Rhomboid Proteolysis is a Rate-Governed Reaction, Yet is Dispensable for *E. coli* Colonization of the Mouse Colon

by

Seth Wayne Dickey

A dissertation submitted to The Johns Hopkins University in conformity with the requirements of the degree of Doctor of Philosophy

Baltimore, Maryland

October, 2013
Abstract

Proteolysis, the controlled dissection of proteins into two parts, is essential to all life. Intramembrane proteolysis, whereby specialized enzymes cut membrane-resident segments, extends this potent chemical transformation to the unique membrane milieu. Intense research on rhomboid proteases, one of three widely conserved intramembrane protease families, has greatly advanced our knowledge of their biology in eukaryotes and their impact on human health and disease. Nonetheless, we have scant information about their roles in bacteria, which constitute the overwhelming diversity on Earth and burden humanity with myriad diseases. In this thesis, I tackle this deficit using the complementary approaches of genetics and biochemistry.

In chapter II, I describe my work using reverse genetics and substrate identification to uncover the role of the *Escherichia coli* rhomboid protease, GlpG, with special focus on the lateral substrate gate. I have found that the lateral gate likely endows GlpG with allosteric regulation, and that GlpG cleaves two *E. coli* proteins, BasS and Rtn. Surprisingly however, GlpG is dispensable for normal growth in pure culture, response to diverse stresses, and colonization of and persistence within the mouse colon.

In chapter III, I sought to generate new leads for bacterial rhomboid protease biology by investigating their biochemistry. I developed a steady-state kinetic assay and complementary thermodynamic methods to examine the kinetics of rhomboid protease cleavage in their natural membrane environment. Unexpectedly, nine diverse bacterial rhomboid proteases show identical and weak affinities ($K_m \sim 0.14$ mole%, $135 \mu$M; $K_d \sim 0.08$ mole%, $191 \mu$M) with four substrate variants. In contrast, the turnover constant ranged 10,000-fold; yet, all rhomboid proteases displayed slow kinetics with the fastest ($k_{\text{cat}} \sim 0.1$ s$^{-1}$) being 1,000-fold slower than trypsin ($k_{\text{cat}} \sim 100$ s$^{-1}$). This kinetic profile exposes an unanticipated similarity to the DNA repair enzymes DNA glycosylases. I conclude with a new model for rhomboid proteolysis as a kinetically driven process, with the rich precedence of DNA glycosylases to guide future research.

Thesis Advisor: Sinisa Urban, PhD
Thesis Reader: Brendan Cormack, PhD
Acknowledgements

“We come unbidden into this life, and if we are lucky we find a purpose beyond starvation, misery, and early death which, lest we forget, is the common lot”

- Abraham Verghese, Cutting for Stone

The capacity to advance on previous work is a characteristic that uniquely describes humanity, and it is the essential drive, mixed with time, that has led to modern societies. As such my thesis is a humble contribution that depends on the hard-earned work of my forebears and contemporaries. Likewise, my 28 years of existence has been shaped by varied intimate interactions with my family and many friends.

First and foremost, I am thankful for the keen guidance given to me by my thesis mentor, Sin Urban. Under his tutelage and example, I have learned the core skills of being a scientist: to approach a project with enthusiasm and tenacity, to immerse myself within the literature, to master any technique necessary, and to communicate my findings effectively.

Members of the Urban Lab, past (Syed Moin and Slavica Djuranovic) and present (Rosanna Baker, Shiv Gandhi, Sangwoo Cho, and John Tran), have provided invaluable discussions and helpful suggestions during my thesis work. I want to give special thanks Rosanna Baker and Sangwoo Cho for our rewarding collaboration described in chapter III of this dissertation.

I thank my thesis committee members: Jeremy Nathans, Brendan Cormack, Peter Maloney, Cynthia Sears, and Jon Lorsch. I value their constructive discussions about my project and my future directions. I am especially grateful to Cynthia Sears and her lab members for their help training me to colonize mice with bacteria, which I describe in chapter II.

I also want to thank my family who have positively shaped who I am. I thank my mother and late father for their long-standing priority on education and their efforts to make that education available to me. I also thank my entire immediate family, including each of my five sisters, for their undying support over the many years.
Finally, I want to thank Iris Hartley, whose love and support in less than a year has transformed my life beyond what I thought possible. I look forward with delight at our promising future together.
# Table of Contents

Abstract ii

Acknowledgements iii

Table of Contents v

List of Tables vi

List of Figures vi

Chapter I: Introduction 1
  Intramembrane Proteases: Ancient Enzymes Serving Key Roles: 2
  Rhomboid Protease Biology: A Tale of Two Domains: 3
  Rhomboid Protease Biochemistry and Mechanism: 8
  References: 12
  Figure Legends: 20

Chapter II: Biological Investigations of the *E. coli* Rhomboid Protease GlpG and its Lateral Gate 25
  Summary: 26
  Introduction: 27
  Results: 28
  Discussion: 35
  Material and Methods: 36
  Acknowledgements: 40
  References: 41
  Figure Legends: 43

Chapter III: Membrane-Immerse Rhomboid Proteolysis is a Rate-Driven Reaction with a Low Affinity for Substrates 58
  Summary: 59
  Introduction: 60
  Results: 61
  Discussion: 67
  Materials and Methods: 69
  Figure Legends: 77
Appendix: *Vibrio cholerae* GlyGly-CTERM Proteins are Substrates of an Atypical
*Vibrio cholerae* Rhomboid Protease

**Introduction:** 89

**Results:** 90

**Discussion:** 91

**Material and Methods:** 91

**References:** 92

**Figure Legends:** 93

**Curriculum Vitae** 98

**List of Tables**

Table 2.1: 46
Table 2.2: 47
Table 2.3: 48
Table 2.4: 49
Table 2.5: 50

**List of Figures**

Figure 1.1: 22
Figure 1.2: 23
Figure 1.3: 24
Figure 2.1: 51
Figure 2.2: 52
Figure 2.3: 53
Figure 2.4: 54
Figure 2.5: 55
Figure 2.6: 56
Figure 2.7: 57
Figure 3.1: 81
Figure 3.2: 82
Figure 3.3: 83
Figure 3.4: 84
Chapter I

Introduction

Figure 1.1 has been adapted from: “Rhomboid protease: a decade of progress on function and mechanism” by S. W. Dickey and S. Urban. 2011. Genome Biology. 12(10): 231
Intramembrane Proteases: Ancient Enzymes Serving Key Roles:

Proteases use water molecules to enzymatically sever peptide bonds within proteins. In cells and multicellular organisms, they serve multifaceted and complex roles, which generalize into two broad categories: 1) degradative, in which cellular or dietary protein is broken down to smaller peptides or amino acids and 2) regulatory, in which specific proteins are cut, often at precise positions, to engender an appropriate and evolved response. These roles can overlap, for example, in the degradative breakdown of cyclins to promote cell cycle progression. The vast majority of proteases live in the aqueous environment as soluble proteins. In contrast, the membrane constitutes an altogether unique environment: it is inherently two-dimensional and excludes water. A consequence of the latter is that the membrane is the living boundary of cells, internal organelles, and the external environment. Proteins that are embedded within membranes through hydrophobic interactions have evolved to monitor this interface, sensing the outside environment, importing nutrients, exporting noxious substances, setting up a transmembrane voltage, and interacting with surfaces and other cells. Although proteolysis in the membrane would be potentially useful for life, it was thought to be impossible because the arid membrane lacks an essential substrate: water.

About 15 years ago this dogma was abruptly overturned (Brown et al., 2000) with the dual observations concerning the first discovered intramembrane protease, the site-2-protease (S2P). S2P harbors its metallo-active-site residues within the middle of hydrophobic transmembrane domains (TMD) (Rawson et al., 1997), and it cuts its substrate SREBP within its own TMD (Sakai et al., 1996). Shortly thereafter two distinct classes of intramembrane proteases were found: the aspartyl presenilin/signal-peptide-peptidase (SPP) family (Fluhrer and Haass, 2007; Weihofen et al., 2002; Wolfe et al., 1999), and the rhomboid serine protease family (Urban et al., 2001). Since their initial discovery, sequencing and bioinformatics has revealed that these enzymes are found nearly ubiquitously throughout evolution (Figure 1.1 for rhomboid proteases), making intramembrane proteases one of the most ancient superfamilies of all membrane proteins (Grigorenko et al., 2002; Kinch et al., 2006; Koonin et al., 2003; Ponting et al., 2002). Consistent with their conservation, members of each of the three classes play important roles in physiology and pathology. In the human brain, presenilin cuts the amyloid precursor protein, APP, to release its byproduct Aβ, which oligomerizes and...
aggregates causing the neurodegenerative Alzheimer’s disease (Hardy and Selkoe, 2002). In addition, presenilin also cleaves Notch (Selkoe and Kopan, 2003) to mediate Notch signaling, which is a crucial pathway in metazoan development, and one that frequently malfunctions in many cancers (Allenspach et al., 2002). Members of S2P family control cholesterol homeostasis in humans, and are virulence factors in many deadly microbial pathogens (Urban, 2009).

**Rhomboid Protease Biology: A Tale of Two Domains:**

Rhomboid proteases were first discovered in the eukaryotic domain and it is here where we understand their biology best (Urban and Dickey, 2011). Many of their important roles concerning human health and disease have emerged. At the same time, however, very little is known about rhomboid protease biology in the bacterial domain, representing a chasm in our ever-expanding knowledge of rhomboid protease function.

Rhomboid was first discovered to be a key gene in patterning the *Drosophila* embryo in the classic screens by Nusslein-Volhard, Eric Wieschaus, and colleagues (Jürgens et al., 1984; Mayer and Nusslein-Volhard, 1988). Subsequently, it was implicated in EGF signaling (Rutledge et al., 1992) and finally it was shown to be a unique serine protease (Urban et al., 2001) that releases EGF from its transmembrane tether, thus initiating cell-cell signaling in *Drosophila* (Figure 1.2A). A variation of this role occurs in *C. elegans* during vulval development in which the rhomboid protease CeROM-1 cleaves a long, though not a short, isoform of EGF (Dutt et al., 2004). Although it is the short isoform sent from the signal-sending anchor cells that initiates EGF signaling in *C. elegans*, CeROM-1 cleavage of the long EGF isoform is thought to amplify the signal.

Later, the mitochondrial-localized rhomboid Pcp1p, one of an ancient rhomboid protease subclass called PARL\(^1\), was found to be crucial for mitochondrial fusion in yeast by cleaving the dynamin-related GTPase Mgm1p (Herlan et al., 2003; McQuibban et al., 2003; Sesaki et al., 2003). Although mammalian PARL rhomboid proteases do not regulate mitochondrial fusion (Cipolat et al., 2006), they are important in removing damaged mitochondria. Consequently, first *Drosophila* (Whitworth et al., 2008) and then human PARL have been implicated in an intriguing pathway known to cause familial Parkinson’s disease (Dodson and Guo, 2007). In healthy mitochondria, PINK1 inserts into the inner mitochondrial membrane where PARL

---

1 PARL: presenilin associated rhomboid-like—a misnomer, as PARL does not associate with presenilin.
cuts it (Deas et al., 2011; Jin et al., 2010), leading to its constitutive degradation. However, in damaged mitochondria, PINK1 stops translocating at the outer mitochondrial membrane (Jin et al., 2010), never reaching the inner membrane and therefore preventing its cleavage by PARL. Intact, PINK1 recruits the E3 ligase PARKIN—another gene linked to familial Parkinson’s disease—to the outer mitochondrial membrane, ultimately resulting in mitophagy and clearance of the damaged mitochondria (Narendra et al., 2008). In fact, mutations in PARL with functional defects have been found in patients with Parkinson’s disease (Shi et al., 2011). These results firmly place mammalian rhomboid proteases in a pathway, the dysfunction of which is known to cause Parkinson’s disease. Furthermore, mice lacking PARL suffer from cachexia, with severe weight loss and muscle atrophy after 4 weeks of age, and die by 12 weeks of age (Cipolat et al., 2006).

In addition to PARL, humans encode 4 secretase-type (i.e. found in an organelle of the secretory pathway or on the plasma membrane) rhomboid proteases. One of these, RHBDL4, localizes to the ER where it participates in the quality control of resident membrane proteins within the larger pathway termed ER-associated-degradation or ERAD (Fleig et al., 2012). Within ERAD, misfolded luminal and membrane proteins are identified, retro-translocated back into the cytosol, ubiquitinated and degraded by the proteasome (Vembar and Brodsky, 2008). For its part, RHBDL4 encodes a ubiquitin-binding motif necessary to detect its targets, resulting in cleavage of their TMDs; moreover, RHBDL4 interacts with the AAA+ ATPase p97, which could provide energy to dislocate membranous stubs created by RHBDL4 and facilitate their destruction by the proteasome.

The roles of the remaining human rhomboid proteases are obscure, but clues including demonstration of activity and target identification exist for human RHBDL2. Although members of the ADAM family of metalloproteases have supplanted the role of EGF shedding in mammalian cells from rhomboid proteases, RHBDL2 can cleave EGF itself (Adrain et al., 2011). Moreover, RHBDL2 also cleaves thrombomodulin (Lohi et al., 2004), and this processing may be important in wound healing (Cheng et al., 2011). B-type eprhins make up additional RHBDL2 substrates (Pascall and Brown, 2004). As a cautionary note, the physiological relevance of these substrates, without clear phenotypes resulting from defective rhomboid protease cleavage, remains in question. In fact, the RHBDL2 provides an excellent example
underscoring the need to authenticate identified-substrates as physiological, and not simply as proteolytic bystanders or overexpression artifacts.

Inactive relatives to rhomboid proteases represent an exciting offshoot with fascinating biology, embodied by the conserved iRhom and Derlin protein families (Adrain and Freeman, 2012; Lemberg, 2013). Derlins are intricately involved in the translocation of misfolded proteins from the ER to the cytosol to mediate their destruction by ERAD (Hebert et al., 2010; Knop et al., 1996; Lilley and Ploegh, 2004). Likewise, iRhoms may also play a role in the degradation of certain substrates through the ERAD pathway (Zettl et al., 2011). However, in a seemingly opposing pathway, mammalian iRhoms promote the forward trafficking of TACE from the ER to the plasma membrane, where TACE releases TNF-α from the membrane to mediate its secretion (Adrain et al., 2012; McIlwain et al., 2012; Siggs et al., 2012). Their fascinating biology aside, however, the focus of this thesis remains on the proteolytic function of rhomboid proteases, and therefore, these proteins, which are incapable of hydrolyzing peptide bonds, will not be discussed further.

Research into a number of pathogenic unicellular parasites has uncovered new and pernicious roles of rhomboid proteases. In the apicomplexa Toxoplasma gondii and Plasmodium spp.—cell-invasive agents that cause toxoplasmosis and malaria, respectively—rhomboid proteases cleave diverse adhesins, which parasites use to adhere to host cells and surfaces during invasion and gliding motility (a unique method for cell locomotion that involves adhesion to a surface followed by translocation of the adhesion complex to the posterior of the cell by actin-myosin motors). An essential late step in both invasion and gliding motility is to dismantle these adhesion junctions. Several lines of evidence suggest that multiple rhomboid proteases play these important roles (Figure 1.2B), including knockouts (Lin et al., 2013; Srinivasan et al., 2009), conditional knockouts (Buguliskis et al., 2010), and substrate-uncleavable mutations (Ejigiri et al., 2012; O’Donnell et al., 2006; Parussini et al., 2012). Indeed, knocking out ROM1 in two plasmodium species that infect mice strongly attenuates their virulence. This result and the fact that at least two groups have failed to knockout ROM4 (Lin et al., 2013; O’Donnell et al., 2006) or create an uncleavable transgenic ROM4 substrate (O’Donnell et al., 2006) in Plasmodium spp, indicating that ROM4 is essential, argue that parasitic rhomboid proteases are targets for novel antimalarials.
In the extracellular parasite, *Entamoeba histolytica*, the sole rhomboid protease, EhROM1, cleaves a conserved family of cell surface lectins (Baxt et al., 2008). Although an exact role of these lectins is lacking, the localization dynamics of EhROM1 is highly suggestive. During infection, *E. histolytica* phagocytoses red blood cells and cellular debris and has a unique mechanism to evade the host immune system: it expels lectins vulnerable to attack by forming a cap at the posterior end of the cell containing the lectins and jettisoning it away (Calderón, 1980). Tantalizingly, EhROM1 localizes to phagocytic vesicles and to the base of the cap. Indeed, EhROM1 is required for efficient phagocytosis and although EhROM1 is not required to form the cap, it may be needed to release it (Baxt et al., 2010). These results imply a role for rhomboid proteases in releasing phagocytic cargo and in an evading the host’s immune system.

In stark contrast to our expanding knowledge of rhomboid protease biology in eukaryotes, there is just a single understood role of a rhomboid protease in bacteria. Magnifying this dearth of knowledge further is the fact that bacteria comprise the overwhelming diversity of cellular life on our planet, and thus encode many more rhomboid enzymes than their eukaryotic counterparts. In the opportunistic pathogen *Providencia stuartii*, the rhomboid proteases, AarA, clips a small N-terminal fragment off TatA (Stevenson et al., 2007), enabling TatA to interact with TatB and TatC to form the twin arginine translocase complex (Figure 1.2C, Fritsch, Krehenbrink, Tarry, Berks, & Palmer, 2012), which exports cofactor-containing and folded proteins from the bacterial cytoplasm. *P. stuartii* cells lacking AarA show a delay in quorum sensing, exhibit a cell chaining phenotype, and fail to grow on MacConkey agar (Rather and Orosz, 1994). However, this role for rhomboid proteases applies only to a small subset of bacteria related to *P. stuartii*: nearly all other bacteria encode a TatA protein already missing the N-terminal extension, thus obviating the need for a rhomboid protease to remove it (Stevenson et al., 2007). In fact, expressing full length *E. coli* TatA rescues AarA-deficient *P. stuartii*. Thus, this one role of a rhomboid protease in *P. stuartii* fails to explain the much broader conservation of rhomboid proteases in bacteria.

Other efforts at understanding bacterial rhomboid protease biology have been reported, resulting in three phenotypes of rhomboid deficient bacteria. *Mycobacterium tuberculosis* and *Mycobacterium smegmatis*, a close relative to and a model of *M. tuberculosis*,
each encode two rhomboid proteases (Kateete et al., 2010). Deletion of the rhomboid protease Ms4904 in *M. smegmatis* results in pleiotropic effects, including reduced biofilm formation and increased susceptibility to novobiocin and ciprofloxacin (Kateete et al., 2012). Interestingly, deleting the other rhomboid protease, Ms5036, rescues the Ms4904 deletion, but itself has no observable phenotype; thus, this hints that these two rhomboid proteases serve in antagonistic pathways. A *Bacillus subtilis* strain lacking its rhomboid protease, YqgP, also shows a cell-chaining phenotype (Mesak et al., 2004) reminiscent of *P. stuartii*, even though *B. subtilis* does not encode for a TatA requiring cleavage by rhomboid proteases. Finally, the only phenotype found for *Escherichia coli* lacking its rhomboid protease GlpG is a slight increase in resistance to cefotaxime, even though this strain was assayed in a high-throughput array format for sensitivity to many small molecules, and for growth in a myriad of conditions with no additional results (Clemmer et al., 2006). Unfortunately, all these examples lack a molecular explanation for their phenotypes, including the substrates that their respective rhomboid proteases cleave and the pathways in which these proteases participate.

The powerful bioinformatic technique partial phylogenetic profiling identified a strong taxonomic co-occurrence in proteobacteria between a subclass of rhomboid proteases, termed rhombosortases, and proteins containing a tripartite C-terminal TMD structure, called a GlyGly-CTERM in proteobacteria (Haft and Varghese, 2011). This co-occurrence implicates the two in a single pathway, and presumably the rhomboid proteases control release of the proteins tethered to the membrane by a GlyGly-CTERM. However, no experimental evidence exists to support this model, which hinders our understanding of the pathway’s biological role.

The widespread conservation of rhomboid proteases in all three evolutionary domains testifies to their utility to cellular life. Understanding their roles in bacteria is particularly key to learning how rhomboid proteases evolved. Albeit controversial, rhomboid proteases may have evolved first in bacteria, and then disseminated to archaea and eukaryotes through multiple horizontal gene transfer events (Koonin et al., 2003). In addition, given the precedence of rhomboid proteases in eukaryotes, uncovering roles of rhomboid protease function in bacteria could open up new avenues for combating deadly bacterial pathogens. With the inexorable rise in antibiotic resistance and the dwindling pipeline for new antibiotic development (Boucher
et al., 2013; Spellberg et al., 2004; Centers for Disease Control and Prevention, 2013), new targets such as rhomboid proteases are critically needed.

**Rhomboid Protease Biochemistry and Mechanism:**

In parallel to the biological research of rhomboid proteases, much effort has been invested into understanding their biochemistry, partly because the questions are intriguing for their own sake, but also because the answers provide context to their biological functions, inform new experiments, and help create assays to find and characterize new drugs targeting rhomboid proteases. Furthermore, of the three enzyme families that perform intramembrane proteolysis, the rhomboid proteases have excelled as the biochemical paradigm to study this important process (Urban, 2010).

Early on, cell-based assays and multiple sequence alignments identified key motifs within rhomboid proteases (Urban et al., 2002) and their substrates (Urban and Freeman, 2003). In rhomboids these include the active site serine and histidine dyad, a WR motif in the large loop 1, and a conserved asparagine that participates in forming the oxyanion hole. A 7-amino-acid sequence within the substrate Spitz TMD encompassing the cleavage site turned out to be necessary and sufficient for proteolysis, with helix destabilizing residues being key.

Further bioinformatic characterization found that rhomboid proteases come in different flavors depending on the number of TMDs and their topology (Koonin et al., 2003; Lemberg and Freeman, 2007). All rhomboid proteases are composed of a 6 TMD core that harbors the catalytic residues and presumably form a conserved architecture. The basic secretase-type rhomboid proteases encode for only the core 6 TMDs with variable N- and C-terminal extramembranous domains. In contrast, the 6+1 secretase-type and the 1+6 PARL-type append an additional TMD on the C- and N-terminus, respectively. The significance of these additional TMDs is unclear, though PARL type rhomboid proteases may remove its extra TMD (Jeyaraju et al., 2011).

A landmark that set the stage for more detailed analyses was the development of an *in vitro* system with purified components (Lemberg et al., 2005; Urban and Wolfe, 2005). Pure rhomboid proteases extracted from membranes into detergent micelles were found to be sufficient for proteolysis with no requirement for any cofactors or prior cleavage of their substrate by another protease, thus setting them apart from the SPP/presenilin and S2P
families whose substrates require a prior cut before intramembrane proteolysis. In addition, one report provided the first hint that rhomboid proteases are sensitive to their environment, be it detergent micelles or pure lipid vesicles.

A watershed for rhomboid proteases came with the publications of three crystal structures of the same homologue, *E. coli* GlpG ([Figure 1.3A](#), Ben-Shem, Fass, & Bibi, 2007; Wang, Zhang, & Ha, 2006; Wu et al., 2006). The atomic-resolution pictures provided an immediate answer to the hydrolysis paradox, finally cementing intramembrane proteolysis as a new paradigm. With six TMDs, GlpG forms a bundle of α-helices encircling the active site serine and histidine dyad. In addition, the periplasmic loop 1 forms an unusual structure partially embedded in the membrane, yet jutting out laterally from the α-helical bundle. Indeed, the active site is recessed within the plane of the membrane (Bondar et al., 2009). However, in the crystal structures it is well hydrated and protected from the hydrophobic membrane; that is, GlpG forms a sink within the membrane. The idea that intramembrane proteases passively import water from the aqueous environment is supported by more-recent crystal structures from a S2P (Feng et al., 2007) and a presenilin homologue (Li et al., 2013), which both create water channels/basins into the active site. In fact, recent molecular dynamics simulations and analyses of GlpG mutants suggest that specific residues in GlpG retain water in the active site ready for hydrolysis (Zhou et al., 2012).

Yet these insights raised another important question: how does the substrate TMD, normally occupying a hydrophobic milieu, access the hydrated active site? Encouraged by two conformations of GlpG in a crystal structure asymmetric unit showing only a rotation of TMD 5 away from the apposing TMD 2, Baker et al undertook a large mutational analysis to search for GlpG variants with increased activity, with the rationale that weakening gating interactions would lead to an increased proteolytic rate (Baker et al., 2007). Indeed, the results honed in on the interface between TMDs 5 and 2 as the lateral substrate gate, formed by series of interdigitating, large and hydrophobic residues ([Figure 1.3B](#)). In fact, mutating just the two residues at the top of the interface to alanine increased the rate 10-fold. This *in vitro* study was later validated *in vivo* in *E. coli* cells (Urban and Baker, 2008).

An unprecedented deconstruction of the GlpG structure was recently reported using a high-throughput method to probe membrane protein thermostability, complemented by
chemical reversible-unfolding-assays (Baker and Urban, 2012). Four regions in particular consolidate the stability of GlpG: two glycines on apposing TMDs 4 and 6 mediate tight packing interactions, and multifaceted hydrogen bonds secure the stability and structure of loop 1 and the cytosolic face of GlpG. The lateral-gate TMDs 5 and 2, however, was the only region free of structural responsibility, supporting the dynamic role inherent to gating function.

Another important question is how rhomboid proteases discriminate substrate from non-substrate TMDs. Spectroscopic studies revealed that, in contrast to soluble proteases, rhomboid proteases achieve substrate specificity through a unique strategy (Moin and Urban, 2012): they sample the secondary structural dynamics of their substrates by exploiting imposed biophysical differences between the membrane and aqueous environments. In the membrane, TMDs are forced to adopt secondary structure, usually α-helices, to satisfy backbone hydrogen bonds in the absence of water (White et al., 2001). However, when exposed to the aqueous inner cavity of rhomboid proteases, only intrinsically unstable α-helices unfold and enter the deeper active site. Consistent with this view, insertion of helix-destabilizing prolines converted non-substrates into substrates, and the cleavage site depth of multiple substrates correlated best with a previously published turn propensity scale (Moin and Urban, 2012; Monné et al., 1999). This scale scores the tendency of a TMD both to break its α-helix and to partition out of the membrane—exactly what this secondary-structural dynamics model would require for cleavage of a TMD by a rhomboid protease.

Layered on top of this more general specificity mechanism, certain rhomboid protease homologues fine-tune specificity by reading the substrate’s sequence (denoted as P1 and P1’ N-terminal and C-terminal to the cut site, respectively, with increasing number moving out). Most rhomboid proteases tested require small P1 and P1’ residues (Akiyama and Maegawa, 2007). Furthermore, Providencia stuartii AarA prefers large hydrophobic amino acids at the P4 and P2’ positions (Strisovsky et al., 2009). Although it was interpreted within the context of soluble proteases as a consensus-binding motif with complementary pockets within the enzyme that mediate high-affinity interactions, no quantitative thermodynamic measurements exists to support this model.

Another fruitful approach has been the search for and the structural characterization of inhibitors. Although rhomboid proteases are recalcitrant to many common serine protease
inhibitors (Urban and Wolfe, 2005), some mechanism-based inhibitors such as isocoumarins (Vinodh Kumar et al., 2010), ß-lactams (Pierrat et al., 2011; Vinodh Kumar et al., 2013), ß-lactones (Wolf et al., 2013), and phosphonates (Xue and Ha, 2012) have been shown to inhibit diverse rhomboid proteases, and more notably have been seen to form covalent bonds with the E. coli GlpG active site serine and histidine residues in multiple crystal structures. These studies have been useful in proving the reactivity of the active site residues and the identity of the bonds that may form the oxyanion hole. A caveat of these complex structures is that the inhibitors are no substitute for natural substrates, and thus cannot be used to infer protease-substrate interactions; these details await a protease-substrate co-crystal structure and the development of quantitative kinetic and thermodynamic assays.

Conspicuously absent from the field are rigorous kinetic and thermodynamic analyses of substrate cleavage by and interaction with rhomboid proteases. In addition the vast majority of biochemical assays for rhomboid proteases use detergent micelles in place of membranes (Baker et al., 2007; Lazareno-Saez et al., 2013; Lemberg et al., 2005; Strisovsky et al., 2009; Vinodh Kumar et al., 2010), where rhomboid proteases have evolved and continue to perform their myriad of biological roles. Without these quantitative measurements in physiologically correct environments, our understanding of rhomboid proteolysis is incomplete. We lack critical information that will guide experiments to uncover biological roles for rhomboid proteases and to search for inhibitors targeting rhomboid proteases as therapeutic agents.

This thesis unfolds in two parts tackling aspects of both the biology and biochemistry of bacterial rhomboid proteases. In Chapter II, I describe my work characterizing the in vivo role of the lateral substrate gate in E. coli, identifying endogenous substrates of E. coli GlpG, and my efforts to uncover its biological role. In Chapter III, I describe my work characterizing the kinetics and thermodynamics of bacterial rhomboid protease cleavage reconstituted in their natural membrane environment, with surprising results showing that diverse rhomboid proteases interact weakly with their substrates and that, overall, rhomboid proteolysis is a rate-driven process. This kinetic analysis has exposed an unforeseen similarity to the DNA repair enzymes DNA glycosylases, which are tasked with identifying and removing rare damaged bases from a vast genome (Friedman and Stivers, 2010; Zharkov et al., 2010). As a possible parallel to DNA glycosylases, my thesis work suggests that an ancient role of
rhomboid proteases may be to monitor the quality of the membrane proteome, aiding in the removal of misfolded proteins from the lipid bilayer.

References:


Ejigiri, I., Ragheb, D.R.T., Pino, P., Coppi, A., Bennett, B.L., Soldati-Favre, D., and Sinnis, P.


**Figure Legends:**

**Figure 1.1: Rhomboid protease diversity**

An unrooted phylogenetic tree from a set of 109 aligned Rhomboid and Derlin (inactive, distant rhomboid homologues) protein sequences. Branches are labeled according to their common characteristics. Refseq accession numbers are given in parentheses following each homologue.
Figure 1.2: Biological roles of rhomboid proteases

(A) *Drosophila* Rhomboid-1 cleaves Spitz to release its EGF domain from the membrane and initiate EGF signaling. (B) Parasitic rhomboid proteases cleave adhesins to resolve adhesion junctions, promoting invasion and gliding motility. (C) *P. stuartii* AarA cleaves a small N-terminal fragment off TatA to allow TatA to form the twin arginine translocase.

Figure 1.3: Mechanism of rhomboid protease cleavage:

(A) Structure of the *E. coli* GlpG (pdb 2nrf). The active serine and histidine (red) are recessed below the plane of the membrane. Gating residues are colored yellow. The dashed lines represent the approximate boundary of the lipid bilayer. (B) Schematic of rhomboid proteolysis. A substrate (purple) TMD spontaneously unwinds and traverses the lateral gate to access the active site of rhomboid proteases.
Figure 1.2

A

EGF

Golgi Lumen

Cytoplasm

Spitz

DmRho1

Cleaved Spitz

B

Receptor

Host Membrane

Adhesion Junction

Parasite Membrane

Adhesin (e.g. EBA-175)

Rhomboid Protease (e.g. PfROM4)

Cleaved Adhesin

C

TatA

AarA

Cleaved TatA

Twin Arginine Translocase
Figure 1.3

A

B
Chapter II

Biological Investigations of the *E. coli* Rhomboid Protease GlpG and its Lateral Gate
**Summary:**

*E. coli* encode their rhomboid protease, GlpG, with large hydrophobic gating residues that restrict activity; in contrast, Sin and I noticed that *H. influenzae* GlpG contains the rarer glycine residue in an analogous position, which, when incorporated into *E. coli* GlpG, increases proteolysis. From this insight, I adopted a reverse genetic and substrate identification approaches to investigate the biological function of *E. coli* GlpG and the reasons it evolved lateral gating residues that limit protease activity. My results reveal that these residues are neither a biophysical necessity nor a substrate determinant; *E. coli* engineered to express a gate-open mutant GlpG from its endogenous locus grow just as well as wild-type *E. coli* in pure culture, and *E. coli* overexpressing a gate-open GlpG mutant do not noticeably acquire new substrates. Furthermore, Sin and I have identified two *E. coli* proteins, Rtn and BasS, as substrates for *E. coli* GlpG. Unexpectedly however, mutant *E. coli* cells lacking rhomboid protease activity are indistinguishable from wild-type cells in every way tested, including growth in culture, tolerance to diverse stresses, a transcriptome profile, and colonization of the mouse colon.
Introduction:

Rhomboid proteases are found widely throughout the bacterial realm (Koonin et al., 2003; Urban and Dickey, 2011); thus, their conservation, extending to the archaeal and eukaryotic domains, hints at some central, ancient, and useful role. But our understanding of their physiological functions in bacteria is minimal (Clemmer et al., 2006; Stevenson et al., 2007), limiting our grasp of their evolution and biology and preventing us from determining if rhomboid proteases are ideal targets for combating pathogenic bacteria.

The bacterium *Escherichia coli* is arguably the best-studied free-living organism in biology. Every sequenced *E. coli* strain encodes a conserved rhomboid protease termed GlpG (for glycerol-6-phosphate G). Echoing its host, GlpG has not disappointed in terms of biochemical research (see chapter I); it has stood as a model for all rhomboid proteases, if not all intramembrane proteases (Urban, 2010). In contrast, the biology of GlpG continues to humble researchers; not only is there no known role for GlpG, but the identity of its physiological substrates remains a mystery as well.

One clue to its biology could be gleaned from the identification of a lateral gating mechanism for GlpG (Baker et al., 2007). Transmembrane (TM) substrates unwind and traverse the crevice between the interface of GlpG TMs 2 and 5, where large hydrophobic residues join together (Moin and Urban, 2012). Interestingly, a single base-pair change, converting the TM 5 tryptophan 236 to glycine (W236G), increases the apparent rate 5-fold (Urban and Baker, 2008). Although rare, nature has not forsaken this fact; the *Haemophilus influenzae* homologue, also termed GlpG, encodes a glycine at the corresponding position.

It is tempting to speculate that this mostly-conserved tryptophan, restraining activity, endows *E. coli* GlpG with allosteric regulation. However, other models must be considered. Perhaps W236 helps form a biophysical septum between the aqueous residues making up the active site and the hydrophobic membrane, or maybe W236 imparts or augments specificity for the unknown physiological substrates of GlpG, as recently concluded by Vinothkumar and Freeman (Vinothkumar and Freeman, 2013).

Ultimately, the lateral-gating mechanism must be placed in the context of the physiological role of GlpG. More specifically, this includes discovering phenotypes of strains lacking GlpG.
activity and revealing the identity of its endogenous substrates. Yet, GlpG, the biochemical paragon of rhomboid proteases, has eluded attempts to uncover any such roles (Clemmer et al., 2006).

In this chapter, I address these three aims: 1) why do E. coli, indeed most bacteria, encode tryptophan in a gating position in which a single base-pair missense mutation would increase the catalytic rate 5-fold? 2) What are the physiological substrates of E. coli and H. influenzae GlpG, and 3) what are the phenotypic consequences of mutating the glpG gene to eradicate rhomboid protease activity?

**Results:**

**E. coli and H. influenzae Genetics: Precisely Engineering Chromosomal Mutants:**

I engineered precise mutations to the E. coli and H. influenzae chromosomes using two rounds of homologous recombination to exchange the glpG gene with desired alleles (Figure 2.1A and Table 1). Because these strains encode glpG with only missense mutations (sometimes with only a single base-pair change) or N-terminal fusions, the chance of polar mutations or other undesired effects are minimized. All strains were verified by PCR and sequencing (Figure 2.1B), and several also by Southern blot (not shown). Furthermore, E. coli cells lacking active GlpG fail to cleave the only known bacterial rhomboid protease substrate, P. stuartii TatA-Flag (Figure 2.1C), thus functionally confirming their lack of rhomboid protease activity.

**The Gating Tryptophan is not a Biophysical Requirement and Does not Enhance Substrate Specificity:**

If the gate separates the aqueous active site from the hydrophobic membrane or if it enhances the proteolytic specificity of GlpG, mutating these residues, which results in hyperactive mutants, should induce a fitness cost to E. coli cells grown in culture. To test this, I competed log-phase cells of Flag-GlpG wild type (WT) with the hyperactive mutant HA-GlpG F153A+W236A (AA, ~10-fold more active than WT) by mixing the strains in equal proportion and diluting the mix into fresh minimal media daily. After purifying genomic DNA from the harvested cells, I used primer sets specific to the sequence encoding the engineered N-terminal epitope tag to differentiate each strain within the population by quantitative PCR (qPCR). Over the course of two weeks and after nearly 100 generations, the mix did not
significantly deviate from a one-to-one ratio (Figure 2.2A); that is, there is no apparent fitness cost of removing the large and greasy gating residues of GlpG to E. coli cells grown in minimal media. This result argues against a biophysical requirement of large gating residues to separate the hydrophilic active site from the hydrophobic membrane, which would otherwise likely be important under any condition. Moreover, from this result and proteomic data below, these large gating residues do not seem to be critical in maintaining specificity.

The Gating Mechanism is Conserved

Only in E. coli GlpG have mutations to gating residues been observed to increase the cleavage rate (Baker et al., 2007; Urban and Baker, 2008); therefore, I tested if similar mutations increase the cleavage rate of two other bacterial rhomboid proteases from Pseudomonas aeruginosa (PaROM) and Vibrio cholerae (VcRho). Because these two rhomboid proteases contained sufficient similarity to GlpG, I developed homology models of their structures (Figure 2.2B and C) based on the published GlpG structure (pdb 2ic8). From these structures, I mutated the residue analogous to GlpG W236 to glycine resulting in PaROM W242G and VcRho W239G. Indeed, akin to GlpG W236G, both PaROM W242G and VcRho W239G cleaved the model substrate GFP-Spitz-Flag faster than their wild-type counterparts in E. coli cells (Figure 2.2D). Therefore, multiple rhomboid proteases encode residues at the lateral substrate gate that restrict catalytic activity.

Although it is unclear why these large gating residues exist, we expect that they regulate proteolytic activity under specific conditions. Further analysis awaits identification of a phenotype and a biological role that GlpG plays.

Bioinformatic and Manual Cleavage Assay Substrate Search:

Most known rhomboid protease substrates are single-pass, type-1 membrane (cytoplasmic C-terminus) proteins with small (Akiyama and Maegawa, 2007) and helix-breaking residues (Moin and Urban, 2012; Urban and Freeman, 2003) near the N-terminal end of the TMD. On this basis, the first approach both Sin Urban and I took towards identifying physiological substrates of GlpG was to predict candidate substrates and test each in an established in vivo cleavage assay (Urban and Baker, 2008). After manually inspecting the transmembrane sequences of experimentally verified (Daley et al., 2005) type-1 membrane proteins in E. coli, we tested all candidates (Table 2.2) for cleavage by a hyperactive GlpG mutant in live E.
coli cells, analyzing cell lysates by western blots for GlpG-dependent cleavage bands. In this way we identified two membrane proteins, Rtn and BasS, as the first-and-only-known E. coli substrates for GlpG substrates (Figure 2.3A).

No experimental topology data for H. influenzae membrane proteins exist. Instead, I analyzed the entire the H. influenzae predicted proteome bioinformatically with TMHMMv2.0 and identified 64 predicted single-pass membrane proteins. Of these, I manually cloned, tagged, and tested 51 candidates, including all type-1 membrane proteins, in the same way as the E. coli candidate substrates by co-expressing each candidate substrate with H. influenzae GlpG in E. coli cells. However, H. influenzae GlpG cleaved none of the 51 candidates (Table 2.3).

**Proteomic Substrate Search:**

Although the bioinformatic substrate search strategy unequivocally expresses and tests each potential substrate directly, it relies on assumptions of GlpG substrates; therefore, I adopted 2D-PAGE/Difference Gel Electrophoresis (DIGE) (Unlü et al., 1997) as a complementary proteomic approach. To my knowledge, this is the first attempt to identify rhomboid protease substrates using proteomics. Initially, I compared whole cell lysates from WT and GlpG null E. coli, however, no differences were detected (data not shown).

The majority of protein present in the gel is likely to be soluble protein that overwhelms any detection of the less-abundant membrane proteins. Since membrane proteins are more likely to be GlpG substrates, I therefore used a sucrose-gradient ultracentrifugation procedure to enrich for membranes from lysed E. coli cells. I verified the membrane fraction by successfully detecting endogenous levels of Flag-GlpG by western blot, which was otherwise undetectable in whole-cell lysates (Figure 2.3B). Notably, this is the first time that endogenous levels of GlpG have been detected in E. coli cells. As before, I analyzed the membrane fractions of WT versus GlpG null E. coli cells by 2D-PAGE/DIGE. Aside from two false-positive spots (where I ultimately found that the Cy3 signal was being quenched), and despite detecting over 1000 unique protein spots, none were significantly and reproducibly different in GlpG null cells compared to WT cells (Figure 2.3C).

The endogenous protein level of GlpG is low. Cleavage of only our most efficient heterologous substrate P. stuartii TatA-Flag can be seen in WT E. coli (Figure 2.1C). Therefore,
to maximize the GlpG specific activity in cells, I overexpressed the gate-open and hyperactive GST-GlpG W236G mutant in *E. coli* and compared the resulting membrane fractions to membranes from cells overexpressing the inactive GST-GlpG S201A mutant by DIGE (Figure 2.3D). Only one spot was unique to GST-GlpG W236G, but mass spectrometry identified this spot as GST from the GST-GlpG fusion. Although unexpected, I subsequently confirmed this difference by western blots probing for GST (data not shown). Thus, GST is being cleaved off GlpG. Importantly, this highlights the ability of this method to detect differences between samples. Taken together, these DIGE results show that GlpG has exquisite substrate specificity and is not used as a general degradative protease of membrane proteins. Moreover, these results argue that the gating residue W236 is not important for substrate specificity because no new substrates were detected despite overexpressing the hyperactive GST-GlpG W236G mutant.

**The Phenotypic Consequences of Mutating the Endogenous glpG Gene in *E. coli***:

As a starting point for finding a phenotypic consequence of removing GlpG activity from *E. coli*, I first tested whether inactivating GlpG in *E. coli* affected growth in cultures. But, GlpG null strains grew at rates indistinguishable from their WT parent strain in minimal media (Figure 2.4A). Consistent with this, removing the *glpG* gene from in *H. influenzae* also failed to affect their growth relative to WT *H. influenzae* (Figure 2.4A). Slight and possibly undetectable differences in growth rates can, over the course of many generations, provide significant negative evolutionary pressure; therefore, I competed GlpG null *E. coli* cells against their isogenic WT strain directly in minimal media. I engineered the competing strains to have differential antibiotic resistance (streptomycin resistance vs nalidixic acid resistance) to differentiate each strain within the population by scoring colonies on appropriate selective plates. In all cases, the population never significantly deviated from an equal ratio over the course of one week and after ~50 generations (Figure 2.4B), consistent with the measured growth rates. Combined, these data show that GlpG activity is dispensable for *E. coli* cells grown *in vitro*.

**Focused Phenotypic Assays:**

Another possibility was that GlpG activity was not required during normal growth, but becomes important under stress conditions. I chose to focus on and characterize the *E. coli*
glpG null strains by subjecting them to a battery of phenotypic assays. For brevity, the results from a selected list are reported here. Please refer to Table 2.4 and 2.5 for an exhaustive list of attempted assays.

Substrate-based phenotypic assays:

Although the biological function of Rtn is unknown, the GlpG substrate BasS and its cognate sensor BasR form a two-component signaling network, which controls responses to a myriad of environmental cues and stresses (Stock et al., 2000). In Salmonella typhimurium, a close relative of E. coli, BasS/BasR (termed pmrB/A in S. typhimurium) responds to cationic antimicrobials such as polymyxin B (Groisman et al., 1997) to alter the composition of the outer membrane, making it less permeable to cationic antimicrobials. Therefore, I tested if GlpG null strains, with a potential defect in BasS/R signaling, are more sensitive to polymyxin B. However, I saw no difference in sensitivity between GlpG null strains and their isogenic parent (Figure 2.5A).

The BasS/R is also activated by increased levels of Fe^{2+} (Hagiwara et al., 2004). Thus, to see if GlpG plays a role in the Fe^{2+} response, I tested if GlpG cleavage of BasS increases in the presence of 400 mM Fe^{2+}. However, GlpG showed no difference in BasS cleavage whether Fe^{2+} was present or not (Figure 2.5B).

Bile salts were recently found to upregulate the expression of BasS/R (Kus et al., 2011). Again, similar to my test for Fe^{2+}, I tested whether growing cells in the presence of bile salts leads to increased cleavage of BasS. However, the presence of bile salts had no significant effect on GlpG cleavage of BasS (Figure 2.5C). In conclusion, GlpG fails to phenocopy known roles of BasS. Given this lack of correlation, BasS is unlikely to be a physiologically consequential substrate of GlpG.

Sensitivity to Cefotaxime:

A large-scale biolog screen found the only phenotype of a GlpG null E. coli strain to be 6% increase in sensitivity to the β-lactam antibiotic cefotaxime (Clemmer et al., 2006). I sought to reproduce this result with my engineered strains, which are more precise GlpG activity ablations, using the same agar disk diffusion assay. Unexpectedly, I found no difference in the zones of clearance between WT and GlpG null E. coli cells (Figure 2.6A). While controlling for other factors, I observed significant variation when testing different agar plates (p = 4.3*10^{-3}).
In fact, not controlling for this factor leads to a spurious significant difference between each strain \((p = 0.048)\). My result, supported by the fact that GlpG null cells showed no difference in sensitivity to other \(\beta\)-lactams, argues that GlpG is not involved in \(\beta\)-lactam antibiotic resistance.

**UV Stress, Acid Tolerance, and Heat Shock:**

We reasoned that GlpG may play a role in a stress response. To test a few common environmental stresses, I determined the sensitivity of GlpG null *E. coli* cells to UV treatment, acidity, and treatment at 60 °C. However, GlpG null cells showed no difference in sensitivity to any of these treatments compared to the WT parent strain (Figure 2.6B).

**Transcriptome Analysis:**

Abrogating GlpG activity in *E. coli* may have triggered a compensatory pathway that rescues any observable phenotypes. To test this idea, I analyzed the transcriptome of GlpG null and WT *E. coli* cells grown in minimal media to log- and stationary-phase. Even though there was a clear difference between log- and stationary-phase cells, WT and GlpG null cells showed indistinguishable transcriptome profiles within each growth phase (Figure 2.6C). Therefore, eliminating GlpG activity in *E. coli* has no effect on transcription, either directly through its own potential signaling network or indirectly by inducing other compensatory pathways.

**GlpG is Dispensable for *E. coli* Colonization of Mice:**

Without any clear indication of GlpG function, I sought an assay that was both a good physiological model and one that subjected bacteria to a myriad of stresses to maximize the chance of observing a phenotype. The streptomycin-treated mouse model is well studied (Chang et al., 2004; Myhal et al., 1982; Sweeney et al., 1996) and has been used to characterize genes necessary for colonization of the mouse colon by analyzing bacteria shed in fecal matter. Mice are initially given streptomycin in their drinking water to clear a niche in which streptomycin-resistant bacteria, introduced into the mouse stomach using an oral gavage needle, can colonize.

To differentiate two strains mixed in competition, I engineered *E. coli* to carry 6 silent mutations in the N-terminal coding region of the chromosomal *glpG* gene (GlpG F7-V12'). Importantly, *E. coli* with these silent mutations make exactly the same GlpG protein as their WT
counterparts. I then designed and validated primer sets that accurately discriminate between this and the WT sequences by quantitative PCR (qPCR).

With the assay established, I competed GlpG WT against catalytically-inactive GlpG S201A+H254A+F7-V12' strains in mice by mixing the two strains in equal proportion and inoculating mice with $10^5$ bacteria by oral gavage. However, over the course of one week, the ratio of the two strains did not significantly deviate from one-to-one, indicating that GlpG is dispensable for *E. coli* colonization of the mouse colon (Figure 2.7A). Although in accordance with the reported literature (Chang et al., 2004; Myhal et al., 1982; Sweeney et al., 1996), $\sim10^5$ bacteria is a seemingly high inoculum, especially considering that the infectious dose of pathogenic *E. coli* O157:H7 is less than 700 (Tuttle et al., 1999). Thus $10^5$ bacteria may overwhelm any differences between WT and GlpG null *E. coli*. Surprisingly, I found that an inoculum of $\sim20$ bacteria resulted in high levels of colonization. When inoculating mice with 40 bacteria of a 1:1 mix, the variation between inoculated mice increased significantly as expected. However, I observed no significant difference in colonization between the WT and GlpG null strains even under these more extreme limiting inoculum conditions (Figure 2.7B).

GlpG may have a non-cell-autonomous role in which the presence of a WT strain would rescue a GlpG null strain in trans and thus explain my results. For that reason, I tested the colonization ability of WT and GlpG null *E. coli* by inoculating each strain separately into mice. I assessed colonization by homogenizing fecal matter and counting colonies on streptomycin-selective agar. However, consistent with the previous result, both strains colonized mice to a level of $\sim10^{10}$ cells/g feces, with no significant differences between each strain (Figure 2.7C).

I next analyzed colonization persistence, which is the ability of bacteria to grow and divide within the colon even when challenged with ectopic bacteria (Chang et al., 2004; Gillor et al., 2009). After colonizing mice with either WT or GlpG null *E. coli*, I added conventional mice—those never treated with antibiotics or exposed to laboratory bacteria—to each cage and removed streptomycin treatment. Endogenous bacteria within the conventional mice can enter colonized mice through the fecal-oral route, thus increasing the difficulty of the colonizing bacteria to persist. Surprisingly, *E. coli* showed robust persistence, maintaining a high level of presence beyond 40 days. However, there was no difference between WT and GlpG null *E.
Finally, I tested the entire fecal-oral route by exposing uncolonized streptomycin-treated or conventional mice (acceptor mice) to previously colonized mice (donor mice). The data from the acceptor mice show high variation levels, which is not surprising because the fecal-oral route introduces an extra step with its own experimental error. Yet, consistent with previous results, *E. coli* GlpG null bacteria showed no significant difference from WT *E. coli* in colonizing either conventional or streptomycin-treated acceptor mice (Figure 2.7E). Therefore, taken together, my data establish that *E. coli* do not require an active GlpG to efficiently colonize mice.

**Discussion:**

In this chapter, I have excluded the hypotheses that the residues lining the lateral substrate gate simply have an important role in separating the aqueous and hydrophobic environments, or that they increase the specificity of GlpG for its natural substrates. The remaining model is that these gating residues enable GlpG to be regulated allosterically, with the expectation that certain conditions, mediated through the lateral gate, promote proteolytic activity. Interestingly, rhomboid proteases harbor variable N- and C-terminal soluble domains. Although the functions of these extra-membranous domains are largely unknown, these provide attractive sites for allosteric binding that could activate rhomboid proteolytic activity. However, a rigorous test of this model requires identification of a phenotype for *E. coli* GlpG, and as such this model remains speculative.

Through thorough manual and high-throughput screens, Sin Urban and I have identified two *E. coli* substrates of GlpG, BasS and Rtn. But are either BasS or Rtn true physiologically consequential substrates? While screening for phenotypes, GlpG null *E. coli* cells failed to recapitulate known BasS roles. Therefore, it seems unlikely that any physiological BasS cleavage by GlpG affects *E. coli* biology. However, because Rtn has no known role itself, its status as a physiological substrate awaits future studies.

Despite subjecting GlpG null *E. coli* cells to a battery of assays including a transcriptome profile and colonization of the mouse colon, the biological role of GlpG is still a mystery. Yet, its conservation and proteolytic activity testifies to its cellular utility; in fact, *H. influenzae* retains an active *glpG* gene even though it encodes a compacted genome and a gene repertoire
of less than half the *E. coli* genome (1789 vs. 4497 for *H. influenzae* Rd and *E. coli* K-12 substr. MG1655). But without leads, how do we design intelligent experiments to tease out the biology of GlpG? In the next chapter, with the aim of better understanding how diverse bacterial rhomboid proteases cleave their substrates, I characterize the steady-state kinetics of their proteolytic mechanism. The results lead to a new model of rhomboid proteolysis, with clear implications to guide future experiments.

**Material and Methods:**

**E. coli and H. influenzae Strains:**

*E. coli* strain K-12 substrain BW25113 and *H. influenzae* strain Rd were used for genetic manipulations and phenotypic assays. *E. coli* was cultured in under aerobic conditions in Lauria-Bertani (LB) broth supplemented with appropriate antibiotics or in M9+ minimal media supplemented with 0.4% glucose (Sambrook and Russell, 2001). *H. influenzae* was cultured in brain heart infusion broth (BHI) supplemented with 10 µg/mL hemin and 2 µg/mL NAD (sBHI) (Poje and Redfield, 2003a) and appropriate antibiotics.

**DNA Constructs:**

Rhomboid proteases were cloned into the pGEX 6P-1 as previously described. UniProt accession numbers for rhomboid proteases are: EcGlpG, P09391; PaROM, H3TJN3; and VcRho, C6YLR7. Thirty *E. coli* candidate substrates in the pET28 vector with GFP fused to the N-terminus were a kind gift from Gunnar von Heijne’s laboratory (Daley et al., 2005). The remaining candidates were cloned from *E. coli* K12 substr. BW25113 genomic DNA into a pET27b vector with a Flag epitope fused to the C-terminus. Those with a cleavage product about the same size as the full-length substrate were cloned with GFP fused to the N-terminus to aid resolution by SDS-PAGE.

**Chromosomal Engineering of E. coli and H. influenzae:**

Allelic exchange of the *E. coli* chromosomal glpG gene was achieved by two successive rounds of homologous recombination with modifications to standard recombineering methods (Datsenko and Wanner, 2000). In the first round of recombineering, glpG was replaced via homologous flanking ends with a cassette that encodes a selection marker, (chloramphenicol acetyltransferase (cat) that confers resistance to chloramphenicol) and a counter-selection marker (either the levansucrase sacB or the ribosomal protein rpsL that confers sensitivity
to sucrose and streptomycin\(^1\), respectively). In the second round, the now chromosomally encoded cassette was replaced with the desired \(glpG\) DNA sequence using the same flanking homology ends by counter-selecting for the desired strain by growth on sucrose or in the presence of streptomycin.

\(H. \textit{influenzae}\) differs from \(E. \textit{coli}\) in two ways: 1) it naturally performs homologous recombination under starvation conditions (Poje and Redfield, 2003b), and 2) it requires \(\sim\)1kb of flanking homology ends. Otherwise the same basic strategy was used to engineer \(H. \textit{influenzae}\). All strains were verified by PCR, and sequencing (Figure 2.1B). For a list of selected engineered strains, see Table 2.1.

\textit{In vitro} Competition Assays:

Each strain was individually grown to log-phase in M9+ minimal media. Adjusting for O.D.\(_{600}\), the cultures were mixed 1:1 and diluted into fresh media. Concurrently, standard mixes of 5:1, 1:1, and 1:5 were made, centrifuged, and the cell pellets were stored at -20 °C until needed. Each day, 1 mL of cells was harvested, and the remaining culture was diluted 1:100 in fresh M9+ minimal media. Genomic DNA was prepared using the DNeasy® Blood and Tissue Kit (Qiagen) following the manufacture’s protocol, and was analyzed by qPCR (see below).

Manual Substrate Screen:

The TMDs of type-1 membrane proteins from \(E. \textit{coli}\) were inspected for the presence of helix-breaking (glycine, prolines, and serines) and small (alanines, glycines, and serines) residues. The \(H. \textit{influenzae}\) predicted proteome was analyzed by TMHMMv2.0, and the resulting data was parsed for type-1 and type-2 (i.e. single-pass) membrane proteins. Following cloning, epitope tagging, and verification by sequencing (see above), each candidate was co-expressed with an active or inactive GlpG (either \(E. \textit{coli}\) or \(H. \textit{influenzae}\) GlpG for \(E. \textit{coli}\) or \(H. \textit{influenzae}\) candidate substrates, respectively) in \(E. \textit{coli}\) cells as previously published (Urban and Baker, 2008). Briefly, cells were grown in LB supplemented with 100 µg/mL ampicillin and 50 µg/mL kanamycin at 37 °C to mid-log phase. Expression was induced at 27 °C with 250 µM IPTG and cells were harvested at 2, 4, or 6 hour time points. Cleavage was assessed by western blot probing with either an αFlag or αGFP antibody.

---

\(^1\) The strain must be streptomycin resistant to use rpsL as a counter-selection marker
Membrane Preparations:

Membrane preparations were prepared by modifying a previously published protocol (Schnaitman, 1970). Briefly, one liter of culture grown to mid-log phase in LB was harvested by centrifugation at 6,000 x g for 15 minutes in a JLA 8.1000 rotor (Beckman). The cell pellet was resuspended in 10 mL ice-cold 1X PBS supplemented with 2X Complete Protease Inhibitors (Roche) and lysed by two passes through a French pressure cell at 16,000 psi. Unbroken cells were removed by centrifugation at 6,000 x g for 10 minutes. The lysate was layered on top of a discontinuous sucrose gradient consisting of 0.8 mL 55% (w/w) sucrose, 1X PBS, 1 mM DTT overlayed with 5 mL 8.8% (w/w) sucrose, 1X PBS, 1 mM DTT. Membranes were collected at the interface of the two sucrose layers after ultracentrifugation in a SW41-Ti rotor at 35,000 rpm for 2 hours and 30 minutes at 4 °C.

2D-PAGE/DIGE:

About 10 µg of sample were labeled with 0.1 nmoles of either Cy5 or Cy3 (GE Healthcare) following the manufacturer’s instructions. The labeled samples were purified with the 2D-Cleanup kit (GE Healthcare) and brought up in U/T Rehydration buffer (7 M urea, 2 M thiourea, 4% w/v CHAPS). IPG strips (11 cm, either pH 3 - 10 or pH 5 - 8, Bio-Rad) were rehydrated with the protein sample overnight and separated using the Ettan™ IPGphor II™ (GE Healthcare) following standard protocols (~17,000 total V-hours). The focused strip was equilibrated SDS equilibration buffer (50 mM Tris pH 8.8, 6 M urea, 30% glycerol, 2% SDS, 0.002% bromophenol blue) twice for 15 minutes each. In the first and second equilibrations, the strips were reduced and alkylated with 65 mM DTT and 135 mM iodoacetamide, respectively. The strips were loaded onto a 2nd dimension 4-20% TGX™ Criterion™ precast gel (Bio-Rad), the final gel was scanned on a Typhoon imager, and analyzed using the DeCyder software.

Agar-Disk Diffusion Assay:

An overnight culture of each strain was diluted 1:10 in to 1X-PBS and spread onto LB agar plates to form a lawn of bacteria after overnight growth. A 0.25” diameter Whatman® 1 filter paper disk was impregnated with 30 µg of cefotaxime and placed on the LB agar plates. After incubating the plates overnight at 37 °C, the diameter of the zone of clearance was measured using Traceable® digital calipers (Fisher Scientific).

Heat Shock:
BW25113 StrR and GlpG S201A+H254A F7-V12’ StrR cells were grown overnight in LB supplemented with 200 µg/mL streptomycin, and the culture was diluted 10-fold in 1X-PBS. Fifty µL diluted cells were placed in thin-wall tubes and treated at 60 °C for 1, 1.75, and 2.25 minutes using a Robocylcer® (Stratagene). The treated cells were immediately serially diluted into fresh 1X PBS and plated on LB agar supplemented with 200 µg/mL streptomycin. The plates were incubated overnight at 37 °C and the colonies were counted to determine the number of viable cells.

**Acid Treatment:**

BW25113 StrR and GlpG S201A StrR cells were grown to O.D.600 ~ 0.3 in LB supplemented with 200 µg/mL streptomycin. Two mLs of each strain were centrifuged and the cell pellet was resuspended in either LB or LB acidified to pH 2.6. Aliquots at time 0, 1, and 3 hours were serially diluted and plated on LB agar supplemented with 200 µg/mL streptomycin, and the plates were incubated overnight at 37 °C to determine the number of viable cells.

**UV Treatment:**

BW25113 StrR and GlpG S201A StrR cells were grown to O.D.600 ~ 0.4 in LB supplemented with 200 µg/mL streptomycin and serially diluted 10^5-fold, from which 50 µL was spread onto LB agar supplemented with 200 µg/mL streptomycin. The plates were either untreated or irradiated with 2,500 or 5,000 µJoules of 254 nm UV light using a UV Stratalinker™ 2400 (Stratagene). The plates were incubated overnight and the number of colonies was counted for each plate to determine viability.

**Transcriptomics:**

BW25113 StrR and GlpG S201A+H254A F7-V12’ StrR were grown to log- and stationary-phase in M9+ minimal media supplemented with 200 µg/mL streptomycin, and RNA was extracted from cells using the RNAeasy® Mini Kit with the RNPreact® Bacteria Reagent (Qiagen). cDNA was prepared using SuperScript® II Reverse Transcriptase (Invitrogen), and was labeled with Cy3 using the One-Color DNA Labeling kit (Nimblegen). The labeled cDNA was analyzed using the 4x72K *E. coli* K12 gene expression arrays (Roche/Nimblegen), which probe 4,254 transcripts, at the in-house microarray core facility.

**Mouse Colonization:**

The streptomycin-treated mouse model was modified from the published literature
(Chang et al., 2004; Myhal et al., 1982; Sweeney et al., 1996). Our protocol was approved by the Johns Hopkins University Animal Care and Use Committee and given the protocol number MO11M55. Four-week-old male C57 BL/6J mice purchased from the Jackson Laboratory were given drinking water supplemented with 5 g/L streptomycin for 48 hours. Stool samples were collected, homogenized with 1X PBS, and plated on LB agar supplemented with 200 µg/mL streptomycin to confirm the absence of streptomycin-resistant bacteria. Conventional mice were kept on untreated in-house drinking water.

Overnight cultures of the indicated E. coli strain were grown in LB supplemented with 200 µg/mL streptomycin. Viable cell counts were determined by serially diluting and plating the overnight cultures on LB agar supplemented with 200 µg/mL streptomycin. The overnight cultures were diluted as needed, and for competition experiments the diluted bacteria were mixed 1:1.

I introduced 100 µL of the bacteria mix were introduced into the mouse stomach by oral gavage, and the mice were returned the mice to their cage with 5 g/L streptomycin drinking water unless indicated. Periodically, I collected stool samples and stored at them -20 °C. I purified genomic DNA from stool using the QIAmp® Stool DNA mini kit (Qiagen) following the manufacturer’s protocol for pathogenic bacteria.

qPCR of Competition Assays:

The E. coli strain GlpG F7-V12’ encodes 6 silent mutations on the GlpG N-terminus; thus this strain can be distinguished from the WT strain using validated primer sets specific for each. Quantitative PCR was performed using the iQ™ SYBR® Green Supermix (Bio-Rad) with the iQ™5 Real-Time PCR Detection System (Bio-Rad). Relative DNA levels of each strain were determined by the comparative C₇ method excluding the internal control because the same sample was used for each primer set (Schmittgen and Livak, 2008).

Viable Cell Count from Stool Samples:

Alternatively, fresh stool samples were homogenized in 1X PBS at 60 mg stool/ mL 1X-PBS. The homogenate was serially diluted and plated on LB agar supplemented with 200 µg/mL streptomycin and colony-forming units were counted the next day.

Acknowledgements:

I thank Cynthia Sears and her laboratory for invaluable help teaching me the mouse
colonization assay, and I thank Shino Shimoji-Krishnan and Jota Bullen from Gerry Hart’s laboratory for teaching me how to perform 2D-PAGE. Sin Urban selected ~35 E. coli candidate substrates, and throughout my work I cloned an additional ~10 candidates from E. coli. Sin Urban tested ~6 of these candidates and discovered Rtn and BasS as GlpG substrates, and I tested all remaining candidates, including all H. influenzae candidates. The microarray facility performed statistical analysis on the microarray data. I performed and analyzed all other experiments.

References:


**Figure Legends:**

**Figure 2.1: Genetic Engineering of E. coli and H. influenzae**

(A) Allelic exchange strategy to engineer the chromosome of E. coli and H. influenzae through two consecutive rounds of homologous recombination. (cat = chloramphenicol acetyltransferase, rpsL = allele of rpsL conferring sensitivity to streptomycin.) (B) Sanger sequencing chromatograms of the glpG locus from GlpG catalytic mutant E. coli. The altered DNA sequence is underlined. (C) Western blot showing cleavage of overexpressed TatA-Flag
by chromosomally-encoded GlpG variants.

**Figure 2.2: In Vivo Analyses of Lateral Gating Residues**

(A) Competition experiment of Flag-GlpG WT and HA-GlpG F153A+W236A (~10-fold hyperactive gating mutant) strains cultured in M9+ minimal media for ~100 generations. Dashed line at 0 represents a 1:1 ratio within the population. Red bar and red dashed lines represent the mean and 95% confidence intervals, respectively. (B) Homology models of PaROM and VcRho were made using SWISS-MODEL with EcGlpG (2ic8) as the template. (C) Overlay of each model with W236 (EcGlpG numbering) shown in sticks. (Blue = EcGlpG, Yellow = PaROM, Red = VcRho). (D) Western blot showing increased cleavage of GFP-Spitz-Flag in *E. coli* cells by the gate-open mutants of HA-EcGlpG, HA-PaROM, and HA-VcRho.

**Figure 2.3: Manual and Proteomic Identification of *E. coli* GlpG Substrates**

(A) Cleavage of overexpressed Rtn-Flag (top) and BasS-Flag (bottom) by overexpressed GST-GlpG in *E. coli* cells. Red star denotes GlpG-dependent cleavage product. (B) Western blot of endogenous epitope-tagged GlpG in membranes from HA-GlpG or Flag-GlpG engineered strains. Red star denotes Flag-GlpG; the other bands were cross-reactive. (C) DIGE analyses of membrane preparations from the indicated engineered strain. Two spots quench the Cy3 signal (white arrows). (D) DIGE analysis of membranes from GST-GlpG overexpressing cells. Two false-positive spots (white arrows) quenched the Cy3 signal and one spot (white star) was specific to active GST-GlpG, but was identified as GST.

**Figure 2.4: Growth Analysis of *E. coli* GlpG Null Strains**

(A) Doubling times of *E. coli* (left) and *H. influenzae* (right) GlpG null strains compared to their WT parent cultured in M9+ minimal media supplemented with 0.4% glucose (*E. coli*) or BHI broth supplemented with Heme and NAD+ (for *H. influenzae*). Means and 95% confidence intervals are shown as red bars and red vertical dashed lines, respectively. (B) Competition of *E. coli* GlpG null strains versus the wild-type parent strain cultured in M9+ minimal media with means (red bars) and 95% confidence intervals (red vertical dashed lines). Horizontal dashed lines denote a 1:1 ratio within the population.

**Figure 2.5: Substrate-based Phenotype Assays for *E. coli* GlpG**

(A) Sensitivity of *E. coli* GlpG null strains and the wild-type parent strain to polymyxin B in a overnight-growth plate assay. Dark gray and clear wells indicate growth and no growth
overnight in M9+ minimal media, respectively. (B) The effect of Fe²⁺ on the cleavage of BasS by GlpG in *E. coli* cells. Black star indicates the BasS cleavage product. AA = GlpG F153A+W236A, U = Uninduced sample. (C) Effect of bile salts on the cleavage of BasS by GlpG in *E. coli* cells. Black star indicates the BasS cleavage product.

**Figure 2.6: Characterization of *E. coli* GlpG Null Cells.**

(A) Cefotaxime sensitivity of *E. coli* GlpG null and its wild-type parent strains by the agar-disk diffusion assay. The nested ANOVA table (right) indicates significant variation from LB agar plates, but no significant differences between strains. (B) Stress (acidity: left, heat: middle, UV: right) tests of *E. coli* GlpG null strains. (C) Volcano plots of RNA expression from *E. coli* GlpG null strains compared to its wild-type parent in log- and stationary-phase grown in M9+ minimal media. No transcript showed more than a 20% difference between strains. Note: see Tables 2.4 and 2.5 for a complete list of assays.

**Figure 2.7: Mouse colonization by *E. coli* GlpG Null Cells**

(A) Colonization of the mouse colon by competing inactive GlpG S201A+H254A+F7-V12’ and WT *E. coli* strains with a high (~10⁵ bacteria) and (B) low (~40 bacteria) inoculum. Horizontal dashed lines represent a 1:1 strain ratio within the bacterial population. (C) Individual colonization of the mouse colon by each strain. (D) Colonization persistence of *E. coli* GlpG null and WT strains. Two days after colonization (arrow), streptomycin treatment was removed and a conventional mouse was added to each cage. (E) Colonization of streptomycin-treated (left) or conventional acceptor mice (right) via the fecal-oral route.
### Table 2.1: Selected list of engineered *E. coli* and *H. influenzae* strains

<table>
<thead>
<tr>
<th><em>E. coli</em> Strain</th>
<th>Strain Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 BW25113</td>
<td>Parent strain</td>
</tr>
<tr>
<td>2 glpG::cat-sacB</td>
<td><em>glpG</em> gene replaced with the <em>cat-sacB</em> cassette</td>
</tr>
<tr>
<td>3 BW25113 StrR</td>
<td>RpsL K43R mutation, confers resistance to streptomycin</td>
</tr>
<tr>
<td>4 glpG::cat-rpsL</td>
<td><em>glpG</em> gene replaced with the <em>cat-rpsL</em> cassette, RpsL K43R</td>
</tr>
<tr>
<td>5 GlpG S201A</td>
<td>Active site serine mutated to alanine</td>
</tr>
<tr>
<td>6 GlpG H254A</td>
<td>Active site histidine mutated to alanine</td>
</tr>
<tr>
<td>7 Flag-GlpG</td>
<td>Flag epitope tag fused the N-terminus</td>
</tr>
<tr>
<td>8 HA-GlpG AA</td>
<td>HA epitope tag fused to the N-terminus, gate-open GlpG</td>
</tr>
<tr>
<td>9 Flag-GlpG StrR</td>
<td>Flag-tag, RpsL K43R mutation</td>
</tr>
<tr>
<td>10 HA-GlpG StrR NalR</td>
<td>HA-tag, RpsL K43R mutation, GyrA S83L mutation, confers resistance to naladixic acid</td>
</tr>
<tr>
<td>11 GlpG F7-V12' StrR</td>
<td>6 silent mutations in GlpG F7-V12, used to specifically detect strain by qPCR</td>
</tr>
<tr>
<td>12 GlpG S201A+H254A F7-V12'</td>
<td>Double active site mutations, silent mutations to F7-V12</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th><em>H. influenzae</em> Strain</th>
<th>Strain Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Rd</td>
<td>Parent strain</td>
</tr>
<tr>
<td>2 RpsL K43R</td>
<td>RpsL K43R mutation, confers resistance to streptomycin</td>
</tr>
<tr>
<td>3 glpG::cat-rpsL</td>
<td><em>glpG</em> gene replaced with the <em>cat-rpsL</em> cassette, RpsL K43R</td>
</tr>
<tr>
<td>4 GlpG S116A StrR</td>
<td>Active site serine mutated to alanine, RpsL K43R</td>
</tr>
<tr>
<td>5 GlpG H169A StrR</td>
<td>Active site histidine mutated to alanine, RpsL K43R</td>
</tr>
</tbody>
</table>
Table 2.2: *E. coli* proteins tested for cleavage by *E. coli* GlpG

<table>
<thead>
<tr>
<th></th>
<th><em>E. coli</em> Protein</th>
<th>Cleavage Observed</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>BasS</td>
<td>+</td>
</tr>
<tr>
<td>2</td>
<td>Rtn</td>
<td>+</td>
</tr>
<tr>
<td>3</td>
<td>AfuB</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>DcuS</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>FepE</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>RseC</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>RstB</td>
<td>-</td>
</tr>
<tr>
<td>8</td>
<td>YgiY</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>YgoP</td>
<td>-</td>
</tr>
<tr>
<td>10</td>
<td>YcbE</td>
<td>-</td>
</tr>
<tr>
<td>11</td>
<td>Wzc</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>YlaB</td>
<td>-</td>
</tr>
<tr>
<td>13</td>
<td>YaiY</td>
<td>-</td>
</tr>
<tr>
<td>14</td>
<td>YciS</td>
<td>-</td>
</tr>
<tr>
<td>15</td>
<td>SohB</td>
<td>-</td>
</tr>
<tr>
<td>16</td>
<td>Cld</td>
<td>-</td>
</tr>
<tr>
<td>17</td>
<td>YjcH</td>
<td>-</td>
</tr>
<tr>
<td>18</td>
<td>YlaC</td>
<td>-</td>
</tr>
<tr>
<td>19</td>
<td>YmfA</td>
<td>-</td>
</tr>
<tr>
<td>20</td>
<td>YobD</td>
<td>-</td>
</tr>
<tr>
<td>21</td>
<td>HemY</td>
<td>-</td>
</tr>
<tr>
<td>22</td>
<td>YccC</td>
<td>-</td>
</tr>
<tr>
<td>23</td>
<td>YhjN</td>
<td>-</td>
</tr>
<tr>
<td>24</td>
<td>YofU</td>
<td>-</td>
</tr>
<tr>
<td>25</td>
<td>YiaW</td>
<td>-</td>
</tr>
<tr>
<td>26</td>
<td>YicC</td>
<td>-</td>
</tr>
<tr>
<td>27</td>
<td>YdaN</td>
<td>-</td>
</tr>
<tr>
<td>28</td>
<td>YedQ</td>
<td>-</td>
</tr>
<tr>
<td>29</td>
<td>YcdR</td>
<td>-</td>
</tr>
<tr>
<td>30</td>
<td>YnjI</td>
<td>-</td>
</tr>
<tr>
<td>31</td>
<td>YqiK</td>
<td>-</td>
</tr>
<tr>
<td>32</td>
<td>YliE</td>
<td>-</td>
</tr>
<tr>
<td>33</td>
<td>CorA</td>
<td>-</td>
</tr>
<tr>
<td>34</td>
<td>Aer</td>
<td>-</td>
</tr>
<tr>
<td>35</td>
<td>WzzE</td>
<td>-</td>
</tr>
<tr>
<td>36</td>
<td>CreC</td>
<td>-</td>
</tr>
<tr>
<td>37</td>
<td>MscL</td>
<td>-</td>
</tr>
<tr>
<td>38</td>
<td>CadC</td>
<td>-</td>
</tr>
<tr>
<td>39</td>
<td>AtoS</td>
<td>-</td>
</tr>
<tr>
<td>40</td>
<td>BaeS</td>
<td>-</td>
</tr>
<tr>
<td>41</td>
<td>HydH</td>
<td>-</td>
</tr>
<tr>
<td>42</td>
<td>QseC</td>
<td>-</td>
</tr>
<tr>
<td>43</td>
<td>HofO</td>
<td>-</td>
</tr>
<tr>
<td>44</td>
<td>EscJ</td>
<td>-</td>
</tr>
<tr>
<td>45</td>
<td>FliF</td>
<td>-</td>
</tr>
<tr>
<td>H. influenzae Protein</td>
<td>Cleavage Observed</td>
<td></td>
</tr>
<tr>
<td>----------------------</td>
<td>-------------------</td>
<td></td>
</tr>
<tr>
<td>HI_0053</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_0094</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_0096</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_0105</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_0106</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_0118</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_0173</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_0187.1</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_0187</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_0199</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_0241</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_0267</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_0330</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_0367</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_0370</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_0437</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_0453</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_0483</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_0484</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_0496</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_0555</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_0602</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_0629</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_0652</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_0673</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_0734</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_0744</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_0775</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_0866</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_0938</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_1092</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_1126</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_1128</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_1131</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_1141</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_1222</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_1266</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_1335</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_1414</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_1495</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_1496</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_1498</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_1514</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_1527</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_1594</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_1605</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_1622</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_1628</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_1631</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_1682</td>
<td>-</td>
<td></td>
</tr>
<tr>
<td>HI_1707</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>
Table 2.4: List of Phenotypic Assays for *E. coli* GlpG Null Cells

<table>
<thead>
<tr>
<th>Phenotype Tested for:</th>
<th>Different than WT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Sensitivity to polymyxin B</td>
<td>-</td>
</tr>
<tr>
<td>2 Effect of Fe(^{2+}) on BasS cleavage</td>
<td>-</td>
</tr>
<tr>
<td>3 Effect of bile salts on BasS cleavage</td>
<td>-</td>
</tr>
<tr>
<td>4 Sensitivity to cefotaxime</td>
<td>-</td>
</tr>
<tr>
<td>5 Tolerance to acidic conditions</td>
<td>-</td>
</tr>
<tr>
<td>6 UV tolerance</td>
<td>-</td>
</tr>
<tr>
<td>7 Heat shock tolerance</td>
<td>-</td>
</tr>
<tr>
<td>8 Mouse colonization</td>
<td>-</td>
</tr>
<tr>
<td>9 Mouse colonization persistence</td>
<td>-</td>
</tr>
<tr>
<td>10 Fecal-oral route colonization</td>
<td>-</td>
</tr>
<tr>
<td>11 <em>E. coli</em> cell size and shape</td>
<td>-</td>
</tr>
<tr>
<td>12 Transcriptome expression levels</td>
<td>-</td>
</tr>
<tr>
<td>13 O-antigen chain length</td>
<td>-</td>
</tr>
<tr>
<td>14 <em>E. coli</em> chaining phenotype</td>
<td>-</td>
</tr>
<tr>
<td>15 Proportion of fimbriated cells</td>
<td>-</td>
</tr>
<tr>
<td>16 Sensitivity to phenethyl alcohol</td>
<td>-</td>
</tr>
<tr>
<td>17 Effect of Indoleacrylic acid on GlpG cleavage</td>
<td>-</td>
</tr>
<tr>
<td>18 Phenethyl alcohol effect on GlpG cleavage</td>
<td>-</td>
</tr>
<tr>
<td>19 Sensitivity to phage</td>
<td>-</td>
</tr>
<tr>
<td>20 Sensitivity to antimicrobial peptides: T2</td>
<td>-</td>
</tr>
<tr>
<td>21 Sensitivity to antimicrobial peptides: T4</td>
<td>-</td>
</tr>
<tr>
<td>22 Sensitivity to antimicrobial peptides: P1vir</td>
<td>-</td>
</tr>
<tr>
<td>23 Sensitivity to antimicrobial peptides: cecropin A</td>
<td>-</td>
</tr>
<tr>
<td>24 Sensitivity to antimicrobial peptides: dermicidin</td>
<td>-</td>
</tr>
<tr>
<td>25 Sensitivity to antimicrobial peptides: LL-37</td>
<td>-</td>
</tr>
</tbody>
</table>
Table 2.5: List of Phenotypic Assays for *H. influenzae* GlpG Null Cells

<table>
<thead>
<tr>
<th>Phenotype Tested for:</th>
<th>Different than WT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1  Growth in BHI media</td>
<td>-</td>
</tr>
<tr>
<td>2  Adhesion to A549 epithelial cells</td>
<td>-</td>
</tr>
</tbody>
</table>
Figure 2.1

A

E. coli or H. influenzae

1st Round

PCR Product

Chromosome

glpE

glpG

glpR

glpE

cat-rpsL

glpR

2nd Round

PCR Product

Chromosome

glpE

cat-rpsL

glpR

B

Wild-Type

Protein: G L S G V V
DNA: GGCGGGCTTTCTGGCGTGGTG

Mutant

Protein: G L A G V V
DNA: GGCGGGCTTGCTGGCGTGGTG

GlpG S201A

GlpG H254A

C

BW25113

glpE::cat-rpsL

Flag-GlpG

HA-GlpG

HA-GlpG S201A/H254A

U 2 6 2 6 2 6 2 6 2 6 2 6

Time post induction of TatA-Flag (hours)

15 kD

TatA-Flag (11.8 kD)

Clavage Product (10.8 kD)
Figure 2.2

A

![Graph showing the log of Flag-GlpG WT to HA-GlpG AA ratio over time for different mixes and days.]

B

- EcGlpG (2ic8 - chain a)
- PaROM
- VcRho

C

D

<table>
<thead>
<tr>
<th></th>
<th>HA-EcGlpG</th>
<th>HA-PaROM</th>
<th>HA-VcRho</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT U 2 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W236G 2 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W242G 2 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>W239G 2 4</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Rhomboid Construct
Hours Post Induction

Full Length GFP-Spltz-Flag

Red = Protease
Green = Substrate

Cleaved Product
Figure 2.3

(A) E. coli GST-GlpG

<table>
<thead>
<tr>
<th></th>
<th>W236G</th>
<th>WT</th>
<th>S201A</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>4</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

GlpG mutant

Hours post induction

Rtn-Flag

Cleaved product

(B) Membrane Samples

HA-GlpG Flag-GlpG

Replicate

30 kD

αFlag Western

(C) GlpG WT - Cy5
GlpG S201A - Cy3
GlpG AA - Cy5
GlpG AA - Cy3

Cy-5

Cy-3

Merge

(D) GST-GlpG W236G - Cy5
GST-GlpG S201A - Cy3
Merge
Figure 2.4

A

E. coli

H. influenzae

B

GlpG WT vs WT

GlpG S201A vs WT

Doub Timp (minutes)

BW25113 GlpG S201A GlpG H254A

Log₂(Flag−GlpG WT / Flag−GlpG WT)

Log₂(HA−GlpG S201A / Flag−GlpG WT)

Doub Dimp (minutes)

GlpG::cat-sacB

E. coli H. influenzae

Rd

E. coli H. influenzae

BW25113 GlpG S201A GlpG H254A
Figure 2.5

A

Strain

replicate 1

BW25113 StrR
GlpG SA+HA StrR
GlpG W236G StrR
BW25113 StrR
GlpG SA+HA StrR
GlpG W236G StrR
polymyxin B (μg/mL): 4 2 1 0.5 0.25 0.125 0.0625 Blank

replicate 2

M9+ + 200 µg/mL Streptomycin

B

<table>
<thead>
<tr>
<th>AA</th>
<th>WT</th>
<th>- (Endogenous GlpG)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>U 0</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>+</td>
</tr>
</tbody>
</table>

Overexpressed GST-GlpG
Hours Post Induction
400 mM FeSO₄
BasS-Flag
Cleaved BasS-Flag

C

<table>
<thead>
<tr>
<th>W236G</th>
<th></th>
<th>WT</th>
<th>S201A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>U 0</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>2 6</td>
<td>2 6</td>
<td>2 6</td>
<td>2 6</td>
</tr>
</tbody>
</table>

Overexpressed GST-GlpG
0.25% Bile Salts
Hours Post Induction
BasS-Flag
Cleaved BasS-Flag
GST-GlpG
Figure 2.6

A

![Zone of Clearance](image)

B

- Acid Tolerance - pH 2.6
- Heat Shock - 60 °C
- UV Tolerance

C

- Log-phase Cultures
- Stationary-phase Cultures

Nested ANOVA Table:

<table>
<thead>
<tr>
<th></th>
<th>degrees of freedom</th>
<th>Sums of Squares</th>
<th>Mean Squares</th>
<th>F-value</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between Strains</td>
<td>1</td>
<td>2.69</td>
<td>2.69</td>
<td>0.85</td>
<td>0.45</td>
</tr>
<tr>
<td>Among Plates</td>
<td>2</td>
<td>6.31</td>
<td>3.15</td>
<td>28.52</td>
<td>0.0043</td>
</tr>
<tr>
<td>Among Disks</td>
<td>4</td>
<td>0.44</td>
<td>0.11</td>
<td>2.37</td>
<td>0.19</td>
</tr>
<tr>
<td>Residuals</td>
<td>8</td>
<td>0.37</td>
<td>0.047</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 2.7

A

Inoculum = $10^5$ bacteria

B

Inoculum = ~40 bacteria

C

Days Post Inoculation

D

Days Post Colonization

Introduce conventional mouse and
Remove streptomycin treatment

E

Strep-Treated Acceptor Mice

Conventional Acceptor Mice

- Log2(Stain 1 / Stain 2)
- Log10(CFU / g feces)

- WT vs WT
- S201A+H254A vs WT
- Replicate 1
- Replicate 2
Chapter III

Membrane-Immerse Rhomboid Proteolysis is a Rate-Driven Reaction with a Low Affinity for Substrates

Summary:

Unlike the well-studied soluble proteases, intramembrane proteases catalyze a hydrolytic reaction in cellular membranes largely devoid of water, driving interest into this enigmatic process. Yet, due to the many technical challenges of working with membrane proteins, a steady-state kinetic dissection of intramembrane proteolysis in membranes has not been achieved. Advancing on a novel pH switch developed in my thesis lab to control activity of rhomboid intramembrane proteases, Sin and I have established a real-time kinetic assay of rhomboid proteolysis reconstituted into their natural membrane environment. By examining nine diverse bacterial rhomboid proteases with four substrate variants, I found that the affinity of rhomboid proteases for their substrates was surprisingly constant and so low that it would be outside any physiological range ($K_M/K_d = 0.14/0.08 \text{ mole\%; } 135/191 \mu \text{M}$). In stark contrast, the turnover constant ($k_{cat}$) spanned 10,000-fold; nevertheless, the proteolytic rates were slow ($k_{cat} < 0.1 \text{ s}^{-1}$, compared with $\sim 100 \text{ s}^{-1}$ for trypsin). Both extreme kinetic parameters held true for cleavage inside living cells, validating the unusual kinetic characteristics of rhomboid proteolysis. These analyses revealed unexpected similarities not to other soluble proteases or membrane proteins, but to a distinct family of DNA repair enzyme: DNA glycosylases. In summary, my data bring into sharp relief that rhomboid proteolysis is fundamentally a rate-driven reaction, and suggest that DNA glycosylases should supplement soluble proteases as precedent for future research into intramembrane proteolysis.
Introduction:

The biological functions of prokaryotic rhomboid proteases remain largely unknown (Urban and Dickey, 2011). In chapter II, I described my use of bacterial genetics in combination with biochemical techniques focusing on the model bacterium *Escherichia coli* to search for substrates and to uncover the roles of the *E. coli* rhomboid protease, *EcGlpG*. After analyzing ~1,000 proteins, Sin and I found two substrates of *EcGlpG*: BasS and Rtn. However, despite applying a diverse array of techniques such as transcriptomics, proteomics, mouse intestinal colonization and targeted tests for known roles of the discovered substrates, *E. coli* strains lacking rhomboid protease activity have in every assay been indistinguishable from the wild-type (WT) parent strain. Yet, the broad dissemination of rhomboid proteases throughout evolution implies they serve important roles in bacterial life.

Confronted with these obstacles, I realized that to design the appropriate experiments to uncover the physiological function of *EcGlpG*, I needed to understand its enzymatic properties better. Concurrently in our lab, Rosanna Baker, while studying the biochemistry of rhomboid proteases, developed a novel pH switch to turn on or off the proteolytic activity of *EcGlpG* in vitro (discussed below), permitting a kinetic analysis of rhomboid proteases in their natural membrane environment. Capitalizing on this advance, I focused my efforts, described in this chapter, to dissect for the first time the enzymatic mechanism of rhomboid proteases through a rigorous kinetic analysis. My goals were not only to understand how rhomboid proteases turnover their substrates but also to gain insight and obtain leads into their biological roles.

Soluble proteases have served as paradigms in enzymology for decades (Fersht, 1999), which has resulted in a sophisticated understanding of how these enzymes target and hydrolyze their diverse sets of substrates. However, intramembrane proteases naturally reside in a fundamentally different setting: the water-deficient, hydrocarbon-rich, two-dimensional membrane bilayer. Although soluble proteases serve as valuable precedents for intramembrane proteases, an irresistible question is how they differ given their distinct biophysical habitats. One example is the recent advance uncovering a surprisingly unique mechanism for substrate specificity (see chapter I for more detail, Moin & Urban, 2012).

Most biochemical analyses of intramembrane proteolysis replace the two-dimensional membrane bilayer with the pseudo-three-dimensional detergent micelle environment to
avoid the technical difficulties of working with membranes (Seddon et al., 2004). Using this setting, two studies have made important steps by reporting steady-state kinetic analyses of a rhomboid protease (Lazareno-Saez et al., 2013) and the mouse homolog of γ-secretase/presenilin (Chávez-Gutiérrez et al., 2012). But these assays are fraught with limitations of working with membrane proteins, including low solubility—which may prevent true enzyme saturation—and use of artificial detergent micelles. To date, no steady-state analysis in membranes for any intramembrane protease has been achieved. Absence of this knowledge precludes a sophisticated model of intramembrane proteolysis that captures the full dynamic interaction of the enzyme and its substrate.

I have developed a real-time kinetic assay of rhomboid proteolysis reconstituted into natural membranes (Rigaud et al., 1995; Seddon et al., 2004). Complementing this, I have developed quantitative equilibrium techniques to measure the affinity of inactive rhomboid mutants and their substrates. After applying these assays to diverse bacterial rhomboid proteases and substrate variants, the results yield a model in which rhomboid proteolysis is primarily governed by the catalytic turnover rate with little, if any, contribution from the affinity of the enzymes for their substrates. Unexpectedly, this kinetic profile has uncovered a remarkable similarity not to other soluble proteases but to the evolutionarily distinct DNA glycosylases.

**Results:**

**Inducible pH Switch:**

The major challenge to establish a kinetic assay of rhomboid proteolysis in membranes is that premature turnover results from the required presence of both the protease and substrate throughout the duration of the lengthy membrane-reconstitution step (Rigaud et al., 1995; Seddon et al., 2004). Sin and Rosanna Baker solved this problem by reconstituting EcGlpG and the model substrate CSpitz-Flag (Urban and Wolfe, 2005) at low pH in which rhomboid proteolysis is off due to protonation of the catalytic histidine (Figure 3.1A). After reconstitution with known amounts of substrate, the reaction was initiated by neutralization, at which time EcGlpG regained full activity (Figure 3.1A).

Importantly, low pH does not adversely affect EcGlpG. Rosanna found that low pH treatment had no effect on the thermostability of EcGlpG in a sensitive assay that our lab
developed recently (Baker & Urban, 2012; my data shown, Figure 3.1B). Further validating the pH switch, Sangwoo Cho in our lab solved the structure of the EcGlpG core domain crystals grown in low pH (Figure 3.1C). In comparison to the structure of EcGlpG in neutral pH, the nucleophilic serine side chain has turned away from the catalytic histidine, consistent with the latter being protonated. However, the Cα RMSD was 0.32 Å between the two structures, revealing no other structural perturbations.

**Real-Time Reaction:**

Sin designed a synthetic substrate, denoted FITC-TatA, that included a fluorescein-isothiocyanate (FITC) fluorophore conjugated to the N-terminus of the first 34 amino acids (including the entire transmembrane domain) from Providencia stuartii TatA (PsTatA), the only published bacterial rhomboid substrate (Stevenson et al., 2007). Crucially, Sin developed a detergent-based method that allowed this substrate to be resuspended at the high concentrations (~800 µM) needed to saturate rhomboid proteases, relative to the low (10 µM) concentration that had been achieved in the literature previously (Lazareno-Saez et al., 2013).

After reconstituting FITC-TatA into membranes, I unexpectedly discovered that pure E. coli lipids strongly quenched fluorescence from FITC-TatA (Figure 3.2A). Coupled with the pH switch, this quenching provided an ideal opportunity to monitor rhomboid proteolysis in real-time reconstituted into pure E. coli membranes (Figure 3.2B). Indeed, EcGlpG relieved the quenching by liberating the fluorophore from the membrane and yielded a fluorogenic signal that increased linearly over time (Figure 3.2C).

Importantly, mutating the EcGlpG active site residues or the substrate P1 residue, a known requirement for rhomboid proteolysis of TatA (Strisovsky et al., 2009), abolished the fluorogenic signal to a background level (Figure 3.2C). Furthermore, pre-incubation of EcGlpG with JLK6, an isocoumarin mechanism-based inhibitor of EcGlpG (Vinothkumar et al., 2010), also eliminated the signal to background (Figure 3.2C). In all cases, the appearance of product on a SDS-PAGE gel perfectly reflected the increased fluorescence (Figure 3.2D). In addition, the FITC fluorophore did not reduce the cleavage efficiency by EcGlpG (Figure 3.2D). And finally, mass spectral analysis of the cleavage product revealed that FITC-TatA is cleaved only after alanine-8, corresponding to the previously reported cleavage site (Stevenson et
al., 2007) (Figure 3.2E). Therefore, our real-time assay in membranes, combined with the inducible pH switch, recapitulates all known features of rhomboid proteolysis, permitting for the first time a rigorous kinetic analysis of its catalytic mechanism in its natural environment.

**E. coli GlpG Kinetic Parameters:**

Applying the inducible real-time assay, I obtained initial rates of EcGlpG cleaving known concentrations of FITC-TatA reconstituted into *E. coli* membranes (Figure 3.3A). A Michaelis-Menten model modified with a Hill coefficient fit the data exceptionally well ($R^2=0.99$) showing clear evidence of saturation. Note that cooperativity was not present when reactions were analyzed by SDS-PAGE, indicating that the Hill coefficient is a feature of the fluorogenic assay conducted in 384-well plates but is not biologically relevant (Figure 3.3B).

After measuring and correcting for the topology of FITC-TatA after insertion into membranes (Figure 3.3A inset), the fit revealed a $K_M$ of 0.14 ± 0.01 mole percent (mole%, relative to the molar lipid content). Considering that the total protein in the *E. coli* membrane is ~1.5 mole% (Kadner, 1996; Schnaitman, 1970a, 1970b), a $K_M$ of 0.14 mole% is extraordinarily high because any substrate of EcGlpG would have to constitute ~10% of the total membrane protein to achieve a concentration near $K_M$, and about all to approach $V_{max}$. Even though $K_M$ can differ from $K_d$ (Fersht, 1999), this suggests that the affinity between EcGlpG and FITC-TatA is low. To test this directly, I developed a Förster-resonance-energy-transfer (FRET; Lakowicz, 2006) based method to measure the $K_d$ in membranes between an EcGlpG inactive catalytic mutant and FITC-TatA. Based on the structure of EcGlpG and our lab’s prior thermostability analyses (Baker and Urban, 2012), I labeled the tryptophan-196-cysteine mutant of inactive EcGlpG with tetramethylrhodamine (EcGlpG-TMR), the FRET acceptor to FITC, and reconstituted EcGlpG-TMR titrated with increasing concentrations of FITC-TatA into membranes. Exciting for FITC fluorescence resulted in a FRET signal dependent on the substrate concentration (Figure 3.3C). A binding isotherm fitted to the data yielded a $K_d$ of 0.08 ± 0.013 mole%, which was strikingly similar to the measured $K_M$. Also consistent with a low affinity interaction, inactive EcGlpG does not impose a dominant-negative effect (Figure 3.3D): it failed to sequester substrate from active EcGlpG even when present at ten-times higher concentrations.

Deriving the catalytic turnover rate, $k_{cat}$, from $V_{max}$ requires knowing the fraction of
enzyme that is active in the purified preparation. The mechanism-based inhibitor JLK6 covalently reacts only with the proteolytically competent fraction of rhomboid. Mass spectral analysis showed that the entire preparation of EcGlpG is labeled by JLK6, demonstrating near 100% activity within the preparation (Figure 3.3E). This results in a $k_{\text{cat}}$ of 0.0063 ± 0.00021 turnovers per second, corresponding to >2.5 minutes per cleavage. This rate is unexpectedly low, considering that other proteases have a $k_{\text{cat}}$ orders of magnitude higher (e.g. trypsin: $k_{\text{cat}}$ ~100 s⁻¹).

**In vivo Estimates of E. coli GlpG Kinetic Parameters:**

Biochemical systems rebuild one aspect of a complex cellular environment with minimal components to study the isolated reaction in a simplified and controlled manner. Yet, due to the minimal approach, many systems imperfectly reflect the desired reaction occurring in cells. Even though I was careful to use the most physiological conditions possible, the surprisingly high $K_M/K_d$ and low $k_{\text{cat}}$ parameters compelled me to measure “apparent” kinetic parameters in living *E. coli* cells.

*Ps* TatA driven from an arabinose promoter accumulates to incredible levels by overexpression, becoming the most abundant membrane protein (Figure 3.3F). This offered the possibility to saturate endogenous EcGlpG by titrating the expression of *Ps* TatA-Flag in living *E. coli* cells, and thus determine apparent *in vivo* kinetic constants.

Indeed, the hyperbolic curve of *Ps* TatA-Flag cleavage is evidence of EcGlpG saturation (Figure 3.3F). Quantifying for how much of the membrane *Ps* TatA-Flag accounted and modeling the data with a Michaelis-Menten equation yielded an apparent $K_M$ of 0.19 ± 0.13 mole %, which was perfectly consistent with the $K_M$ found *in vitro*.

I next engineered *E. coli* strains to fuse a Flag tag to the N-terminus of the chromosomally encoded EcGlpG (see chapter II and Table 2.1), which permitted me to determine the levels of endogenous EcGlpG relative to FITC-TatA, and thus calculate an apparent $k_{\text{cat}}$ as well. Amazingly, I found the $k_{\text{cat}}$ of EcGlpG cleavage *in vivo* to be 0.0069 ± 0.0009 s⁻¹, which is indistinguishable from the $k_{\text{cat}}$ measured *in vitro* (Figure 3.3F). Granted that the *in vivo* $K_M$ and $k_{\text{cat}}$ are not true kinetic parameters, their close agreement to the actual *in vitro* parameters validates my steady-state kinetic assay, and verifies that rhomboid proteolysis by EcGlpG is a slow process with weak affinity for its substrate.
Membranes limit the rate of catalysis:

The membrane is the unique environment for proteolysis, and my kinetic system allows me to ask how membrane affects the kinetics of EcGlpG. Therefore, I removed the membrane and extended the kinetic analysis to detergent-solubilized rhomboid proteases in a more-familiar 3-dimensional environment. In this case, the $K_m$ remains high at $135 \pm 16.6 \mu M$ (Figure 3.4A). Consistent with this high $K_m$, FITC-TatA failed to co-precipitate with a His-tagged inactive EcGlpG from solution using Ni$^{2+}$-NTA resin (Figure 3.4B). Independently, gel filtration of both FITC-TatA and inactive EcGlpG yielded a small elution peak of FITC absorbance corresponding to the complex (Figure 3.4C). Ignoring dissociation during the gel filtration run, the apparent $K_d$ would be extraordinarily high at $817 \pm 76 \mu M$. However, because both co-precipitation and gel filtration promote dissociation by removing the complex from individual components, I used equilibrium gel filtration (Hummel and Dreyer, 1962) to determine a proper $K_d$ between inactive EcGlpG and FITC-TatA (Figure 3.4D). I injected FITC-TatA mixed with inactive EcGlpG onto two tandem gel filtration columns pre-equilibrated with FITC-TatA. Because the complex drags FITC-TatA out of the free pool to elute with the complex, in the FITC-TatA elution profile a peak followed by a trough at the complex and FITC-TatA elution times, respectively, indicates binding. The peak area is a direct measurement of the complex concentration. Indeed, a reproducible peak followed by a trough was observed, and corresponded to a $K_d$ of $191 \pm 29.4 \mu M$, which is in excellent agreement with the measured $K_m$ in detergent micelles. Therefore, in both membrane and detergent micelle environments, EcGlpG interacts weakly with FITC-TatA.

Interestingly, the $k_{cat}$ in detergent micelles increased 6.5-fold to $0.0404 \pm 0.00336 s^{-1}$; therefore, the membrane environment slows catalysis (Figure 3.4A).

The lateral gate is the rate-limiting step:

The lateral gate appears to be rate limiting for rhomboid proteases in detergent-based (Baker et al., 2007) and in vivo cleavage assays (Urban and Baker, 2008). But a limitation on these important analyses is the inability to differentiate the affinity ($K_m$) from the catalytic rate ($k_{cat}$) effects; that is, the mutation may have increased the affinity for substrate. My real-time kinetic assay can distinguish between these alternatives within a natural membrane environment. Furthermore, to test whether peptide bond hydrolysis or gating is rate limiting, I
compared solvent isotope to gate-open mutant effects. The former arises if deuterium oxide, substituted for water, affects the overall rate. Because proteolysis involves many hydrogen atoms, hydrolysis rate-limiting proteolytic reactions tend to have large solvent isotope effects of ≥2 (Elrod et al., 1980). However, EcGlpG cleavage of FITC-TatA showed a mild solvent isotope effect of 1.26 ± 0.07 (Figure 3.5A). Hydrogen-deuterium exchange at other non-catalytic sites in proteins can cause such mild effects and consistent with such an equilibrium effect, the solvent isotope effect was linearly dependent on the fraction of deuterium oxide present (Figure 3.5B). Unlike the mild solvent isotope effect, EcGlpG F153A+W236A, a gate-open mutant, turned over substrate 3.15 ± 0.41-fold faster than EcGlpG WT (Figures 3.5C & D). Importantly, the gating mutation only affected $k_{\text{cat}}$, with no effect on $K_{M}$. Taken together, these results show that gating, not hydrolysis of the peptide bond, limits the rate of intramembrane proteolysis in membranes by EcGlpG but has no effect on $K_{M}$.

**Nine Diverse Bacterial Rhomboid Proteases:**

Rhomboid orthologs differ widely in their specific activity (Urban and Wolfe, 2005). To understand the importance and evolutionary plasticity of each kinetic parameter, I tested 9 diverse bacterial rhomboid enzymes, including *P. stuartii* AarA (Figure 3.6A), the natural protease of *PsTatA* (Stevenson et al., 2007). Overall, these enzymes share <3% sequence identity and were cloned from gram-positive, gram-negative, and thermophilic bacteria. Remarkably the $K_{M}$ of all enzymes was statistically indistinguishable from EcGlpG, with no more than a 4-fold difference (Figures 3.6B-D). In stark contrast to this similarity, the $k_{\text{cat}}$ ranged ~10,000-fold. Validating this surprisingly expansive span, the slowest enzyme reacted completely with JLK6, excluding low fractional activity as a cause of the low $k_{\text{cat}}$ (Figure 3.6E). Interestingly, increasing the reaction temperature for AqROM to 85 °C, the physiological temperature for the host *Aquifex aeolicus*, increased $k_{\text{cat}}$ by >150-fold, with no effect on $K_{M}$ (Figures 3.6B-D). Similarly, the $k_{\text{cat}}$ of PaROM was increased 5-fold by replacing the pure *E. coli* lipids with DMPC lipids (Figures 3.6C-D). Therefore, in bacteria, in which rhomboid proteases are thought to have arisen (Koonin et al., 2003), evolution has tailored $k_{\text{cat}}$, not $K_{M}$, to presumably meet the specific needs of cells, arguing that rhomboid proteolysis is a rate-driven reaction.

**Four Substrate Variants:**
Next, to understand how known substrate features affect rhomboid proteolysis, I analyzed the kinetics of PsAarA cleaving three substrate variants. Sin and I chose to mutate two critical regions of PsTatA, the P4-P2’ motif (FITC-TatA I5A+F10A) and part of the helix-destabilizing motif (FITC-TatA G11V+S12V). In addition we designed a FITC-Spitz substrate that is based on an independent eukaryotic rhomboid protease substrate from Drosophila. In all cases, $k_{\text{cat}}$ was significantly reduced compared to FITC-TatA (Figures 3.7A-C). In fact, FITC-TatA I5A+F10A severely compromised proteolysis, reducing $k_{\text{cat}} >100$-fold. However, contrary to the P4-P2’ motif being important for binding, the $K_m$ of FITC-TatA I5A+F10A was reduced, indicating a 4-fold increase in affinity, which itself is likely due to a slower turnover of the complex. Therefore the P4-P2’ motif allows for optimal catalysis ($k_{\text{cat}}$) by PsAarA, with the changes to the apparent affinity ($K_m$) being a secondary effect.

Discussion:
Rhomboid Proteases are Rate-Governed Enzymes:

My kinetic analysis of rhomboid proteolysis has led to the unexpected conclusion that rhomboid intramembrane proteolysis is a rate-driven reaction regulated by the lateral substrate gate. Three major results support this: 1) The $k_{\text{cat}}$ spans four orders of magnitude with nine diverse bacterial enzymes and four substrate variants, 2) the lateral substrate gate, not the hydrolysis itself, limits the rate of rhomboid proteolysis, and 3) the $K_m$ does not differ among each rhomboid protease and three substrate variants. Even in the sole case when the $K_m$ did change, the $K_m$ was unexpectedly reduced ~4-fold, likely due to the much larger reduction in $k_{\text{cat}}$ of ~100-fold.

Rhomboid Proteases Display Non-Physiological Affinity for their Substrates

A recent mutational analysis (Strisovsky et al., 2009) and a spate of EcGlpG crystal structures with bound mechanism-based inhibitors (Vinothkumar and Freeman, 2013) has led, through the lens of soluble protease precedent, to a model of substrate specificity for rhomboid proteases. This model posits that rhomboid proteases form high affinity interactions directly with a conserved primary sequence motif in their substrates, in which small P1 and large hydrophobic P4 and P2’ residues occupy analogous enzyme subsites. However, my data, the first quantitative measurements of rhomboid protease-substrate interactions, defy this model. First, the $K_m$ for diverse rhomboid proteases and substrate variants is incredibly
high—to the point that any substrate would have to constitute all of the membrane proteome to approach the $V_{\text{max}}$ of a reaction. Second, mutating the P4-P2' motif to disallowed alanines resulted in an increased affinity, which is likely secondary to the 50-fold greater effect on $k_{\text{cat}}$. My results suggest an alternative interpretation of the sequence-motif preference: preferred residues optimize $k_{\text{cat}}$, that is, their sterics ensure proper alignment of the scissile bond along the active site cleft of rhomboid proteases.

**The Kinetic Profile Uncovers Similarities to DNA Glycosylases:**

Rate-driven catalysis has been observed before in a class of DNA repair enzymes, the DNA glycosylases (Figure 3.7D). DNA glycosylases rely on bases flipping out from the DNA helix and have the considerable challenge of removing only the damaged bases from DNA (Friedman and Stivers, 2010). Although they show high affinity for DNA through interactions with the phosphate backbone (akin to rhomboid proteases showing high affinity for the membrane), any additionally affinity for damaged bases is modest and insufficient to achieve their exquisite specificity (Zharkov et al., 2010). Instead, DNA glycosylases utilize a kinetic mechanism to discriminate between damaged and undamaged bases. These bases first interact with DNA glycosylases at a site distinct from the active site to form an interrogation complex (IC). Damaged bases then advance to the active site to form the excision complex (EC) in which they are removed from the DNA helix. Undamaged bases avoid excision in two ways: 1) they spend more time in a normal base-pair (Parker and Stivers, 2011), and therefore less time in the IC, which effectively reduces their chance of advancing to the EC and 2) their steric and electrostatic clashes hinder access to the active site (reducing $k_{\text{cat}}$). With my data in the context of a recent dynamic-specificity mechanism (Moin and Urban, 2012), this attractive model applies to rhomboid proteases. Rhomboid protease substrates contain analogous features to damaged DNA bases: 1) small P1 and P1' residues allow access to the active site just as damaged bases fit in the excision complex of DNA glycosylases and 2) helix-breaking residues facilitate interaction with rhomboid proteases through spontaneous unwinding (Moin and Urban, 2012), and thus spend more time in a possible interrogation complex of rhomboid proteases. A presumed interrogation complex, though hypothetical, for rhomboid proteases would likely involve the lateral gate as this controls the rate of the reaction and this is where a substrate would first productively contact a rhomboid protease. However, evidence for such an
interaction awaits a structure of the substrate-protease complex and sophisticated methods to probe their dynamics throughout the rhomboid proteolytic mechanism.

**Ancient Roles: a New Hope?:**

What leads does this connection give to the biology of bacterial rhomboid proteases? From this newfound kinship to DNA glycosylases, it seems that rhomboid proteases are kinetically armed to play a quality-control role, in which improperly folded membrane proteins, due to stress or other spontaneous events, more easily unwind after encountering a rhomboid protease, and thus accelerate their rate of cleavage. This process then acts to clear the cell of potentially toxic denatured membrane proteins, a problem made worse by the natural hydrophobicity and tendency for aggregation of membrane proteins. In fact the known roles of human RHBDL4 (Fleig et al., 2012), Derlins (Knop et al., 1996), and the proteolytically inactive iRhoms (Zettl et al., 2011) in ER-associated degradation support this notion that rhomboid proteases participate in membrane protein quality control. Perhaps a unifying feature of rhomboid proteases—possibly all rhomboid proteins including iRhoms and Derlins—is not to simply bind transmembrane segments as suggested recently (Adrain and Freeman, 2012), but to recognize ones that are damaged or unwind easily.

**A Wide-Utility Assay for Intramembrane Proteases:**

To date, there are only two groups that have analyzed the steady-state kinetics of intramembrane protease (Chávez-Gutiérrez et al., 2012; Lazareno-Saez et al., 2013) and both studies are severely limited by technical issues common to membrane proteins, including low solubility, low abundance, and use of detergent micelles replacing natural membranes. My work, enabled by the pH switch developed by Sin and Rosanna, is the first kinetic assay for intramembrane proteases in membranes. Importantly, this technique can be extended to other intramembrane proteases: in addition to low pH, the metallo site-2-proteases and the signal peptide peptidases could also be inhibited by chelating zinc or adding reversible inhibitors, respectively, during the membrane reconstitution step, and membranes should robustly quench a FITC fluorophore placed in an analogous position to FITC-TatA for other substrates.

**Materials and Methods:**

**Rhomboid Protease Purification:**
Each rhomboid protease was fused with an N-terminal hemagglutinin (HA) tag and expressed and purified as described previously (Urban and Wolfe, 2005). Briefly, glutathione-S-transferase (GST) fusion proteins were expressed in C43 (DE3) *E. coli* cells, purified with glutathione-sepharose resin (GE Healthcare), and eluted on-column with PreScission™ protease. HA-EcGlpG was quantified relative to a BSA standard by Coomassie staining; all other HA-rhomboid proteases were quantitated relative to HA-EcGlpG by western blot using infrared fluorescence after probing with αHA (Roche, 11-867-423-001).

**Substrate Preparation:**

C-terminal Flag-tagged recombinant substrate APP+Spi7-Flag was expressed and purified from *E. coli* as described (Baker et al., 2007). Substrates were synthesized by standard solid-state Fmoc chemistry with a FITC fluorophore conjugated to the N-terminus through a β-alanine linkage and amidated at the C-terminus. FITC-TatA encodes residues 1 – 33 of *P. stuartii* TatA and FITC-Spitz encodes residues 135 – 168 of *Drosophila* Spitz. The synthetic substrates were resuspended to a final concentration ~800 µM in 50 mM Tris pH 7.4, 150 mM NaCl, 1 mM DTT, and 0.2% Sarkosyl. CysTatA-His including residues 1 – 33 of *P. stuartii* TatA with a cysteine inserted after methionine-1 and a His6 tag fused to the C-terminus was cloned into a pET27b vector such that the vector-encoded pelB leader peptide was removed. CysTatA-His was expressed in ∆GlpG (DE3) cells grown in M9+ minimal media with 250 µM IPTG shaking at 37 °C for 3.5 hours. The cells were lysed by passage twice through a French pressure cell (16,000 psi), and membranes were collected by ultracentrifugation at 60,000 rpm for 30 minutes in a 70-Ti rotor (Beckman Coulter). Membranes were solubilized with 0.4% sarkosyl, and CysTatA-His was purified with a Ni²⁺-affinity resin. Purified protein was concentrated to ~ 800 µM and dialyzed against 50 mM Tris pH 7.4, 150 mM NaCl, 0.2% sarkosyl.

**E. coli** Liposome Preparation:

*E. coli* polar lipid extract (Avanti Polar Lipids) was dried using a rotary evaporator in a 50 mL round bottom flask to form a lipid film. After drying completely overnight under vacuum, the lipid film was resuspended to 10 mg/mL in 10 mM HEPES pH 7, 10 mM NaCl, 1 mM DTT, briefly sonicated in a temperature-controlled cup-horn sonicator (Branson), and extruded through 200 nm pore filters to form liposomes of defined size.
Inducible Reconstitutions and Real-Time Cleavage Assays:

In 50 mM Na-Acetate pH 4.0 and 150 mM NaCl, 3 µg of *E. coli* polar lipid extract (Avanti Polar Lipids) was mixed with 12.5 - 1600 pmoles of FITC-TatA and 0.05 – 500 pmoles of purified enzyme to a total volume of 25 µL and incubated at room temperature for 10 minutes. The mix was diluted 20-fold with 12.5 mM Na-Acetate pH 4.0 and 37.5 mM NaCl, reducing the detergent below its the critical micelle concentration, and incubated at room temperature for 15 minutes to promote protein insertion into the lipid vesicles. The vesicles were collected by ultracentrifugation at 600,000 x g for 30 minutes at 23 °C in a TLA 120.1 rotor (Beckman-Coulter), and the supernatant was aspirated off. The pellets were resuspended in 50 mM Tris pH 7.4, 150 mM NaCl, and 1 mM DTT, immediately placed into a pre-warmed 384-well clear-bottom microplate. Fluorescence was monitored in real-time using a Synergy H4 Hybrid plate reader (BioTek) using monochromators with excitation and emission wavelengths of 485 ± 20 and 528 ± 20 nm, respectively. Alternatively, time points were quenched with 2X Tricine sample buffer (Life Technologies), separated on a 16% SDS-polyacrylamide gel with tricine running buffer (Life Technologies) and scanned using a Typhoon scanner with the blue laser and FITC emission filters (GE Healthcare).

**CysTatA-His Cleavage Assay**

1600 pmoles of CysTatA-His and 5 pmoles of HA-EcGlpG were reconstituted into *E. coli* membranes at low pH, neutralized after ultracentrifugation and time points were quenched with 2X SDS sample buffer. The samples were resolved by 16%/6 M urea-SDS-PAGE with tricine running buffer and probed for αHis (Qiagen).

**Mass Spectrometry:**

FITC-TatA cleaved in proteoliposomes was filtered through a 30,000 MW cutoff centrifugal filter (Millipore) following the manufacturer’s protocol. The sample was purified using a C18 ZipTip (Millipore), spotted with a sinapinic acid (Laser Biolabs), and analyzed by MALDI-TOF using a Voyager DE-STR mass spectrometer calibrated with bradykinin fragment 1-7, angiotensin I, ACTH fragment 18-39, and insulin oxidized B chain (Sigma-Aldrich).

HA-EcGlpG and HA-PaROM were pre-treated with 20% DMSO or 20% DMSO plus 0.4 mM (HA-EcGlpG) or 2 mM (HA-PaROM) JLK6 at 37 °C for 3 hours. The samples were purified with a C4 ZipTip (Millipore) and spotted with a α-cyano-4-hydroxycinnamic acid matrix (Laser...
Biolabs) and analyzed by MALDI-TOF using a Voyager DE-STR mass spectrometer (Applied Biosystems) calibrated with insulin, cytochrome C, apomyoglobin, and aldolase (Sigma).

**FRET Analysis of Protease-Substrate Affinity in Membranes**

Purified N-terminally His-tagged EcGlpG H254A+C104A+W196C protein was labeled at the single, engineered cysteine residue with the FRET acceptor tetramethylrhodamine-5-maleimide (TMR) (Life Technologies), purified from free dye using Ni\(^{2+}\) affinity resin, and dialyzed against 1X PBS, 0.1% DDM, and 15% glycerol. Reconstitutions into *E. coli* liposomes were performed as described above, except that 30 pmoles of TMR-labeled protein was mixed with 25 – 800 pmoles of the FITC- TatA ligand. Following reconstitution, the samples were incubated at 37°C in a Synergy H4 Hybrid plate reader. Fluorescence was measured at 5 nm interval emission wavelengths from 510 - 600 nm, with excitation at 485 nm. The FRET interaction was calculated at the 580 nm emission wavelength by subtracting the TMR-labeled GlpG protein alone and the FITC-TatA alone samples from the sample containing both macromolecules.

**In Vivo TatA-Flag Titration:**

Expression of TatA-Flag (full-length protein with a C-terminal Flag tag) in log-phase (*O.D._{600} ~ 0.3*) *E. coli* K12 BW25113 cells was titrated using arabinose-mediated induction from a pBAD plasmid (Life Technologies). Cultures were induced at 37 °C in a shaking incubator for 2 hours. Cleavage by endogenous GlpG was quantified by resolving cell lysates for each titration on 16% SDS-polyacrylamide gels with tricine running buffer, followed by western blot with αFlag (Sigma), and imaging with an Odyssey infrared scanner (LiCor Biosciences). To quantify expression levels of TatA-Flag, cells were lysed using a French pressure cell (2 passes at 16,000 psi), and the lysate was clarified to remove unbroken cells at 9,000 x g for 8 minutes in a JLA 8.1000 rotor (Beckman-Coulter). Membranes were collected by ultracentrifugation at 50,000 rpm for 1 hour in a MLA-55 rotor (Beckman-Coulter). Peripheral and contaminating soluble proteins were removed with a 100 mM Na-carbonate pH 11.0 wash at 4 °C for 30 minutes. Total membrane protein from each titration were separated on a 4-20% SDS-polyacrylamide gel with tris-glycine running buffer, stained with a colloidal Coomassie blue dye (LiCor Biosciences), and quantified with an Odyssey infrared scanner. TatA-Flag expression was converted to molar ratio of membrane proteins by correcting signals for
molecular weight, then converted to mole% of membranes considering that total membrane protein concentration in an *E. coli* cell is ~1.5 mole% (Kadner, 1996; Schnaitman, 1970a, 1970b). An apparent $K_M$ was estimated as follows:

$$\frac{[\text{TatAFlag}]_{\text{Product}}}{V_{\text{max}}[\text{TatAFlag}]_{\text{Total}}} = \frac{[\text{TatAFlag}]_{\text{Total}}}{K_M + [\text{TatAFlag}]_{\text{Total}}}$$

To estimate $k_{\text{cat}}$ *in vivo*, membranes from Flag-*EcGlpG* cells (Table 2.1), with endogenous expression of Flag-*EcGlpG*, were prepared by sucrose gradient centrifugation (see Methods and Materials in chapter II) and Flag-*EcGlpG* was quantified by western blot using purified Flag-*EcGlpG* as a standard. In parallel, BW25113 Str$^R$ cells expressing saturating amounts of TatA-Flag, induced at 1 mM arabinose for 2 hours at 37 °C, were analyzed for TatA-Flag cleavage by western blot probing with αFlag. Simultaneously, membranes were prepared from these cells and total TatA-Flag was quantified relative to BSA standards by Coomassie staining. $k_{\text{cat}}$ was calculated as follows:

$$k_{\text{cat}} = \frac{\text{TatAFlag}(\text{pmoles})}{\text{FlagEcGlpG}(\text{pmoles})} \times \frac{\text{FractionCleaved}}{7200s}$$

**Detergent Micelle Cleavage Assay:**

FITC-TatA in 50 mM Tris pH 7.4, 150 mM NaCl, 1 mM DTT, 0.15% Sarkosyl, 0.1% (w/v) DDM was incubated at 37 °C for 10 minutes. The mix was centrifuged at 13,000 rpm in a microcentrifuge for 10 minutes to remove any insoluble substrate, and mixed with 83 nM HA-*EcGlpG* in reaction tubes pre-warmed to 37 °C. Reactions were quenched and analyzed on 16% SDS-polyacrylamide gels with tricine running buffer as above. Initial substrate concentrations were calibrated using FITC-TatA standards analyzed in parallel. Cleavage reactions of APP+Spi7-Flag were performed as described previously and imaged with an Odyssey infrared scanner (LiCor Biosciences) (Baker et al., 2007).

**Co-Precipitation Analysis**

Ni$^{2+}$-NTA resin (Qiagen) was equilibrated in 2 mM phosphate buffer pH 7.4, 25 mM Tris pH 7.4, 120 mM NaCl, 0.15% Sarkosyl, 0.1% DDM, 1 mM DTT, 2.5 μM His-EcGlpG C104A+H254A and 70 μM FITC-TatA were incubated with the equilibrated resin for 3 hours at
room temperature. The resin was washed, and bound proteins were eluted with SDS-sample buffer for 30 minutes. FITC-TatA was detected as above using a 16% SDS-polyacrylamide gels with tricine running buffer and imaging with a Typhoon scanner (GE Healthcare). His-EcGlpG C104A+H254A was detected by western blot with the mouse α-His antibody (Qiagen), and imaged with an Odyssey infrared scanner (LiCor Biosciences).

**HPLC Equilibrium Gel Filtration Analysis:**

Two Superdex 200 PC 3.2/30 high-resolution gel filtration columns (GE Healthcare) were connected in tandem to a ProStar 410 HPLC system (Agilent/Varian). The columns were equilibrated with 9.5 μM FITC-TatA in 2 mM phosphate buffer pH 7.4, 25 mM Tris pH 7.4, 120 mM NaCl, 0.15% Sarkosyl, 0.1% DDM, 1 mM DTT. 20 μM HA-EcGlpG S201A+H254A was preincubated with 9.5 μM FITC-TatA, and 50 μL was injected using an Autosampler 410 onto the columns at a flow rate of 0.05 mL/minute. Elution was monitored by absorbance at all wavelengths simultaneously between 200 to 600 nm using a PDA 330 detector. The FITC-TatA peak was quantified relative to known FITC-TatA concentration standards that were run on the columns equilibrated with 9.5 μM FITC alone in buffer.

Conventional gel filtration analysis of the HA-EcGlpG/FITC-TatA complex was conducted under similar conditions, except that the columns were equilibrated with buffer lacking FITC-TatA (2 mM phosphate buffer pH 7.4, 25 mM Tris pH 7.4, 120 mM NaCl, 0.15% Sarkosyl, 0.1% DDM, 1 mM DTT).

**Fitting and Statistical Analysis**

All data were analyzed and graphed using the R language and environment. Initial rates were extracted from real-time curves between 4 and 14 minutes using the slope (m) in a linear model: \( y = mx + b \). Initial rates versus substrate concentration of reconstituted reactions were modeled using the Hill-modified Michaelis-Menten equation:

\[
V_0 = \frac{V_{\text{max}} \left[ S_0 \right]^h}{K_M^h + \left[ S_0 \right]^h}
\]

Importantly, cooperativity was not observed with reconstituted reactions analyzed by SDS-PAGE, indicating that cooperativity is not a true feature of the enzyme reaction. P-values in pairwise comparisons were derived from multiple non-linear regression analysis and \( k_{\text{cat}} \).
p-values were corrected for multiple comparisons using the Bonferroni method ($K_m$ p-values were not corrected because none achieved significance).

**Acknowledgements:**

Sin Urban designed the synthetic substrates and worked out the method to resuspend the substrates at a high (~800 µM) concentration. Sin Urban designed and Rosanna Baker performed experiments establishing the pH switch inducible reconstitution. Sangwoo Cho crystallized GlpG at pH 4.5 and solved its structure. Rosanna Baker designed and I performed the detergent-based kinetic assay, whereas I developed and performed the real-time kinetic assay in membranes. I developed and performed all remaining assays and remaining experiments.

**References:**


**Figure Legends:**

**Figure 3.1: Inducible pH Switch to Control Rhomboid Proteolysis**
Western blot (left) and quantification (right) of CSpitz-Flag cleavage by HA-EcGlpG reconstituted into E. coli membranes at low (pH 4.0) or neutral pH (pH 7.4). The reconstitutions were neutralized and incubated at 37 °C for an additional 1 hour. Thermostability analysis of HA-EcGlpG with or without pH 4.0 pretreatment, under conditions similar to reconstitution at low pH. The data were fit to a two-state Boltzmann curve to derive transition midpoints (T_m).

Structure of ΔN-EcGlpG crystals grown at pH 4.0 overlayed with a crystal structure at neutral pH. Note the overall similarity (Ca = 0.32 Å), except that the catalytic serine side-chain at low pH has moved away from the catalytic histidine (inset).

**Figure 3.2: Inducible, Real-Time Reaction for Rhomboid Proteolysis in Membranes**

(A) FITC-TatA fluorescence from 510 – 700 nm with excitation at 485 ± 20 nm in detergent micelles (blue) or reconstituted into E. coli membranes (red). Note the robust quenching by E. coli membranes. (B) Schematic of the inducible, real-time fluorescence assay for rhomboid proteolysis. (C) Real-time progress curves and of FITC-TatA or FITC-TatA A8V (bottom-left) cleavage by HA-EcGlpG (top-left), HA-EcGlpG S201A+H254A (top-right), or JLK6-pretreated HA-EcGlpG (bottom-right). (D) Fluorescent scans of tricine SDS-PAGE gels from (C) and a western blot of unlabeled CysTatA-His cleavage by HA-EcGlpG with quantification (below) revealing that the FITC fluorophore does not interfere with cleavage. (E) Mass spectrum of the soluble FITC-TatA cleavage product showing that FITC-TatA is cleaved at the correct site by EcGlpG in membranes.

**Figure 3.3: Kinetic and Thermodynamic Analysis of EcGlpG Cleavage**

(A) Michaelis-Menten plot from real-time data of FITC-TatA cleavage by HA-EcGlpG in E. coli membranes. Top inset: Roughly half of the C-terminal cysteine of FITC-TatA is labeled with the membrane-impermeable IR800CW-maleimide dye, which shows topological insertion is random. Bottom inset: Parameter estimates for K_m and V_max', and R^2 value of the model fit. (B) Michaelis-Menten plot from tricine SDS-PAGE data of FITC-TatA cleavage by HA-EcGlpG in E. coli membranes. Inset: Parameter estimates for K_m and V_max', and R^2 value of the model fit. (C) A binding isotherm from FRET interactions between FITC-TatA and His-EcGlpG C104A+H254A+W196C-TMR. Inset: Estimate for K_d and the excitation and emission wavelengths used to detect the FRET interaction. (D) Initial rates of FITC-TatA cleavage with or without adding 10-times more inactive HA-EcGlpG S201A+H254A than HA-EcGlpG WT.
shows that HA-EcGlpG S201A+H254A does not act as a dominant negative. (E) Mass spectra of HA-EcGlpG purified protein pre-treated with DMSO or the mechanism-based inhibitor JLK6. The complete spectral shift revealed that nearly 100% of purified HA-EcGlpG is active. (F) In vivo analysis of full-length TatA-Flag cleavage by endogenous GlpG. Top: western blot of TatA-Flag cleavage by endogenous GlpG showed increased cleavage by increasing substrate expression. Middle: western blot of endogenous Flag-EcGlpG in E. coli membranes compared to purified Flag-EcGlpG. Note that the cross-reactive, lower band served as a loading control. Bottom: A pseudo-Michaelis-Menten plot of Tat-Flag cleavage by endogenous EcGlpG in living E. coli cells. A Coomassie-stain of membrane preparations (inset) showed that at 1,000 µM arabinose, TatA-Flag accumulates to become the most abundant membrane protein.

**Figure 3.4: Kinetic and Thermodynamic Analyses of HA-EcGlpG and FITC-TatA in Detergent Micelles**


**Figure 3.5: Kinetic Effects of Gate-Open Mutations and Solvent Isotopes**

(A) Kinetic solvent isotope effect of FITC-TatA cleavage by HA-EcGlpG WT and the gate-open HA-EcGlpG F153A+W236A reconstituted into E. coli membranes. (B) The kinetic isotope effect dependence on fraction of D₂O fitted to the data with a linear model fitted to the data. Dashed lines represent 95% confidence intervals of the linear fit. (C) Left: Michaelis-Menten plots of FITC-TatA cleavage by HA-EcGlpG WT and the gate-open HA-EcGlpG F153A+W236A mutant reconstituted into E. coli membranes. Right: Quantification of gate-open HA-EcGlpG F153A+W236A mutant Kₐ and kₐₑₐₜ parameters relative to HA-EcGlpG WT.

**Figure 3.6: Kinetic Characterization of Diverse Bacterial Rhomboid Proteases**

(A) Michaelis-Menten plot of FITC-TatA cleavage by PsAarA. Inset: Parameter estimates and R² value for the model fit. (B) Michaelis-Menten plots of FITC-TatA cleavage by 9 diverse
bacterial rhomboid proteases. (C) Base-10 logarithm plot of kinetic parameter estimates for 8 bacterial rhomboid proteases relative to HA-EcGlpG (D) Phylogenetic tree of the surveyed rhomboid proteases and their derived kinetic constants. (E) Overlayed mass spectra of HA-PaROM pre-treated with DMSO or the mechanism-based inhibitor JLK6. Complete shift of the spectrum showed nearly 100% activity for the HA-PaROM purified protein preparation.

**Figure 3.7: Kinetic Analyses of Substrate Variants and a Model for Rhomboid Proteolysis**

(A) Michaelis-Menten plot of FITC-substrate cleavage by PsAarA. Inset: Parameter estimates and $R^2$ value for the model fit. (B) Base-10 logarithm plot of kinetic parameter estimates for cleavage of 3 substrate variants relative FITC-TatA by HA-PsAarA. (C) FITC-substrates sequence alignment showing substrate differences and their kinetic parameter estimates. (D) A model for rhomboid proteolysis in membranes and their kinetic constants with allusion to DNA glycosylase-catalyzed removal of damaged DNA bases. “I”, “E”, and “S” refer to interrogation, excision, and scission complexes, respectively.
Figure 3.1

A. Reconstitution pH

<table>
<thead>
<tr>
<th>Hours post reconstitution pH 7.4</th>
<th>0</th>
<th>1</th>
<th>0</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substrate</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Product</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HA-EcGipG</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

B. Relative Activity

<table>
<thead>
<tr>
<th>pH 7.4</th>
<th>pH 4.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.2</td>
</tr>
<tr>
<td>0.4</td>
<td>0.6</td>
</tr>
<tr>
<td>0.8</td>
<td></td>
</tr>
</tbody>
</table>

C. Relative Light Scatter (%)

- HA-EcGipG - Control
  $T_m = 72.0 \pm 0.12^\circ$ C
- HA-EcGipG - pH 4.0 Treated
  $T_m = 73.1 \pm 0.08^\circ$ C

Cα RMSD = 0.32 Å
Figure 3.2
Figure 3.3

A

B

C

D

E

F

Figure 3.3
Figure 3.4

A

B

C

D

Kd = 135 ± 16.6 µM
Kd, App = 817 ± 76 µM
Kd = 191 ± 29.4 µM

Vmax = 3.34*10^-3 ± 2.78*10^-4 µM^-1*s^-1

h = 2.31 ± 0.52
kcat = 0.0407 ± 0.00336 s^-1

Elution (minutes)

Absorbance (mA.U.)

FITC-TatA [µM]

KM = 135 ± 16.6 µM

λ = 280 nm (protein)
λ = 485 nm (FITC)
λ = 485 nm (FITC)

V0 (µM / minute)

FITC-TatA + EcGlpG S201A+H254A

FITC-TatA + EcGlpG S201A+H254A

FITC-TatA Alone

λ = 485 nm (FITC)

Figure 3.4

FITC-TatA + EcGlpG C104A+H254A

Input
Elution
10%
100%

FITC-TatA

His-EcGlpG

Kd = 191 ± 29.4 µM

Kd = 191 ± 29.4 µM

BSA

Kd = 191 ± 29.4 µM
Figure 3.5

A

Initial rates (RFU/minute)

E. coli GlpG WT
E. coli GlpG F153A+W236A

B

Initial rates (RFU/minute)

n (fraction D₂O)

R² = 0.88

C

Vₐ (RFU/minute)

FITC–TatA [mole %]

E. coli GlpG F153A+W236A
E. coli GlpG WT

kₑ/kₑ₀ = 3.15 ± 0.41

kₑ/kₑ₀ = 3.15 ± 0.41

kₑ/kₑ₀ = 1.26 ± 0.07

H₂O

H₂O

D₂O

kₑ/kₑ₀ = 1.34 ± 0.03

Ratio of E. coli GlpG F153A+W236A to E. coli GlpG WT

p = 5.1*10⁻⁷

p = 0.76

p = 5.1*10⁻⁷

p = 0.76
Figure 3.6

A

B

C

D

E

k_{cat} (s^{-1})

K_M (mole %)

Ratio Relative to EcGlpG (log_{10})

b_{li} (mole %)

h_{li} (m^2)

Intensity (%)

Mass (m/z)
Figure 3.7
Appendix

*Vibrio cholerae* GlyGly-CTERM Proteins are Substrates of an Atypical *Vibrio cholerae* Rhomboid Protease
Introduction:

A recent report discovered a strong phylogenetic co-occurrence between a subfamily of bacterial rhomboid proteases, subsequently termed rhombosortases (TIGR03902), and proteins sharing a conserved C-terminal transmembrane region, subsequently termed the GlyGly-CTERM (TIGR03501)—a tripartite structure consisting of a glycine-rich sequence, a transmembrane domain (TMD), and a cluster of basic residues (Haft and Varghese, 2011). Similar conserved tripartite structures exist elsewhere: the LPXTG (Schneewind et al., 1993) and PEP-CTERM (Haft et al., 2006) carry their eponymous motif, in addition to a TMD and a series of basic residues.

LPXTG containing proteins are acted on by transpeptidases, termed sortases, that use a conserved cysteine residue to attach LPXTG containing proteins to the peptidoglycan between the threonine and glycine residues (Gaspar et al., 2005). As such, this system acts to “sort” substrates to the peptidoglycan layer in gram-positive bacteria. Although lacking sequence homology with LPXTG sortases, the presumed PEP-CTERM sortases share a similar invariant nucleophilic cysteine residue (Haft et al., 2006). However, whether it acts as transpeptidase or a protease depends on the nucleophile that resolves the acyl-enzyme intermediate (i.e. a peptide amine or a water molecule for a transpeptidase and protease, respectively).

Another shared feature of these three systems is a many-to-one relationship between the motif containing proteins and their corresponding sortase: in each case, when a sortase is found encoded in a genome, the chances are high that the same genome encodes more than one proteins with the corresponding C-terminal domain. This is consistent with the sortase being the enzymatic component of the system, suggesting that rhombosortases could act enzymatically on GlyGly-CTERM proteins. As rhomboid proteases are well-described serine proteases, the expectation is that rhombosortases cleave GlyGly-CTERM proteins near the GlyGly motif. In fact, the authors who describe the rhombosortase/GlyGly-CTERM system explicity predict this: “The extensive set of species showing co-occurrence, despite sporadic distribution, strongly suggests a direct functional connection: cleavage of GlyGly-CTERM protein tail regions by rhombosortase (Haft and Varghese, 2011).”

Missing from this impressive bioinformatic discovery, however, is direct experimental
evidence that rhombosortases cleave or in any other way act on GlyGly-CTERM containing proteins. Having already screened for *E. coli* and *H. influenzae* rhomboid protease substrates (see Chapter II), I was well positioned to test this hypothesis using the same cleavage assay in *E. coli* cells.

Here, I test for cleavage of three GlyGly-CTERM containing proteins from *Vibrio cholerae*, a human pathogen that has caused 7 different pandemics in recent history. Interestingly, *V. cholerae* encodes two rhomboid proteases: a rhombosortase, termed VcROM, and the rhomboid protease published as VcRho (Baker and Urban, 2012), which has high similarity to *E. coli* GlpG.

**Results:**

**VcROM, but not VcRho nor EcGlpG Cleave GlyGly-CTERM Containing Proteins**

I cloned 3 proteins (VC_A0820, VC_A1254, and VC_A2198) that contained the GlyGly-CTERM motifs from *V. cholerae* serotype O1 strain O395 genomic DNA into a pET27b vector, in a way removing the pelB leader sequence, and with a Flag tag fused to the C-terminus for detection. VC_A0820 and VC_A1254 encode trypsin-like domains of variable length, whereas VC_A2198 contains a lamin tail domain and an exo/endonuclease/phosphatase domain (Figure A.1), consistent with diverse enzymatic domains found attached to a GlyGly-CTERM motif (Haft and Varghese, 2011).

When each candidate substrate was co-expressed with the putative rhombosortase VcROM in *E. coli* cells, a distinct band corresponding to the size of the predicted product was present on a western blot specific to the Flag tag (Figure A.2). Notably, the non-rhombosortases VcRho and EcGlpG failed to cleave VC_A1254-Flag and VC_A2198-Flag, despite being expressed at similar levels to VcROM. Furthermore, VcRho and EcGlpG cleavage of VC_A0820 was drastically reduced compared to cleavage by VcROM. Taken together, these results confirm that rhombosortases selectively cleave GlyGly-CTERM proteins from *Vibrio cholerae*.

**The Signature Histidine in VcROM is Dispensable for Proteolysis**

Nearly all rhomboid proteases encode a tyrosine or phenylalanine (Y205 for EcGlpG) four residues C-terminal to the active site serine (S201 for EcGlpG). In *E. coli* GlpG crystal structures (Ben-Shem et al., 2007; Wang et al., 2006; Wu et al., 2006), Y205 stacks with
and stabilizes the catalytic histidine, H254. Unlike most rhomboid proteases however, rhombosortases almost invariably encode an atypical histidine in place of Y205 (Haft and Varghese, 2011). The authors of the bioinformatic study concluded that this histidine is an essential feature of a rhombosortase (Haft and Varghese, 2011).

To test whether this histidine, (H145 in VcROM), is required for cleavage of GlyGly-CTERM proteins, I mutated this residue in VcROM to tyrosine (H145Y) or alanine (H145A). The H145Y mutation had no effect on cleavage of VC_A0820-Flag and VC_A1254-Flag, but did decrease the cleavage of VC_A2198-Flag about two-fold (Figure A.3). Similarly, the H145A mutation had no effect on the cleavage of VC_A0820-Flag and decreased the cleavage of VC_A2198-Flag two-fold. However, the H145A mutation abolished cleavage by of VC_A1254-Flag. Therefore, these results show that the signature rhombosortase histidine is not essential for cleavage of GlyGly-CTERM containing proteins, and that the more typical tyrosine functionally substitutes for the conserved histidine.

Discussion:

Rhombosortases Enzymatically Act on GlyGly-CTERM Proteins

Here, I provide clear evidence that three GlyGly-CTERM containing proteins are enzymatically hydrolyzed by rhombosortases. Furthermore, I showed that only the rhombosortase VcROM, and not the non-rhombosortase VcRho and EcGlpG, efficiently cleaves the GlyGly-CTERM proteins selectively.

The Signature Histidine may be a Founder Effect

The invariant histidine near the active site serine is found in nearly all rhombosortases (Haft and Varghese, 2011). My data suggests that this histidine is not a functional requirement for rhombosortases to cleave GlyGly-CTERM proteins. Why then, does this histidine pervade rhombosortases yet rarely appear in other rhomboid proteases family members? Perhaps, this histidine represents a founder effect, whereby an ancient rhombosortase encoded a rare histidine that was retained by its descendants. Whatever the case, this histidine residue alone does not explain the specificity of rhombosortases for GlyGly-CTERM proteins for which VcRho and EcGlpG could not cleave. The basis for this selectivity awaits future studies.

Material and Methods:

Plasmids and Strains
VcROM (Uniprot: Q9KQL7) and VcRho (Uniprot: Q9KVP2) were cloned into a pGEX 6p-1 expression vector as described for EcGlpG (Urban and Wolfe, 2005). *V. cholerae* substrates (VC_A0820: Q9KSQ6; VC_A1254: Q9KRJ1; VC_A2198: A5F515) were amplified from genomic DNA (ATCC) with specific primers that fused a Flag-tag fused to their C-terminal coding region and cloned into a pET 27b vector using the NdeI and NotI restriction sites. This cloning strategy removed the pelB leader sequence present in pET27b. All constructs were verified by sequencing the entire ORF.

A ∆GlpG strain with a kanamycin-resistance cassette replacing the endogenous glpG gene (Baba et al., 2006) was obtained from the Coli Genetic Stock Center (Yale University). I cured this strain of its kanamycin-resistance cassette by FLP mediate recombination, and lysogenized it with DE3 to promote protein expression using the T7 RNA polymerase (creating the ∆glpG (DE3) strain).

**In Vivo Cleavage Assay**

The cleavage assays were performed as described previously (Urban and Baker, 2008), except that ∆glpG (DE3) cells were transformed with each plasmid, and proteins were co-expressed for the amount of time indicated following induction with IPTG. In addition, lysates of cells expressing VC_A0820-Flag and VC_A1254-Flag were separated on 10% NuPAGE gels (Life Technologies) with 1X MES running buffer at 200V for 35 minutes.

**References:**


Figure Legends:

Figure A.1: Domain Schematic and Sequence Motifs of the GlyGly-CTERM Proteins

The schematic shows the predicted domain structure (by PFAM) and the GlyGly-CTERM sequence of VC_A0820, VC_A1254, and VC_A2198. The glycine-rich region and the basic residue cluster are colored red and blue, respectively.

Figure A.2: Cleavage of GlyGly-CTERM proteins from V. cholerae

Western blot of cleavage occurring in E. coli cells. Co-expressed proteins were separated by SDS-PAGE, revealing efficient cleavage of GlyGly-CTERM proteins by VcROM, but not by VcRho nor EcGlpG. (U: uninduced sample before addition of IPTG). Red star indicates non-rhombosortase cleavage band that is eliminated by co-expression of a rhomboid
protease. The >3.5 kD marker indicates the 3.5 kD marker migrated beyond the bottom of the region of the αFlag western blot shown, whereas the 10 kD marker (not shown) was far above.

**Figure A.3: H145 mutations has little effect on GlyGly-CTERM cleavage**

Western blot of GlyGly-CTERM protein cleavage in *E. coli* cells co-expressing VcROM H145Y and H145A mutants. (U: uninduced sample before addition of IPTG) The >3.5 kD marker indicates the 3.5 kD marker migrated beyond the bottom of the region of the αFlag western blot shown, whereas the 10 kD marker (not shown) was far above.
**Figure A.1**
Figure A.2
Figure A.3
CURRICULUM VITAE FOR Ph.D. CANDIDATES
The Johns Hopkins University School of Medicine

Seth Wayne Dickey
Name
November 1, 2013
Date of Version

Educational History:
Ph.D. expected 2013 Program in BCMB Johns Hopkins School of Medicine
Mentor: Sinisa Urban
B.S. 2007 Biochemistry Michigan State University

Academic and other honors:
2013 International Travel Award International Proteolysis Society
2012 Collaborative Research Travel Grant Burroughs Wellcome Fund
2012 Outstanding Journal Club Presentation Department of Molecular Biology and Genetics

Peer Reviewed Publications:


Posters, Abstracts, and Oral Presentations:


Inventions:
2012 An inducible reconstitution and real-time quantitative kinetic system for the analysis of catalysis and inhibition in membranes.

Service and leadership:
2008 - 2009 Tutorial for Biochemical and Biophysical Principles: Led a group discussion of course topics once a week to supplement lectures.