTatA concentration in the presence of ecGlpG-FL (black dots) or the ecGlpG-MD (open dots). The data were obtained from upper gel and show Mean Grey Value (MGV) per minute. Inset represents Woolf-Augustinsson-Hofstee plot of the same data with $r^2=0.90$ for ecGlpG-FL and $r^2=0.92$ for ecGlpG-MD.

swapping in the hinge region may affect the energy needed to switch from one oligomeric state to another; therefore modifying the environmental conditions and the hinge loop may alter the equilibrium between monomers and dimers[41; 50; 51]. We were able to shift the equilibrium towards dimeric state with an increased temperature and with substitution in the hinge loop residue Asn 33 to Pro.

The intriguing remaining question is the biological relevance of domain-swapped dimers of the cytoplasmic domain. Here we showed that when the cytoplasmic domain is separated from the membrane domain, 20% is in a domain-swapped dimeric conformation. Besides, the ratio of monomer/dimer can be significantly altered by the experimental conditions. Preliminary studies from the Goto group[52] showed that the presence of detergents, structurally similar to lipids of the membrane, facilitated the conversion between the two oligomeric states. As domain swapping often occurs via partially unfolded state [52 53], the membrane lipids in close proximity to this domain could help in disrupting the large hydrophobic core in the monomeric form of the cytoplasmic domain to promote such an extended structural rearrangement.

Recently, it was shown that several prokaryotic rhomboids, including ecGlpG-FL, exist as dimers *in vitro* [40]. In this paper, we also demonstrated that both the membrane and the cytoplasmic domains are able to dimerize on their own. Based on these results, dimerization of ecGlpG-FL probably occurs through both domains. It is tempting to speculate that a metamorphic switch between two oligomeric forms of cytoplasmic domain of ecGlpG-FL could be a regulatory process.

In this study, to assess the possible role of the cytoplasmic domain in the regulation of catalytic activity, we developed optimized kinetic assays using two model substrates – water-

soluble fluorescently labeled casein and the membrane protein psTatA. The advantage of our fluorescent assay is its high reproducibility and ability to follow real time kinetics of the reaction. Also the optimization of all parameters makes this assay a powerful tool for deciphering catalytic activity of rhomboids. The function of ecGlpG-cyto was the subject of discussion in previous studies [24; 25; 36]. Fluorescent and gel-based assays were also used to monitor the rhomboids activity [20; 37; 43], but they were performed at a single substrate concentration, and an excess of the enzyme. Another approach to assess the activity of rhomboid protease was undertaken using rhodamine-tagged fluorophosphonate – the chemical probe that reacts specifically with serine hydrolases. Using this compound, the accessibility of the active site can be monitored but not the steady-state kinetics [25]. These studies, performed on various rhomboids, showed that removal of the cytoplasmic domain results in decrease of activity. One hypothesis to explain this phenomenon is a possible interaction between the two domains [20; 36]. To reveal the putative residues of the cytoplasmic domain involved in this interaction, 10 ecGlpG-FL mutants were generated on the basis of an alignment of the 32 amino acid sequences of prokaryotic rhomboids. The catalytic parameters were assessed for all mutants, in parallel with the purified membrane domain of ecGlpG. Surprisingly, both assays showed almost no difference in $K_M$ and $V_{max}$ values for the membrane domain in comparison to ecGlpG-FL. The catalytic parameters for all mutants were not altered as well. This contradiction with the previous results could be explained by the fact that conditions for steady-state kinetics were not satisfied in most studies. We also optimized the purification protocol as we confirmed that gel filtration, a purification step widely used in previous studies, reduces the activity of ecGlpG, probably due to delipidation [47]. We were able to avoid this step by incorporating an immobilized chymotrypsin step in the purification of the

membrane domain, thus enabling a direct comparison of the ecGlpG-FL and ecGlpG-MD activities.

Previously studies also proposed that the cytoplasmic domain could modulate the activity of the enzyme via its interaction with the membrane domain. This interaction could alter the position of the membrane domain within the lipid bilayer or tilt its orientation and thus increase active site accessibility[36]. From the kinetic studies presented here, we can conclude that the cytoplasmic domain neither affects the structural integrity nor accessibility of the active site during the catalysis, but an interaction between the two domains remains possible.

The natural substrate for *E. coli* rhomboid is yet to be identified; therefore our comprehension of ecGlpG-cyto biological function is incomplete. One possible role of ecGlpG-cyto could be substrate binding and thus serving as an exosite in a similar manner to the cytoplasmic domain of mammalian rhomboid RHBDL4 [35]. The presence of one or even several exosites - the secondary binding site that captures the substrate and directs its proper orientation - was proposed to be a common feature of the rhomboid family [35; 54]. Another possible role of cytoplasmic domain could be binding site for other cytoplasmic proteins. Similar to ecGlpG, the human erythrocyte anion exchanger (AE1) consists of a membrane domain and soluble cytoplasmic domain and exists as a dimer. Furthermore both domains dimerize and function independently [55]. The membrane domain transports bicarbonate while the cytoplasmic domain links the protein to the cytoskeleton. In addition, the cytoplasmic domain mediates the formation of higher oligomers, which is a regulatory process in cell aging. The shift towards the higher oligomeric forms caused by alterations in the cytoplasmic domain, like sulfhydryl oxidation or a change in the interaction with hemoglobin or the cytoskeleton. Further analysis is needed for the identification of ecGlpG-cyto binding partners.

Domain swapping is an elegant mechanism to create new structural features of the same protein which can regulate molecular functions via signal sensing, allostery [56], or even cause disease related aggregation [57]. For ecGlpG-cyto, upon domain swapping, a newly exposed surface is created. This event could facilitate dimerization, or expose a new exosite for substrate and/or regulatory molecule binding, thus providing a role for this complicated and energetically demanding mechanism of domain swapping. It remains to be determined if this domain swapping feature has a functional role. In the future, it would also be interesting to measure the kinetics of ecGlpG once its native substrate(s) is(are) discovered. Our kinetic assay could be used to assess the catalytic parameters of other rhomboids including eukaryotes, which opens future avenues in intramembrane proteases research.

## 4-4. Materials and Methods

### 4-4.1. Crystallization, structure determination and refinement of ecGlpG-cyto.

For crystallization purposes, different truncations of the cytoplasmic domain of ecGlpG were cloned and purified; the intact cytoplasmic domain (ecGlpG$^{1-91}$), ecGlpG$^{1-83}$, ecGlpG$^{1-74}$ and ecGlpG$^{1-60}$. Each purified protein was concentrated to 20mg/ml and initial crystallization screening was carried out using a Gryphon crystallization robot (Art Robbins Instruments, USA). ecGlpG$^{1-74}$ was the only construct yielding well diffracting crystals. Selenomethionine substituted ecGlpG$^{1-74}$ was produced in BL21 cells and purified as previously described. Selenomethionine-substituted ecGlpG$^{1-74}$ was crystallized at room temperature using hanging drop vapour diffusion and crystals were obtained in 0.49 M sodium phosphate monobasic monohydrate and 0.91 M potassium phosphate dibasic, pH 6.9. Data for a multiple-wavelength

anomalous dispersion experiment were collected at two wavelengths from a single crystal on the beamline 12.3.1 at the Advanced Light Source (Berkeley, U.S.A). Intensities were processed and scaled using XDS package (Kabsch, 2010). AutoSol Wizard from Phenix (Echols *et* al., 2010) was used to generate experimental phases. Two selenium sites were found by HYSS and phases were then calculated with Phaser and improved via density modification (RESOLVE). Automated model building was carried out with RESOLVE and 63 out of 74 residues were successfully fitted in the density with their side chains.

Selenomethionine-substituted ecGlpG[1-74] crystals also grew in 100 mM sodium acetate trihydrate pH 4.6, 0.2 M ammonium sulfate, 25% PEG 4000. Diffraction data were recorded to a resolution of 1.35 Å at CLS beamline 08ID-1 (Saskatoon, Canada). Intensity data were processed with XDS package [58] and Friedel pairs were kept separate as anomalous intensity difference was expected at a wavelength of 0.9795 Å. Refinement process consisted of energy minimization and individual anisotropic B factor refinement for protein atoms and isotropic B-factor refinement for solvent atoms using the program PHENIX [59] Experimental phases restraints were used throughout the refinement. Model building was carried out with Coot [60] between refinement cycles. The current refined atomic model comprises 67 residues and 66 water molecules, lacking the N-terminal residues 68-74. The quality of the model was checked using the MolProbity server [61]. The structure factors and coordinates were deposited in the Protein Data Bank (accession code 4HDD).


*4-4.2. ecGlpG proteolytic activity assay using BODIPY FL casein as a substrate.*

The reaction mixture contained 0.179 to 8.95 μM of BODIPY FL casein (Invitrogen, Inc.), reaction buffer (50mM Tris-HCl, pH 8.0, 150 mM NaCl, 10% Glycerol, 0.1% DDM) and 0.179

µM of ecGlpG. The total volume of reaction, the concentration of detergent, enzyme concentration and the time of reaction were optimized. The substrate was mixed with the reaction buffer and incubated at 37°C for 1 hour in the dark. The reaction was started with the protease. Fluorescence emission at 513 nm was measured at 37°C every 5 min during 2 hours in FluoStar fluorescence microplate reader with an excitation wavelength of 503 nm. Fluorescence detection of each substrate concentration without enzyme was used as a negative control. The linear correlation between the emitted fluorescence and the amount of product was verified. SigmaPlot was used for data analysis as well as statistical analysis (One-way ANOVA). Hill coefficient was determined by fitting the V vs substrate concentration data with Hill equation. To convert the fluorescence units of generated product into µM, the cleavage reaction was performed under the same conditions, using the highest and the lowest substrate concentrations; the cleaved products were resolved with SDS-PAGE, visualized with Luminescent Image Analyzer, the amount of appeared product in µM was calculated and compared to fluorescence produced for the same substrate concentration.

 *4-4.3. ecGlpG proteolityc activity assay with psTatA substrate.*

ecGlpG cleavage reaction mixture consisted of 0.5 to 10 µM of psTatA substrate mixed with 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 10% Glycerol, 0.1% DDM buffer and 0.3 µM of ecGlpG enzyme. The substrate was mixed with the reaction buffer first and incubated at 37°C for 1 hour for equilibration. The cleavage reaction was started with ecGlpG. The concentration of ecGlpG and the time of incubation were selected by carrying out cleavage reactions over several time intervals with various amounts of protease and with substrate concentration at saturation level. Enzyme concentration that gave linear product formation over time was then used for the kinetic

studies. After incubation for 2 hours at 37°C, the reactions were stopped by adding 4X SDS-Sample buffer. The same concentrations of substrate with no protease were used as negative controls and showed no degradation after the time of reaction. Protein samples were revealed with Western-blot using anti-His antibodies (Sigma-Aldrich, USA). The assay was repeated at least 3 times to ensure reproducibility. For all images, digitization was carried out with ImageQuant LAS 4000 (GE Healthcare, USA). Quantification was carried out with ImageQuant software.

*4-4.4. Expression and purification of ecGlpG-FL*

The gene of ecGlpG was cloned into pBAD-Myc/HisA plasmid (Invitrogen, Canada), having C-terminal tobacco etch virus (TEV) protease cleavage site, Myc-epitope and His$_6$-tag. TOP10 chemically competent cells were transformed with plasmids bearing genes of either wild type or the mutants of ecGlpG. The purification for all proteins was performed in the same manner. The protein was induced with 0.002% arabinose and expressed at 24°C for 6 hours in LB media. The cells were harvested esuspended in 50 mM Tris-HCl pH 8.0, 150 mM NaCl and lysed using an EmulsiFlex (Avestin Inc, Ottawa, Canada) apparatus at 15,000 psi. Unbroken cells were removed with a centrifugations step at 16,500 rpm (22,320 $g$), and the membranes were isolated by ultracentrifugation at 35000 rpm (95,800 $g$) for 2h. The membranes were solubilized in 50 mM Tris-HCl pH 8.0, 300mM NaCl, 10 mM imidazole, 20% glycerol, 1% (w/v) DDM and applied onto a Ni-NTA column (Qiagen, Ontario, Canada). The proteins were eluted with 250-500 mM of imidazole, 50 mM Tris-HCl pH 8.0, 300 mM NaCl, 20% glycerol, 0.1% DDM. The His-tag was removed by TEV protease (1 mg per 100 mg of protein, overnight, 16°C) and subsequent Ni-NTA column was performed to remove

uncleaved protein and TEV protease. The flow-through was collected and concentrated using 30,000 Da MWCO concentrators (Millipore, USA). The protein samples were kept at -80$^{\circ}$C.

*4-4.5. Purification of ecGlpG-MD.*

TOP10 cells were transformed with pBad Myc/HisA plasmid bearing the gene of ecGlpG minus 6 amino acids on the C-terminus with TEV cleavage site and His6-tag upstream of the N-terminus of ecGlpG. The protein was expressed and membranes were isolated in the same manner as the full length protein. After the first Ni-NTA column the N-terminal domain and few amino acids on the C-terminus were removed using immobilized chymotrypsin (ProteoChem, USA) (0.05 μg of chymotrypsin per 1 μg of protein) at room temperature overnight. Chymotrypsin was separated from the protein sample by centrifugation at 2000 rpm for 2 min. The cytoplasmic domain was removed by Ni-NTA column, the flow-through containing the membrane domain was collected and concentrated.

*4-4.6. Purification of ecGlpG-cyto.*

The coding sequence for cytoplasmic domain of ecGlpG was amplified from the same construct as described above using PCR. The obtained gene contained residues 1 to 74 of N-terminal part of ecGlpG. pBAD Myc/HisA vector was used for protein expression in BL21 cells. The protein was induced with 0.2% arabinose and expressed for 4 hours at 37$^{\circ}$C. The cells were homogenized in 50 mM Tris-HCl pH 8.0, 150 mM NaCl and lysed using an EmulsiFlex apparatus (Avestin Inc, Ottawa, Canada) 15,000 psi. After centrifugation at 16,500 rpm (22,320 *g*), the supernatant was collected and applied onto a Ni-NTA column. The protein was eluted with a gradient of 20-350mM of imidazole in 20 column volumes, 50

mM Tris-HCl pH 8.0, 300 mM NaCl. The His-tag was removed by TEV protease (1 mg per 100 mg of protein, overnight, 16$^{\circ}$C) and a subsequent Ni-NTA column was performed to separate uncleaved protein, contaminants and TEV protease. The flow-through was collected and concentrated, using 3,000 Da MWCO concentrator (Millipore, USA). The protein was stored at -80$^{\circ}$C.

### 4-4.7. Cross-linking of the membrane and cytoplasmic domains of ecGlpG.

The protein sample was dialysed against PBS, pH 7.4 buffer. DSP and DTSSP (Pierce Protein Research Products, Thermo Fisher Scientific, Rockford, USA) were dissolved in DMSO and water respectively to 25 mM and added to the protein sample to the final concentration of 4 mM. The sample without cross-linker was treated as a control. The protein concentrations used in the sample were either 1 or 0.1 mg/ml. For the latter concentration the protein sample was concentrated to 1 mg/ml prior loading on SDS-PAGE gel. The reaction mixture was incubated at room temperature for 30 min. 1 M Tris pH 7.4 was added to the samples to final concentration of 50 mM to stop the reaction. To cleave DSP and DTSSP the samples were incubated with 1 M DTT (the final concentration 10-50 mM) at 37$^{\circ}$C for 30 min. All samples were then subjected to SDS-PAGE under non-reducing conditions.

### 4-4.8. Providencia stuartii TatA purification.

*Providencia stuartii* TatA (*ps*TatA) gene was supplied as a gift from Dr. Matthew Freeman (MRC, Cambridge). It was cloned in pBAD vector and Flag-tag was introduced at the N-terminus with the help of PCR. GlpG knockout cells from Keio library (glpG::Kn)[62] were used to express psTatA. The protein was induced with 0.02% arabinose and expressed at 24$^{\circ}$C

for 6 hours. The membrane fraction was isolated, solubilized in 50 mM Tris-HCl pH 8.0, 150 mM NaCl, 20% glycerol, 1% DDM and applied on ANTI-FLAG M2 Affinity Gel (Sigma-Aldrich, USA). The protein was eluted with 0.1 M Glycine, pH 3.5, 0.1% DDM and pH was immediately adjusted with 1 M Tris-HCl, pH 7.5, 0.1% DDM. The protein samples were concentrated and flash-frozen instantly.

*4-4.9. Sequence alignment of prokaryotic rhomboids.*

Three transmembrane helices/regions prediction servers (TMHMM, HMMTOP, and TMPRED) were used to predict the membrane and cytoplasmic domains of the prokaryotic rhomboids. N- and C-terminus cytoplasmic domains predicted by the three servers, are highlighted in pink and orange respectively, the membrane domains are highlighted in blue. For *E. coli* and *H. influenza*e, the known structures of the rhomboids were used to determine the two domains (PDB codes: 2IC8 and 2NR9, respectively). For *P. aeruginosa*, *P. stuartii* and *B. subtilis_YqgP*, the various domains are the ones previously predicted[18]. For the clarity of the sequence alignment, *B. subtilis* that contains a N- and C-terminus cytoplasmic domains, is not represented as the N-terminus cytoplasmic domain is longer than the other ones (175 residues). For both domains, the secondary structure elements of ecGlpG are represented at the top of the sequence alignment. All the prokaryotic rhomboids are predicted to contain six or seven transmembrane segments. A black star at the top of the sequence alignment represents the end of the structured sequence which was reconstructed from the X-ray dataset in this study. More particularly, for ecGlpGN, the two domains A and B, involved in 3D domain swapping, are represented at the top of the sequence alignment with the same color code as Figure 4-1. The hinge residue asparagine 33 is highlighted in yellow and a yellow star

is positioned on the top of the sequence alignment. In the membrane domain (blue), the highly conserved residues are highlighted in red with the WR motif located in loop1 and the catalytic dyad (S201 and H254). The catalytic residues are marked with a red star at the top of the sequence alignment.

**Accession numbers:** Coordinates and structure factors have been deposited in the Protein Data Bank with accession number, PDB ID: 4HDD.

## 4-5. References

1.    Freeman, M. (2008). Rhomboids: 7 years of a new protease family. *Semin Cell Dev Biol*.
2.    Urban, S. & Dickey, S. W. (2011). The rhomboid protease family: a decade of progress on function and mechanism. *Genome Biol* **12**, 231.
3.    Bier, E., Jan, L. Y. & Jan, Y. N. (1990). rhomboid, a gene required for dorsoventral axis establishment and peripheral nervous system development in Drosophila melanogaster. *Genes Dev* **4**, 190-203.
4.    Brown, K. E., Kerr, M. & Freeman, M. (2007). The EGFR ligands Spitz and Keren act cooperatively in the Drosophila eye. *Dev Biol* **307**, 105-13.
5.    Cipolat, S., Rudka, T., Hartmann, D., Costa, V., Serneels, L., Craessaerts, K., Metzger, K., Frezza, C., Annaert, W., D'Adamio, L., Derks, C., Dejaegere, T., Pellegrini, L., D'Hooge, R., Scorrano, L. & De Strooper, B. (2006). Mitochondrial rhomboid PARL regulates cytochrome c release during apoptosis via OPA1-dependent cristae remodeling. *Cell* **126**, 163-75.
6.    Frezza, C., Cipolat, S., Martins de Brito, O., Micaroni, M., Beznoussenko, G. V., Rudka, T., Bartoli, D., Polishuck, R. S., Danial, N. N., De Strooper, B. & Scorrano, L. (2006). OPA1 controls apoptotic cristae remodeling independently from mitochondrial fusion. *Cell* **126**, 177-89.
7.    Whitworth, A. J., Lee, J. R., Ho, V. M., Flick, R., Chowdhury, R. & McQuibban, G. A. (2008). Rhomboid-7 and HtrA2/Omi act in a common pathway with the Parkinson's disease factors Pink1 and Parkin. *Dis Model Mech* **1**, 168-74; discussion 173.
8.    Alexander, C., Votruba, M., Pesch, U. E., Thiselton, D. L., Mayer, S., Moore, A., Rodriguez, M., Kellner, U., Leo-Kottler, B., Auburger, G., Bhattacharya, S. S. & Wissinger, B. (2000). OPA1, encoding a dynamin-related GTPase, is mutated in autosomal dominant optic atrophy linked to chromosome 3q28. *Nat Genet* **26**, 211-5.
9.    Walder, K., Kerr-Bayles, L., Civitarese, A., Jowett, J., Curran, J., Elliott, K., Trevaskis, J., Bishara, N., Zimmet, P., Mandarino, L., Ravussin, E., Blangero, J., Kissebah, A. & Collier, G. R. (2005). The mitochondrial rhomboid protease PSARL is a new candidate gene for type 2 diabetes. *Diabetologia* **48**, 459-68.
10.   Hatunic, M., Stapleton, M., Hand, E., DeLong, C., Crowley, V. E. & Nolan, J. J. (2009). The Leu262Val polymorphism of presenilin associated rhomboid like protein (PARL) is associated with earlier onset of type 2 diabetes and increased urinary microalbumin creatinine ratio in an Irish case-control population. *Diabetes Res Clin Pract* **83**, 316-9.
11.   Yan, Z., Zou, H., Tian, F., Grandis, J. R., Mixson, A. J., Lu, P. Y. & Li, L. Y. (2008). Human rhomboid family-1 gene silencing causes apoptosis or autophagy to epithelial cancer cells and inhibits xenograft tumor growth. *Mol Cancer Ther* **7**, 1355-64.
12.   Zou, H., Thomas, S. M., Yan, Z. W., Grandis, J. R., Vogt, A. & Li, L. Y. (2009). Human rhomboid family-1 gene RHBDF1 participates in GPCR-mediated transactivation of EGFR growth signals in head and neck squamous cancer cells. *FASEB J* **23**, 425-32.
13.   Brossier, F., Jewett, T. J., Sibley, L. D. & Urban, S. (2005). A spatially localized rhomboid protease cleaves cell surface adhesins essential for invasion by Toxoplasma. *Proc Natl Acad Sci U S A* **102**, 4146-51.
14.   Baker, R. P., Wijetilaka, R. & Urban, S. (2006). Two Plasmodium rhomboid proteases preferentially cleave different adhesins implicated in all invasive stages of malaria. *PLoS pathogens* **2**, e113.

15.    Mesak, L. R., Mesak, F. M. & Dahl, M. K. (2004). Expression of a novel gene, gluP, is essential for normal Bacillus subtilis cell division and contributes to glucose export. *BMC Microbiol* **4**, 13.

16.    Stevenson, L. G., Strisovsky, K., Clemmer, K. M., Bhatt, S., Freeman, M. & Rather, P. N. (2007). Rhomboid protease AarA mediates quorum-sensing in Providencia stuartii by activating TatA of the twin-arginine translocase. *Proc Natl Acad Sci U S A* **104**, 1003-8.

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

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

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

20.    Wang, Y., Zhang, Y. & Ha, Y. (2006). Crystal structure of a rhomboid family intramembrane protease. *Nature*, 1-5.

21.    Ben-Shem, A., Fass, D. & Bibi, E. (2007). Structural basis for intramembrane proteolysis by rhomboid serine proteases. *Proc Natl Acad Sci U S A* **104**, 462-6.

22.    Wu, Z., Yan, N., Feng, L., Oberstein, A., Yan, H., Baker, R. P., Gu, L., Jeffrey, P. D., Urban, S. & Shi, Y. (2006). Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry. *Nat Struct Mol Biol* **13**, 1084-91.

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

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

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

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

27.    Lemberg, M. K., Menendez, J., Misik, A., Garcia, M., Koth, C. M. & Freeman, M. (2005). Mechanism of intramembrane proteolysis investigated with purified rhomboid proteases. *Embo J* **24**, 464-72.

28.    Maegawa, S., Ito, K. & Akiyama, Y. (2005). Proteolytic action of GlpG, a rhomboid protease in the Escherichia coli cytoplasmic membrane. *Biochemistry* **44**, 13543-52.

29.    Tsruya, R., Schlesinger, A., Reich, A., Gabay, L., Sapir, A. & Shilo, B. Z. (2002). Intracellular trafficking by Star regulates cleavage of the Drosophila EGF receptor ligand Spitz. *Genes & development* **16**, 222-34.

30.    Shen, J. & Prywes, R. (2004). Dependence of site-2 protease cleavage of ATF6 on prior site-1 protease digestion is determined by the size of the luminal domain of ATF6. *The Journal of biological chemistry* **279**, 43046-51.

31. Erez, E. & Bibi, E. (2009). Cleavage of a multispanning membrane protein by an intramembrane serine protease. *Biochemistry* **48**, 12314-22.

32. Sheiner, L., Dowse, T. J. & Soldati-Favre, D. (2008). Identification of trafficking determinants for polytopic rhomboid proteases in Toxoplasma gondii. *Traffic* **9**, 665-77.

33. Lohi, O., Urban, S. & Freeman, M. (2004). Diverse substrate recognition mechanisms for rhomboids; thrombomodulin is cleaved by Mammalian rhomboids. *Curr Biol* **14**, 236-41.

34. Cheng, T. L., Wu, Y. T., Lin, H. Y., Hsu, F. C., Liu, S. K., Chang, B. I., Chen, W. S., Lai, C. H., Shi, G. Y. & Wu, H. L. (2011). Functions of rhomboid family protease RHBDL2 and thrombomodulin in wound healing. *J Invest Dermatol* **131**, 2486-94.

35. Fleig, L., Bergbold, N., Sahasrabudhe, P., Geiger, B., Kaltak, L. & Lemberg, M. K. (2012). Ubiquitin-Dependent Intramembrane Rhomboid Protease Promotes ERAD of Membrane Proteins. *Molecular cell* **47**, 558-69.

36. Del Rio, A., Dutta, K., Chavez, J., Ubarretxena-Belandia, I. & Ghose, R. (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. *Journal of molecular biology* **365**, 109-22.

37. Xue, Y. & Ha, Y. (2011). The catalytic mechanism of rhomboid protease GlpG probed by 3,4-dichloroisocoumarin and diisopropyl fluorophosphonate. *J Biol Chem*.

38. Bennett, M. J., Schlunegger, M. P. & Eisenberg, D. (1995). 3D domain swapping: a mechanism for oligomer assembly. *Protein Sci* **4**, 2455-68.

39. Barrientos, L. G. & Gronenborn, A. M. (2002). The domain-swapped dimer of cyanovirin-N contains two sets of oligosaccharide binding sites in solution. *Biochem Biophys Res Commun* **298**, 598-602.

40. Sampathkumar, P., Mak, M. W., Fischer-Witholt, S. J., Guigard, E., Kay, C. M. & Lemieux, M. J. (2012). Oligomeric state study of prokaryotic rhomboid proteases. *Biochimica et biophysica acta* **1818**, 3090-3097.

41. Rousseau, F., Schymkowitz, J. W., Wilkinson, H. R. & Itzhaki, L. S. (2001). Three-dimensional domain swapping in p13suc1 occurs in the unfolded state and is controlled by conserved proline residues. *Proc Natl Acad Sci U S A* **98**, 5596-601.

42. MacArthur, M. W. & Thornton, J. M. (1991). Influence of proline residues on protein conformation. *J Mol Biol* **218**, 397-412.

43. Strisovsky, K., Sharpe, H. J. & Freeman, M. (2009). Sequence-specific intramembrane proteolysis: identification of a recognition motif in rhomboid substrates. *Mol Cell* **36**, 1048-59.

44. Porcelli, I., de Leeuw, E., Wallis, R., van den Brink-van der Laan, E., de Kruijff, B., Wallace, B. A., Palmer, T. & Berks, B. C. (2002). Characterization and membrane assembly of the TatA component of the Escherichia coli twin-arginine protein transport system. *Biochemistry* **41**, 13690-7.

45. Clemmer, K. M., Sturgill, G. M., Veenstra, A. & Rather, P. N. (2006). Functional characterization of Escherichia coli GlpG and additional rhomboid proteins using an aarA mutant of Providencia stuartii. *J Bacteriol* **188**, 3415-9.

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

47. Lemieux, M. J., Reithmeier, R. A. & Wang, D. N. (2002). Importance of detergent and phospholipid in the crystallization of the human erythrocyte anion-exchanger membrane domain. *J Struct Biol* **137**, 322-32.

48.    Lee, J. R., Urban, S., Garvey, C. F. & Freeman, M. (2001). Regulated intracellular ligand transport and proteolysis control EGF signal activation in Drosophila. *Cell* **107**, 161-71.

49.    Bennett, M. J., Choe, S. & Eisenberg, D. (1994). Domain swapping: entangling alliances between proteins. *Proc Natl Acad Sci U S A* **91**, 3127-31.

50.    Schymkowitz, J. W., Rousseau, F., Wilkinson, H. R., Friedler, A. & Itzhaki, L. S. (2001). Observation of signal transduction in three-dimensional domain swapping. *Nat Struct Biol* **8**, 888-92.

51.    Liu, Y., Gotte, G., Libonati, M. & Eisenberg, D. (2001). A domain-swapped RNase A dimer with implications for amyloid formation. *Nat Struct Biol* **8**, 211-4.

52.    Sherratt, A. (2012). Beyond the Active Site of the Bacterial Rhomboid Protease:Novel Interactions at the Membrane to Modulate Function. PhD degree University of Ottawa.

53.    Liu, L., Byeon, I. J., Bahar, I. & Gronenborn, A. M. (2012). Domain swapping proceeds via complete unfolding: a 19F- and 1H-NMR study of the Cyanovirin-N protein. *Journal of the American Chemical Society* **134**, 4229-35.

54.    Reddy, T. & Rainey, J. K. (2012). Multifaceted substrate capture scheme of a rhomboid protease. *The journal of physical chemistry. B* **116**, 8942-54.

55.    Casey, J. R. & Reithmeier, R. A. (1991). Analysis of the oligomeric state of Band 3, the anion transport protein of the human erythrocyte membrane, by size exclusion high performance liquid chromatography. Oligomeric stability and origin of heterogeneity. *The Journal of biological chemistry* **266**, 15726-37.

56.    Macol, C. P., Tsuruta, H., Stec, B. & Kantrowitz, E. R. (2001). Direct structural evidence for a concerted allosteric transition in Escherichia coli aspartate transcarbamoylase. *Nature structural biology* **8**, 423-6.

57.    Yamasaki, M., Li, W., Johnson, D. J. & Huntington, J. A. (2008). Crystal structure of a stable dimer reveals the molecular basis of serpin polymerization. *Nature* **455**, 1255-8.

58.    Kabsch, W. (2010). Xds. *Acta crystallographica. Section D, Biological crystallography* **66**, 125-32.

59.    Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Davis, I. W., Echols, N., Headd, J. J., Hung, L. W., Kapral, G. J., Grosse-Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C. & Zwart, P. H. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta crystallographica. Section D, Biological crystallography* **66**, 213-21.

60.    Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). Features and development of Coot. *Acta crystallographica. Section D, Biological crystallography* **66**, 486-501.

61.    Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S. & Richardson, D. C. (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic acids research* **35**, W375-83.

62.    Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L. & Mori, H. (2006). Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular systems biology* **2**, 2006 0008.

# Chapter 5

# General Discussion

**The section 5-3. is a review originally published in Biochemistry and Cell Biology.**

**Christelle Lazareno-Saez, Cory L. Brooks, M. Joanne Lemieux.**

**5-1. Summary of significant findings**

The narrow link between rhomboid proteases and pseudoproteases with a diverse array of diseases suggest that these intramembrane proteins could be an interesting therapeutic target. However, this requires a good understanding of the catalytic mechanism, substrate entry and regulation of such membrane proteins. This work provided significant insight into the substrate access to the active site as well as its cleavage by rhomboids proteases. In addition, this thesis provides a complete functional and structural study about an extra-membranous domain of rhomboids, which was suggested to play a role in regulation.

In Chapter 2, we provided more details about the active site mechanism and substrate gating in hiGlpG. Our new X-ray structure of hiGlpG present significant disorder in the region of helix 5 and its two flanking loops (loop 4 and loop5), this disorder was supported with combined B-factor analysis and real-space correlation plots of this region of the structure. A mutagenesis study combined with an *in vitro* cleavage assay suggests that only helix 5 and loop 5 are mobile and work jointly during substrate gating. This result seems to unify the controversy about substrate entry in ecGlpG. While further study is required, we also proposed a reaction pathway based on the substrate access to the active site. A primary common serine peptidase inhibitors screening was performed to determine which inhibitor and in which conditions to use for enzyme:inhibitor co-crystallization or crystal soaking. Both TPCK and DCI provided an inhibitory effect on both hiGlpG and ecGlpG-FL.

Cytoplasmic domains on C-or N-terminal ends of rhomboid-family proteins are of particular interest as little is known about their role and their implication in regulation have been hypothetized. In Chapter 3, I present the results of purification and crystallization of the full-length ecGlpG-FL. Its over-expression and purification protocol in DDM, which avoids

proteolysis and provide a good protein yield is detailed. 9Å resolution diffracting crystals were obtained, which contradicts the hypothesis that high flexibility in the cytoplasmic domain of ecGlpG-FL would avoid crystal growth. Extensive resolution optimizations using basic methods (additives screening, temperature variation and cryoprotectants), detergent exchanges at different steps of the purification and crystallization protocols using detergents and synthetic amphiphile molecule mixture, bicelles or even designing constructs to remove small flexible parts of ecGlpG-FL were performed. While, I was unable to observe substantial diffraction improvement, I proposed more drastic methods (shorten cytoplasmic and membrane domains linker or antibody:enzyme co-crystallization) to reach this goal.

In chapter 4, another strategy to obtain structural insights into ecGlpG cytoplasmic domain was performed. Structural details of this domain were obtained with a 1.35 Å resolution crystal structure of this extra-membranous domain. Our data support the concept that this domain undergoes a domain swapping between two monomers. We provided evidences that this domain-swapped dimer is not an artefact of crystallization and occur in full-length protein in solution. In addition, we developed a much needed steady-state kinetic assays using two model substrates, either soluble or membrane protein, as the physiological substrate of this enzyme remains unknown. The apparent kinetic parameters of ecGlpG membrane domain and ecGlpG-FL did not significantly differ. These results associated with a mutagenesis study suggest that ecGlpG-cyto does not affect the structural integrity or the accessibility of the active site of the enzyme. The exact role of this domain in proteolytic activity regulation remains yet to be fully determined and understood. More studies are necessary to determine the exact role of this domain and also the importance of the energy consuming domain swapping.

In subsequent sections, remaining questions about each of these conclusions will be discussed in the context of actual knowledge provided by three new structure of inhibitor :enzyme complexes.


**5-2. Overview regarding the catalytic mechanism of rhomboids**

Even if rhomboids are intramembrane serine peptidases, structural data highlighted two main differences between intramembrane and soluble serine peptidases: the catalytic dyad active site and the attack of the substrate was proposed to be on its si-face **(Figure 5-1 A)**. In addition, an initial study defined a possible oxyanion hole for ecGlpG, primarily composed by the main chain amide of S201 and the main chain amide of L200, the side chain amide of N154 and the side chain of H150[1]. Key steps in obtaining more data on rhomboid catalytic mechanism require an enzyme :substrate or enzyme :inhibitor complex structure.

However, little was known about rhomboid protease inhibition and some controversy existed on the ability of TPCK and DCI to inhibits them. Data presented in chapter 2 confirm the inhibiton of hi- and ec-GlpGs by these two inhibitors. The work in this thesis on obtaining inhibitor bound rhomboids structure was interrupted by the publications of such structures containing 7-amino-4-chloro-3-methoxy-isocoumarin, Cbz-Ala$^P$(O-iPr)F (CAPF) and disopropylfluorophosphonate (DFP) serine peptidase inhibitors[2; 3; 4] **(Figure 5-1)**. In the first structure, the ring of the isocoumarin is doubly covalently bond to the catalytic serine S201 and histidine H254 to form an alkylated acyl enzyme **(Figure 5-1 B)**.


However, the covalent bond between H254 and the peptide does not occur naturally and it may have impact the conformation of the histidine or the inhibitor. In the other structures, irreversible tetrahedral intermediates, only covalently bond to the catalytic serine, might be observed **(Figure**

**5-1 C and 5-1 D)**. In ecGlpG :DFP complex, three phosphorus atoms are covalently bonded to S201 whom hydroxyl oxygen replace the fluorine in DFP **(Figure 5-1 C)**. In all ecGlpG:inhibitor complex structures, only small conformational changes were observed on the other residues except for ecGlpG :DFP complex where the catalytic dyad member H254 undergoes a 90° rotation interacting with helix 5 residues. This rotation is also observed in Ser-His catalytic dyad serine peptidases. Our results in chapter 2 confirmed the oxyanion hole of hiGlpG being composed by the main-chain NH from S116 and the protonated $N^{\varepsilon}$ from the neighbouring H65. These structures also supported the composition of the oxyanion hole of ecGlpG-FL **(Figure 5-1 A)**. However, a controversy arose regarding the attack of the scissile bond of the substrate. The cleavage stereochemistry is well known and the nucleophilic attack of the peptide bond must occur at Bürgi–Dunitz angle $(107°)$[5] to the carbonyl. 7-amino-4-chloro-3-methoxy-isocoumarin:enzyme and (CAPF):enzyme structures suggested a si-face and re-face attack, respectively. However, combining our results in chapter 2 and results from the previously described studies tend to support the si-face attack. This attack does not occur on the si-face for common soluble serine peptidase but it does for some Ser-Lys catalytic dyad serine peptidase[6]. Even if such data represent key tools in investigating the catalytic mechanism of rhomboids, more inhibitor :enzyme or even substrate :enzyme structures are needed .

## 5-3. Overview regarding the model of substrate gating for prokaryotic rhomboids

I published a review in 2011 which discuss this question regarding the data obtained in chapter 2. Here I present this review followed by an up to date analysis.

Rhomboids are a family of intramembrane serine peptidases found in all kingdoms[7] and are distributed in different membranes in the cell[8; 9]. They have a remarkable mechanism for peptide
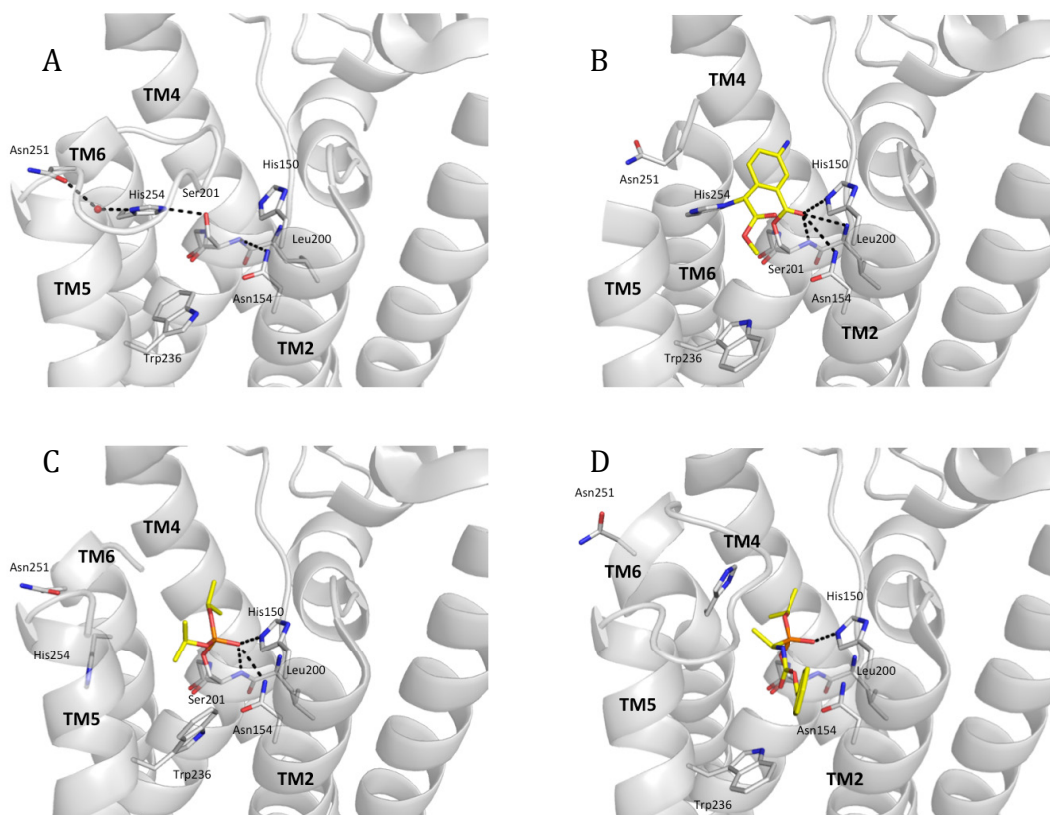
**Figure 5-1. Catalytic site of apo ecGlpG or complexed with inhibitors**

ecGlpG is represented in cartoon as apo form (PDB ID 2XOV) **A)**, complexed with 7-amino-4-chloro-3-methoxy-isocoumarin (PDB ID 2XOW) **B)**, disopropylfluorophosphonate (DFP (PDB ID 3TXT) **C)** and ) Cbz-Ala$^P$(O-iPr)F (CAPF) (PDB ID 3UBB) **D)**. The residues of interest are represented in gray sticks on ecGlpG. The inhibitors are also represented in yellow sticks. The bonds defining the oxyanion hole are shown.

hydrolysis executed in or adjacent to the lipid bilayer[10]. Substrate proteolysis results in the release of a soluble fragment, which acts in cellular signalling processes[11]. Although the general role of rhomboids is known, the specific functions of many of these enzymes remain to be discovered. Nevertheless, rhomboids have been directly linked to various disease states such as parasite invasion [12; 13; 14], Parkinson's disease[15], type II diabetes[16], respiratory defects[17], and epithelial cancer[18], which underlines the necessity for a detailed understanding of rhomboid structure and function. For a comprehensive review, see[19].

So far, the structures of two rhomboid orthologs (GlpG) have been solved: those of *Escherichia coli* (ecGlpG)[1; 20; 21] and *Haemophilus influenzae* (hiGlpG)[22] (**Figure 5-2 a, 5-2b, Table 5-1**). These two structures revealed a solvent-filled active site with a Ser–His catalytic dyad. The rhomboids ecGlpG and hiGlpG have six transmembrane segments and structurally are very similar (root mean square deviation of 1.09 Å between *hi*GlpG 2NR9.pdb and *ec*GlpG 2IC8.pdb). The main structural differences were found in loop 4, loop 5, and helix 5 (**Figure 5-2a and 5-2b**). One emerging issue was the identification of the mobile element in the protein, which allows substrate docking. It has been proposed that only the movement of loop 5 (also called the cap)[23]  or helix 5[24] allows the substrate to access the catalytic dyad. A structure of ecGlpG co-crystallized with a coumarin inhibitor has been published and indicates movements in loop 5 and to a lesser degree in helix 5 and loop 4[2]. Also, **the work presented in chapter 2 supports the cooperation of loop 5, helix 5, and loop 4 to act as part of the substrate entry gate.** Loop 5 and helix 5 interact mainly with the neighbouring helix 4 (containing the catalytic serine residue) and helix 2. Using the available rhomboid structures, this review highlights key differences observed in the regions of the substrate entry gate in terms of orientation, flexibility, and interaction networks and also discusses some functional and regulatory implications.

(a)

Loop5

Helix5

Loop4    Helix2

(b)

Loop5

Helix5

Loop4    Helix2

(c)



Relative B-factor dierence

2NR9 (*hi*GlpG)
2IRV chain B (*ec*GlpG open)
2NRF chain A (*ec*GlpG open)

2IC8
(*ec*GlpG →
closed)

*ec*GlpG residues

loop 4    helix 5    loop 5

215  GERDPQSGIYLQRGLIIFALIWIVAGWFDLFGMSMA  2IC8 (*ec*GlpG closed)
130  DKLNHHLFDLPEGFFTMLLVGIALGFISPLFGVEMG  2NR9 (*hi*GlpG)

132
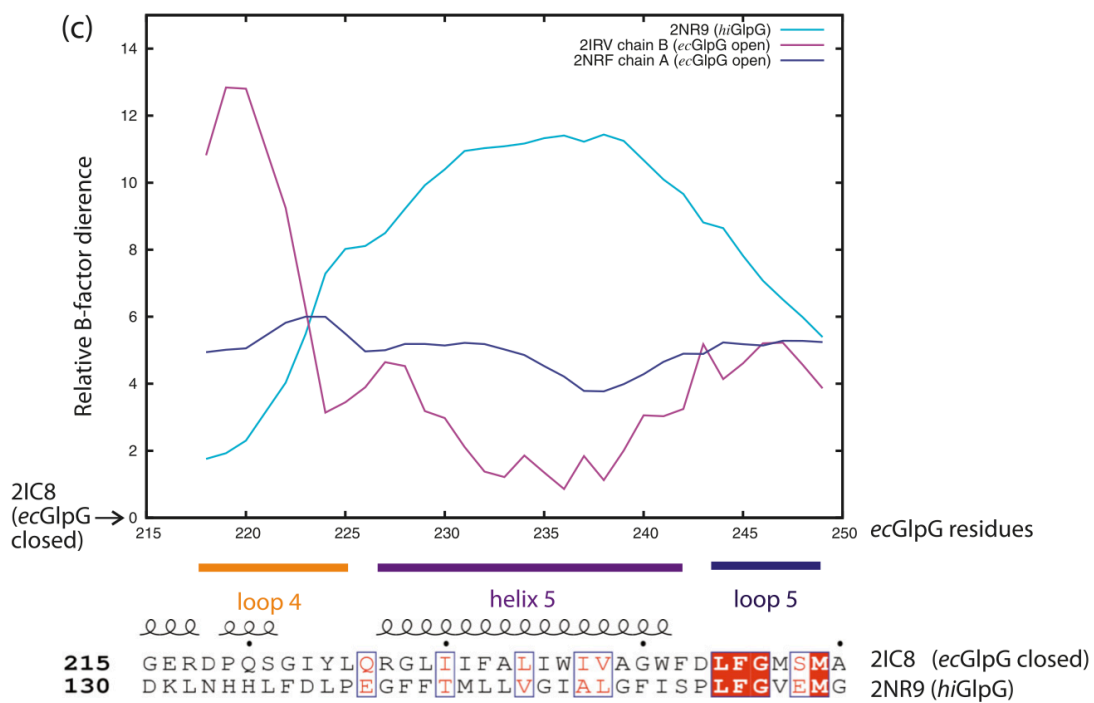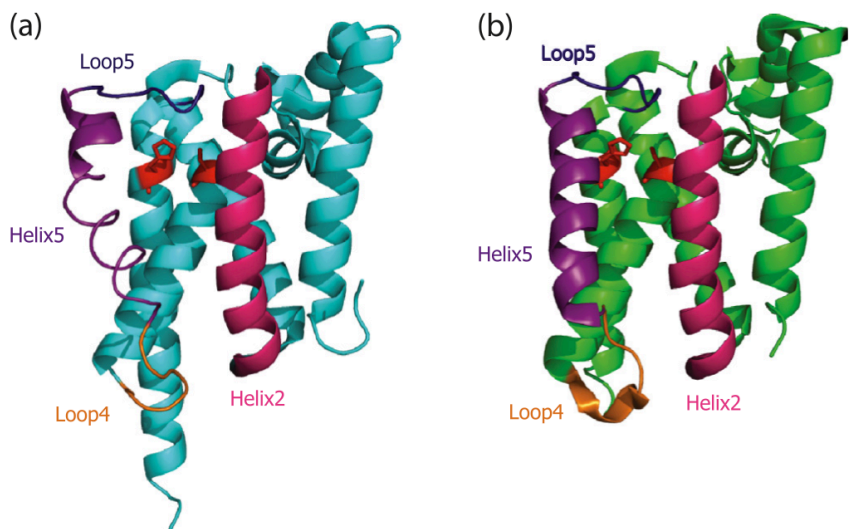
**Figure 5-2.Characteristics of the substrate entry gate for hiGlpG and ecGlpG.** (*a*) A cartoon representation of *hi*GlpG highlights the secondary structure elements for helix 2 (pink), loop 4 (orange), helix 5 (purple), loop 5 (dark blue), and the catalytic dyad (red) formed by Ser116 and His169. (*b*) A cartoon representation of the closed conformation of *ec*GlpG shows the secondary structure elements helix 2 (pink), loop 4 (orange), helix 5 (purple), and loop 5 (dark blue) and the catalytic dyad (red) formed by Ser201 and His254. (*c*) Representation of the relative B-factor difference in function of the residue position in the substrate entry gate (loop 5 and helix 5) and its anchor (loop 4). This difference was calculated with the following formula[32], which limits bias from the refinement procedure:

$$[(B_i - \langle B_i \rangle) - (B_{2IC8} - \langle B_{2IC8} \rangle)]/(\langle B_i \rangle - \langle B_{2IC8} \rangle)$$, where *i* corresponds to 2NR9 (*hi*GlpG, in cyan), chain B of 2IRV (open conformation of *ec*GlpG, in pink), or chain A of 2NRF (open conformation of *ec*GlpG, in blue). $\langle B \rangle$ refers to the mean value for the whole structure. The reference for the three different calculations is the closed conformation of *ec*GlpG (2IC8). A structural alignment was performed prior to calculation and only main-chain atoms were used. On the *x*-axis, the residues of *ec*GlpG and their position in the elements of secondary structure are represented. For *hi*GlpG (2NR9), the corresponding residues were used (not represented on the *x*-axis): N133 to M164 from *hi*GlpG correspond to D218 to M249 from *ec*GlpG. A structural alignment of regions containing loop 4, helix 5, and loop 5 in *ec*GlpG and *hi*GlpG is presented. Structural alignment was carried out with EXPRESSO[33] (available from http://tcoffee.vital-it.ch/cgi-bin/Tcoffee/tcoffee_cgi/index.cgi) with display by ESPript[34] (available from http://espript.ibcp.fr/ESPript/ESPript).

| Protein | PDB code | Relevance | Color in figures | Reference |
|---|---|---|---|---|
| ecGlpG | 2IC8 | Closed conformation | Green | Wang et al. 2006 |
| | 2NRF chain A | Open conformation | Blue | Wu et al. 2006 |
| | 2IRV chain B | Open conformation | Pink | Ben-Shem et al. 2007 |
| | 3B45 | Intermediate conformations | Not represented | Wang et al. 2007 |
| | 2O7L | Intermediate conformations | Not represented | Wang et al. 2007 |
| | 2NRF chain B | Intermediate conformations | Not represented | Wu et al. 2006 |
| | 2NRF chain A | Intermediate conformations | Not represented | Ben-Shem et al. 2007 |
| hiGlpG | 2NR9 | Open and highly flexible conformation | Cyan | Lemieux et al. 2007 |
| | 3ODJ | Loop 5 and helix 5 disordered | Not represented | Brooks et al. 2011 |

**Table 5-1. Structures of ecGlpG** (at the time of review publication)
Please see **Table 1-1** for an up to date list.

*5-3.1 Different conformations of the substrate entry gate in ecGlpG*

The first structure of a rhomboid from *E. coli* (ecGlpG) was significant, as it provided the first structural characterization of this new class of membrane protein[21]. In a short period of time, various groups presented several ecGlpG structures[1; 20; 21; 25] crystallized in different space groups. Among these structures, the main differences are located in the substrate entry gate (loop 5 and helix 5). Three of the structures exhibit the largest degree of structural heterogeneity and can be classified into two types: one closed conformation and two open conformations **(Table 5-1)**. The other structures of ecGlpG adopt diverse intermediate conformations for helix 5 and a disordered loop 5, and this may reflect transient movements. Although some of the conformational variability can be attributed to crystallization artefacts, the possibility of these movements occurring in vivo must be considered.

The relative B-factor difference, using the structure of the closed ecGlpG state as the control, was calculated to measure overall flexibility in the open structures **(Figure 5-1c)**. The B-factor difference for hiGlpG was assessed but will be discussed in the subsequent section. For ecGlpG, the two open conformations, 2NRF:chain A and 2IRV:chain B, show a higher relative B-factor than the closed state, 21C8 (control and therefore zero relative difference), suggesting a greater relative flexibility in this region compared with the closed conformation. For loop 4, we see a higher relative B-factor difference in the open form 2IRV:chain B compared with the open form 2NRF:chain A. In helix 5, this difference is reversed, with a low point observed near Ile237 for 2IRV:chain B. The switch in relative difference suggests a rotation point may exist in the open conformations between these two regions that intersect on the graph. In loop 5, the relative B-factor difference is comparably high for both open conformations. It is tempting to speculate that these structural changes may reflect those occurring during substrate docking. However, a

structure of an enzyme–substrate complex is needed to resolve this question.

A closer look at residues involved in intramolecular interactions explains the differences observed in relative B-factors. The closed conformation provides an extended hydrophobic interaction network linked to residues in the substrate entry gate (Trp236, Phe232, and Leu229 on helix 5 and Ser248 and Phe245 of loop 5) **(Figures 5-3 and 5-4a)**. These amino acids interact primarily with residues on helix 2, helix 4, and loop 1 (see detailed description below). In the open conformations of *ec*GlpG, many of these stabilizing interactions are displaced, resulting in a more relaxed structure compared with the closed one. One open conformation of ecGlpG (2NRF:chain A) displays a kink in helix 5 below residue Phe232 that results in an opening or shift of helix 5 near loop 5 **(Figure 5-3a and 5-3b)**. This opening has destabilized the hydrophobic interaction network at the top of helix 5, diminishing the strength of interaction between Trp236 and helix 2 and helix 4. However, the bottom of helix 5 and loop 4 are similar to the closed conformation, conserving the interactions of Phe232 and Leu229 with the protein core. In loop 5, Ser248 and Phe245 are completely reoriented and do not interact with the protein core **(Figure 5-3c)**. In the second open conformation of ecGlpG (**Figure 5-3d**, 2IRV:chain B), helix 5 moves away from the central helical bundle at a 90° angle compared with the direction of movement of helix 5 in the first open state. Furthermore, the helix is shifted up in comparison with the closed conformation. These movements result in a reorganization of the hydrophobic interaction network in helix 5 **(Figure 5-3d and 5-3e)**. The only remaining original interactions between helix 5 and helix 2 from the closed state occur from residues Phe232 and Trp157 **(Figures 5-3e and 5-4a)**. Moreover, new interactions between helix 5 and helix 4 are observed through Trp241, Ile237, and Ala233 on helix 5. This may have a link to the previously observed decrease of the relative B-factor difference close to residue Ile235 in helix 5. In comparison with
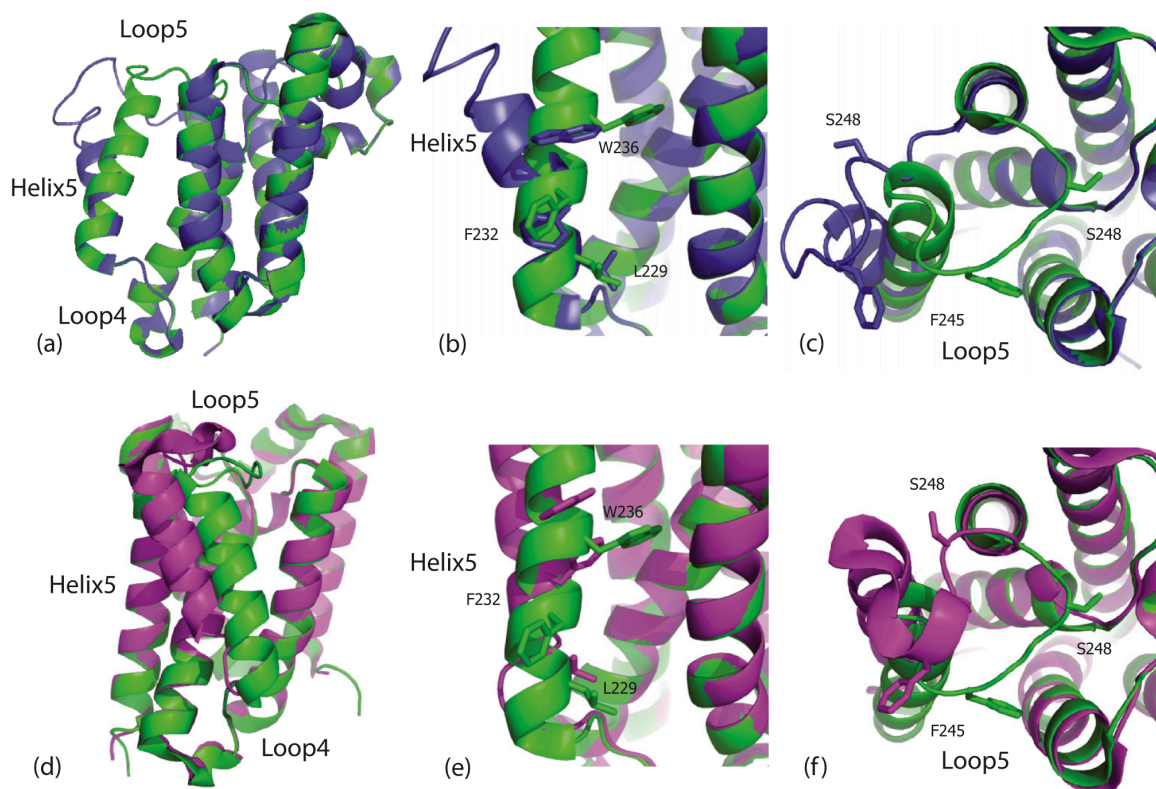
**Figure 5-3.Comparison of ecGlpG conformations**. Three different conformations of *ec*GlpG are represented in cartoon (2IC8 in green, 2NRF:chain A in blue, and 2IRV:chain B in pink), with residues involved in intramolecular interactions shown in sticks. These three structures exhibit the largest degree of structural heterogeneity in the predicted substrate entry gate (loop 5 and helix 5) among *ec*GlpG structures. In this figure, two structures, 2NRF:chain A and 2IRV:chain B, called open conformations, are compared with 2IC8, which represents the closed state of the substrate entry gate. (*a*) Superimposition of 2NRF:chain A with 2IC8 reveals a shift of helix 5 and loop 5. The residues involved in hydrophobic interactions are shown for (*b*) helix 5 and (*c*) loop 5. (*d*) For the other open conformation, superimposition of 2IRV:chain B with 2IC8 reveals a shift of helix 5 and loop 5. The residues involved in hydrophobic interactions in (*e*) helix 5 and (*f*) loop 5 for 2IRV:chain B are represented. They are, respectively, represented in the same orientations. In *a* and *d*, the two different superimpositions of molecules with 2IC8 as reference are rotated by 90°. The remaining pairs, *b* and *e* and *c* and *f*, are represented in similar orientations. These comparisons highlight the movement of key residues in both loop 5 and helix 5.

137
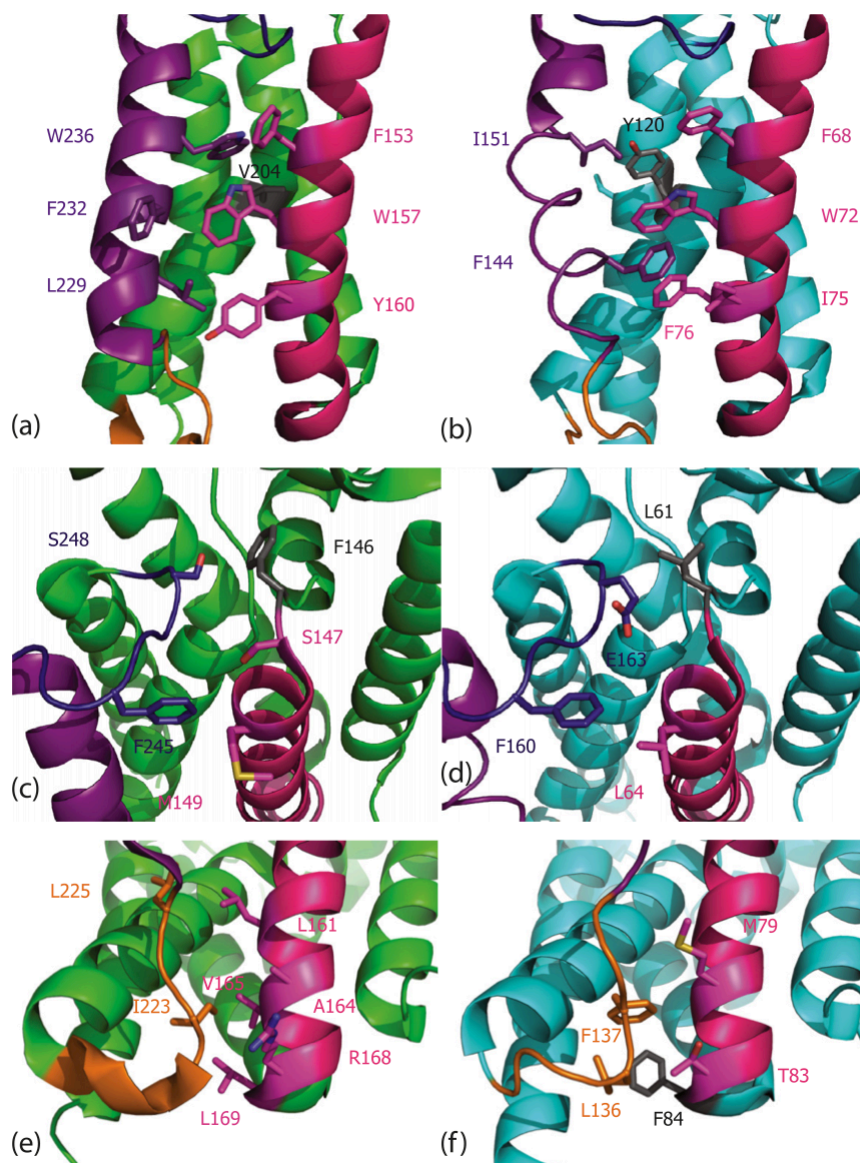
**Figure 5-4. Hydrophobic interaction network in helix 5, loop 5, and loop 4 for ecGlpG and hiGlpG.** *ec*GlpG and *hi*GlpG are represented in cartoon (green and cyan, respectively). The residues involved in hydrophobic interactions are represented in sticks, with nitrogen in blue, oxygen in red, and sulfur in yellow. The color code for loop 5, helix 5, and loop 4 is the same as in **Figure 5-1a and 5-1b**. Grey was used for residues from other parts of the enzyme for clarity. The helix 5 area is observed for (*a*) *ec*GlpG and (*b*) *hi*GlpG. In *a* and *b*, helix 1 and loop 1 were removed for clarity. The loop 5 area is observed for (*c*) *ec*GlpG and (*d*) *hi*GlpG. Finally, loop 4 is represented for (*e*) *ec*GlpG and (*f*) *hi*GlpG.

138

the closed conformation of ecGlpG, Ser248 and Phe245 in loop 5 are completely reoriented and do not interact with the protein core **(Figure 5-3f)**. However, in loop 4 the interactions remain unchanged in the second open conformation of ecGlpG. The remainder of this review will focus primarily on structures of the two rhomboid orthologs by comparing hiGlpG with the closed conformation of ecGlpG, keeping in mind the key structural observations of the open conformations of ecGlpG.

5-3.2. Structural differences in the substrate entry gate between hiGlpG and ecGlpG

5-3.2.1. Helix 5 in the substrate entry gate

It has been proposed that in order for substrates to gain access to the rhomboid peptidase's buried active site, helix 5 moves in ecGlpG[24] and hiGlpG, according to the data presented in chapter 2, during substrate docking. In these studies, mobility and mutagenesis of residues in the substrate gate demonstrated that the interactions between helix 5 and the protein core are crucial for enzyme function. Helix 5 interacts with key features in the protein core such as helix 4, where the catalytic serine is located, as well as helix 2. Thus a detailed analysis of the interactions between helix 5 and the protein core may provide crucial insights into how the substrate gate operates in these two proteins.

A major difference between the closed state of ecGlpG and hiGlpG is the position of helix 5 relative to the rest of the protein core **(Figure 5-2a and 5-2b and Figure 5-4a and 5-4b)**. A kink in the middle of helix 5 of hiGlpG divides this helix into two halves: the top half (residues 150 to 156) and the bottom half (residues 142 to 149) **(Figure 5-2a)**. The top half of helix 5 in hiGlpG is shifted in comparison with the closed state of ecGlpG, in the same direction as one of the open

conformations of ecGlpG (2IRV:chain B). In addition, the three conformations of ecGlpG present a similar axis for the bottom half of helix 5, which differs from the axis of the bottom half of helix 5 in hiGlpG. Altogether, the structure observed for hiGlpG represents an intermediate conformation between the closed and open conformations of ecGlpG. However, the particular orientation of helix 5 in hiGlpG creates a bow opening from the bottom to the top of helix 5, which is different than the conformations of ecGlpG. Furthermore, the overall structure of helix 5 in hiGlpG is unusual because it is partially unwound, suggesting conformational flexibility **(Figure 5-2a)**. The new structure of hiGlpG (3ODJ.pdb), detailed in chapter 2, underlines the flexibility of helix 5 and loop 5 by showing disorder for the substrate entry gate. Compared with the ecGlpG structures, the relative B-factor difference for loop 5 and helix 5 is higher in the hiGlpG structure, highlighting the flexibility of the substrate entry gate **(Figure 5-2c)**.

In the closed state of ecGlpG, helix 5 is associated with the protein core owing to a network of hydrophobic interactions mediated by large aromatic residues **(Figure 5-4a)**. The major interactions occur between helices 2 and 5 via pairs Trp157–Phe232 and Tyr160–Leu229. Phe153 and Trp157 on helix 2, Trp236 on helix 5, and Val204 on helix 4 form a group of residues linking helix 5 to the central core. Mutagenesis of some of these residues of ecGlpG significantly increased the activity of ecGlpG[24], revealing the importance of disrupting interactions between helix 5 and helix 2 for gating activity. Interestingly, the interaction network between helix 5 and helix 2 is strikingly different when comparing ecGlpG and hiGlpG **(Figures 5-4a and 5-4b)**. In both proteins, these interactions are mediated by large hydrophobic residues. However, as described above, there is a clear difference in the orientation of the partially unwound helix 5 of hiGlpG. Contrary to both the open and closed ecGlpG conformations, the

interactions between helix 5 and the protein core of hiGlpG are not along the length of helix 5 but rather at the edges of the helix. In particular, at the bottom of helix 5 in hiGlpG, Phe144 participates in a group of three interactions with Trp72, Ile75, and Phe76 on helix 2 **(Figure 5-4b)**. In addition, Ile151 on helix 5 interacts with Tyr120 on helix 4, providing stabilization to the top of helix 5 **(Figure 5-4b)**. The mutagenesis assay, performed in the work presented in chapter 2, has confirmed the importance of these residues in hiGlpG function. Thus the key differences between E. coli and H. influenzae rhomboids are characterized by residues located on helix 5 and identical or similar residues found on helix 2. In hiGlpG, compared with ecGlpG, helix 5 demonstrates an open and more flexible conformation with an overall weaker hydrophobic interactions network. These differences may have functional implications for enzyme kinetics for the different rhomboid orthologs.

### 5-3.2.2. Loop 5 in the substrate entry gate

Loop 5 is thought to act as a cap, providing access to the buried active site residues [25]. As mentioned above, our new hiGlpG structure is disordered in this region but in our original hiGlpG structure, loop 5 is ordered[22]. The conformation of loop 5 in hiGlpG is similar to that of the closed state of ecGlpG **(Figure 5-4c and 5-4d)** and contrary to the two open structures of ecGlpG **(Figure 5-3c and 5-3f)**. In the closed state of ecGlpG, Ser248 on loop 5 interacts with Phe146 on loop 1 and there are two interactions between Phe245 on loop 5 and Ser147 and Met149 on helix 2. Similarly, in hiGlpG, Phe160 and Glu163 form hydrophobic interactions with Leu64 and Leu61 on helix 2 and loop 1, respectively. Despite the similarities between loop 5 in hiGlpG[22] and the closed ecGlpG structure[21], the calculation of the relative B-factor shows an increased mobility of loop 5 in hiGlpG **(Figure 5-1c)**. All the conformations of ecGlpG show

loop 5 in different orientations or with a lack of electron density. For structures presenting electron density in this region, the relative B-factor is higher in hiGlpG than in all of the ecGlpG structures **(Figure 5-1c and Table 5-1)**. Furthermore, our new structure of hiGlpG showing a lack of electron density in loop 5 supports high mobility in this region

### 5-3.2.3. Loop 4 anchors the substrate entry gate

Loop 4 is located at the bottom of the substrate entry gate (helix 5 and loop 5) and thus acts as an "anchor" for the substrate entry gate of hiGlpG and the closed conformation of ecGlpG. In ecGlpG, the middle of loop 4 interacts with the protein core through hydrophobic interactions between Ile223 from loop 4 and Ala164, Val165, and Leu169 from helix 2 **(Figure 5-4e)**. At the C-terminal end of loop 4, hydrophobic interactions exist between Leu225 from loop 4 and Leu161 from helix 2. In hiGlpG, however, only hydrophobic interactions in the middle of loop 4 are observed. Phe137 from loop 4 interacts with Met79, Thr83, and Phe84 from the protein core **(Figure 5-4f)**. Also, Leu136 from loop 4 interacts with Thr83 from helix 2 and Phe84 from loop 2. In chapter 2, we presented some mutations of residues in loop 4 of hiGlpG demonstrate that interactions in this region are necessary for substrate cleavage As all ecGlpG structures present similar characteristics in loop 4, one can notice that the C-terminal end of hiGlpG is less stabilized than that of its ortholog. Furthermore, the relative B-factor difference at the C-terminal end of loop 4 is higher in hiGlpG than in the open and closed conformations of ecGlpG **(Figure 5-1c)**. This, along with the lack of interactions at the C-terminal end of loop 4 in hiGlpG, suggests that in hiGlpG the increased mobility of the substrate entry gate is linked to this flexible anchor at the end of loop 4.

*5-3.3. Concluding perspectives*

This detailed comparison of ecGlpG and hiGlpG crystal structures indicates the main differences were in loop 5, helix 5, and the C-terminal end of loop 4. Loop 5 and helix 5 are believed to play a role in substrate docking in the buried active site. The divergent conformations of these regions support this hypothesis. The ecGlpG and hiGlpG structures also show different intramolecular associations between helix 5 and helix 2. This association is weaker in hiGlpG than in ecGlpG. We hypothesize that this will affect the rate of substrate cleavage. The substrates for ecGlpG and hiGlpG are currently unknown and furthermore the cellular function of these enzymes is unclear. The apparent increased mobility of the hiGlpG substrate gate may in turn reflect an increased rate of proteolytic activity compared with ecGlpG. Presently, a proper kinetic assay is unavailable to assess cleavage rates. It would be interesting to mutate helix 5 of hiGlpG into that found in ecGlpG and assess its cleavage kinetics. Information about the structure and activity of the mutant would provide insight into the structural differences observed between these two rhomboid peptidases.

Notably, four additional X-ray structures of ecGlpG with or without common serine peptidase inhibitor were obtained after the publication of this review. All of these structures are in an intermediate conformation presenting a lift out of the loop 5 and a small motion of helix 5[2; 3; 4]. More precisely, the three structures of ecGlpG membrane domain in complex with inhibitors present small structural changes. Loop 5 is either disordered or tilted out from the active site and only small motions are observed for helix 5, even if one inhibitor was almost as large as a real peptide substrate (CAPF)[2; 3; 4]. This inhibitor actually extended just under loop 5 between helix 2 and helix5[4]. In addition, all of the three structures were solved in the same space group R32

143

**(Table 1-1),** which might interfere with helix 5 ability to move after soaking the crystals in the inhibitors solutions[26]. In conclusion, these new structures seem to support the model proposed in the review: loop 5 and helix 5 act cooperatively to form part of the substrate entry gate. However, a substrate:enzyme complex structure is needed for more details.

## 5-4. Overview regarding cytoplasmic domains of rhomboids.

The 6 TMS minmal core of rhomboid-family proteins have been extensively structurally determined. However, the extra TMS present or cytoplasmic domains possibly present at the N- and/or C-terminal ends of some rhomboids have remain poorly characterized. In chapter 4, we present the first X-ray structure of such cytoplasmic domain, from ecGlpG. The structure reveals a domain swapping between two monomers. These extra-membranous domains are of particular importance as they have been suggested to play a role in substrate recognition[27; 28] and enzyme trafficking[9]. Nevertheless, other roles might be hypothesized such as a regulation or oligomerization (supported by our data from chapter 4). However, much more information is needed to understand such domains from a functional and structural point of view. In addition, one may underline that their roles might be a key for diversification and specificity of rhomboids.

As described in chapter 4, among 32 sequences of prokaryotic rhomboids from NBI database, similar N-terminal cytoplasmic domain (90 residues approximatively) were observed in 29 of them. Alignment of these domains **(Figure 4-10)** revealed their high similarity and that they are predicted to be structured. In addition, we generated a globular monomer model from our domain swapped dimer structure. Its superimposition with the NMR structures of the cytoplasmic domains of *P. aeruginosa*[29] and ecGlpG[30] reveal an overall similar fold with an r.m.s.d. of 4.6 Å

144

and 1.2 Å , respectively (**Figure 4-4A and 4-4B**). Taken together, it is tempting to hypothesize that prokaryotic rhomboids containing such domain present a similar structure, which may undergo domain swapping. However, other high-resolution structures of N-terminal cytoplasmic domains from other prokaryotic rhomboids are needed as well as studies about their dimerization process. In addition, it would[31] be interesting to test the catalytic activity of prokaryotic rhomboids using our optimized functional assays, detailed in chapter 4, with and without the presence of the cytoplasmic domain. Furthermore, even if *P. stuartii* rhomboid does not contain a N-terminal cytoplasmic domain, our optimized functional assay developed with a water-soluble molecule and its own physiological substrate TatA would be relevant. Such study is undergoing in the laboratory.

In eukaryotic rhomboids, such extra-membranous domains may exist. Consequently, for rhomboid proteases, it would be of particular interest to focus on disease-related RHBDL2 and RHBDL4, the domains of which are potentially linked to substrate recognition. RHBDL4 cytoplasmic domain has been predicted to be largely unstructured and therefore being a real challenge for structural characterization. However, a structural prediction of these domains might highlight some structured parts suitable for crystallization. Pseudoproteases have also been intensively studied due to their narrow link to diseases such as cancer and cystic fibrosis. It would be essential to analyse in the same way their extra-membranous domains. While not all Derlins have been predicted to possess such extra-membranous domain, iRhoms present an extremely long N-terminal domain. The high conservation of this domain for iRhoms most probably implies functional significance. Especially, N-terminal cytoplasmic domain mutations of RHBDF2 were identified in several cases of inherited TOC. Therefore, it is of particular

importance to gain more structural information about them. It would be interesting to investigate the putative oligomerization of these domains via, potentially, domain swapping.

Finally, in chapter 4, our data did not allow us to determine if the cytoplasmic and membrane domains of ecGlpG were interacting and the structure of the full-length protein might be a way to answer this question. However, the problems to obtain well-diffracting crystals developed in chapter 3, are hard to overcome. In addition, it would be possible to crystallize the dimer of ecGlpGFL, which may restrain the membrane and cytoplasmic domains from interacting.

## 5-5. Concluding perspectives

Rhomboids have been linked to various human diseases, including epithelial cancer, type II diabetes and many others. Unfortunately, the exact mechanism by which they cause disease is still unknown. Critical to a detailed understanding of these enzymes is to ascertain their regulation and molecular mechanism. By studying different constructs of ecGlpG and hiGlpG, we gained valuable insights into their structure and proteolytic catalysis. The next goal would be to adapt these techniques to human rhomboids. New structural and functional details of these rhomboids is the first step towards understanding how these proteases cause disease and will provide insight into specific drug design.

## 5-6. References

1.    Ben-Shem, A., Fass, D. & Bibi, E. (2007). Structural basis for intramembrane proteolysis by rhomboid serine proteases. *Proc Natl Acad Sci U S A* **104**, 462-6.
2.    Vinothkumar, K. R., Strisovsky, K., Andreeva, A., Christova, Y., Verhelst, S. & Freeman, M. (2010). The structural basis for catalysis and substrate specificity of a rhomboid protease. *EMBO J* **29**, 3797-809.
3.    Xue, Y. & Ha, Y. (2012). Catalytic mechanism of rhomboid protease GlpG probed by 3,4-dichloroisocoumarin and diisopropyl fluorophosphonate. *J Biol Chem* **287**, 3099-107.
4.    Xue, Y., Chowdhury, S., Liu, X., Akiyama, Y., Ellman, J. & Ha, Y. (2012). Conformational change in rhomboid protease GlpG induced by inhibitor binding to its S' subsites. *Biochemistry* **51**, 3723-31.
5.    Burgi, H. B. & Dunitz, J. D. (1974). Stereochemistry of reaction paths at carbonyl centers. *Tetrahedron* **30**, 10.
6.    Botos, I., Melnikov, E. E., Cherry, S., Tropea, J. E., Khalatova, A. G., Rasulova, F., Dauter, Z., Maurizi, M. R., Rotanova, T. V., Wlodawer, A. & Gustchina, A. (2004). The catalytic domain of Escherichia coli Lon protease has a unique fold and a Ser-Lys dyad in the active site. *J Biol Chem* **279**, 8140-8.
7.    Lemberg, M. K. & Freeman, M. (2007). Cutting proteins within lipid bilayers: rhomboid structure and mechanism. *Mol Cell* **28**, 930-40.
8.    Cipolat, S., Rudka, T., Hartmann, D., Costa, V., Serneels, L., Craessaerts, K., Metzger, K., Frezza, C., Annaert, W., D'Adamio, L., Derks, C., Dejaegere, T., Pellegrini, L., D'Hooge, R., Scorrano, L. & De Strooper, B. (2006). Mitochondrial rhomboid PARL regulates cytochrome c release during apoptosis via OPA1-dependent cristae remodeling. *Cell* **126**, 163-75.
9.    Sheiner, L., Dowse, T. J. & Soldati-Favre, D. (2008). Identification of trafficking determinants for polytopic rhomboid proteases in Toxoplasma gondii. *Traffic* **9**, 665-77.
10.   Strisovsky, K., Sharpe, H. J. & Freeman, M. (2009). Sequence-specific intramembrane proteolysis: identification of a recognition motif in rhomboid substrates. *Mol Cell* **36**, 1048-59.
11.   Lee, J. R., Urban, S., Garvey, C. F. & Freeman, M. (2001). Regulated intracellular ligand transport and proteolysis control EGF signal activation in Drosophila. *Cell* **107**, 161-71.
12.   Urban, S. & Freeman, M. (2003). Substrate specificity of rhomboid intramembrane proteases is governed by helix-breaking residues in the substrate transmembrane domain. *Mol Cell* **11**, 1425-34.
13.   Brossier, F., Jewett, T. J., Sibley, L. D. & Urban, S. (2005). A spatially localized rhomboid protease cleaves cell surface adhesins essential for invasion by Toxoplasma. *Proc Natl Acad Sci U S A* **102**, 4146-51.
14.   Howell, S. A., Hackett, F., Jongco, A. M., Withers-Martinez, C., Kim, K., Carruthers, V. B. & Blackman, M. J. (2005). Distinct mechanisms govern proteolytic shedding of a key invasion protein in apicomplexan pathogens. *Mol Microbiol* **57**, 1342-56.
15.   Whitworth, A. J., Lee, J. R., Ho, V. M., Flick, R., Chowdhury, R. & McQuibban, G. A. (2008). Rhomboid-7 and HtrA2/Omi act in a common pathway with the Parkinson's disease factors Pink1 and Parkin. *Dis Model Mech* **1**, 168-74; discussion 173.
16.   Hatunic, M., Stapleton, M., Hand, E., DeLong, C., Crowley, V. E. & Nolan, J. J. (2009). The Leu262Val polymorphism of presenilin associated rhomboid like protein (PARL) is

associated with earlier onset of type 2 diabetes and increased urinary microalbumin creatinine ratio in an Irish case-control population. *Diabetes Res Clin Pract* **83**, 316-9.

17.    Gottlieb, E. (2006). OPA1 and PARL keep a lid on apoptosis. *Cell* **126**, 27-9.

18.    Yan, Z., Zou, H., Tian, F., Grandis, J. R., Mixson, A. J., Lu, P. Y. & Li, L. Y. (2008). Human rhomboid family-1 gene silencing causes apoptosis or autophagy to epithelial cancer cells and inhibits xenograft tumor growth. *Mol Cancer Ther* **7**, 1355-64.

19.    Urban, S. (2010). Taking the plunge: integrating structural, enzymatic and computational insights into a unified model for membrane-immersed rhomboid proteolysis. *Biochem J* **425**, 501-12.

20.    Wu, Z., Yan, N., Feng, L., Oberstein, A., Yan, H., Baker, R. P., Gu, L., Jeffrey, P. D., Urban, S. & Shi, Y. (2006). Structural analysis of a rhomboid family intramembrane protease reveals a gating mechanism for substrate entry. *Nat Struct Mol Biol* **13**, 1084-91.

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

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

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

24.    Baker, R. P., Young, K., Feng, L., Shi, Y. & Urban, S. (2007). Enzymatic analysis of a rhomboid intramembrane protease implicates transmembrane helix 5 as the lateral substrate gate. *Proc Natl Acad Sci U S A* **104**, 8257-62.

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

26.    Brooks, C. L. & Lemieux, M. J. (2013). Untangling structure-function relationships in the rhomboid family of intramembrane proteases. *Biochim Biophys Acta* **1828**, 2862-72.

27.    Fleig, L., Bergbold, N., Sahasrabudhe, P., Geiger, B., Kaltak, L. & Lemberg, M. K. (2012). Ubiquitin-dependent intramembrane rhomboid protease promotes ERAD of membrane proteins. *Mol Cell* **47**, 558-69.

28.    Lohi, O., Urban, S. & Freeman, M. (2004). Diverse substrate recognition mechanisms for rhomboids; thrombomodulin is cleaved by Mammalian rhomboids. *Curr Biol* **14**, 236-41.

29.    Del Rio, A., Dutta, K., Chavez, J., Ubarretxena-Belandia, I. & Ghose, R. (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. *Journal of molecular biology* **365**, 109-22.

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

31.    Erez, E. & Bibi, E. (2009). Cleavage of a multispanning membrane protein by an intramembrane serine protease. *Biochemistry* **48**, 12314-22.

32.    Girard, E., Marchal, S., Perez, J. *et al.*, (2010). Structure-function perturbation and dissociation of tetrameric urate oxidase by hydrostatic pressure. *Biophys. J.* **98**, 2365-73.

33.    Armougom, F., Moretti, S., Poirot O., Audic, S., Dumas, P., Schaeli, B., Keduas, V., Notredame, C. (2006). Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. *Nucleic Acids Res.,* **34**, W604-8

34.  Gouet, P., Courcelle, E., Stuart, D.I., Metoz, F. (1999) ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **15**, 305-8.

**Bibliography**

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

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W.*, et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr *66*, 213-221.

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

Akiyama, K., Yamauchi, S., Nakato, T., Maruyama, M., Sugahara, T., and Kishida, T. (2007). Antifungal activity of tetra-substituted tetrahydrofuran lignan, (-)-virgatusin, and its structure-activity relationship. Biosci Biotechnol Biochem *71*, 1028-1035.

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

Alexander, C., Votruba, M., Pesch, U.E., Thiselton, D.L., Mayer, S., Moore, A., Rodriguez, M., Kellner, U., Leo-Kottler, B., Auburger, G.*, et al.* (2000). OPA1, encoding a dynamin-related GTPase, is mutated in autosomal dominant optic atrophy linked to chromosome 3q28. Nat Genet *26*, 211-215.

Armougom, F., Moretti, S., Poirot, O., Audic, S., Dumas, P., Schaeli, B., Keduas, V., and Notredame, C. (2006). Expresso: automatic incorporation of structural information in multiple sequence alignments using 3D-Coffee. Nucleic Acids Res *34*, W604-608.

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

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

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

Barrett, A.J., Rawlings, N.D., and O'Brien, E.A. (2001). The MEROPS database as a protease information system. J Struct Biol *134*, 95-102.

Barrientos, L.G., and Gronenborn, A.M. (2002). The domain-swapped dimer of cyanovirin-N contains two sets of oligosaccharide binding sites in solution. Biochem Biophys Res Commun *298*, 598-602.

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

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

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

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

Blaydon, D.C., Etheridge, S.L., Risk, J.M., Hennies, H.C., Gay, L.J., Carroll, R., Plagnol, V., McRonald, F.E., Stevens, H.P., Spurr, N.K.*, et al.* (2012). RHBDF2 mutations are associated with tylosis, a familial esophageal cancer syndrome. Am J Hum Genet *90*, 340-346.

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

Botos, I., Melnikov, E.E., Cherry, S., Tropea, J.E., Khalatova, A.G., Rasulova, F., Dauter, Z., Maurizi, M.R., Rotanova, T.V., Wlodawer, A.*, et al.* (2004). The catalytic domain of Escherichia coli Lon protease has a unique fold and a Ser-Lys dyad in the active site. J Biol Chem *279*, 8140-8148.

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

Brooks, C.L., and Lemieux, M.J. (2013). Untangling structure-function relationships in the rhomboid family of intramembrane proteases. Biochim Biophys Acta *1828*, 2862-2872.

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

Brown, K.E., Kerr, M., and Freeman, M. (2007). The EGFR ligands Spitz and Keren act cooperatively in the Drosophila eye. Dev Biol *307*, 105-113.

Burgi, H.B., Dunitz, J.D., Lehn, J.M., Wipff G. (1974). Stereochemistry of reaction paths at carbonyl centres. Tetrahedron *30*, 1563-1572.

Carvalho, P., Goder, V., and Rapoport, T.A. (2006). Distinct ubiquitin-ligase complexes define convergent pathways for the degradation of ER proteins. Cell *126*, 361-373.

Casey, J.R., and Reithmeier, R.A. (1991). Analysis of the oligomeric state of Band 3, the anion transport protein of the human erythrocyte membrane, by size exclusion high performance liquid chromatography. Oligomeric stability and origin of heterogeneity. J Biol Chem *266*, 15726-15737.

Chae, P.S., Gotfryd, K., Pacyna, J., Miercke, L.J., Rasmussen, S.G., Robbins, R.A., Rana, R.R., Loland, C.J., Kobilka, B., Stroud, R.*, et al.* (2010a). Tandem facial amphiphiles for membrane protein stabilization. J Am Chem Soc *132*, 16750-16752.

Chae, P.S., Rasmussen, S.G., Rana, R.R., Gotfryd, K., Chandra, R., Goren, M.A., Kruse, A.C., Nurva, S., Loland, C.J., Pierre, Y.*, et al.* (2010b). Maltose-neopentyl glycol (MNG) amphiphiles for solubilization, stabilization and crystallization of membrane proteins. Nat Methods *7*, 1003-1008.

Chae, P.S., Wander, M.J., Bowling, A.P., Laible, P.D., and Gellman, S.H. (2008). Glycotripod amphiphiles for solubilization and stabilization of a membrane-protein superassembly: importance of branching in the hydrophilic portion. Chembiochem *9*, 1706-1709.

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

Cheng, T.L., Wu, Y.T., Lin, H.Y., Hsu, F.C., Liu, S.K., Chang, B.I., Chen, W.S., Lai, C.H., Shi, G.Y., and Wu, H.L. (2011). Functions of rhomboid family protease RHBDL2 and thrombomodulin in wound healing. J Invest Dermatol *131*, 2486-2494.

Christianson, J.C., Olzmann, J.A., Shaler, T.A., Sowa, M.E., Bennett, E.J., Richter, C.M., Tyler, R.E., Greenblatt, E.J., Harper, J.W., and Kopito, R.R. (2012). Defining human ERAD networks through an integrative mapping strategy. Nat Cell Biol *14*, 93-105.

Cipolat, S., Rudka, T., Hartmann, D., Costa, V., Serneels, L., Craessaerts, K., Metzger, K., Frezza, C., Annaert, W., D'Adamio, L.*, et al.* (2006). Mitochondrial rhomboid PARL regulates cytochrome c release during apoptosis via OPA1-dependent cristae remodeling. Cell *126*, 163-175.

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

Davis, I.W., Leaver-Fay, A., Chen, V.B., Block, J.N., Kapral, G.J., Wang, X., Murray, L.W., Arendall, W.B., 3rd, Snoeyink, J., Richardson, J.S.*, et al.* (2007). MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. Nucleic Acids Res *35*, W375-383.

Del Rio, A., Dutta, K., Chavez, J., Ubarretxena-Belandia, I., and Ghose, R. (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.

Dominguez, M., Wasserman, J.D., and Freeman, M. (1998). Multiple functions of the EGF receptor in Drosophila eye development. Curr Biol *8*, 1039-1048.

Dutt, A., Canevascini, S., Froehli-Hoier, E., and Hajnal, A. (2004). EGF signal propagation during C. elegans vulval development mediated by ROM-1 rhomboid. PLoS Biol *2*, e334.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr D Biol Crystallogr *66*, 486-501.

Erez, E., and Bibi, E. (2009). Cleavage of a multispanning membrane protein by an intramembrane serine protease. Biochemistry *48*, 12314-12322.

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

Faham, S., and Bowie, J.U. (2002). Bicelle crystallization: a new method for crystallizing membrane proteins yields a monomeric bacteriorhodopsin structure. J Mol Biol *316*, 1-6.

Feldman, A.R., Lee, J., Delmas, B., and Paetzel, M. (2006). Crystal structure of a novel viral protease with a serine/lysine catalytic dyad mechanism. J Mol Biol *358*, 1378-1389.

Fleig, L., Bergbold, N., Sahasrabudhe, P., Geiger, B., Kaltak, L., and Lemberg, M.K. (2012). Ubiquitin-dependent intramembrane rhomboid protease promotes ERAD of membrane proteins. Mol Cell *47*, 558-569.

Freeman, M. (2004). Proteolysis within the membrane: rhomboids revealed. Nat Rev Mol Cell Biol *5*, 188-197.

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

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

Frezza, C., Cipolat, S., Martins de Brito, O., Micaroni, M., Beznoussenko, G.V., Rudka, T., Bartoli, D., Polishuck, R.S., Danial, N.N., De Strooper, B.*, et al.* (2006). OPA1 controls apoptotic cristae remodeling independently from mitochondrial fusion. Cell *126*, 177-189.

Girard, E., Marchal, S., Perez, J., Finet, S., Kahn, R., Fourme, R., Marassio, G., Dhaussy, A.C., Prange, T., Giffard, M.*, et al.* (2010). Structure-function perturbation and dissociation of tetrameric urate oxidase by high hydrostatic pressure. Biophys J *98*, 2365-2373.

Gottlieb, E. (2006). OPA1 and PARL keep a lid on apoptosis. Cell *126*, 27-29.

Gouet, P., Courcelle, E., Stuart, D.I., and Metoz, F. (1999). ESPript: analysis of multiple sequence alignments in PostScript. Bioinformatics *15*, 305-308.

Ha, Y. (2007). Structural principles of intramembrane proteases. Curr Opin Struct Biol *17*, 405-411.

Hatunic, M., Stapleton, M., Hand, E., DeLong, C., Crowley, V.E., and Nolan, J.J. (2009). The Leu262Val polymorphism of presenilin associated rhomboid like protein (PARL) is associated with earlier onset of type 2 diabetes and increased urinary microalbumin creatinine ratio in an Irish case-control population. Diabetes Res Clin Pract *83*, 316-319.

Howell, S.A., Hackett, F., Jongco, A.M., Withers-Martinez, C., Kim, K., Carruthers, V.B., and Blackman, M.J. (2005). Distinct mechanisms govern proteolytic shedding of a key invasion protein in apicomplexan pathogens. Mol Microbiol *57*, 1342-1356.

Hunte, C., and Michel, H. (2002). Crystallisation of membrane proteins mediated by antibody fragments. Curr Opin Struct Biol *12*, 503-508.

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

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

Kabsch, W. (2010). Xds. Acta Crystallogr D Biol Crystallogr *66*, 125-132.

Kateete, D.P., Katabazi, F.A., Okeng, A., Okee, M., Musinguzi, C., Asiimwe, B.B., Kyobe, S., Asiimwe, J., Boom, W.H., and Joloba, M.L. (2012). Rhomboids of Mycobacteria: characterization using an aarA mutant of Providencia stuartii and gene deletion in Mycobacterium smegmatis. PLoS One *7*, e45741.

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

Lazareno-Saez, C., Brooks, C.L., and Lemieux, M.J. (2011). Structural comparison of substrate entry gate for rhomboid intramembrane peptidases. Biochem Cell Biol *89*, 216-223.

Lee, J.R., Urban, S., Garvey, C.F., and Freeman, M. (2001). Regulated intracellular ligand transport and proteolysis control EGF signal activation in Drosophila. Cell *107*, 161-171.

Lemberg, M.K. (2013). Sampling the membrane: function of rhomboid-family proteins. Trends Cell Biol *23*, 210-217.

Lemberg, M.K., and Freeman, M. (2007a). Cutting proteins within lipid bilayers: rhomboid structure and mechanism. Mol Cell *28*, 930-940.

Lemberg, M.K., and Freeman, M. (2007b). Functional and evolutionary implications of enhanced genomic analysis of rhomboid intramembrane proteases. Genome Res *17*, 1634-1646.

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

Lemieux, M.J. (2007). Eukaryotic major facilitator superfamily transporter modeling based on the prokaryotic GlpT crystal structure. Mol Membr Biol *24*, 333-341.

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

Lemieux, M.J., Reithmeier, R.A., and Wang, D.N. (2002). Importance of detergent and phospholipid in the crystallization of the human erythrocyte anion-exchanger membrane domain. J Struct Biol *137*, 322-332.

Li, Y.M., Lai, M.T., Xu, M., Huang, Q., DiMuzio-Mower, J., Sardana, M.K., Shi, X.P., Yin, K.C., Shafer, J.A., and Gardell, S.J. (2000). Presenilin 1 is linked with gamma-secretase activity in the detergent solubilized state. Proc Natl Acad Sci U S A *97*, 6138-6143.

Liu, L., Byeon, I.J., Bahar, I., and Gronenborn, A.M. (2012). Domain swapping proceeds via complete unfolding: a 19F- and 1H-NMR study of the Cyanovirin-N protein. J Am Chem Soc *134*, 4229-4235.

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

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

Luo, Y., Pfuetzner, R.A., Mosimann, S., Paetzel, M., Frey, E.A., Cherney, M., Kim, B., Little, J.W., and Strynadka, N.C. (2001). Crystal structure of LexA: a conformational switch for regulation of self-cleavage. Cell *106*, 585-594.

Maassen, J.A., Janssen, G.M., and t Hart, L.M. (2005). Molecular mechanisms of mitochondrial diabetes (MIDD). Ann Med *37*, 213-221.

MacArthur, M.W., and Thornton, J.M. (1991). Influence of proline residues on protein conformation. J Mol Biol *218*, 397-412.

Macol, C.P., Tsuruta, H., Stec, B., and Kantrowitz, E.R. (2001). Direct structural evidence for a concerted allosteric transition in Escherichia coli aspartate transcarbamoylase. Nat Struct Biol *8*, 423-426.

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

Maegawa, S., Koide, K., Ito, K., and Akiyama, Y. (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.

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

Mesak, L.R., Mesak, F.M., and Dahl, M.K. (2004). Expression of a novel gene, gluP, is essential for normal Bacillus subtilis cell division and contributes to glucose export. BMC Microbiol *4*, 13.

Moin, S.M., and Urban, S. (2012). Membrane immersion allows rhomboid proteases to achieve specificity by reading transmembrane segment dynamics. Elife (Cambridge) *1*, e00173.

Newstead, S., Ferrandon, S., and Iwata, S. (2008). Rationalizing alpha-helical membrane protein crystallization. Protein Sci *17*, 466-472.

Ni, Y., Lempp, F.A., Mehrle, S., Nkongolo, S., Kaufman, C., Falth, M., Stindt, J., Koniger, C., Nassal, M., Kubitz, R*., et al.* (2014). Hepatitis B and D viruses exploit sodium taurocholate co-transporting polypeptide for species-specific entry into hepatocytes. Gastroenterology *146*, 1070-1083.

Oda, Y., Okada, T., Yoshida, H., Kaufman, R.J., Nagata, K., and Mori, K. (2006). Derlin-2 and Derlin-3 are regulated by the mammalian unfolded protein response and are required for ER-associated degradation. J Cell Biol *172*, 383-393.

Paetzel, M., Dalbey, R.E., and Strynadka, N.C. (1998). Crystal structure of a bacterial signal peptidase in complex with a beta-lactam inhibitor. Nature *396*, 186-190.

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

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

Porcelli, I., de Leeuw, E., Wallis, R., van den Brink-van der Laan, E., de Kruijff, B., Wallace, B.A., Palmer, T., and Berks, B.C. (2002). Characterization and membrane assembly of the TatA component of the Escherichia coli twin-arginine protein transport system. Biochemistry *41*, 13690-13697.

Radisky, E.S., Lee, J.M., Lu, C.J., and Koshland, D.E., Jr. (2006). Insights into the serine protease mechanism from atomic resolution structures of trypsin reaction intermediates. Proc Natl Acad Sci U S A *103*, 6835-6840.

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

Reddy, T., and Rainey, J.K. (2012). Multifaceted substrate capture scheme of a rhomboid protease. J Phys Chem B *116*, 8942-8954.

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

Sampathkumar, P., Mak, M.W., Fischer-Witholt, S.J., Guigard, E., Kay, C.M., and Lemieux, M.J. (2012). Oligomeric state study of prokaryotic rhomboid proteases. Biochim Biophys Acta *1818*, 3090-3097.

Sanders, C.R., and Prosser, R.S. (1998). Bicelles: a model membrane system for all seasons? Structure *6*, 1227-1234.

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

Sheiner, L., Dowse, T.J., and Soldati-Favre, D. (2008). Identification of trafficking determinants for polytopic rhomboid proteases in Toxoplasma gondii. Traffic *9*, 665-677.

Shen, J., and Prywes, R. (2004). Dependence of site-2 protease cleavage of ATF6 on prior site-1 protease digestion is determined by the size of the luminal domain of ATF6. J Biol Chem *279*, 43046-43051.

Sherratt, A.R., Braganza, M.V., Nguyen, E., Ducat, T., and Goto, N.K. (2009). Insights into the effect of detergents on the full-length rhomboid protease from Pseudomonas aeruginosa and its cytosolic domain. Biochim Biophys Acta *1788*, 2444-2453.

Sibley, L.D. (2013). The roles of intramembrane proteases in protozoan parasites. Biochim Biophys Acta *1828*, 2908-2915.

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

Smith, T.S., Southan, C., Ellington, K., Campbell, D., Tew, D.G., and Debouck, C. (2001). Identification, genomic organization, and mRNA expression of LACTB, encoding a serine beta-lactamase-like protein with an amino-terminal transmembrane domain. Genomics *78*, 12-14.

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

Southan, C. (2001b). A genomic perspective on human proteases as drug targets. Drug Discov Today *6*, 681-688.

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

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

Tsruya, R., Schlesinger, A., Reich, A., Gabay, L., Sapir, A., and Shilo, B.Z. (2002). Intracellular trafficking by Star regulates cleavage of the Drosophila EGF receptor ligand Spitz. Genes Dev *16*, 222-234.

Tsruya, R., Wojtalla, A., Carmon, S., Yogev, S., Reich, A., Bibi, E., Merdes, G., Schejter, E., and Shilo, B.Z. (2007). Rhomboid cleaves Star to regulate the levels of secreted Spitz. EMBO J *26*, 1211-1220.

Ulas, U.H., Unlu, E., Hamamcioglu, K., Odabasi, Z., Cakci, A., and Vural, O. (2006). Dysautonomia in fibromyalgia syndrome: sympathetic skin responses and RR interval analysis. Rheumatol Int *26*, 383-387.

Urban, S. (2009). Making the cut: central roles of intramembrane proteolysis in pathogenic microorganisms. Nat Rev Microbiol *7*, 411-423.

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.

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

Urban, S., and Dickey, S.W. (2011). The rhomboid protease family: a decade of progress on function and mechanism. Genome Biol *12*, 231.

Urban, S., and Freeman, M. (2002). Intramembrane proteolysis controls diverse signalling pathways throughout evolution. Curr Opin Genet Dev *12*, 512-518.

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

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

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

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

Vagin, A., Teplyakov, A. (1997). MOLREP: an automated program for molecular replacement. . J Appl Crystallogr *30*, 1022-1025.

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

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

Walder, K., Kerr-Bayles, L., Civitarese, A., Jowett, J., Curran, J., Elliott, K., Trevaskis, J., Bishara, N., Zimmet, P., Mandarino, L.*, et al.* (2005). The mitochondrial rhomboid protease PSARL is a new candidate gene for type 2 diabetes. Diabetologia *48*, 459-468.

Wan, C., Fu, J., Wang, Y., Miao, S., Song, W., and Wang, L. (2012). Exosome-related multi-pass transmembrane protein TSAP6 is a target of rhomboid protease RHBDD1-induced proteolysis. PLoS One *7*, e37452.

Wang, B., Heath-Engel, H., Zhang, D., Nguyen, N., Thomas, D.Y., Hanrahan, J.W., and Shore, G.C. (2008). BAP31 interacts with Sec61 translocons and promotes retrotranslocation of CFTRDeltaF508 via the derlin-1 complex. Cell *133*, 1080-1092.

Wang, Y., and Ha, Y. (2007). Open-cap conformation of intramembrane protease GlpG. Proc Natl Acad Sci U S A *104*, 2098-2102.

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

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

Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F., and Jones, D.T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. J Mol Biol *337*, 635-645.

Wasserman, J.D., Urban, S., and Freeman, M. (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.

Whitworth, A.J., Lee, J.R., Ho, V.M., Flick, R., Chowdhury, R., and McQuibban, G.A. (2008). Rhomboid-7 and HtrA2/Omi act in a common pathway with the Parkinson's disease factors Pink1 and Parkin. Dis Model Mech *1*, 168-174; discussion 173.

Wojnarowicz, P.M., Provencher, D.M., Mes-Masson, A.M., and Tonin, P.N. (2012). Chromosome 17q25 genes, RHBDF2 and CYGB, in ovarian cancer. Int J Oncol *40*, 1865-1880.

Wolfe, M.S. (2009). Intramembrane proteolysis. Chem Rev *109*, 1599-1612.

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

Xue, Y., Chowdhury, S., Liu, X., Akiyama, Y., Ellman, J., and Ha, Y. (2012). Conformational change in rhomboid protease GlpG induced by inhibitor binding to its S' subsites. Biochemistry *51*, 3723-3731.

Xue, Y., and Ha, Y. (2012). Catalytic mechanism of rhomboid protease GlpG probed by 3,4-dichloroisocoumarin and diisopropyl fluorophosphonate. J Biol Chem *287*, 3099-3107.

Yamasaki, M., Li, W., Johnson, D.J., and Huntington, J.A. (2008). Crystal structure of a stable dimer reveals the molecular basis of serpin polymerization. Nature *455*, 1255-1258.

Yan, Z., Zou, H., Tian, F., Grandis, J.R., Mixson, A.J., Lu, P.Y., and Li, L.Y. (2008). Human rhomboid family-1 gene silencing causes apoptosis or autophagy to epithelial cancer cells and inhibits xenograft tumor growth. Mol Cancer Ther *7*, 1355-1364.

Ye, Y., Shibata, Y., Yun, C., Ron, D., and Rapoport, T.A. (2004). A membrane protein complex mediates retro-translocation from the ER lumen into the cytosol. Nature *429*, 841-847.

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

Younger, J.M., Chen, L., Ren, H.Y., Rosser, M.F., Turnbull, E.L., Fan, C.Y., Patterson, C., and Cyr, D.M. (2006). Sequential quality-control checkpoints triage misfolded cystic fibrosis transmembrane conductance regulator. Cell *126*, 571-582.

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

Zou, H., Thomas, S.M., Yan, Z.W., Grandis, J.R., Vogt, A., and Li, L.Y. (2009). Human rhomboid family-1 gene RHBDF1 participates in GPCR-mediated transactivation of EGFR growth signals in head and neck squamous cancer cells. FASEB J *23*, 425-432.