Characterisation of two putative rhomboid proteases in *Shigella sonnei* physiology and pathogenesis.



Guangyu Liu

Sir William Dunn School of Pathology

Lincoln College

University of Oxford

A thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Trinity Term 2018

Acknowledgment

I would like to thank Professor Christoph Tang and Dr Rachel Exley for their guidance and support throughout my DPhil. They taught me how to start from ideas to well-planned experiments. I am also grateful for the past and present members of the Tang lab, with special thanks to Miranda Rogers, who is my first student and did a wonderful project on characterising the growth and oxidative stress response of *S. sonnei* lacking GlpG and Rhom7.

I am thankful to the people who contributed to this work: Dr Stephen Beaton helped measure the hydrogenase-2 activity in *Shigella sonnei* lacking GlpG and Rhom7. Dr Kvido Strisovsky provided the pKS508 plasmid, which encodes the artificial substrate. Professor Matthew Freeman, Professor Fraser Armstrong, Dr Rhiannon Evans, Dr Adam Grieve, and Dr Stephen Beaton have been a great source for knowledge, discussions and suggestions on experiments.

I also would like to thank to those who agreed to act as my examiners at various stages of my DPhil: Professor Matthew Freeman, Professor Judy Armitage and my *viva voce* Professor Ben Berks and Professor Ian Henderson.

Finally, I would like to acknowledge my funding source, the E P Abraham Departmental Studentship.

The following publications were achieved during my DPhil studies:

Liu, G., Tang, C.M., and Exley, R.M. (2015). Non-pathogenic *Neisseria*: members of an abundant, multi-habitat, diverse genus. Microbiology *161*, 1297-1312.

Wormann, M.E., Horien, C.L., Johnson, E., Liu, G., Aho, E., Tang, C.M., and Exley, R.M. (2016). *Neisseria cinerea* isolates can adhere to human epithelial cells by type IV pilus-independent mechanisms. Microbiology *162*, 487-502.

Characterisation of two putative rhomboid proteases in *Shigella sonnei* physiology and pathogenesis.

Guangyu Liu

DPhil Pathology, Trinity Term 2018, Lincoln College

Abstract

Shigella sonnei is a human-specific pathogen and a major cause of bacillary dysentery. Rhomboids are evolutionarily conserved serine intramembrane proteases that control a variety of important cellular processes, including regulating apoptosis in mammals, development in *Drosophila*, and invasion of eukaryotic parasites. However, functions of bacterial rhomboids remain largely unexplored.

This work identified two active rhomboids: GlpG and Rhom7, in *S. sonnei*. Rhom7 possesses a novel domain composition among rhomboids characterised so far, consisting of seven transmembrane domains (TMDs) and a putative C-terminal WD40 domain. However, neither the 7th TMD nor the C-terminal domain is required for the proteolytic activity of Rhom7. To identify physiological substrates of GlpG and Rhom7, the cleavage propensity of TMDs derived from all predicted type I & III membrane proteins in *S. sonnei* were assessed. Candidate substrates were further validated by testing the cleavage of the full-length substrates by GlpG and/or Rhom7, revealing that HybA and FdoH are physiological substrates for GlpG, and that HybA and FdnH are substrates for Rhom7 but only when Rhom7 is expressed from a plasmid.

Further analyses of these substrates revealed that they are all components of multi-protein respiratory complexes (HybA from hydrogenase-2, FdoH from formate dehydrogenase-O, and FdnH from formate dehydrogenase-N). I propose that their cleavage by rhomboids follows a 'complex protection' model, which states that only the uncomplexed substrates are cleaved by rhomboids, while the substrates residing in their native complexes are protected. As rhomboids are known to display propensity to cleave TMDs with low helical stability. I proposed that discrimination of uncomplexed and complexed substrates is likely to be mediated by the presence of helix-destabilising residues in the substrate TMDs, which render uncomplexed substrates prone to rhomboid cleavage while the stability of substrate TMDs is maintained by packing with TMDs from their binding partner(s) in the complex. Consistent with that, we identified helix-destabilising proline residues in HybA, FdoH, and FdnH. Moreover, substitutions of these residues with alanine abolished rhomboid cleavage regardless of complex integrity. These findings lead to the proposal that rhomboids might act as a quality control mechanism scrutinising membrane complex integrity, whose role might only become important under stresses.

Table of contents

Chapter 1. Introduction 16

1.1. Rhomboid proteases 16

1.1.1. Regulated intramembrane proteolysis and an overview of rhomboid proteases 16

1.1.2. Structure and enzymology of rhomboid proteases 17

1.1.3. Biological functions of prokaryotic rhomboid proteases 22

1.2. Overview of *Shigella* spp. and shigellosis 25

1.2.1. Clinical manifestations of *Shigella* infection 28

1.2.3. Epidemiology 28

1.3. *Shigella* pathogenesis 30

1.3.1. Survival in the stomach and small intestine 30

1.3.2. Interactions with microfold cells (M cells) and macrophages 32

1.3.3. Intracellular replication and intercellular spread 36

1.3.4. *Shigella* clearance by neutrophils and environmental dissemination 40

1.3.5. Oxidative stress response in *Shigella* pathogenesis 41

1.4. Hydrogenases 42

1.4.1. Overview of bacterial hydrogenases 42

1.4.2. Hydrogenase-1 (Hyd-1) 43

1.4.3. Hydrogenase-2 (Hyd-2) 44

1.4.4. Hydrogenase-3 (Hyd-3)/formate hydrogenlyase-1 (FHL) 47

1.4.5. Hydrogenase-4 (Hyd-4)/formate hydrogenlyase (FHL-2) 48

1.5. Formate dehydrogenases 49

1.5.1. Formate dehydrogenase-N (FDH-N) 49

1.5.2. Formate dehydrogenase-O (FDH-O) 51

1.6. Project Aims 52

Chapter 2. Materials and Methods 53

2.1. Strains and plasmids 53

2.2. Molecular techniques 59

2.2.1. Polymerase Chain Reaction (PCR) 59

2.2.2. Plasmid and genomic DNA extraction 60

2.2.3. Gel electrophoresis and purification of DNA 60

2.2.4. Restriction digestion and ligation 60

2.2.5. Gibson assembly 61

2.2.6. Transformation 61

2.2.7. Electroporation 61

2.3. Plasmid and strain construction 62

2.3.1. Construction of pCONJ4s and its derivatives 62

2.3.2. Generation of *S. sonnei* markerless mutants by conjugation 64

2.3.3. Construction of pBAD33 and pUC19 derivatives for *S. sonnei* complementation 65

2.3.4. Bioinformatic identification of substrates 67

2.3.5. Construction of the transmembrane domain library for rhomboid substrate screen 67

2.4. SDS-PAGE and western blot analysis 68

2.5. Rhomboid-dependent protein cleavage assays 68

2.6. Analysis of type three secretion system secretion 69

2.7. Hydrogenase activity assays 69

2.8. Biolog phenotypic microarrays 71

2.9. Oxidative stress assays 71

2.10. Quantification and Statistical analysis 72

Chapter 3. Identification and biochemical characterisation of rhomboid proteases in *S. sonnei* 73

3.1. Identification of Rhom7, a potential active rhomboid protease 73

3.2. Development of an assay to measure rhomboid activity using a TatA derived substrate in *S. sonnei* CS14 78

3.3. Serine 133 and Histidine 187 constitute the catalytic dyad of Rhom7 82

3.4. The C-terminal domain and 7th transmembrane domain of Rhom7 do not affect Rhom7-mediated cleavage of the TatA derived artificial substrate 86

3.5. Conclusion and discussion 87

Chapter 4. Phenotypic characterisation of *S. sonnei* lacking GlpG and Rhom7 91

4.1. Biolog high-throughput phenotypic screen of *S. sonnei* CS14 Δ*glpG*Δ*rhom7* 91

4.2. Growth of *S. sonnei* CS14Δ*glpG*Δ*rhom7* in stress conditions 92

4.3. Deletion of *glpG* and *rhom7* does not impair T3SS secretion in *S. sonnei* CS14 95

4.4. Conclusion and discussion 97

Chapter 5. Identification of GlpG and Rhom7 substrates 100

5.1. Design of a genome-wide screen to search for substrates of GlpG and Rhom7 100

5.2. Construction of artificial substrates containing candidate substrate transmembrane domains 102

5.3. Results of genome-wide substrate screen and preliminary validation 105

5.4. Further investigation of HybA and HybO as substrates of GlpG and Rhom7 114

5.5. Identification of HybA residues key to GlpG cleavage 118

5.6. Conclusion and discussion 122

Chapter 6. Impact of GlpG on Hyd-2 activity in *S. sonnei* 126

6.1. Impact of GlpG cleavage on Hyd-2-mediated hydrogen-dependent growth of *S. sonnei* CS14 126

6.2. Impact of GlpG cleavage on Hyd-2-dependent hydrogen uptake in *S. sonnei* CS14 130

6.3. Conclusion and discussion 134

Chapter 7. GlpG acts as a quality control enzyme cleaving uncomplexed substrates 135

7.1. HybA cleavage by GlpG is enhanced by deletion of HybB. 135

7.2. FdoH cleavage by GlpG is only observed upon deletion of FdoI. 141

7.3. FdnH cleavage by Rhom7 is only observed upon deletion of *fdnI*. 143

7.4. Proline-to-alanine mutations in rhomboid substrates abolish cleavage regardless of complex protection. 146

7.5. Conclusion and discussion 152

Chapter 8. Discussion and future directions 155

Appendices 165

References 186

Table of Figures

Figure 1.1. Structure of the core domain of *E. coli* GlpG. 18

Figure 1.2. Positions of P4, P1, and P2' residues of *P. stuartii* TatA relative to the AarA cleavage site. 19

Figure 1.3. Proposed kinetic scheme of GlpG-mediated intramembrane proteolysis and its inhibition by a susbtrate-mimicking peptide aldehyde inhibitor. 21

Figure 1.4. Schematic representation of the *Shigella* T3SS. 26

Figure 1.5. Stages of *S. flexneri* infection upon reaching the colonic epithelium. 33

Figure 1.6. Detection of and responses to *S. flexneri* in a macrophage. 35

Figure 1.7. *S. flexneri*-host interactions in the cytosol of epithelial cells. 39

Figure 1.8. Proposed quaternary structure of a dimer of *E. coli* Hyd-2. 45

Figure 1.9. The predicted structure of the *E. coli* Hyd-3/FHL complex. 48

Figure 1.10. Crystal structure of FDH-N and its metabolic interplay with NarGHI. 51

Figure 2.1. Map of pCONJ4s. 63

Figure 3.1. Sequence, topology and domain analyses of *S. sonnei* CS14 GlpG (SSON_3661). 74

Figure 3.2. Sequence, topology and domain analyses of *S. sonnei* CS14 Rhom7 (SSON_0610). 76

Figure 3.3. Genetic organisation of *S. sonnei* CS14 *glpG* and *rhom7* and comparison with their counterparts in *E. coli* MG1655. 78

Figure 3.4. Representations of the artificial substrate and the *in vivo* cleavage assay to detect rhomboid activity in *S. sonnei*. 80

Figure 3.5. *S. sonnei* CS14 GlpG is an active rhomboid. 82

Figure 3.6. *S. sonnei* CS14 Rhom7 is an active rhomboid protease, with Ser^{133} and His^{187} required for its activity. 83

Figure 3.7. Neither S133A or H187A substitution of Rhom7 affects its expression. 85
Figure 3.8. The 7th TMD and the CTD of Rhom7 are dispensable for Rhom7 activity. 86
Figure 4.1. GlpG and Rhom7 are not required for aerobic growth of *S. sonnei* in LB. 91
Figure 4.3. GlpG and Rhom7 do not impact *S. sonnei* survival in oxidative stress induced by H_2O_2 or PQ. 94
Figure 4.4. T3SS-mediated secretion of Ipa proteins of *S. sonnei* is independent of rhomboid proteases. 96
Figure 5.1. The workflow for the bioinformatic identification of *S. sonnei* candidate rhomboid substrates. 101
Figure 5.2. Schematic representation of the artificial substrate used for screening. 102
Figure 5.3. Cleavage of candidate substrates by GlpG.....107
Figure 5.4. Cleavage of candidate substrates by Rhom7.....110
Figure 5.5. Confirmation of cleavage of six candidate substrates by GlpG or Rhom7. 112
Figure 5.6. Venn diagram showing the putative substrates of GlpG and/or Rhom7 identified and alignment of amino acid sequence of TMD regions cloned into the artificial substrates. 113
Figure 5.7. Schematic representation of full-length and cleaved tagged HybO and HybA. 114
Figure 5.8. Full-length HybO is not amenable to cleavage by endogenous or plasmid-encoded GlpG or Rhom7. 115
Figure 5.9. Endogenous GlpG cleaves full-length HybA but Rhom7 cleavage of HybA is dependent upon expression from a plasmid. 117
Figure 5.10. Endogenous GlpG cleaves HybA when HybA is tagged with a C-terminal Halo-tag. 118

Figure 5.11. HybA Gly^{296} is critical for cleavage by the *Shigella* rhomboids, GlpG and Rhom7. 120

Figure 5.12. Pro^{300} of HybA is critical for cleavage by the *Shigella* rhomboids, GlpG and Rhom7. 121

Figure 6.1. Schematic representations of the loci encoding Hyd-1, Hyd-2, and Hyd-3 in *E. coli* MG1655 and *S. sonnei* CS14. 128

Figure 6.2. Hyd-2 contributes to *S. sonnei* growth in the presence of 5% H_2 but GlpG and cleavage of HybA have no effect. 130

Figure 6.3. Neither GlpG nor HybA cleavage impacts Hyd-2-dependent H_2 uptake. 133

Figure 7.1. HybA cleavage by GlpG is increased upon deletion of *hybB*, or HybA over-expression. 137

Figure 7.2. Schematic representation of the 'complex protection' model. 138

Figure 7.3. GlpG does not cleave HybO even upon deletion of *hybB*, or HybO over-expression. 140

Figure 7.4. Structure of the formate dehydrogenase-N (FDH-N) complex and a schematic representation of the predicted organisation of the formate dehydrogenase-O (FDH-O) complex. 141

Figure 7.5. Endogenous GlpG cleaves full-length FdoH only in the absence of FdoI. 142

Figure 7.6. Plasmid-encoded Rhom7 cleaves full-length FdnH only in the absence of FdnI. 144

Figure 7.7. FdnH is not a substrate of GlpG. 145

Figure 7.8. Schematic representation of the potential hypotheses accounting for the 'complex protection' model. 147

Figure 7.9. Mutation of the TMD Pro^{300} to alanine in HybA abolishes cleavage of uncomplexed HybA by GlpG. 148

Figure 7.10. Mutation of TMD-embedded Pro^{259} to alanine in FdoH is sufficient to abolish cleavage of uncomplexed FdoH by GlpG. 150

Figure 7.11. Mutation of TMD-embedded Pro^{259} to alanine in FdnH is sufficient to abolish cleavage of uncomplexed FdnH by plasmid-encoded Rhom7. 151

Figure 8.1. Quality control of the αβTCR by assembly-dependent membrane integration 159

Figure 8.2. Alignment of amino acid sequences of TMD regions of FdoH, FdnH, and HybA. 161

Figure 4.2. Characterisation of *S. sonnei* rhomboid mutants by Biolog phenotypic microarrays. 185

List of Tables

Table 1.1. Composition of hydrogenases and formate dehydrogenases in *E. coli*. 43

Table 2.1. Strains and plasmids used in this work 53

Table 5.1. The sequence of the predicted TMDs and additional amino acids of putative rhomboid substrates from *S. sonnei* CS14 used as substrates for *in vivo* screening. 103

Table 2.2. Primers used in this work 165

Chapter 1. Introduction

The aim of this thesis is to characterise the biological functions of two rhomboid proteases in *Shigella sonnei*. I begin the introduction with an overview of our current knowledge of rhomboid proteases with particular focus on prokaryotic enzymes. This is followed by an overview of *Shigella* in terms of 1) the disease it causes, 2) interactions between *Shigella* and the human host, and 3) the molecular mechanisms underlying the pathology. The final sections of the introduction focus on hydrogenases and formate dehydrogenases, as during the course of this work, these were identified as rhomboid substrates.

1.1. Rhomboid proteases

1.1.1. Regulated intramembrane proteolysis and an overview of rhomboid proteases

Regulated intramembrane proteolysis (RIP) is a conserved mechanism that controls a wide range of important cellular processes in all kingdoms of life (Chun *et al.,* 2007; Knopf and Adam, 2012; Konovalova *et al.,* 2014; Lu *et al.,* 2014; Parente *et al.,* 2014). RIP involves two fundamental components: an intramembrane protease (IMP), and a substrate with at least one transmembrane domain (TMD). There are several defining features of IMPs: i) they are all polytopic membrane proteins, ii) they all have a membrane-buried active site, and iii) they usually cleave the TMD of substrates. IMPs can be categorised into four classes based on their catalytic mechanisms (*i.e.*, serine (Urban *et al.,* 2001), metallo- (Rawson *et al.,* 1997), aspartate (De Strooper *et al.,* 1999), and glutamate proteases (Manolaridis *et al.,* 2013)), and they are phylogenetically unrelated to the soluble protease counterparts (Freeman, 2014).

Intramembrane serine proteases are also known as rhomboids. The first protein of this family, rhomboid-1, was identified as a protein required for establishment of the dorsoventral axis and the peripheral nervous system development during *Drosophila* embryogenesis (Bier *et al.,* 1990). Later, rhomboid-1 was found to cleave and activate the EGF receptor ligand, Spitz (Wasserman *et al.,* 2000) and was subsequently shown to be an intramembrane serine protease (Lee *et al.,* 2001 ; Urban *et al.,* 2001). Spitz is a type I integral membrane protein, which is defined by the presence of a signal peptide, a single TMD, an extracellular N-terminus, and a cytosolic C-terminus (Lee *et al.,* 2001 ; Urban *et al.,* 2001). Gly^{143} in the N-terminal half of the TMD, which reduces the stability of transmembrane helix, was also shown to be critical for Spitz cleavage by rhomboid-1 *in vivo* (Urban and Freeman, 2003). These features of Spitz were then used as criteria for rational searches for novel rhomboid substrates, leading to the identification of rhomboid substrates from a range of evolutionarily distant organisms, including several adhesins in *Toxoplasma gondii* (Urban and Freeman, 2003), TVAG_280090, a protein with unknown function, in *Trichomonas vaginalis* (Riestra *et al.,* 2015), Gurken and Keren in *Drosophila* (Urban *et al.,* 2002), and thrombomodulin in mammals (Lohi *et al.,* 2004). Due to the success of this approach, type I integral membrane protein was used as a selection criterion for rhomboid substrate identification in this work.

1.1.2. Structure and enzymology of rhomboid proteases

To date, only two rhomboids have been structurally determined at atomic resolution by X-ray crystallography: *Escherichia coli* GlpG and its homologue in *Haemophilus influenzae*, both of which share a common fold (Lemieux *et al.,* 2007; Wang *et al.,* 2006). The structure of the membrane-bound domain of *E. coli* GlpG (Wang *et al.,* 2006; Wu *et al.,* 2006) demonstrated that the core domain of GlpG consists of six transmembrane (TM) helices (**Figure 1.1**) (Wang *et al.,* 2006). Ser^{201} and His^{254} that form the Ser-His catalytic dyad are

located in TM4 (α4), and TM6 (α4), respectively (Wang *et al.,* 2006; Wu *et al.,* 2006). The catalytic dyad is embedded about 10 Å below the membrane surface (Wang *et al.,* 2006; Wu *et al.,* 2006). Since proteolysis involves hydrolysis of the peptide bond, water is accessible to the catalytic dyad in the lipid bilayer *via* a central cavity formed by the TM helices (Wang *et al.,* 2006; Wu *et al.,* 2006). The L5 loop that connects TM5 (α5) and TM6 (α6) is flexible and was proposed to act as a cap for regulating water access to the active site (**Figure 1.1**) (Wu *et al.,* 2006). Furthermore, TM5 was shown to be able to move laterally and hence was proposed to be the gate for substrate entry to the active site (Wu *et al.,* 2006).

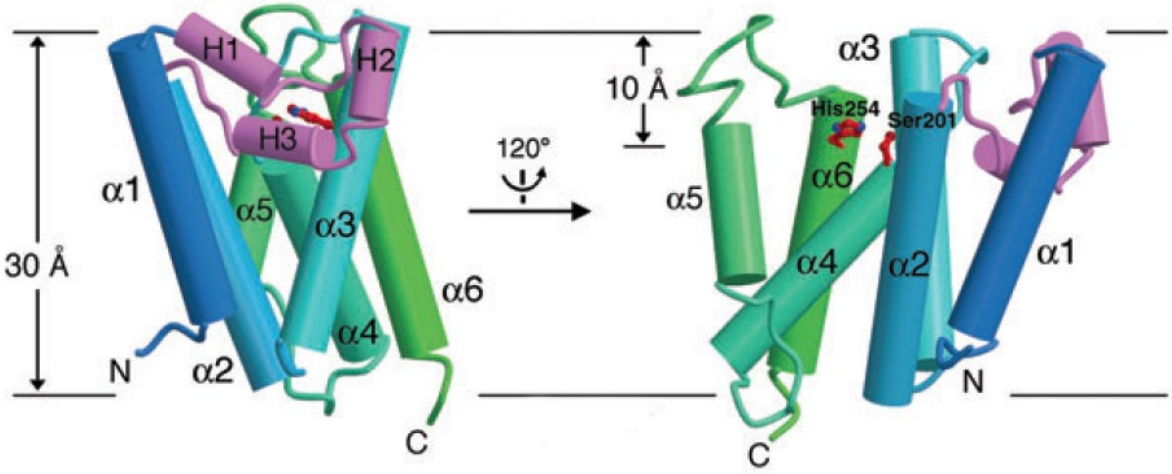


Figure 1.1. Structure of the core domain of *E. coli* GlpG. *TM1-6 are depicted as α1-6. Three α-helices that form the L1 loop (in purple) are designated as H1-3. The catalytic dyad formed by* Ser^{201} *and* His^{254} *(in red) is embedded within the lipid bilayer.*

In addition to structural information, biochemical analyses have yielded insights into the mechanism of rhomboid proteolysis. Strisovsky *et al.* performed mutagenesis of 14 residues (P7-P7') surrounding the cleavage site of a known rhomboid substrate, TatA, and tested its cleavage by the rhomboid AarA from *Providencia stuartii*. Three positions (P1, P4, and P2') were shown to be critical in defining the cleavage site (**Figure 1.2**). The P1 residue can only

be a small, aliphatic residue and this requirement extends to rhomboids from divergent evolutionary backgrounds, suggesting some conserved features in the rhomboid fold might influence substrate specificity (Strisovsky *et al.,* 2009).

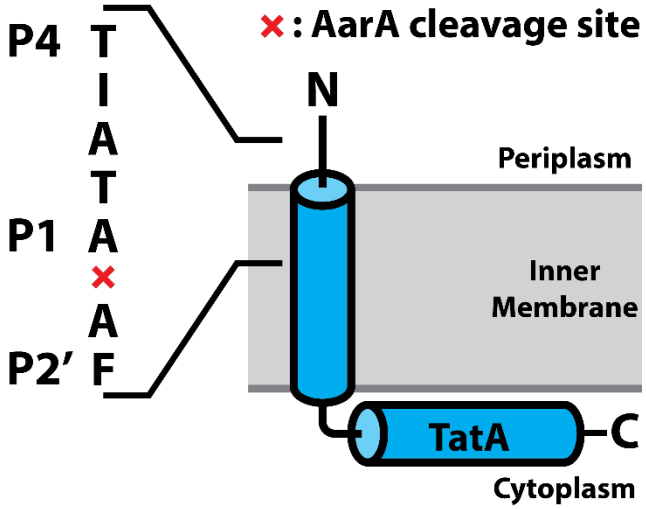


Figure 1.2. Positions of P4, P1, and P2' residues of *P. stuartii* TatA relative to the AarA cleavage site.

Using an *in vitro* system consisting of FITC-TatA as an artificial substrate, the thermodynamic and kinetic features of GlpG have also been determined (Dickey *et al.,* 2013). Strikingly, GlpG displays no physiological affinity to the substrate with an extremely high K_M of around 135 μM, meaning the substrate concentration would need to be one TM segment per 350 phospholipids to achieve half V_{max} of GlpG, which is non-physiologically high. GlpG-mediated intramembrane proteolysis is slow as the k_{cat} of GlpG *in vitro* is roughly 0.006 cuts per second, meaning a single cleavage event only occurs every 2.5 minutes when GlpG is saturated with substrate; the rate-limiting step is gate opening of the enzyme within the membrane rather than hydrolysis (Dickey *et al.,* 2013). A high K_M was observed for nine other rhomboids with low sequence similarity, suggesting that low affinity to substrate is an evolutionarily conserved feature for rhomboids (Dickey *et al.,* 2013). In

contrast, the k_{cat} values of the rhomboids were markedly different (by up to 10,000-fold) (Dickey *et al.,* 2013). These studies led to the conclusion that rhomboid proteolysis is not thermodynamically-governed, but is a kinetically-driven reaction, and the specificity (defined by k_{cat}/K_M) of rhomboid proteolysis is entirely k_{cat}-dependent (Dickey *et al.,* 2013). The low catalytic efficiency of rhomboids is in agreement with their conserved His-Ser catalytic dyad, which is less efficient than the His-Ser-Asp triad commonly found in soluble serine proteases (Dickey *et al.,* 2013).

A wide range of substrate-mimicking peptide inhibitors of rhomboids have provided insights into the molecular mechanisms of rhomboid catalysis (Cho *et al.,* 2016; Vinothkumar *et al.,* 2013; Zoll *et al.,* 2014). Using a peptide aldehyde inhibitor mimicking a *Drosophila* rhomboid substrate Gurken, Cho *et al.* postulated that GlpG-mediated intramembrane proteolysis follows a two-stage catalytic mechanism (Cho *et al.,* 2016). Essentially, in a proteolytic event, GlpG can exist in two forms: the interrogation complex and the scission complex (**Figure 1.3**). The interrogation complex (the Michaelis complex in kinetic terms) is formed after a substrate TMD binds laterally to a gate-open GlpG (Dickey *et al.,* 2013; Urban and Baker, 2008). To transition from the interrogation complex to the scission complex, unwinding of the substrate TMD is required and this unwinding is driven by the helix-destabilising residues (**Figure 1.3**) (Cho *et al.,* 2016; Moin and Urban, 2012). The oxyanion of the catalytic tetrahedral intermediate is stabilised by unusual tripartite interactions with the side chains of His^{150}, Asn^{154}, and the backbone of Ser^{201} (Cho *et al.,* 2016). Interestingly, the structure of GlpG covalently linked to the peptide aldehyde inhibitor also revealed information on GlpG substrate specificity compared to *P. stuartii* AarA. Both rhomboids prefer a small aliphatic residue in the P1 position (Cho *et al.,* 2016; Strisovsky *et al.,* 2009). However, in contrast to AarA which prefers hydrophobic P4

residues, GlpG only has mild preference for hydrophobic P4 residues (Cho *et al.,* 2016). Moreover, GlpG interacts extensively with P2 and P3 residues, for which AarA displays no restrictions (Cho *et al.,* 2016; Strisovsky *et al.,* 2009).

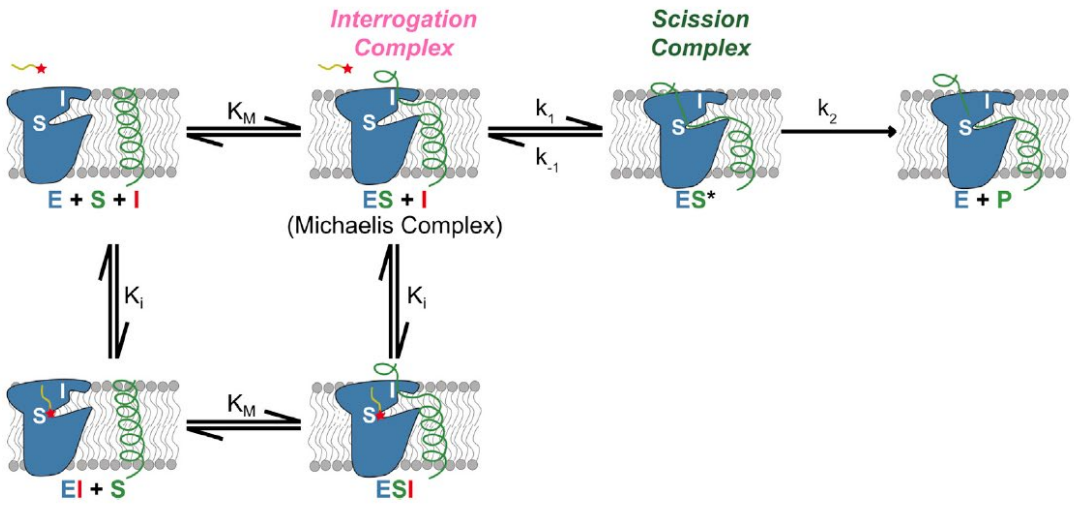


Figure 1.3. Proposed kinetic scheme of GlpG-mediated intramembrane proteolysis and its inhibition by a susbtrate-mimicking peptide aldehyde inhibitor. *GlpG (E) is drawn in blue, with the uncleaved substrate TMD (S) and the cleaved substrate TMD (P) in green, and the substrate-mimicking peptide aldehyde inhibitor (I in red) depicted as a red star with a yellow tail. The scission site (white S) in GlpG marks the catalytic centre and the inhibitor binding site. The interrogation site (white I) indicates the peripheral gate-open site to which substrate TMDs first dock. This figure is adapted from (Cho et al., 2016).*

However, a significant limitation shared by all these studies was the use of substrates known for other rhomboids and/or peptide inhibitors derived from these substrates due to the fact that physiological substrates of GlpG still remain enigmatic.

1.1.3. Biological functions of prokaryotic rhomboid proteases

Despite the extensive work on prokaryotic rhomboid structure and enzymology, surprisingly little is known about their physiological roles. *Providencia stuartii* is a Gram-negative opportunistic pathogen that affects patients with long-term urinary catheters (Warren, 1986). In *P. stuartii*, the rhomboid AarA is essential for quorum sensing (QS), which allows bacteria to communicate and coordinate gene expression in a population-dependent manner (Miller and Bassler, 2001). In *P. stuartii*, QS is mediated by an extracellular factor prone to protease degradation and controls the expression of a 2'-*N*-acetyltransferase, conferring resistance to aminoglycosides (Rather *et al.,* 1997). AarA is required for the production of the QS-mediating extracellular factor (Rather *et al.,* 1999). Of note, AarA is functionally interchangeable with rhomboid-1 from *Drosophila*, highlighting the conserved role of rhomboid family members in extracellular signalling (Gallio *et al.,* 2002).

The physiological substrate of AarA was identified serendipitously in a study to identify rhomboid-like proteins using fragments of *P. mirabilis* genomic DNA to functionally complement a *P. stuartii* Δ*aarA* mutant (Stevenson *et al.,* 2007). The *tatA* allele from *P. mirabilis* was sufficient to complement Δ*aarA* defects, suggesting that TatA functions downstream of AarA in QS (Stevenson *et al.,* 2007). TatA is a membrane protein component of the Twin-arginine translocation (Tat) system, which exports fully-folded proteins with a twin-arginine motif in their signal peptide (Palmer and Berks, 2012). *P. stuartii* TatA is synthesised as a precursor with an N-terminal extension, which is absent in *P. mirabilis* TatA (Stevenson *et al.,* 2007). AarA activates TatA by removing the N-terminal extension (Stevenson *et al.,* 2007). A model was therefore proposed in which AarA mediates QS directly by processing TatA, thereby allowing the translocation of either the extracellular

signal itself or factors required to generate this signal outside the inner membrane (Stevenson *et al.,* 2007).

In *Bacillus subtilis*, the rhomboid YqgP is encoded in the glucokinase operon consisting of *yqgP*, *yqgQ*, and *glcK*, where *glcK* encodes glucokinase (Skarlatos and Dahl, 1998). A Δ*yqgP* mutant of *B. subtilis* exhibits cell division defects and lower sporulation frequency compared to the wild-type (Mesak *et al.,* 2004). Moreover, the *yqgP* mutant is unable to export glucose, suggesting a role for YqgP in glucose metabolism, consistent with its chromosomal location in the glucokinase operon (Mesak *et al.,* 2004). However, the physiological substrate of YqgP remains to be determined.

Mycobacterium tuberculosis encodes two putative rhomboids Rv0110 and Rv1337, both of which are conserved in *Mycobacteria*, including the non-pathogenic *Mycobacterium smegmatis* (Kateete *et al.,* 2012; Kateete *et al.,* 2010). *M. smegmatis* Rv1337 is able to complement the Δ*aarA* QS-defect in *P. stuartii* whereas Rv0110 is not (Kateete *et al.,* 2012). Moreover, a strain of *M. smegmatis* lacking the gene encoding the Rv1337 has decreased ability to form biofilms and increased susceptibility to certain antimicrobials (Kateete *et al.,* 2012). To date, studies on the role of rhomboids in pathogenic mycobacteria are still lacking, and similar to YqgP, the physiological substrates of Rv0110 and Rv1337 are unknown.

The biological function of GlpG in *E. coli* remains enigmatic. A high-throughput phenotypic screen (Biolog) has been performed on *E. coli* lacking GlpG, and the only phenotype revealed was a slight increase in resistance to the β-lactam antibiotic cefotaxime (Clemmer *et al.,* 2006). More recently, a *glpG* mutant of extraintestinal pathogenic *E. coli* (ExPEC)

was shown to be attenuated during intestinal colonisation in a murine model, although the underlying mechanism remains unknown (Russell *et al.,* 2017).

From the perspective of substrates, GlpG was also shown to complement the Δ*aarA* phenotype in *P. stuartii* (Clemmer *et al.,* 2006). Interestingly, *P. stuartii* TatA only functionally complements *E. coli* TatA in the presence of GlpG (Fritsch *et al.,* 2012). Moreover, GlpG was demonstrated to cleave truncated versions of *E. coli* MdfA, a multi-pass integral membrane protein both *in vitro* and *in vivo*. However, full-length MdfA is not cleaved by GlpG, suggesting GlpG might target unfolded multi-pass membrane proteins (Erez and Bibi, 2009).

1.2. Overview of *Shigella* spp. and shigellosis

Shigella is a genus of Gram-negative, facultative anaerobic, non-spore-forming, non-motile bacilli belonging to the family Enterobacteriaceae. *Shigella* spp. are human-specific pathogens causing shigellosis, a severe form of diarrhoea caused by bacterial invasion of the host colonic mucosa. *Shigella* are highly adapted to the human host as illustrated by the low infectious dose; as few as 10-500 colony-forming units (cfu) of *Shigella* spp. can cause symptomatic infection (Dupont, 1989). The genus *Shigella* consists of four species: *Shigella boydii, Shigella dysenteriae, Shigella flexneri,* and *Shigella sonnei* (Kotloff *et al.,* 2017). The best-characterised member of the *Shigella* genus is *S. flexneri*.

The virulence of all *Shigella* spp. critically depends on the Type III Secretion System (T3SS), which is a syringe-like protein complex that injects a wide range of effector proteins into host cells (Abrusci *et al.,* 2013; Sansonetti *et al.,* 1982). In *Shigella*, the T3SS is composed of several modules including the export apparatus (EA), the inner membrane ring, the outer membrane ring, the needle, and the needle tip complex. Each of these modules consists of multiple protein subunits (**Figure 1.4**) (Shen and Blocker, 2016). Contact with the eukaryotic host cell membrane initiates secretion, generating pores formed by IpaB and IpaC in the host target cell membranes. Unfolded protein effectors are translocated directly from the bacterial cytoplasm into eukaryotic cells through the needle and the pore complex **(Figure 1.4)** (Magdalena *et al.,* 2002; Shen and Blocker, 2016). *Shigella* require the T3SS to both enter cells and to escape from vacuoles once inside host cells, enabling them to replicate in the host cytosol (Ray *et al.,* 2010). Therefore, the T3SS system is crucial for invasion, manipulation of host cell physiology, and host innate immunity evasion (Phalipon and Sansonetti, 2007).

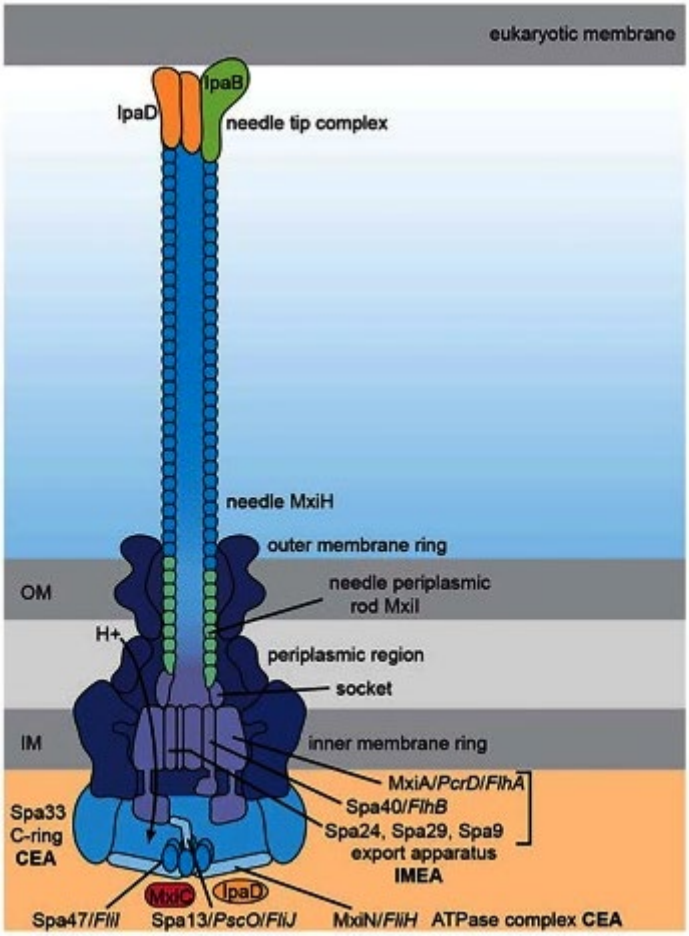


Figure 1.4. Schematic representation of the *Shigella* T3SS. *Needle components in blue and green, transmembrane base dark purple, inner membrane export apparatus (IMEA) light purple, cytoplasmic export apparatus (CEA) components (C-ring and ATPase complex) in light blues. A curved arrow shows the proposed direction of proton flow during protein export, which may be coupled to ATP hydrolysis. T3SS components in Shigella are designated in standard font with their homologues in Pseudomonas and/or Salmonella in italics. This figure is adapted from (Shen and Blocker, 2016).*

Another major virulence factor of *Shigella* is the lipopolysaccharide (LPS). LPS from enteric bacteria consists of three parts: lipid A, an inner core, and the O-antigen. The O-antigen is subject to glycosylation and acetylation, giving rise to different serotypes of *Shigella* (Kenne

et al., 1978). Modifications on LPS are important for *S. flexneri* colonisation in the rabbit ligated ileal loop model (West *et al.,* 2005). One example of that is illustrated by O-antigen glucosylation, which shortens the length of the LPS molecule by half and facilitates *Shigella* virulence by enhancing T3SS access to the host cell membrane (West *et al.,* 2005).

Shigella spp. share a characteristic form of pathogenesis involving acute inflammatory destruction of host colonic mucosa, leading to diarrhoea containing blood and mucus, also known as shigellosis (Takeuchi *et al.,* 1965). However, the advance of genome sequencing has led to the conclusion that *Shigella* spp. are in fact pathotypes of *Escherichia coli* (Pupo *et al.,* 2000; Zhang and Lin, 2012). Current models propose that *Shigella* evolved from an *E. coli* ancestor by first acquiring a large (≈ 220kb) virulence plasmid (pINV) *via* horizontal gene transfer (HGT), which harbours a pathogenicity island encoding the T3SS and a number of virulence factors that are required for pathogenicity (Sansonetti *et al.,* 1981, 1982). In the case of *Shigella sonnei*, the virulence plasmid also encodes enzymes involved in the biosynthesis of a group 4 capsule (Caboni *et al.,* 2015). This group 4 capsule also exerts significant impact on the virulence of *S. sonnei*. Uncapsulated *S. sonnei* are hyperinvasive but display reduced resistance to complement-mediated killing, and decreased ability to spread from the gut to peripheral organs in a rabbit ligated ileal loop model (Caboni *et al.,* 2015).

In addition to the virulence plasmid, the *Shigella* chromosome contains two pathogenicity islands (PAIs). SHI-1 encodes *Shigella* enterotoxin 1 (ShET1) as well as a cytopathic protease and a mucinase (Al-Hasani *et al.,* 2001; Rajakumar *et al.,* 1997) . SHI-2 contains the *iucABCD* operon, which is involved in the synthesis and transport of the iron siderophore aerobactin, in addition to conferring immunity against colicins I and V (Vokes *et al.,* 1999).

1.2.1. Clinical manifestations of *Shigella* infection

The faecal-oral route is the predominant mode of transmission for *Shigella* spp. (Kotloff *et al.,* 2017). Shigellosis typically has an incubation period from one to four days, except for *S. dysenteriae* type 1, which has an incubation period up to eight days (Levine *et al.,* 1973). The first symptoms of shigellosis usually include malaise, headache, fever, anorexia, and vomiting, followed several hours later by watery diarrhoea. In healthy individuals, symptoms are usually mild and abate in a few days. In other cases, disease progresses to dysentery within hours to days, manifested by frequent watery stools containing mucus and blood, along with lower abdominal cramp (Kotloff *et al.,* 2017). On rare occasions, shigellosis induces severe intestinal complications, including rectal prolapse, toxic megacolon, intestinal obstruction, and perforation. Complications are more commonly associated with *S. dysenteriae* type 1 (Avital *et al.,* 1982). Moreover, *Shigella* occasionally traverse the gasterointestinal tract, and cause arthritis, osteomyelitis, meningitis, and splenic abscess.

1.2.3. Epidemiology

In 2010, approximately 188 million cases of shigellosis occurred worldwide, including 62.3 million cases in children aged <5 years (Kotloff *et al.,* 2017). In terms of mortality, *Shigella* is the second most common cause of diarrhoeal death after rotavirus, responsible for around 164,300 deaths every year, 54,900 of which are among children aged <5 years (Troeger *et al.,* 2017). In addition to children aged <5 years, two other populations are at risk of shigellosis (Kotloff *et al.,* 2017). Men who have sex with men (MSM) as an epidemiological niche for *Shigella* emerged in the 1970s and was best illustrated by an outbreak caused by *S. flexneri* which spread intercontinentally within MSM (Baker *et al.,* 2015). Travellers represent the second at-risk population. Around 536 million people travel to low-income

and middle-income countries annually and 10–40% develop diarrhoea, of which around 2–9% is caused by *Shigella* based on culture (Shah *et al.,* 2009).

S. dysenteriae was the first member of the genus identified in 1897, but nowadays this species is rarely isolated from patients (Havelaar *et al.,* 2015; Qiu *et al.,* 2015). *S. flexneri* currently causes most cases of bacillary dysentery in low-income settings accounting for up to 62% of all *Shigella* infections in parts of Asia and sub-Saharan Africa. *S. flexneri* has been confined to the endemic regions with limited global spread and has acquired antimicrobial resistance traits on multiple independent occasions over the past 30 years (Connor *et al.,* 2015). In contrast, *S. sonnei* emerged from Europe in the late 1900s, spread globally *via* travellers (Holt *et al.,* 2012) and is now the prevailing *Shigella* species in transitional or high-income countries responsible for up to 80% of all *Shigella* infections in Europe and North America (Havelaar *et al.,* 2015).

Several explanations have been proposed for the dominance of *S. sonnei* in high-income settings. Firstly, people from countries with poor sanitation are more likely to have immunity against *S. sonnei* due to exposure to water contaminated with *Pleisiomonas shigelloides* O17 (Sack *et al.,* 1994), which possesses an O-antigen capsule almost identical to that of *S. sonnei* (Shepherd *et al.,* 2000). Second, *S. sonnei* can evade chlorination and other sanitation processes by surviving intracellularly in the free-living amoeba *Acanthamoebae castellanii*; *S. flexneri* kills *A. castellanii* upon phagocytosis and hence cannot exploit this protective niche (Jeong *et al.,* 2007). Third, compared to *S. flexneri, S. sonnei* more frequently carries mobile genetic elements conferring antimicrobial resistance (Thompson *et al.,* 2015). Finally, *S. sonnei*, but not *S. flexneri*, possesses a chromosomal Type six secretion system (T6SS), which allows the bacterium to kill closely related enteric bacteria competing for the

same niche in gut (Anderson *et al.,* 2017). Thus, *S. sonnei* could out-compete *S. flexneri* in regions where both species circulate.

1.3. *Shigella* pathogenesis

The *Shigella* infection cycle consists of (i) survival in the stomach and small intestine prior to reaching the colonic epithelium, (ii) crossing the colonic epithelial barrier and encountering macrophages, (iii) intracellular replication and intercellular spread within the colonic epithelial cells, and (iv) clearance and dissemination to the environment. Each of these stages will be discussed in the following sections.

1.3.1. Survival in the stomach and small intestine

Shigella encounters three major challenges during its passage through the gastrointestinal tract; the acidic environment in the stomach, bile salts in the small intestine, and host microbiota in the colon.

Gastric acid in the stomach and volatile fatty acids produced within the intestine are two major sources of acid stress that enteric bacteria encounter (Giannella *et al.,* 1972). *Shigella* has evolved several strategies to counter acid stress. The stationary-phase sigma factor RpoS plays a crucial role in acid resistance in *S. flexneri* by inducing expression of several genes that enhance acid resistance (AR) (Kanjee and Houry, 2013). This includes the periplasmic ATP-independent chaperones such as HdeA and HdeB (Waterman and Small, 1996), and the cytoplasmic glutamic-acid dependent acid resistance (GDAR) pathway (Kanjee and Houry, 2013; Waterman and Small, 1996). HdeA and HdeB form dimers at neutral pH but transform into partially unfolded monomers at low pH. HdeA and HdeB bind to acid-

denatured substrate proteins and release substrate proteins in a refolding-competent conformation upon pH increases (Kern *et al.*, 2007; Tapley *et al.*, 2010). The GDAR pathway increases cytoplasmic pH by converting glutamate into GABA and CO_2 at the expense of protons (Capitani *et al.,* 2003; Lin *et al.,* 1995). In addition to these AR systems, hydrogenase-1, a multicomponent membrane complex that oxidises H_2 to protons in the periplasm, has also been implicated in acid survival of *S. flexneri*. A *S. flexneri* mutant lacking hydrogenase-1 is more sensitive to acid stress than a mutant in the GDAR pathway. The mechanism underlying hydrogenase 1-dependent acid resistance is not entirely understood but it was proposed that protons generated in the periplasm might reduce proton influx from the environment (McNorton and Maier, 2012).

In the small intestine, bile is an antimicrobial detergent-like mixture of compounds (Merritt and Donaldson, 2009; Sistrunk *et al.,* 2016). In the case of *S. flexneri*, bile resistance mainly stems from the AcrAB efflux pump, which actively pumps out bile salts; the O-antigen of LPS, which reduces bile salts penetration of the outer membrane; and biofilm formation, which restricts bacterial exposure to bile salts (Nickerson *et al.,* 2017). Exposure to bile salts such as deoxycholate also primes *S. flexneri* for adherence and invasion (Nickerson *et al.,* 2017; Pope *et al.,* 1995). More specifically, exposure to bile salts upregulates the *ospE1/ospE2* genes, which encode two T3SS effectors. OspE1 and OspE2 localise on the outer membrane of *S. flexneri*, and mediate enhanced adherence to epithelial cells (Faherty *et al.,* 2012). Furthermore, binding of deoxycholate to IpaD can enhance epithelial cell invasion. Binding induces a conformational change in IpaD that leads to recruitment of IpaB to the needle tip, facilitating the formation of the needle tip complex required for T3SS secretion, leading to enhanced invasion (Barta *et al.,* 2012; Dickenson *et al.,* 2011; Stensrud *et al.,* 2008).

An essential function of a healthy and stable microbiota is to provide the host with colonisation resistance against invading pathogens (Vollaard and Clasener, 1994). Perturbation of the gut microbiota by streptomycin treatment significantly increases murine susceptibility to *Salmonella* infection (Bohnhoff *et al.,* 1954). Similar observations have been made with other pathogens, such as *Clostridium difficile* and *Vibrio cholerae* (Bassis *et al.,* 2014; Freter, 1955). A variety of mechanisms contribute to colonisation resistance, including direct antagonism of pathogens by the microbiota, pathogen exclusion through occupancy of overlapping gut niches, and indirect killing of pathogens *via* induction of host immune responses (Mullineaux-Sanders *et al.,* 2018). Direct antagonism can occur through contact-dependent killing *via* type VI secretion systems (T6SS) (Hood *et al.,* 2010). T6SS-mediated interbacterial antagonism has been demonstrated in multiple Gram-negative pathogens including *Salmonella* Typhimurium and *V. cholerae,* as well as commensals from the *Bacteroides* genus (Chatzidaki-Livanis *et al.,* 2016; Sana *et al.,* 2016). On the other hand, T6SS can also be employed by pathogens such as *S. sonnei* to overcome colonisation resistance at least partially, by killing *E. coli* and *S. flexneri* competitors in a murine colonisation model (Anderson *et al.,* 2017).

1.3.2. Interactions with microfold cells (M cells) and macrophages

The lumen of the colon is generally considered anaerobic. In this environment, *S. flexneri* is primed for invasion by assembling its T3SS but in an inactive state (Marteyn *et al.,* 2010). This is due to repressed transcription of *spa32* and *spa33*, which encode T3SS components key to regulation of secretion, by FNR (fumarate and nitrate reduction), an oxygen-sensing transcription factor (Marteyn *et al.,* 2010). However, as *S. flexneri* reaches the mucosal surface of the colon, a relatively aerobic zone due to local vascular supply, the expression

of *spa32* and *spa33* is induced, allowing secretion and consequent invasion **(Figure 1.5)** (Marteyn *et al.,* 2010).

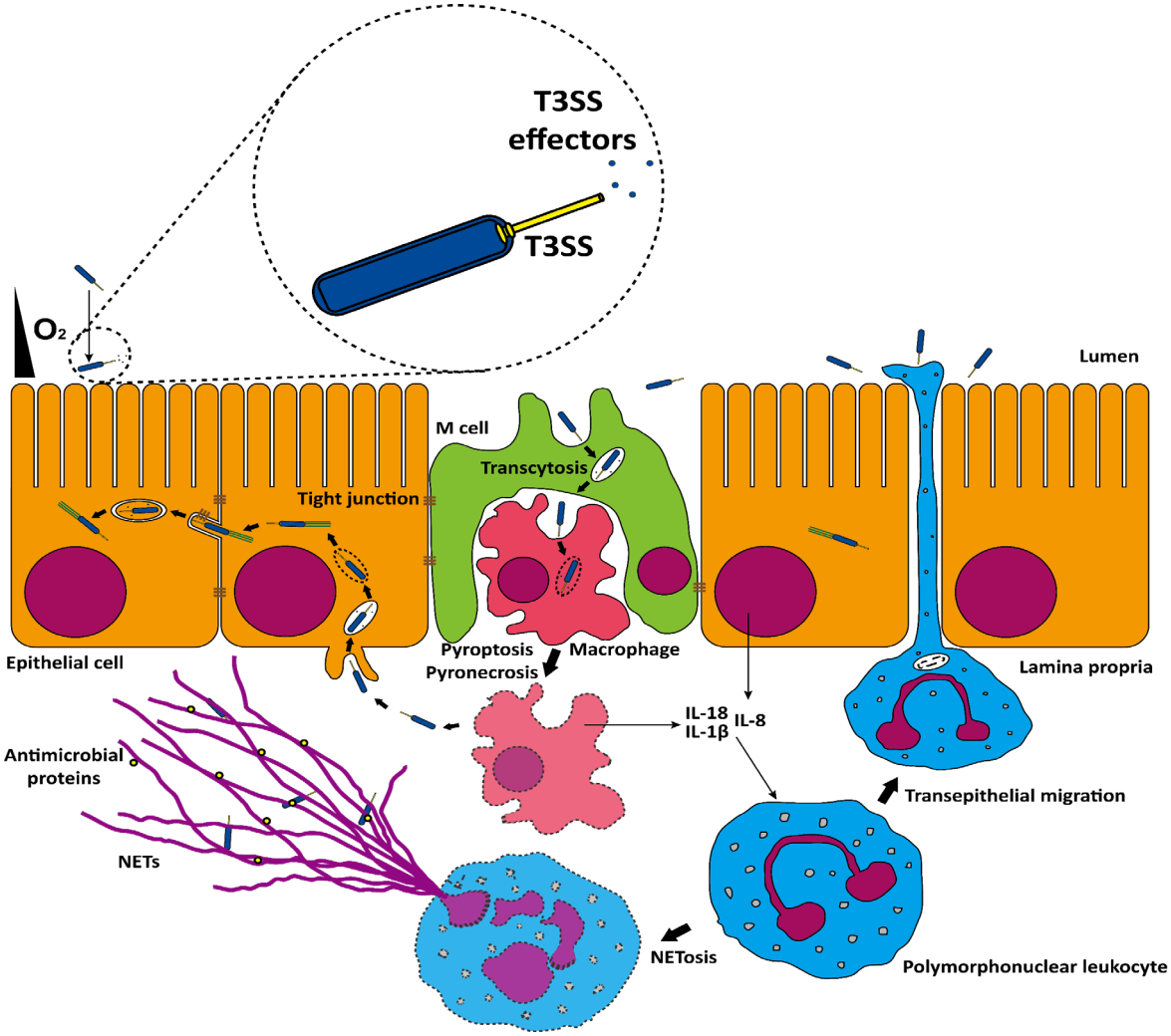


Figure 1.5. Stages of *S. flexneri* infection upon reaching the colonic epithelium. *S. flexneri are depicted as blue rods with assembled T3SS in yellow. T3SS-secreted factors are shown as blue circles. S. flexneri spread intra- and intercellularly by forming actin tails (shown as green lines). Colonic epithelial cells are shown in orange, M cell in green, macrophage in pink, and polymorphonuclear leukocytes (PMNs)/neutrophils in blue with grey granules. Neutrophil extracellular traps (NETs) are shown as purple webs decorated with antimicrobial proteins shown as yellow circles. Nuclei in all cell types are in purple. This Figure is adapted from Guangyu Liu's Part II Dissertation 439062.*

Using an *icsA* mutant deficient in intercellular spread, *S. flexneri* invasion was pinpointed to M cells, a specialised form of epithelial cells acting as sentinels by taking up antigens and microorganisms from the gut lumen and delivering them to antigen presenting cells (APCs) in the lamina propria (Neutra *et al.,* 1996; Owen and Jones, 1974; Sansonetti and Phalipon, 1999). *S. flexneri* that transcytose through M cells are phagocytosed by macrophages resident in the lamina propria (Schroeder and Hilbi, 2008). However, IpaB, a T3SS effector, induces phagosomal rupture by forming an ion channel on the phagosomal membrane, allowing *S. flexneri* to escape the phagosome and reach the cytoplasm of the macrophages (High *et al.,* 1992; Senerovic *et al.,* 2012). Inside the cytoplasm, *S. flexneri* is detected by host pattern recognition receptors (PRRs) (**Figure 1.5**). NLRC4 (NLR family CARD-containing protein 4) and NLRP3 (NLR family, pyrin domain containing 3) are cytosolic PRRs that belong to the NOD-like receptor (NLR) family. NLRC4 recognises a sequence motif present on flagellin, a highly conserved pathogen-associated molecular pattern (PAMP). As *S. flexneri* does not express flagellin, cytoplasmic bacteria are detected by a related sequence present in MxiI, a rod component of T3SS, through Naip2, another NLR family member (Suzuki *et al.,* 2014). Naip2 in turn recruits NLRC4 to assemble the NLRC4 inflammasome (Suzuki *et al.,* 2014). This activates caspase-1 *via* and results in pyroptosis and maturation of IL-1β and IL-18, two cytokines crucial for the characteristic inflammation of shigellosis (**Figure 1.6**) (Sansonetti *et al.,* 2000; Suzuki *et al.,* 2014; Suzuki *et al.,* 2007). NLRP3 is also implicated in macrophage responses to *Shigella*, although the exact PAMP recognised by this PRR remains unknown. First, NLRP3 assembles the NLRP3 inflammasome which also mediates IL-1β maturation (Willingham *et al.,* 2007). At later time points (6 hours after initiation of infection), it triggers an ASC (Apoptosis-associated Speck-like protein containing a CARD)-dependent, but caspase-1 and inflammasome-independent, form of cell death known as pyronecrosis (**Figure 1.6**) (Rathinam *et al.,* 2012;

Sander *et al.,* 2011; Willingham *et al.,* 2007). Both forms of macrophage cell death allow *Shigella* to escape macrophage killing, and subsequently to invade colonic epithelial cells from the basolateral side (**Figure 1.5**).

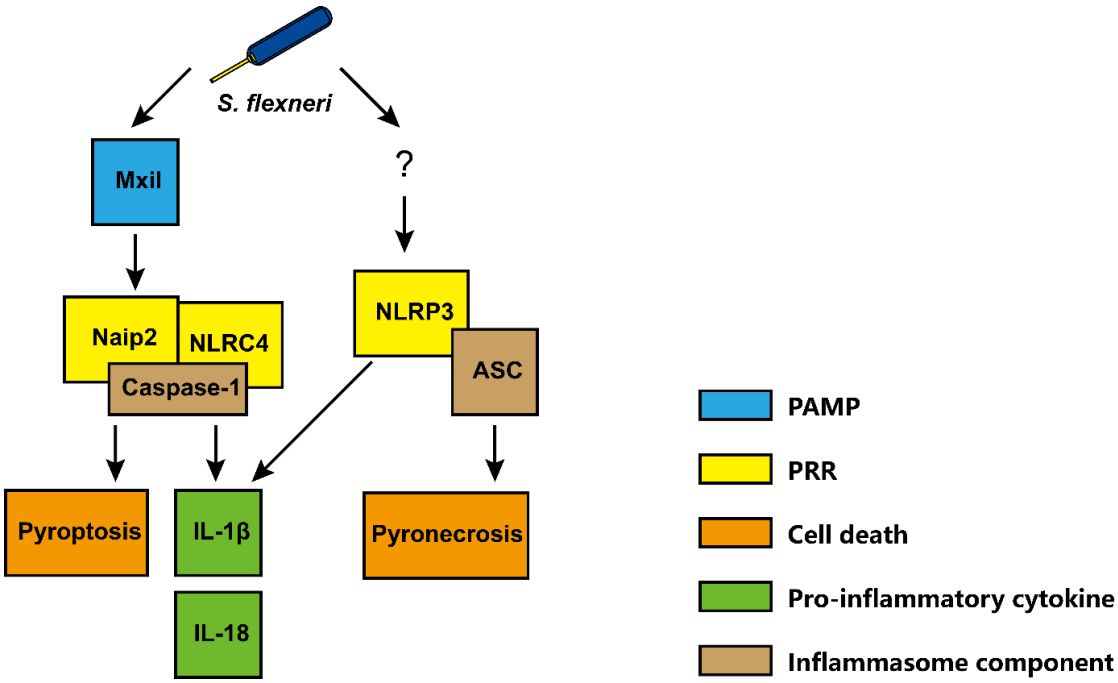


Figure 1.6. Detection of and responses to *S. flexneri* in a macrophage. *In the cytoplasm of macrophages, S. flexneri is detected by host PRRs (in yellow) through recognition of Shigella-derived PAMPs (in blue). PAMP-bound PRRs recruit other inflammasome componnents (in brown), leading to assembly of NLRC4 and NLRP3 inflammasomes, and ultimately resulting in release of pro-inflammatory cytokines (in green) and macrophage cell death pathways (in orange). This Figure is adapted from Guangyu Liu's Part II Dissertation 439062.*

1.3.3. Intracellular replication and intercellular spread

S. flexneri invasion of the colonic epithelium from the basolateral side occurs in a T3SS-dependent manner (Cossart and Sansonetti, 2004) **(Figure 1.5)**. First, IpaB and IpaC form a pore at the epithelial cell membrane, permitting secretion of T3SS effectors into the cell (Tran Van Nhieu *et al.,* 1999). Secreted IpaA then binds to vinculin to induce F-actin depolymerisation, recycling actin for the formation of membrane ruffles required for bacterial engulfment (Bourdet-Sicard *et al.,* 1999). Furthermore, IpgB1, a GTP-exchange factor (GEF) secreted by the T3SS, localises to the host plasma membrane and binds to ELMO (engulfment and cell motility) and exploits the RhoG-ELMO-Dock180 pathway to enhance membrane ruffling (Costa and Lesser, 2014; Handa *et al.,* 2007; Huang *et al.,* 2009). Following entry into the epithelial cell, *S. flexneri* escape the phagocytic vacuole in less than 10 minutes (Ray *et al.,* 2010), and enter the cytosol in an IpaB-dependent manner (High *et al.,* 1992), and replicate in the cytosol. *Shigella* can then disseminate intra- and intercellularly *via* a surface autotransporter IcsA (Bernardini *et al.,* 1989; Egile *et al.,* 1999). IcsA recruits the neural Wiskott–Aldrich syndrome protein (N-WASP), and consequently the Arp2/3 complex, which mediates the nucleation of F-actin to form a tail to propel *Shigella* in the host cytosol **(Figure 1.5)** (Pollard and Borisy, 2003; Suzuki *et al.,* 1998). The host cell senses vacuole lysis through galectin-8, which recognises host glycans that become exposed due to membrane rupture and recruits a cargo receptor NDP52 to initiate autophagy (Thurston *et al.,* 2012). Autophagy is usually an effective mechanism to target bacteria to autolysosomes for degradation (Huang and Brumell, 2014). However, in the case of *S. flexneri*, although IcsA is recognised by Atg5, a protein crucial for autophagosome formation (Mizushima and Komatsu, 2011), the bacterium secrete IcsB, which binds to IcsA at the same region as Atg5, shielding IcsA from Atg5 recognition (Deretic and Levine, 2009; Kayath *et al.,* 2010; Ogawa *et al.,* 2005). In addition, IcsB also recruits transducer of

CDC42-dependent actin assembly 1 (Toca-1), prohibiting the recruitment of the NDP52 and LC3 adaptor proteins that are crucial for autophagosome formation (Baxt and Goldberg, 2014). Finally, IcsB inhibits septins, which are host GTP-binding proteins that trigger autophagy by forming cage-like structures around bacteria along with polymerised actin (Mostowy *et al.,* 2010; Sirianni *et al.,* 2016).

The epithelial cell cytosol serves as a replicative niche for *S. flexneri*. *Shigella* are able to use pyruvate as the major energy source to proliferate rapidly with a generation time of approximately 37 ± 4 minutes (Kentner *et al.,* 2014). On the other hand, the cytosol is also a battlefield where a plethora of host and bacterial factors form a complex antagonistic network **(Figure 1.7)**. Nucleotide-binding oligomerisation domain (NOD) proteins NOD1 and NOD2 respond to peptidoglycan (PGN) released from intracellular *S. flexneri* and initiate nuclear factor-κB (NF-κB) -dependent and mitogen-activated protein kinase (MAPK)-dependent inflammation (Nigro *et al.,* 2008; Philpott *et al.,* 2014). NODs work in concert with NOD-like receptor NLRX1, and Toll-like receptor 4 (TLR4) which detects bacterial LPS, to amplify the NF-κB pathway (Hoshino *et al.,* 1999; Tattoli *et al.,* 2008). Retinoic acid inducible gene I (RIG-I), which senses 5'-ppp-containing RNA, is induced by interferon γ (IFNγ) and restricts *S. flexneri* intracellular replication *via* an unknown mechanism (Jehl *et al.,* 2012). To enhance intracellular replication, *S. flexneri* reduces host inflammation to delay the recruitment of polymorphonuclear leukocytes (PMNs)/neutrophils. Six T3SS effectors have been reported to inhibit the pro-inflammatory, NF-κB pathway. OspF is a phosphothreonine lyase that irreversibly dephosphorylates MAPKs, preventing histone H3 phosphorylation in a subset of NF-κB-responsive genes (Arbibe *et al.,* 2007; Li *et al.,* 2007). OspG impairs the ubiquitination and hence degradation of IκBα, an inhibitor of NF-κB (Kim *et al.,* 2005). OspI deamidates an E2 ubiquitin-

conjugating enzyme Ubc13, inhibiting TRAF6-NF-κB signalling (Sanada *et al.,* 2012). OspZ is a methyltransferase that methylates a Zinc-coordinating cysteine in TAB2 and TAB3, two ubiquitin chain-sensing proteins involved in nuclear translocation of the p65 subunit of NF-κB (Newton *et al.,* 2010; Zhang *et al.,* 2012). IpaH family proteins are E3 ubiquitin ligases. IpaH4.5 ubiquitinates the p65 subunit of NF-κB (Wang *et al.,* 2013); IpaH9.8 has a dual role. It polyubiquitinates NEMO, which is crucial for NF-κB activation, and inhibits U2AF, a splicing factor crucial for IL-8 expression (Ashida *et al.,* 2010; Okuda *et al.,* 2005). IpgD also acts as a phosphatidylinositol 4,5-bisphosphate phosphatase, producing phosphatidylinositol 5-phosphate to block connexin hemichannels, through which epithelial cells release ATP to induce inflammation (Puhar *et al.,* 2013).

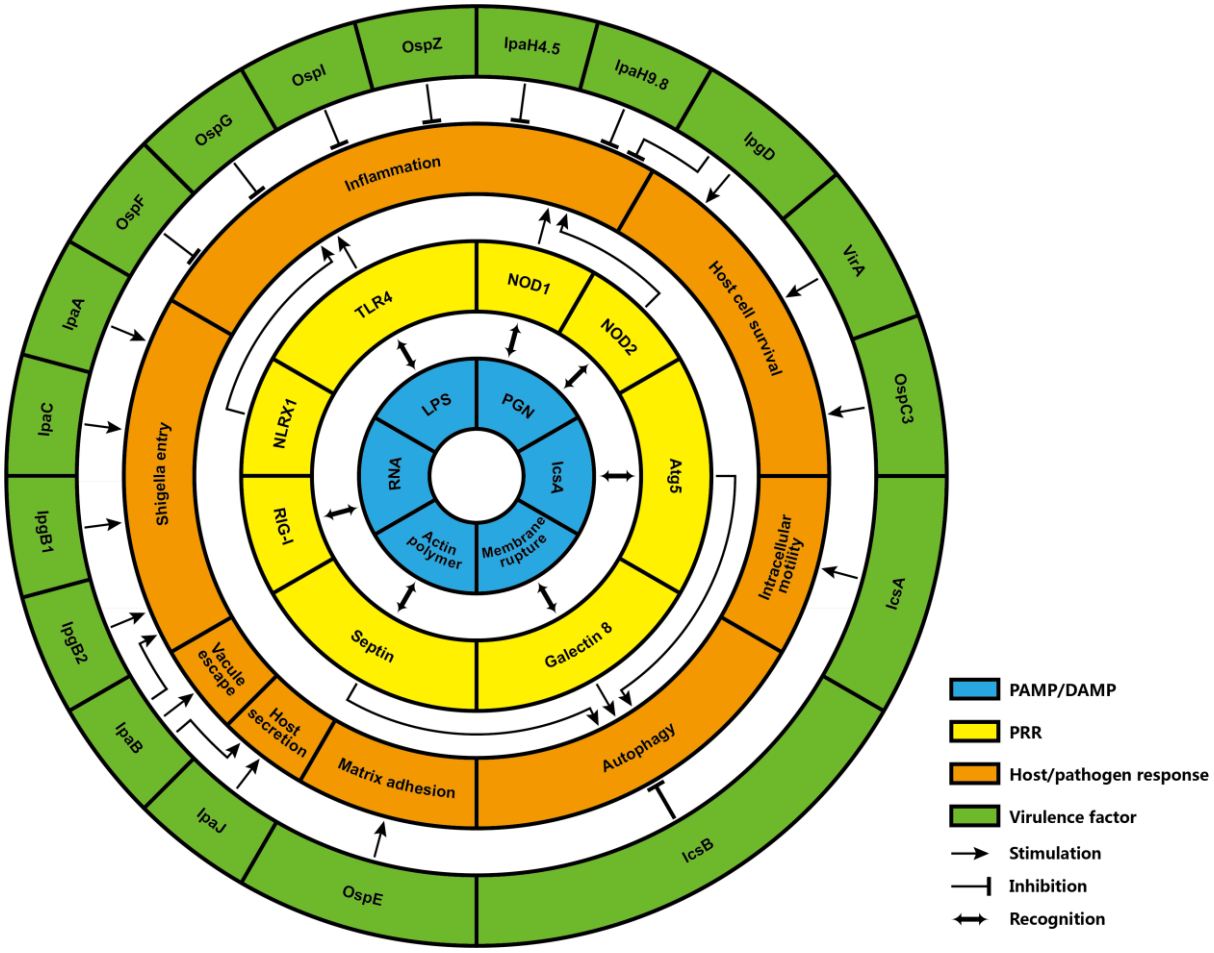


Figure 1.7. *S. flexneri*-host interactions in the cytosol of epithelial cells. *S. flexneri expresses virulence factors (green ring) to subvert host defences. Host detection of S. flexneri relies on the recognition of bacterial danger- and pathogen-associated molecular patterns (DAMPs/PAMPs, blue ring) by host PRRs (yellow ring) (Akira et al., 2006). Host responses (orange ring) to counteract S. flexneri are mainly inflammation and autophagy. This Figure is adapted from Guangyu Liu's Part II Dissertation 439062.*

1.3.4. *Shigella* clearance by neutrophils and environmental dissemination

Shigella-infected epithelial cells markedly increase expression of the potent chemokine, IL-8 (Pedron *et al.,* 2003). Together with IL-1β and IL-18, which are released by macrophages undergoing IpaB-mediated pyroptosis, IL-8 elicits a neutrophil infiltrate to the site of infection (Sansonetti *et al.,* 1999; Singer and Sansonetti, 2004).

Neutrophils play a conflicting role during *Shigella* infection. On the one hand, *S. flexneri* has evolved to bind to the cationic granular proteins released by neutrophils, modifying bacterium's surface charge and rendering it hyperinvasive (Eilers *et al.,* 2010). On the other hand, unlike in macrophages, *S. flexneri* is unable to escape from neutrophil phagosomes, and consequently the bacteria are rapidly killed (Weinrauch *et al.,* 2002). Also, neutrophils can kill *S. flexneri* extracellularly with neutrophil extracellular traps (NETs). NETs consist of processed chromatin, decorated with granular antimicrobial proteins, and arise after PMNs undergo NETosis, a caspase-independent form of cell death characterised by nuclear membrane dissolution before loss of plasma membrane integrity (Brinkmann and Zychlinsky, 2012; Steinberg and Grinstein, 2007). Interactions with neutrophils are generally considered as the final stage of *Shigella* infection in the host as the resultant inflammatory responses ultimately lead to bacterial clearance.

1.3.5. Oxidative stress response in *Shigella* pathogenesis

The molecular basis of oxidative stress stems from the oxidation of proteins, lipids, carbohydrates and nucleic acids by partially reduced reactive oxygen species (ROS) including superoxide (O_2^-), hydrogen peroxide (H_2O_2), and hydroxyl radical HO· (Imlay, 2013). The source of ROS can be internal or external. ROS generated internally through adventitious oxidation of redox enzymes is rendered non-toxic by basal levels of superoxide dismutases (SODs) (Imlay, 2008). More importantly, *Shigella* encounters external ROS at multiple stages of its infection. First, *Shigella* phagocytosed by macrophages or PMNs are attacked by ROS generated by host NADPH oxidase, which catalyses the formation of superoxide (Goldblatt, 2014; Slauch, 2011). In addition, bacteriocidal antibiotics such as β–lactams, quinolones, aminoglycosides have been shown to trigger ROS generation in target bacterial cells and contribute to antibiotic lethality (Belenky *et al.,* 2015; Dwyer *et al.,* 2014). *Shigella* mitigates these ROS insults by mounting defence mechanisms including superoxide dismutases, catalases, and peroxidases (Franzon *et al.,* 1990; Mishra and Imlay, 2012). These mechanisms are crucial for *Shigella* pathogenesis, as exemplified by the fact that *S. flexneri* lacking SodB (an iron-containing superoxide dismutase) is extremely sensitive to macrophage killing and almost avirulent in a rabbit ileal-loop model (Franzon *et al.,* 1990).

1.4. Hydrogenases

1.4.1. Overview of bacterial hydrogenases

In addition to the cellular immune subversion and survival strategies discussed in section 1.3, the fitness of *Shigella* inside a human host is also greatly influenced by its metabolic capacity. In the cytoplasm of a host epithelial cell, *Shigella* exploits pyruvate from host glycolysis as the major energy source by converting it to acetyl-CoA, and ultimately acetate and ATP (Kentner *et al.,* 2014). Whilst in the colon, the majority of the carbohydrate sources are complex polysaccharides that are indigestible by the host. Many members of the human microbiota, such as Bacteroides and Clostridiales, have enzymes dedicated for digesting these complex polysaccharides, however, *Shigella* is not capable of breaking down complex sugars (Kamada *et al.,* 2013; Marteyn *et al.,* 2010). To survive in the colon, Enterobacteriacae, such as *Shigella, Salmonella*, and *E. coli*, utilise a variety of electron acceptors, such as fumarate and nitrate, as well as various alternative electron donors, including formate and hydrogen (Unden *et al.,* 2014). Hydrogen (H_2) gas is abundant in human colon, making up to around 20% of flatus from individuals (Steggerda, 1968). The primary source of colonic H_2 is the re-oxidation of reduced pyridine (NADH) and flavin ($FADH_2$) nucleotides generated from fermentation of dietary substances by the strict anaerobes *Clostridium* spp. and *Bacteroides* spp. (Carbonero *et al.,* 2012; Wolf *et al.,* 2016). H_2 has a negative redox potential (H^+/H_2: $E^{o\prime} = -414$ mV) under standard conditions at pH = 7, and therefore can be classified as a high-energy compound and an excellent energy source (Pinske and Sawers, 2016). *S. flexneri* and *E. coli*, can consume H_2 during respiration using enzymes known as hydrogenases (McNorton and Maier, 2012). Hydrogenases can be divided into two classes, [FeFe]- and [NiFe]-containing, based on the nature of the metal composition at their catalytic sites (Lubitz *et al.,* 2014). In bacteria, [FeFe]-hydrogenases are found in anaerobes and are especially prevalent among the fermentative organisms (*e.g.*

Firmicutes). In contrast, [NiFe]-hydrogenases are frequently present in Cyanobacteria, Proteobacteria, and Actinobacteria (Peters *et al.,* 2015). *E. coli* has three well-characterised hydrogenases: hydrogenase-1, hydrogenase-2, and hydrogenase-3 (**Table 1.1**), which are also conserved in *S. flexneri* and have homologues in *S. sonnei* (section 6.1 in this work) but have not been extensively characterised functionally in either organism (McNorton and Maier, 2012; Pinske and Sawers, 2016). The following sections provide an overview of hydrogenase complex structure and function largely based on studies in *E. coli*.

Complex name	Subunit name/number of TMDs				Complex composition	References
	α	β	γ	δ		
Hydrogenase						
Hyd-1	HyaB/0	HyaA/1	HyaC/4	N/A	$\alpha_2\beta_2\gamma_2$ (predicted)	(Volbeda *et al.,* 2013)
Hyd-2	HybC/0	HybO/1	HybB/10	HybA/1	αβγδ (predicted)	(Beaton *et al.,* 2018)
Formate dehydrogenase						
FDH-N	FdnG/0	FdnH/1	FdnG/4	N/A	$\alpha_3\beta_3\gamma_3$	(Jormakka *et al.,* 2002)
FDH-O	FdoG/0	FdoH/1	FdoG/4	N/A	$\alpha_3\beta_3\gamma_3$ (predicted)	(Abaibou *et al.,* 1995)

Table 1.1. Composition of hydrogenases and formate dehydrogenases in *E. coli*. *Hyd-3 and Hyd-4 are not included in this figure but discussed separately (section 1.4.4 and 1.4.5) as its complex composition bears little resemblance to the enzyme complexes shown in this table.*

1.4.2. Hydrogenase-1 (Hyd-1)

Hydrogenase-1 (Hyd-1) is a periplasm-facing, inner membrane [NiFe]-enzyme encoded by the *hyaABCDEF* operon (**Table 1.1**) (Menon *et al.,* 1990). The operon encodes a catalytic α-subunit HyaB, an electron-transferring β-subunit HyaA which possesses a Tat signal peptide, a membrane-bound γ-subunit HyaC, an endoprotease HyaD, a Tat-translocation

chaperone HyaE, and a protein of unknown function, HyaF (Ballantine and Boxer, 1985; Dubini and Sargent, 2003; Forzi and Sawers, 2007; Sawers and Boxer, 1986). Among subunits of Hyd-1, HyaA is a type I protein and hence could potentially be a rhomboid substrate (Nesbit *et al.,* 2012; Volbeda *et al.,* 2013).

Most hydrogenases are susceptible to losing their catalytic activity upon exposure to O_2. Hyd-1 is therefore unique compared with the other [NiFe]-hydrogenases in *E. coli* as it is O_2-tolerant (Pinske and Sawers, 2016). The physiological function of Hyd-1 remains unknown (Pinske *et al.,* 2012). Hyd-1 poorly couples H_2 oxidation to reduction of fumarate (Laurinavichene and Tsygankov, 2001), which suggests Hyd-1 plays a less significant role in anaerobic respiration than the other hydrogenases. In addition, Hyd-1 does not significantly oxidise H_2 during glucose fermentation (Pinske *et al.,* 2012; Redwood *et al.,* 2008). Moreover, Hyd-1 is co-expressed with the high-affinity cytochrome *bd*-II oxidase (Atlung T and L., 1994; Nesbit *et al.,* 2012). This co-expression, in combination with the high oxygen affinity of cytochrome *bd*-II oxidase, suggests that, Hyd-1 can scavenge O_2 with cytochrome *bd*-II oxidase under *in vivo* fermenting conditions (Nesbit *et al.,* 2012; Volbeda *et al.,* 2013).

1.4.3. Hydrogenase-2 (Hyd-2)

Hydrogenase-2 (Hyd-2) is a periplasm-facing, inner membrane-bound enzyme, encoded by the *hybOABCDEFG* operon that is only expressed under anaerobic conditions (Dubini *et al.,* 2002; Menon *et al.,* 1994; Sargent *et al.,* 1998). Hyd-2 consists of a catalytic α-subunit HybC, an electron-transferring β-subunit HybO (Tat signal-bearing), an extra electron-transferring δ-subunit HybA (Tat signal-bearing, no counterpart present in Hyd-1), a membrane-bound γ-subunit HybB (**Figure 1.8**) (Beaton *et al.,* 2018). Other genes in the operon encode

proteins involved in Hyd-2 biogenesis, including a system-specific metalloprotease HybD, a Tat-translocation 'proof-reading' chaperone HybE, a nickel insertase HybF, and a putative $Fe(CN)_2(CO)$ cofactor insertase HybG (Blokesch *et al.,* 2001; Blokesch *et al.,* 2004; Dubini and Sargent, 2003; Forzi and Sawers, 2007). Structural characterisation combined with homology modelling of the Hyd-2 complex suggested that it exists as a HybOABC dimer *in vivo* (**Figure 1.8**) (Beaton *et al.,* 2018). HybO and HybA are both type I membrane proteins (Beaton *et al.,* 2018), fulfilling the criteria for putative rhomboid substrates.

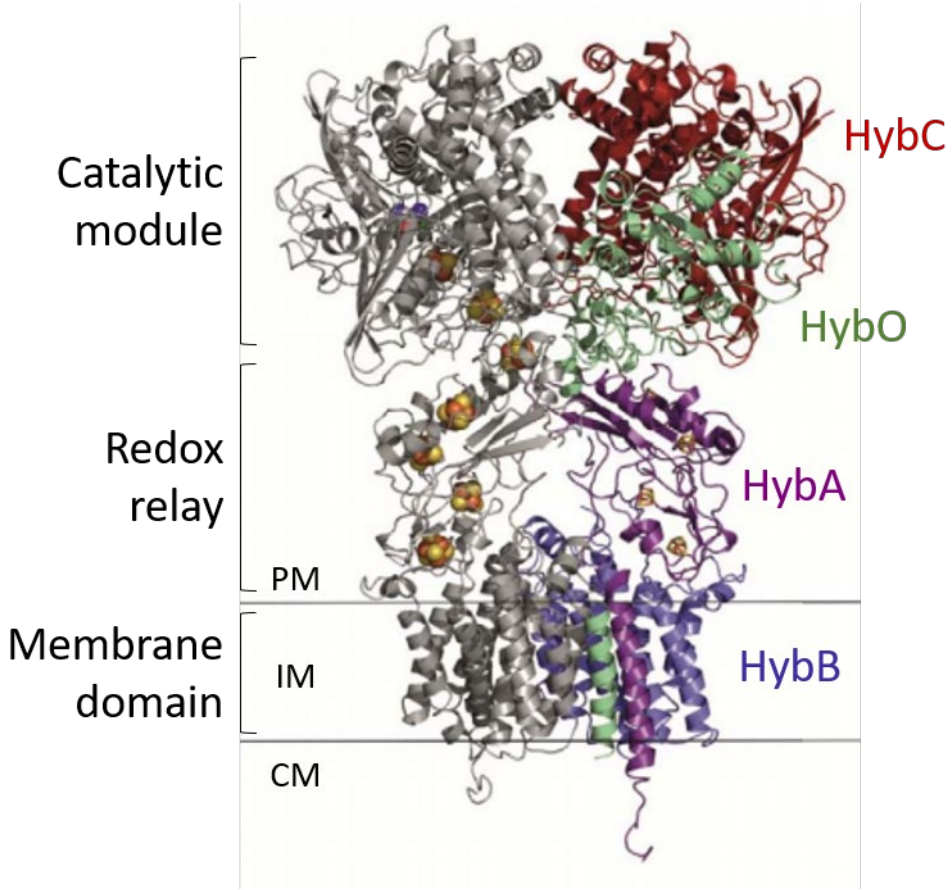


Figure 1.8. Proposed quaternary structure of a dimer of *E. coli* Hyd-2. *The crystal structure of $(HybOC)_2$ is combined with homology models of HybAB to give rise to a predicted quaternary structure of the E. coli Hyd-2 complex as a dimer (one Hyd-2 is shown in grey while the other is illustrated with four colours corresponding to four subunits, respectively). Subunits of Hyd-2 are indicated with the same colour as in the structure model.*

The linker connecting the globular domain and the TMD of HybO is not included in the model due to its flexibility. PM: periplasm, IM: inner membrane, CM: cytoplasm. This figure is adapted from (Beaton et al., 2018).

Hyd-2 is an 'uptake' hydrogenase mainly mediating H_2 oxidation (Pinske *et al.,* 2015). However, during anaerobic respiration, Hyd-2 is able to conserve redox energy in the form of proton motive force (pmf) and, in addition, can also generate H_2 under fermentative conditions where reducing equivalents are in excess (Pinske *et al.,* 2015). This unique property could arise from the extra electron-relaying subunit HybA in Hyd-2 (Dolla *et al.,* 2000). HybA is a Tat-substrate with a single C-terminal TMD anchoring it to the inner membrane. Although HybA is essential for Hyd-2 function, the membrane localisation of HybOC is not HybA-dependent, suggesting the assembly of the Hyd-2 complex does not follow a linear pathway (Dubini *et al.,* 2002; Hatzixanthis *et al.,* 2003). HybB is a large integral membrane protein, predicted to contain 10 TMDs with both its N- and C-termini in the cytoplasm (Menon *et al.,* 1994), in contrast to the four-TMD cytochrome *b* HyaC in Hyd-1 (Volbeda *et al.,* 2013). Sequence analysis suggests that HybB lacks the ability to bind haem as it lacks histidine residues required as axial ligands for the haem iron (Dubini *et al.,* 2002). Instead, HybB was proposed to act a conformational proton pump to generate pmf as electrons travel through the Hyd-2 complex to the quinone pool during H_2 oxidation (Beaton *et al.,* 2018; Pinske *et al.,* 2015).

Hyd-2 has also been implicated in virulence of pathogenic enteric bacteria. A *Salmonella* Typhimurium *hybA* mutant exhibits reduced growth compared to the wild-type in the early phase of gut colonisation in a murine model, but this difference is abolished upon antibiotic

treatment, suggesting H_2 derived from the host microbiota is used as an important fuel in a Hyd-2-dependent manner by enteric pathogens (Maier *et al.,* 2013).

1.4.4. Hydrogenase-3 (Hyd-3)/formate hydrogenlyase-1 (FHL)

The Hyd-3/FHL complex is the major contributor to H_2 evolution during glucose fermentation by enteric bacteria (Ballantine and Boxer, 1985; Redwood *et al.,* 2008). During anaerobic fermentative growth (*i.e.* in the absence of all exogenous electron acceptors), enteric bacteria such as *Shigella* and *E. coli* perform mixed-acid fermentation, converting glucose to formate, succinate, acetate, lactate, and ethanol. As the environmental pH drops, for example during passage through the stomach, extracellular formate is transported back into the cytoplasm *via* the formate transporter FocA (Beyer *et al.,* 2013). When its concentration reaches a certain threshold, the membrane-bound FHL complex converts formate into CO_2 and H_2 (Sawers, 1994).

The FHL complex is composed of FdhF (a molybdenum-dependent formate dehydrogenase), HycE, the catalytic α-subunit of Hyd-3, three iron-sulphur proteins; HycB, HycF, and HycG, and two polytopic integral membrane proteins HycC (8-TMD) and HycD (14-TMD), which anchor the core complex to the cytoplasmic side of the membrane **(Figure 1.9)** (McDowall *et al.,* 2014). The reaction that FHL catalyses can be split into two sequential steps. The formate dehydrogenase step, in which a formate molecule is converted into one CO_2, one proton, and two electrons, is catalysed by FdhF. Electrons derived from formate oxidation are then transported to HycE for proton reduction and H_2 evolution (Böhm *et al.,* 1990; Sawers *et al.,* 1985).

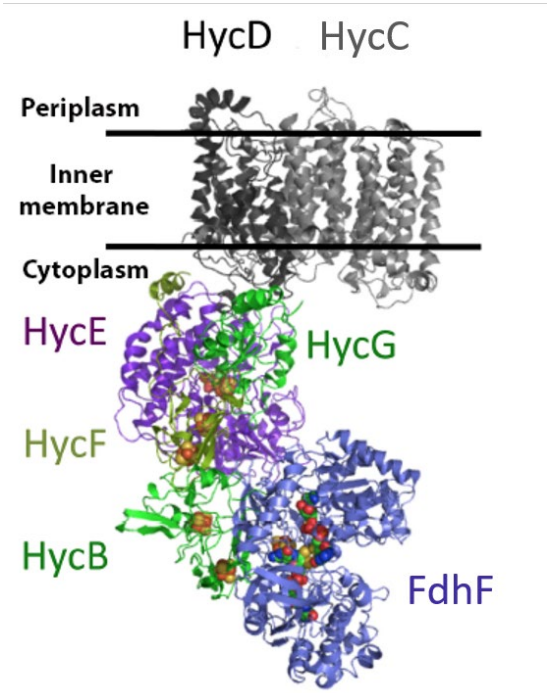


Figure 1.9. The predicted structure of the *E. coli* Hyd-3/FHL complex. *The membrane-bound components of Hyd-3 are shown in black and gray, the small subunits in dark green, and the HycF electron-transferring subunit in light green. The [NiFe] active site of Hyd-3 is located in the purple HycE subunit. The formate dehydroganse FdhF subunit is in blue. This figure is adapted from (Pinske and Sawers, 2016).*

1.4.5. Hydrogenase-4 (Hyd-4)/formate hydrogenlyase (FHL-2)

The Hyd-4/FHL-2 is a protein complex encoded by the *hyfABCDEFGHIRfocB* operon (Andrews et al., 1997). While HyfR and FocB are proposed to be involved in transcriptional regulation of the operon (Andrews et al., 1997; Skibinski et al., 2002) and formate uptake (Andrews et al., 1997), respectively. Seven of the putative proteins encoded by the operon are orthologs of the FHL complex, whereas three putative membrane proteins (HyfD, HyfE, and HyfF) possess no orthologs in the FHL complex but are homologous to membrane subunits of complex I (Andrews et al., 1997; Pinske and Sawers, 2016). The identity of the

formate dehydrogenase subunit of FHL-2 remains undefined, although both FdhF and YdeP, a homolog of FdhF, have been suggested to play this role (Bagramyan et al., 2002; Sargent, 2016).

Hyd-4 has been shown to mediate H_2 production when the enzyme is overexpressed (Kuniyoshi et al., 2015) and when the environment is slightly alkaline (pH ≈ 7.5) (Bagramyan et al., 2002). However, this phenomenon could also be attributable to Hyd-2-mediated H_2 evolution (Pinske et al., 2015; Trchounian et al., 2013). Therefore, the physiological role of Hyd-4/FHL-2 unknown, which is largely due to the failure of identifying a physiological condition under which the Hyd-4 operon is robustly expressed (Self et al., 2004; Skibinski et al., 2002). In addition, the *in vitro* biochemical activity of Hyd-4 still remains uncharacterised (Pinske and Sawers, 2016).

1.5. Formate dehydrogenases

1.5.1. Formate dehydrogenase-N (FDH-N)

In addition to FdhF, which is part of the Hyd-3 complex, *E. coli* and *Shigella* contain two other formate dehydrogenase complexes: FDH-N and FDH-O. FDH-N is a periplasm-facing, inner membrane-bound enzyme encoded by the *fdnGHI* operon (Berg and Stewart, 1991). The expression of this operon is induced by nitrate under anaerobic conditions and its main function is to generate pmf by coupling formate oxidation to nitrate or nitrite reduction (Berg and Stewart, 1990; Darwin *et al.,* 1993; Enoch and Lester, 1975; Li and Stewart, 1992). Formate is provided either from the environment or mixed-acid fermentation, mainly through the activity of the pyruvate formate lyase, which converts pyruvate into formate and acetyl-CoA (Lehtio and Goldman, 2004). FDH-N contains the catalytic α-subunit, FdnG, an

electron-transferring β-subunit FdnH with type I topology, and a membrane-bound γ-subunit FdnI with four TMDs (**Figure 1.10A**) (Jormakka *et al.,* 2003; Jormakka *et al.,* 2002). The α- and β-subunits are located in the periplasm. FdnH contains a single C-terminal TMD, providing the membrane localisation with the four TMDs in FdnI.

Quinones are the electron acceptors used by FDH-N in the reaction: $HCO_2^- + H_2O + Q + 2H^+_{in} \rightarrow HCO_3^- + QH_2 + 2H^+_{ex}$ (Unden *et al.,* 2014). Combining the structural and topological information from analysis of FDH-N and the nitrate reductase complex NarGHI, it is likely that they form a redox loop, and couple electron flow from formate to nitrate, with pmf generated through two consecutive steps: (i) two protons are removed from the cytoplasm by FdnI to reduce menaquinone into menaquinol, (ii) the menaquinol then diffuses across the membrane and releases the protons into the periplasm, while the menaquinol becomes oxidised in NarI. The net effect of these two steps is equivalent to $4H^+$ translocation across the bacterial inner membrane per molecule of formate oxidised (**Figure 1.10B**) (Jones, 1980). However, this redox loop formed by FDH-N and NarGHI might not represent the full picture in the host gut, particularly under inflammatory conditions, for example, it was demonstrated that, in a mouse DSS-induced colitis model, inflammation-derived nitrate enhanced the expression of FDH-N in enteric bacteria such as *E. coli* (Winter *et al.,* 2013). FDH-N then allowed *E. coli* to utilise microbiota-derived formate using O_2 as the terminal electron acceptor, driving the expansion of *E. coli* in the inflamed gut (Hughes *et al.,* 2017).

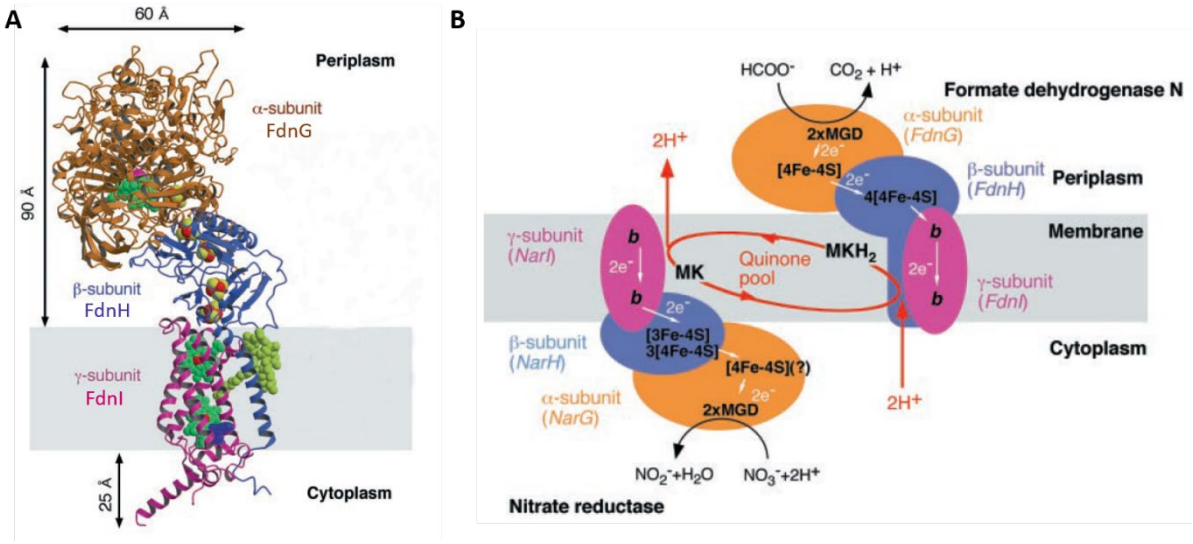


Figure 1.10. Crystal structure of FDH-N and its metabolic interplay with NarGHI. ***(A)*** *The crystal structure of a FdnGHI trimer in the membrane.* ***(B)*** *Redox loop formed by FdnGHI and the nitrate reductase NarGHI. Both* ***A*** *and* ***B*** *are adapted from (Jormakka et al., 2002).*

1.5.2. Formate dehydrogenase-O (FDH-O)

Formate dehydrogenase-O (FDH-O) also converts formate into CO_2 and two protons (Abaibou *et al.,* 1995; Pommier *et al.,* 1992), and is encoded by the *fdoGHI-fdhE* operon (Abaibou *et al.,* 1995). The operon encodes the catalytic α-subunit FdoG, an electron-transferring β-subunit FdoH, a membrane-bound γ-subunit FdoI predicted to have four TMDs (**Table 1.1**), and FdhE, an iron-binding rubredoxin necessary for the biogenesis of both FDH-N and FDH-O (Luke *et al.,* 2008). The expression of FDH-O is induced by nitrate under aerobic conditions (Abaibou *et al.,* 1995). FDH-O is very similar in sequence, subunit composition and prosthetic groups to FDH-N, suggesting a similar overall structure and biochemical activity. The orientation of FdoGHI complex was first proposed to be inverse to FdnGHI (Benoit *et al.,* 1998), with FdoG, FdoH, and the formate site located in the cytoplasm. However, later studies demonstrated that this topology model is incorrect, and

complicated by the involvement of the Tat pathway and fusion tags (Stanley *et al.,* 2002). The physiological function of FDH-O remains elusive. It has been proposed that FDH-O facilitates the transition of *E. coli* from aerobic to anaerobic environments to utilise formate before the expression of FDH-N is switched on (Abaibou *et al.,* 1995). In a mouse DSS-induced colitis model, while FDH-N was shown to drive the expansion of *E. coli* in the inflamed gut, FDH-O was dispensable for this process (Hughes *et al.,* 2017).

1.6. Project Aims

The field of bacterial rhomboids is still in its infancy. Their wide distribution in bacteria strongly suggests functional importance. However, the failure to identify a clear phenotype of an *E. coli glpG* mutant suggests GlpG does not contribute to growth or stress response under standard laboratory conditions. *S. sonnei* is a pathogen with ever increasing medical importance and so far, no reports on rhomboid functions in *Shigella* spp.. This project aims to identify and characterise rhomboid proteases in *S. sonnei* and determine their function. The specific aims of my thesis were: 1) identification of rhomboids in *S. sonnei*, 2) functional characterisation of these rhomboids, 3) identification of rhomboid substrates and/or rhomboid mutant phenotypes, and 4) elucidation of the roles of rhomboids in *S. sonnei* physiology and/or pathogenesis.

Chapter 2. Materials and Methods

2.1. Strains and plasmids

Strains and plasmids used in this work are listed in Table 2.1 unless otherwise specified. *E. coli* was grown at 37°C with agitation at 180 r.p.m. in LB medium (4 g NaCl, 4 g yeast extract, and 8 g Tryptone per one litre medium), or on LB agar plates containing 1.5 % bacteriological agar (w/v). Antibiotics were used at the following concentrations: carbenicillin (carb), 100 μg/ml; streptomycin (strep), 50 μg/ml; chloramphenicol, 20 μg/ml. For routine aerobic growth, *S. sonnei* strain CS14 and derivatives were grown at 37°C with agitation at 180 r.p.m in LB broth. Anaerobic growth was performed in a Whitley A35 anaerobic station (5% H_2, 10% CO_2, 85% N_2 or 10% H_2, 10% CO_2, 80% N_2) at 37°C with agitation at 180 r.p.m in relevant media, described for each assay. Bacterial growth was assessed by measurements of optical density at 600 nm (OD_{600}) using a SpetraMax M5 spectrophotometer. Viable cells (Colony Forming Units (CFU)/ml) were enumerated by serially diluting 20 μl bacterial culture in 180 μl of PBS or media (as specified) in a 96-well plate. An aliquot of 10 μl from each well was plated onto LB agar and incubated at 37°C overnight before counting.

Table 2.1. Strains and plasmids used in this work

Strain name/genotype	**Source or reference**
DH5α, *E. coli* cloning strain	Lab collection
SM10 λpir,	Simon *et al.,* 1993
CS14, *S. sonnei* wild-type	(Holt *et al.,* 2012)
CS14Δ*glpG*Δ*rhom7*	This study
CS14Δ*glpG*Δ*rhom7* + *glpG*	This study

CS14 *hybA-sfCherry-3xFLAG*	This study
CS14Δ*glpG hybA-sfCherry-3xFLAG*	This study
CS14Δ*rhom7 hybA-sfCherry-3xFLAG*	This study
CS14Δ*glpG*Δ*rhom7 hybA-sfCherry-3xFLAG*	This study
CS14Δ*glpG*Δ*rhom7* + *glpG hybA-sfCherry-3xFLAG*	This study
CS14 *hybO-sfCherry-3xFLAG*	This study
CS14Δ*glpG*Δ*rhom7 hybO-sfCherry-3xFLAG*	This study
CS14 *hybA*$_{G296F}$-*sfCherry-3xFLAG*	This study
CS14Δ*glpG hybA*$_{G296F}$-*sfCherry-3xFLAG*	This study
CS14Δ*rhom7 hybA*$_{G296F}$-*sfCherry-3xFLAG*	This study
CS14Δ*glpG*Δ*rhom7 hybA*$_{G296F}$-*sfCherry-3xFLAG*	This study
CS14 *hybA*$_{P300A}$-*sfCherry-3xFLAG*	This study
CS14Δ*glpG hybA*$_{P300A}$-*sfCherry-3xFLAG*	This study
CS14Δ*rhom7 hybA*$_{P300AF}$-*sfCherry-3xFLAG*	This study
CS14Δ*glpG*Δ*rhom7 hybA*$_{P300AF}$-*sfCherry-3xFLAG*	This study
CS14Δ*hyaA-F*Δ*hycE*Δ*spa33*	This study
CS14Δ*hyaA-F*Δ*hybO-G*Δ*hycE*Δ*spa33*	This study
CS14Δ*hyaA-F*Δ*hycE*Δ*spa33*Δ*rhom7*	This study
CS14Δ*hyaA-F*Δ*hycE*Δ*spa33*Δ*glpG*Δ*rhom7*	This study
CS14Δ*hyaA-F*Δ*hycE*Δ*spa33*Δ*rhom7hyaA*$_{G296F}$	This study
CS14Δ*hyaA-F*Δ*hycE*Δ*spa33*Δ*rhom7 hybA-sfCherry-3xFLAG*	This study
CS14Δ*hyaA-F*Δ*hycE*Δ*spa33*Δ*glpG*Δ*rhom7 hybA-sfCherry-3xFLAG*	This study

CS14Δ*hyaA-F*Δ*hycE*Δ*spa33*Δ*rhom7 hybA*$_{G296F}$*-sfCherry-3xFLAG*	This study
CS14Δ*rhom7 hybA-sfCherry-3xFLAG 3xHA-glpG*	This study
CS14Δ*rhom7 hybA-sfCherry-3xFLAG 3xHA-glpG*$_{S201A}$	This study
CS14Δ*rhom7*Δ*hybB hybA-sfCherry-3xFLAG 3xHA-glpG*	This study
CS14Δ*rhom7*Δ*hybB hybA-sfCherry-3xFLAG 3xHA-glpG*$_{S201A}$	This study
CS14Δ*rhom7 hybA*$_{P300A}$*-sfCherry-3xFLAG 3xHA-glpG*	This study
CS14Δ*rhom7 hybA*$_{P300A}$*-sfCherry-3xFLAG 3xHA-glpG*$_{S201A}$	This study
CS14Δ*rhom7*Δ*hybB hybA*$_{P300A}$*-sfCherry-3xFLAG 3xHA-glpG*	This study
CS14Δ*rhom7*Δ*hybB hybA*$_{P300A}$*-sfCherry-3xFLAG 3xHA-glpG*$_{S201A}$	This study
CS14Δ*rhom7 hybA-sfCherry-3xFLAG 3xHA-glpG pKS508 hybA-sfCherry-3xFLAG*	This study
CS14Δ*rhom7 hybA-sfCherry-3xFLAG 3xHA-glpG*$_{S201A}$ *pKS508 hybA-sfCherry-3xFLAG*	This study
CS14Δ*rhom7 fdoH-sfCherry-3xFLAG 3xHA-glpG*	This study
CS14Δ*rhom7 fdoH-sfCherry-3xFLAG 3xHA-glpG*$_{S201A}$	This study
CS14Δ*rhom7*Δ*fdoI fdoH-sfCherry-3xFLAG 3xHA-glpG*	This study
CS14Δ*rhom7*Δ*fdoI fdoH-sfCherry-3xFLAG 3xHA-glpG*$_{S201A}$	This study
CS14Δ*rhom7 fdoH*$_{P259A}$*-sfCherry-3xFLAG 3xHA-glpG*	This study
CS14Δ*rhom7 fdoH*$_{P259A}$*-sfCherry-3xFLAG 3xHA-glpG*$_{S201A}$	This study
CS14Δ*rhom7*Δ*fdoI fdoH*$_{P259A}$*-sfCherry-3xFLAG 3xHA-glpG*	This study

CS14Δ*rhom7*Δ*fdoI* *fdoH*$_{P259A}$*-sfCherry-3xFLAG 3xHA-glpG*$_{S201A}$	This study
CS14Δ*rhom7 hybA-halo-3xFLAG 3xHA-glpG*	This study
CS14Δ*rhom7 hybA- halo-3xFLAG 3xHA-glpG*$_{S201A}$	This study
Plasmids	
pBAD33	Guzman *et al.,* 1995
pBAD33 *3xHA-glpG*	This study
pBAD33 *3xHA-glpG*$_{S201A}$	This study
pBAD33 *3xHA-glpG*$_{S201A\ H254A}$	This study
pBAD33 *rhom7*	This study
pBAD33 *rhom7*$_{S133A}$	This study
pBAD33 *rhom7*$_{S133A\ H187A}$	This study
pBAD33 *3xHA-rhom7*$_{\Delta TM7}$	This study
pBAD33 *3xHA-rhom7*$_{\Delta TM7\Delta CTD}$	This study
pUC19	(Yanisch-Perron *et al.,* 1985)
pUC19 *3xHA-glpG*	This study
pUC19 *3xHA-glpG*$_{S201A\ H254A}$	This study
pKS508	(Strisovsky *et al.,* 2009)
pKS508 TM-CjrB	This study
pKS508 TM-BcsB	This study
pKS508 TM-CcmD	This study
pKS508 TM-DgcJ	This study
pKS508 TM-DjlA	This study

pKS508 TM-DjlB	This study
pKS508 TM-ElaB	This study
pKS508 TM-FdnH	This study
pKS508 TM-FdoH	This study
pKS508 TM-YebO	This study
pKS508 TM-FliO	This study
pKS508 TM-Flk	This study
pKS508 TM-FtsH	This study
pKS508 TM-HflK	This study
pKS508 TM-HyaA	This study
pKS508 TM-HybA	This study
pKS508 TM-HybO	This study
pKS508 TM-LapA	This study
pKS508 TM-MxiJ	This study
pKS508 TM-NrfF	This study
pKS508 TM-PpdC	This study
pKS508 TM-RS17255	This study
pKS508 TM-RxsB	This study
pKS508 TM-SecG	This study
pKS508 TM-SohB	This study
pKS508 TM-TcdA	This study
pKS508 TM-TorS	This study
pKS508 TM-YajC	This study
pKS508 TM-YbdJ	This study

pKS508 TM-YbjT	This study
pKS508 TM-YgaM	This study
pKS508 TM-YgiM	This study
pKS508 TM-YhcB	This study
pKS508 TM-YhdP	This study
pKS508 TM-YhhM	This study
pKS508 TM-YibN	This study
pKS508 TM-YjeT	This study
pKS508 TM-YmcD	This study
pKS508 TM-YnaJ	This study
pKS508 TM-YqiK	This study
pKS508 TM-YqjD	This study
pKS508 TM-YtjB	This study
pKS508 TM-YtjC	This study
pKS508 TM-ZipA	This study
pKS508 *hybA-sfCherry-3xFLAG*	This study
pKNG101	Kaniga *et al.,* 1991
pCONJ4s	This study
pCONJ4s Δ*glpG*	This study
pCONJ4s $glpG_{S201A}$	This study
pCONJ4s *3xHA-glpG*	This study
pCONJ4s Δ*rhom7*	This study
pCONJ4s Δ*spa33*	This study
pCONJ4s Δ*hyaA-F*	This study

pCONJ4s Δ*hybO-G*	This study
pCONJ4s Δ*hycE*	This study
pCONJ4s Δ*hybB*	This study
pCONJ4s Δ*fdoI*	This study
pCONJ4s Δ*fdnI*	This study
pCONJ4s *hybA*$_{G296F}$	This study
pCONJ4s *hybA-sfCherry-3xFLAG*	This study
pCONJ4s *hybA*$_{G296F}$-*sfCherry-3xFLAG*	This study
pCONJ4s *hybA*$_{P300A}$-*sfCherry-3xFLAG*	This study
pCONJ4s *fdoH-sfCherry-3xFLAG*	This study
pCONJ4s *fdoH* $_{P259A}$-*sfCherry-3xFLAG*	This study
pCONJ4s *fdnH-sfCherry-3xFLAG*	This study
pCONJ4s *fdnH* $_{P259A}$-*sfCherry-3xFLAG*	This study
pCONJ4s *hybO-sfCherry-3xFLAG*	This study
pCONJ4s *hyaA-sfCherry-3xFLAG*	This study

2.2. Molecular techniques

2.2.1. Polymerase Chain Reaction (PCR)

PCR for DNA amplification was performed with Q5 High Fidelity DNA Polymerase (New England Biolabs) following the manufacturer's protocol using 50 µg of genomic DNA or purified plasmid as template per 50 µl reaction. 'Colony PCR' was performed using Taq DNA polymerase (Sigma Aldrich) according to the manufacturer's protocol using a small number of bacteria from a colony as template per 20 µl reaction. All primers used are documented in **Table 2.2.** in the Appendices.

2.2.2. Plasmid and genomic DNA extraction

Plasmid DNA was extracted from liquid cultures of *E. coli* grown overnight at 37°C in LB broth, using the GenElute™ Plasmid Miniprep Kit (Sigma-Aldrich) according to the manufacturer's protocol. Genomic DNA of *S. sonnei* CS14 was prepared using the Wizard® Genomic DNA Purification Kit (Promega) following the manufacturer's protocol for Gram-negative bacteria.

2.2.3. Gel electrophoresis and purification of DNA

DNA was analysed by electrophoresis in 1% (w/v) agarose (Invitrogen) gels in 1x TAE buffer (40 mM Tris, 20 mM acetic acid, 1 mM EDTA) containing 1x SYBR safe DNA Gel Stain (Invitrogen). DNA samples were separated by electrophoresis in 1x loading buffer containing 0.04% bromophenol blue (w/v), 0.04% xylene cyanol (w/v), 5% glycerol (v/v). Hyperladder™ 1kb (Bioline) was used as the size marker. Gels were imaged using a Gel Doc™ XR+ Imager (Bio-Rad) and the Image Lab 5.2.1 software. For DNA purification by gel extraction, DNA bands were visualised on a Safe Imager™ 2.0 (Invitrogen), excised and DNA purified using the Wizard® SV Gel and PCR Clean-Up System (Promega) following the recommended protocol.

2.2.4. Restriction digestion and ligation

Restriction digestion was performed at 37°C with appropriate restriction endonucleases and their corresponding buffers (New England Biolabs) following standard procedures. Ligation reactions were performed using Quickstick ligase (Bioline) at room temperature for 15 minutes.

2.2.5. Gibson assembly

Purified DNA fragments were assembled using the NEBuilder® HiFi DNA Assembly Master Mix (New England Biolabs) for one hour at 50°C following the recommended protocol.

2.2.6. Transformation

Chemically competent cells of *E. coli* were prepared according to published protocols (Inoue *et al.,* 1990). Briefly, *E. coli* was grown in LB to an OD_{600} of around 0.5. Bacteria were then pelleted by centrifugation and resuspended in RF1 solution (15% w/v glycerol, 100 mM RbCl, 50 mM $MnCl_2$, 30 mM CH_3COOK, 10 mM $CaCl_2$, pH 5.8), incubated on ice for 45 minutes, pelleted and resuspended in RF2 solution (15% w/v glycerol, 10 mM RbCl, 10 mM MOPS, 75 mM $CaCl_2$, pH 6.8) for 15 minutes. Competent cells were then aliquoted into 50 µl aliquots and stored at -80°C until required. Transformation was performed by mixing 50 µl of chemically competent *E. coli* cells with plasmid DNA on ice for 10 minutes before heat shock at 42°C for 45 seconds followed by two minutes on ice. One ml of LB broth was added, and bacteria were incubated at 37°C with agitation at 180 r.p.m for 40 minutes before plating onto LB agar plates containing appropriate antibiotics.

2.2.7. Electroporation

Electrocompetent cells of *S. sonnei* were prepared as follows. *S. sonnei* was grown in LB to an OD_{600} of approximately 0.5. One millilitre of culture was pelleted by centrifugation and pellets were washed with one ml of 10% (w/v) glycerol three times, then resuspended in 50 µl of 10% (w/v) glycerol and stored at -80°C until required. Electroporation was performed using a Pulse Controller Plus (Bio-Rad) (800 Ω, 25 µF, 2.5 kV) in 2 mm gap electroporation cuvettes (Molecular BioProducts). Electroporated bacteria were then incubated in LB at

37°C with agitation at 180 r.p.m for 40 minutes before plating onto LB agar containing appropriate antibiotics.

2.3. Plasmid and strain construction

2.3.1. Construction of pCONJ4s and its derivatives

To achieve markerless mutagenesis in *Shigella*, a novel suicide vector, pCONJ4s (**Figure 2.1**), which only replicates in hosts supplying the π protein encoded by the *pir* gene, was designed to deliver mutagenesis constructs from bacterial donors to recipients through bacterial mating. pCONJ4s was constructed using the NEBuilder HiFi DNA Assembly Cloning Kit (New England Biolabs) as follows. The plasmid backbone of pKNG101 (Kaniga *et al.,* 1991), was amplified with primers pGL225/pGL224 and assembled with the *bla* gene from pGEM-T Easy (Promega) using pGL222/pGL223, which encodes the ampicillin/carbenicillin resistance cassette, generating pCONJ1. pCONJ1 was linearised by *Spe*I digestion and a *Pme*I restriction site was added through assembly with a primer heterodimer formed by pGL185/pGL186, yielding pCONJ2. Subsequently, the backbone of pCONJ2 was amplified using pGL154/pGL97 and assembled with '*superfolder' GFP* (*sfGFP*) (Pedelacq *et al.,* 2006) amplified from pNCC1 *sfGFP* (Wormann *et al.,* 2016) using primers pGL98/pGL155, generating pCONJ3. pCONJ3 was then linearised by *Eco*RV digestion (New England Biolabs) and assembled with the *sacB* counter-selection marker from pKNG101 using pGL247/pGL248, which encodes the SacB protein that kills bacterial hosts grown on sucrose, giving rise to pCONJ4. Finally, pCONJ4 was digested by *Afl*II and assembled with a primer heterodimer formed by pGL342/pGL343, to remove 95 bp of unnecessary sequence to form pCONJ4s. pCONJ4s was used to achieve targeted mutagenesis in *Shigella* through homologous recombination following conjugation as described in Chapter 2.3.2. To this end, the upstream and downstream regions

(approximately 500 bp each) flanking each target gene were amplified by PCR with the relevant primers and assembled into pCONJ4s at the *Pme*I site by Gibson assembly. For constructs to introduce a chromosomal sfCherry-3xFLAG tag, the sfCherry ORF was synthesised (ThermoFisher Scientific), PCR amplified by pGL627/pGL628, and assembled with the upstream and downstream regions flanking the gene of interest into pCONJ4s. pCONJ4s and its derivatives were generated and maintained in *E. coli* SM10 λ*pir* (Simon *et al.,* 1983) and constructs were verified by sequencing with pGL715/pGL716 (Source Bioscience).

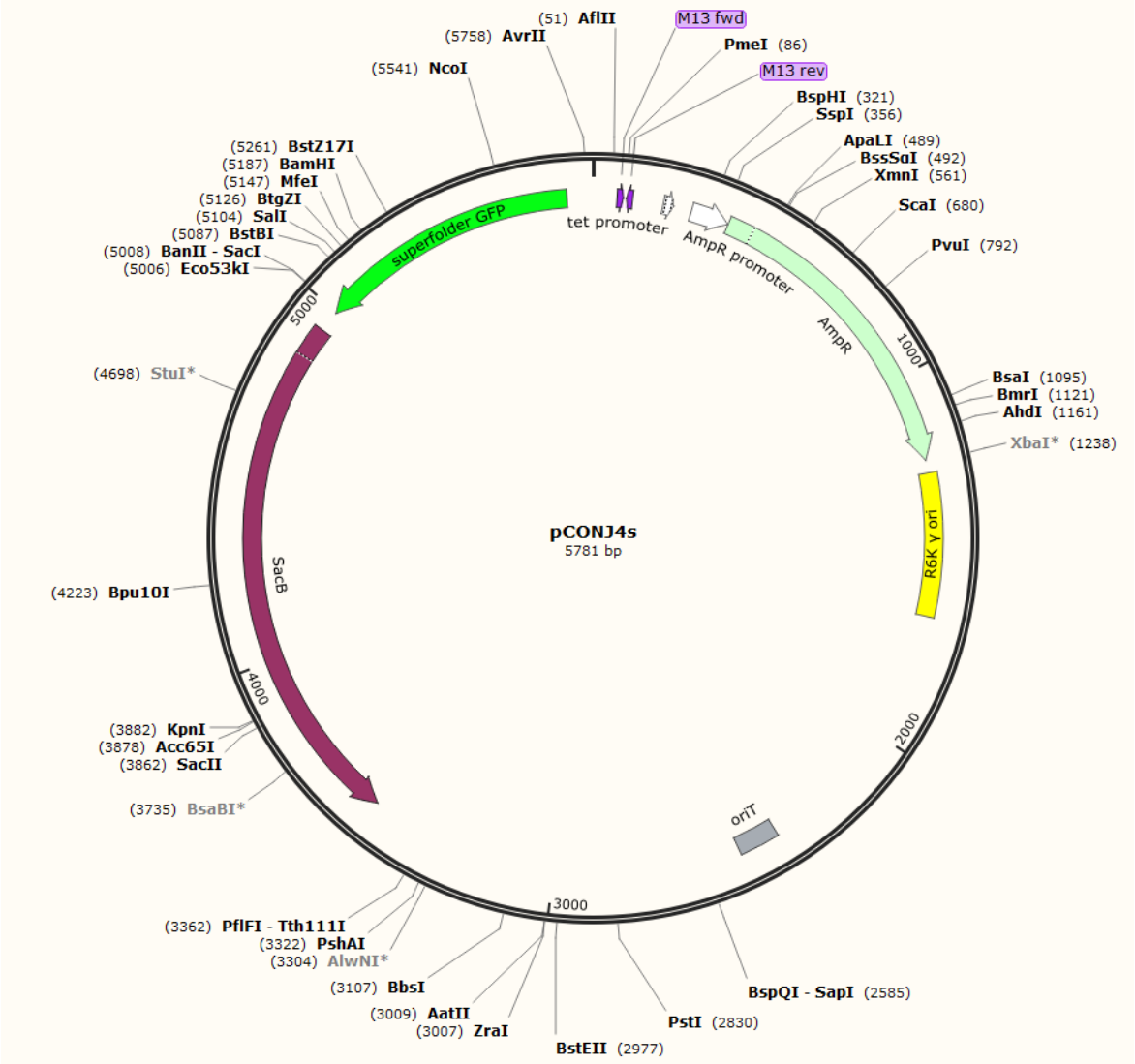


Figure 2.1. Map of pCONJ4s. pCONJ4s is the vector used for generating constructs for markerless mutagenesis through homologous recombination. It has a R6K γ origin, which

can only replicate in hosts possessing the *pir* gene; an *oriT* sequence that allows transfer of pCONJ4s through conjugation; an ampicillin-resistance cassette; a *sacB* gene conferring sucrose sensitivity; a *superfolder gfp* gene; and a *Pme*I site used for inserting DNA fragments required for mutagenesis.

2.3.2. Generation of *S. sonnei* markerless mutants by conjugation

All *S. sonnei* CS14 markerless mutants in this study were generated by conjugation. First, 20 µl of stationary phase SM10 λpir harbouring a pCONJ4s derivative containing a specific mutagenesis construct was mixed with 20 µl of stationary phase *S. sonnei* CS14 recipient in an Eppendorf tube, and the entire volume was spotted onto LB agar. Plates were incubated at room temperature until the spots dried and then at 37°C for 4 hours, after which the bacteria were scraped from the plate, resuspended in 1 ml of LB, and 100 µl was plated onto LB agar plates containing carbenicillin and streptomycin to select for *S. sonnei* that have plasmid integration to the chromosome. Plates were then incubated at 37°C for 14 hours, and colony PCR was performed to identify clones in which plasmids recombined at the desired locus. Next, a single colony was used to inoculate liquid LB containing carbenicillin and streptomycin and the culture was grown to stationary phase to eliminate potential *E. coli* (streptomycin sensitive) carryover. Bacteria were collected by centrifugation, washed and diluted five-fold into 5 ml of LB without antibiotics, then grown at 37°C for 4 hours, during which chromosomally integrated plasmids in a subpopulation of bacteria will excise stochastically through homologous recombination, and are lost. Depending on the flanking region in which the excision event occurs, the plasmid integrant can either revert to wild-type or the mutant of interest. To select for bacteria that lacked integrated pCONJ4s, 100 µl of 50-fold diluted culture was plated onto salt-free LB agar containing 10% (w/v) sucrose and incubated for 14 hours at 37°C. As pCONJ4s carries a *gfp* gene on the backbone, GFP-

positive clones indicate clones retaining pCONJ4s but surviving sucrose killing and hence were not further screened. Colony PCR was thus performed on GFP-negative clones to identify and confirm clones in which the desired mutations were introduced.

2.3.3. Construction of pBAD33 and pUC19 derivatives for *S. sonnei* complementation

pBAD33 (Guzman *et al.,* 1995) was linearised by digestion with *Sph*I and *Xba*I. A DNA fragment containing the *glpG* open reading frame, and 19 bp of 5' UTR, which includes the predicted ribosomal binding site, was amplied by PCR from *S. sonnei* CS14 genomic DNA using pGL271/pGL272. The product was digested with *Sph*I and *Xba*I, and ligated into *Sph*I/*Xba*I linearised pBAD33, generating pBAD33*glpG*. To introduce a triple HA (3xHA) tag at the N-terminus of GlpG, pBAD33*glpG* was then used as the template for PCR amplification of two fragments using pGL413/pGL363 and pGL412/pGL362, one of which encodes the 3xHA tag. These two fragments were then assembled using the NEBuilder HiFi DNA Assembly Cloning Kit, generating pBAD33 *3xHA-glpG*. To introduce a Ser201Ala (S201A) mutation in 3xHA-GlpG, pBAD33 *3xHA-glpG* was used as the template for PCR amplification of two fragments using pGL353/pGL363 and pGL354/pGL362, which were assembled to generate pBAD33 *3xHA-glpG*$_{S201A}$. To introduce a second His254Ala (H254A) mutation in 3xHA-GlpG$_{S201A}$, pBAD33 *3xHA-glpG*$_{S201A}$ was used as the template for PCR amplification of two fragments using pGL583/pGL363 and pGL582/pGL362, which were assembled to generate pBAD33 *3xHA-glpG*$_{S201A\ H254A}$.

To generate vectors that allow inducible expression of *S.sonnei* Rhom7 and derivatives, a DNA fragment containing the *S. sonnei rhom7* (SSON_0610) open reading frame and putative 5' UTR including the ribosomal binding site was amplied by PCR from *S. sonnei*

CS14 genomic DNA using pGL255/pGL256, digested with *Sph*I and *Xba*I, and ligated into *Sph*I/*Xba*I linearised pBAD33, generating pBAD33 *rhom7*. pBAD33 *rhom7* was used as the template for construction of pBAD33 *rhom7*$_{H187A}$, pBAD33 *rhom7*$_{S133A}$, pBAD33 *rhom7*$_{\Delta TM7\Delta CTD}$, pBAD33 *rhom7*$_{\Delta CTD}$, and pBAD33 *3xHA-rhom7* by PCR amplification of pairs of fragments using primer pairs (pGL305/pGL363, pGL306/pGL362), (pGL360/pGL363, pGL361/pGL362), (pGL377/pGL363, pGL378/pGL362), (pGL257/pGL363, pGL446/pGL362), and (pGL415/pGL363, pGL414/pGL362), which were Gibson assembled. Similarly, pBAD33 *3xHA-rhom7* was used as the template for PCR amplification of pairs of fragments using (pGL305/pGL363, pGL306/pGL362), (pGL360/pGL363, pGL361/pGL362), (pGL377/pGL363, pGL378/pGL362), and (pGL257/pGL363, pGL446/pGL362), which were Gibson assembled to generate pBAD33 *3xHA-rhom7*$_{H187A}$, pBAD33 *3xHA-rhom7*$_{S133A}$, pBAD33 *3xHA-rhom7*$_{\Delta TM7\Delta CTD}$, pBAD33 *3xHA-rhom7*$_{\Delta CTD}$, respectively.

To generate vectors for high-level expression of *S. sonnei* GlpG and derivatives, pUC19 was chosen as the backbone. Although pUC19 uses an IPTG-inducible *lac* promoter, its high copy-number allows high expression even in the absence of IPTG. pUC19 was linearised by digestion with *Sph*I and *Hind*III. Two DNA fragments comprising the ORF of *3xHA-glpG* and *3xHA-glpG*$_{S201A\ H254A}$ were PCR amplified using pGL809/pGL810 from pBAD33 *3xHA-glpG* and pBAD33 *3xHA-glpG*$_{S201A\ H254A}$, respectively, and assembled into *Sph*I/*Hind*III linearised pUC19, giving rise to pUC19 *3xHA-glpG* and pUC19 *3xHA-glpG*$_{S201A\ H254A}$, respectively. All plasmids generated in the section were verified by sequencing (Source Bioscience).

2.3.4. Bioinformatic identification of substrates

As the *S. sonnei* CS14 genome has not been annotated, the genome of its close relative, *S. sonnei* SS046 (Holt *et al.,* 2012) was used for bioinformatic identification of rhomboid substrates. Type I and Type III proteins are proteins with a single TMD, a periplasmic N-terminus and a cytoplasmic C-terminus, with or without a signal peptide, respectively. The predicted proteome of *S. sonnei* SS046 from PubMed was first analysed using TMHMM Server v. 2.0 using default parameters to identify proteins with type I and type III topology (Krogh *et al.,* 2001). HHpred was used to identify structural homologues of the predicted type I and type III proteins. Amino acid sequences of the TMD of the candidates were used as queries for HHpred using the *E. coli* K-12 proteome (07_Mar version) as the HHpred server does not include a *Shigella* proteome, using the PDB_mmCIF70_3_Jul database and default parameters (Soding *et al.,* 2005). Orthologues (>90% sequence identity) of proteins that were also present in *S. sonnei* CS14 were subsequently analysed by Phobius 1.01 (Kall *et al.,* 2007) and TOPCONS 2.0 to exclude proteins with multiple transmembrane domains (Tsirigos *et al.,* 2015).

2.3.5. Construction of the transmembrane domain library for rhomboid substrate screen

pKS508 (gift from Dr. K. Strisovsky, Czech Republic) was linearised by digestion with *Kpn*I and *Xba*I, and Gibson assembled with annealed primers encoding TMDs plus 16 additional amino acids of predicted type I and III membrane proteins. The primers used for constructing the library of artificial substrates are shown in Table 2.2 in the Appendix. All plasmids were verified by sequencing (Source Bioscience).

2.4. SDS-PAGE and western blot analysis

Whole cell lysates were prepared by boiling bacteria for 10 minutes in SDS-PAGE sample buffer (50 mM Tris, 0.75% (w/v) SDS, 10% (v/v) glycerol, 25 mM EDTA, 1% (v/v) β-mercaptoethanol, 0.11 mM bromophenol blue) at a dilution of 333 µl per OD_{600} unit. 20 µl of lysate was separated by electrophoresis (SDS-PAGE) on 12% polyacrylamide gels at 150 V for 75 minutes in SDS-PAGE buffer (25 mM Tris, 192 mM glycine, 0.1% (w/v) SDS, pH 8.5). Proteins were transferred to nitrocellulose membranes (constant voltage, 25V, 20 minutes) using the Trans-Blot Turbo system (Bio-Rad). Tagged proteins were detected using anti-FLAG M2 antibodies (Sigma, 1:1000) and goat anti-mouse IgG-HRP (Bio-Rad, 1:10,000), or anti-HA antibodies (Cell Signaling Technology, 1:1000) and goat anti-rabbit IgG-HRP (Santa Cruz Biotechnology), or anti-V5 antibodies (Santa Cruz Biotechnology, 1:1000) and rabbit anti-goat IgG-HRP (DAKO, 1:2000). RecA was detected using anti-RecA (*E. coli*) antibodies (Abcam, 1:5000) and goat anti-rabbit IgG-HRP. Bands were detected by chemiluminescence (ECL, GE Healthcare). Band intensity was quantified using Photoshop CC (Lynda.com) as follows. A box of defined dimensions that covers the bands to be quantified was defined and the total number of pixels within each box was measured as a representation of the relative amount of protein within this band.

2.5. Rhomboid-dependent protein cleavage assays

Cleavage assays were performed in *S. sonnei* CS14. Each rhomboid protease was expressed either using pBAD33 with arabinose induction, in which case endogenous *glpG* and/or *rhom7* loci were deleted (details in section 2.3.2); or from their endogenous locus in which case the wild-type gene was replaced with an allele fused to a triple-HA tag. In parallel, substrates were expressed using pKS508 or as fusion proteins from their endogenous locus. For assays performed under aerobic conditions, 50 µl of overnight culture was added to 5

ml LB and bacteria were grown at 37°C, 180 r.p.m. for 75 minutes before addition of L-arabinose at a final concentration of 0.2% (w/v). Bacteria were then grown for a further two hours and harvested for SDS-PAGE and western blot analysis. Assays performed under anaerobic conditions were performed in the same way except that 100 µl of overnight culture was added to 5 ml of LB, and bacteria were grown for one hour anaerobically before addition of L-arabinose, after which bacteria were grown for a further two hours and harvested for SDS-PAGE and western blot analysis.

2.6. Analysis of type three secretion system secretion

Overnight cultures of *S. sonnei* grown aerobically in LB, were used to inoculate 10 ml Tryptic Soy Broth (TSB), and grown at 37°C with agitation at 180 r.p.m until an OD_{600} of 1.5. Bacteria were harvested by centrifugation at 21,100 x g, washed and re-suspended in PBS to an OD_{600} of 5. Congo Red (Sigma) was added at 200 µg/ml final concentration to the bacterial suspension, which was then incubated at 37°C for 15 minutes to induce secretion. Bacteria were then incubated on ice for 5 minutes, centrifuged at 21,100 x g, and 150 µl of supernatant was mixed with an equal volume of 2x SDS-PAGE sample buffer and heated at 100°C for 10 minutes prior to electrophoresis (20 µl) on 10% acrylamide gels. Proteins were stained using SilverXpress™ Silver Staining Kit (Thermo Fisher Scientific) following the manufacturer's protocol.

2.7. Hydrogenase activity assays

To measure hydrogenase activity, the growth of wild-type and mutant strains in M9 minimal (6.78 g/L Na_2HPO_4, 3 g/L KH_2PO_4, 1 g/L NH_4Cl, 0.5 g/L NaCl) medium supplemented with fumarate and casamino acids, as previously described (Dubini *et al.,* 2002), was measured

in anaerobic conditions with hydrogen. Specifically, one ml of overnight culture of *S. sonnei* grown aerobically in LB at 37°C with agitation at 180 r.p.m was centrifuged at 21,100 x g for one minute and the bacterial pellets were washed twice with, and resuspended in, one ml of M9 minimal medium supplemented with 0.5% fumarate, 12.5 μg/ml nicotinic acid 0.2% casamino acids. The number of bacteria was quantified by OD_{600}. The bacterial suspensions were then transferred to a Whitely A35 anaerobic workstation supplied with 5% H_2, 10% CO_2 and 85% N_2 and diluted to an OD_{600} of 0.01 in 50 ml of supplemented M9 minimal medium as defined above that had been pre-incubated in 5% H_2, 10% CO_2, 85% N_2 overnight. Bacteria were grown at 37°C with agitation at 180 r.p.m, and OD_{600} was measured every hour for 24 hours.

Hydrogen uptake was measured by Gas Chromatography. *S. sonnei* strains were prepared as described above, then diluted in 20 ml of supplemented minimal medium to an OD_{600} of 0.01 in glass gas chromatography vials in a Whitley A35 anaerobic workstation. Vials were sealed with #25 suba-seal septa (SIGMA-ALDRICH) and parafilm, before removal from the anaerobic chamber. The headspace was purged with 10% Hydrogen, 90% Argon (BOC) for 10 mins (100 scc/min total flow rate) through a mass flow controller (Sierra Instruments) using a Sterican neural therapy needle (0.80 x 120 mm, 21 G x 4¾'', B Braun) at room temperature for 5 mins before transfer to a shaking incubator at 37°C, 250 r.p.m. Samples (15 μl) of the headspace gas were taken periodically using a Hamilton syringe and the amount and composition of the gas sampled was analysed using a Gas Chromatograph System 7890A (Agilent Technologies) with ShinCarbonST 100/120 Column, and N_2 as the carrier gas, at 32°C. Each run was stopped approximately 3 minutes post-injection to ensure the capture of the entire H_2 peak. Prior to measurements, a 15 μl air injection was applied to verify the gas chromatograph. At each time point, two independent samples of headspace

were analysed for each test sample and one sample was analysed for the negative control. After each injection, the suba seal was reinforced with parafilm. The amount of H_2 in the headspace was determined by calculating the area underneath the H_2 peak, which was quantified by Agilent EZChrom *Elite*. H_2 uptake at a given time point T_x is calculated by the formula: H_2 uptake T_x = amount of H_2 in the headspace at T_0 - amount of H_2 in the headspace at T_x.

2.8. Biolog phenotypic microarrays

Biolog phenotypic microarrays including more than 950 nutritional and stress conditions from PM1-PM10 (Bochner *et al.,* 2001) were used to compare the growth of three strains of *S. sonnei*: CS14 wild-type, CS14Δ*glpG*Δ*rhom7, and* CS14Δ*glpG*Δ*rhom7* + *glpG.* Growth and analyses were performed by the company (https://biolog.com/).

2.9. Oxidative stress assays

For H_2O_2 stress, 500 µl of overnight culture of *S. sonnei* CS14 grown aerobically in LB was used to inoculate 50 ml LB, which was incubated at 37°C until the OD_{600} reached 0.5. Five ml of culture was aliquoted into 50 ml falcon tubes, and bacteria were incubated with 200 µM H_2O_2 (the 'adaptation' dose) (Daugherty *et al.,* 2012) for 20 minutes before the addition of H_2O_2 to a final concentration of 2 mM (the 'killing' dose) and incubation for a further 15 minutes. Viable cells of both wild-type and mutants undergoing the 'adaptation' dose only and the 'adaptation' plus the 'killing' doses were enumerated by plating onto LB agar and counting colonies after overnight incubation at 37°C.

For paraquat (PQ) stress, 500 µl of overnight culture of *S. sonnei* CS14 grown aerobically in LB was used to inoculate 50 ml M9 minimal medium (supplemented with 0.4% glucose, 12.5 µg/ml nicotinic acid) and grown to an OD_{600} of 0.5, then 5 ml of culture was aliquoted into 50 ml falcon tubes (Chattopadhyay *et al.,* 2003; Minton *et al.,* 1990). The bacteria were then incubated with 200 µM PQ (the 'adaptation' dose) for 15 minutes before the addition of PQ to a final concentration of 10 mM (the 'killing' dose) and incubated for another 45 minutes before viable cell enumeration using the same method as for H_2O_2 stress above.

2.10. Quantification and Statistical analysis

Western blot band intensity was quantified by Photoshop CC (2017) using methods provided by Lynda.com. GC measurements were quantified by Agilent EZChrom Elite. Two-way ANOVA was used to calculate p values for OD_{600} and hydrogenase activity measurements. One-way ANOVA was used to calculate p values for CFU counts. P values below 0.05 were considered statistically significant.

Chapter 3. Identification and biochemical characterisation of rhomboid proteases in *S. sonnei*

3.1. Identification of Rhom7, a potential active rhomboid protease

Based on existing knowledge, *E. coli* GlpG and *P. stuartii* AarA are the architype for prokaryotic rhomboid proteases with six TMDs and seven TMDs, respectively. These characterised rhomboids share only 25% amino acid sequence identity, suggesting different evolution origins. We sought to identify rhomboid proteases in *S. sonnei* strain CS14 (Holt *et al.,* 2012) by searching for homologues using the amino acid sequences of *E. coli* GlpG and *P. stuartii* AarA as queries. BLASTp searches were performed using the sequences of *E. coli* MG1655 GlpG (accession no. YP_026220.1) and *P. stuartii* AarA (accession no. AAA61597.1). The *S. sonnei* SS046 proteome (taxid:300269) was used for initial analysis because the proteome of *S. sonnei* CS14 is not available. Proteins were considered to be homologues if they share > 60% query coverage in the rhomboid fold region with E value < 10^{-3}. In *S. sonnei* SS046, BLASTp analysis revealed one homologue of GlpG: (SSON_3661), and one homologue of AarA: (SSON_0610). Both proteins were confirmed to be present in CS14 based on its whole genome sequence (unpublished data from our lab).

SSON_3661 (276 amino acids long) differs from *E. coli* MG1655 GlpG ($GlpG_{Ec}$) by a single amino acid (Ala^{130} in *S. sonnei* is Thr^{130} in *E. coli*) so was designated GlpG (Figure **3.1A**). To gain further insights into of the topology of GlpG, the protein was analysed using Phobius 1.01, a TMD and signal peptide prediction server (Kall *et al.,* 2007). GlpG was predicted to contain six TMDs and a cytosolic domain, which is consistent with $GlpG_{Ec}$ (Figure **3.1B**) (Maegawa *et al.,* 2005). In addition, BLASTp analysis predicted that GlpG possesses a

Rhomboid-N domain (Pfam12122, amino acids 1-84) and a rhomboid core domain (COG0705, amino acids 95-276) (Figure **3.1C**).

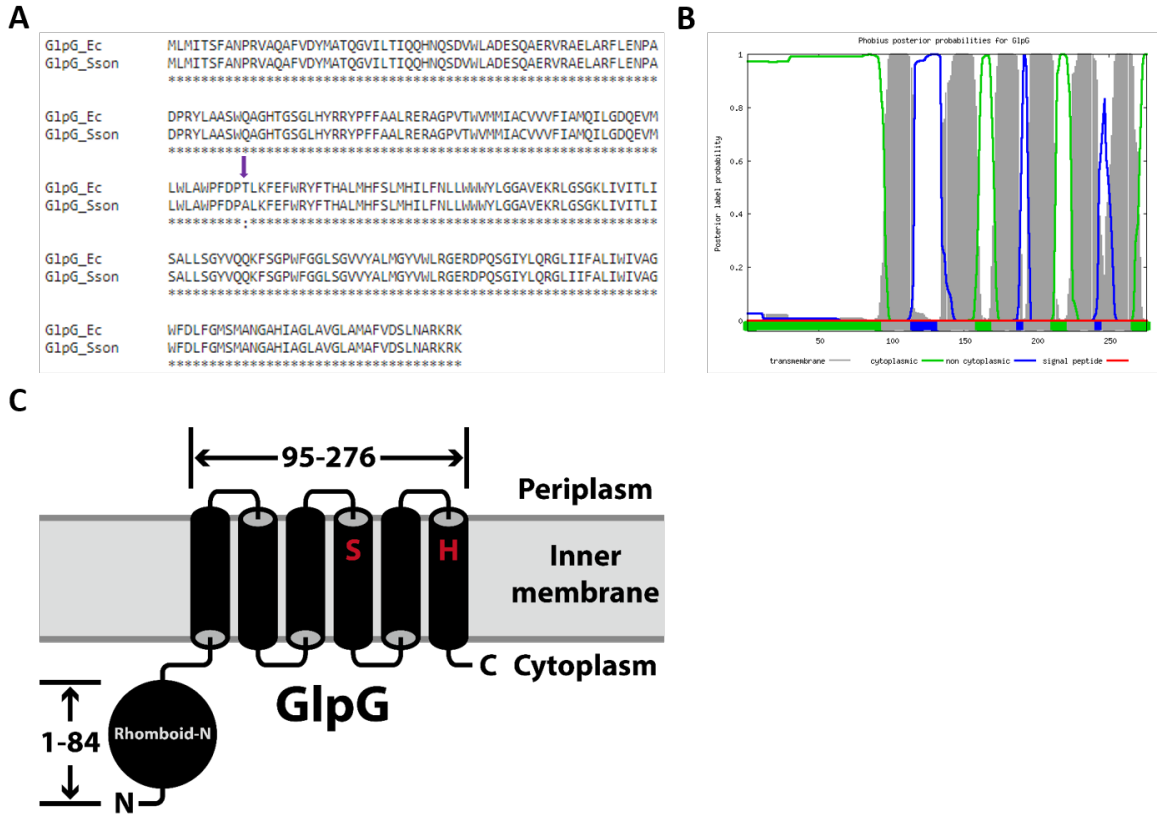


Figure 3.1. Sequence, topology and domain analyses of *S. sonnei* CS14 GlpG (SSON_3661). ***(A)*** *Amino acid sequence alignment of S. sonnei CS14 GlpG with E. coli GlpG. The single amino acid difference between the two rhomboids at position 130 is highlighted by a purple arrow.* ***(B)*** *Predicted topology of CS14 GlpG.* ***(C)*** *Schematic representation of CS14 GlpG topology and domain organisation with respect to the bacterial inner membrane. Putative catalytic serine and histidine residues are shown in red. Amino acids predicted to form the Rhomboid-N domain (1-84) and the rhomboid core domain (95-276) are also shown. Alignments were performed by Clustal Omega multiple sequence alignment program with default parameters. Domain predictions are based on BLASTp analyses, while TMD and topology predictions were performed with Phobius.*

SSON_0610 (625 amino acids long) shares 33% amino acid identity with *P. stuartii* AarA ($AarA_{Ps}$) (**Figure 3.2A**). As $GlpG_{Ec}$ is the only rhomboid with a structually defined catalytic site (formed by Ser^{201} and His^{254}), the protein sequence of SSON_0610 was aligned with the sequence of $GlpG_{Ec}$ and as a result, Ser^{133} and His^{187} were predicted to be the catalytic residues of SSON_0610 (**Figure 3.2B**). Phobius 1.01 analysis suggested that SSON_0610 possesses seven TMDs and a periplasmic domain (**Figure 3.2C**). Therefore, SSON_0610 was designated as 'Rhom7'. In addition, BLASTp analysis predicted that Rhom7 possesses a rhomboid core domain (COG0705, amino acids 7-187) and a WD40 domain (cd00200, amino acids 297-597). Of note, different predictions on the C-terminal boundary of the rhomboid core domain of Rhom7 were given by BLASTp (amino acid 187) and by Phobius (amino acid 201). After incorporating the alignment results of Rhom7 with $GlpG_{Ec}$, the Phobius result was believed to be more accurate (**Figure 3.2D**).

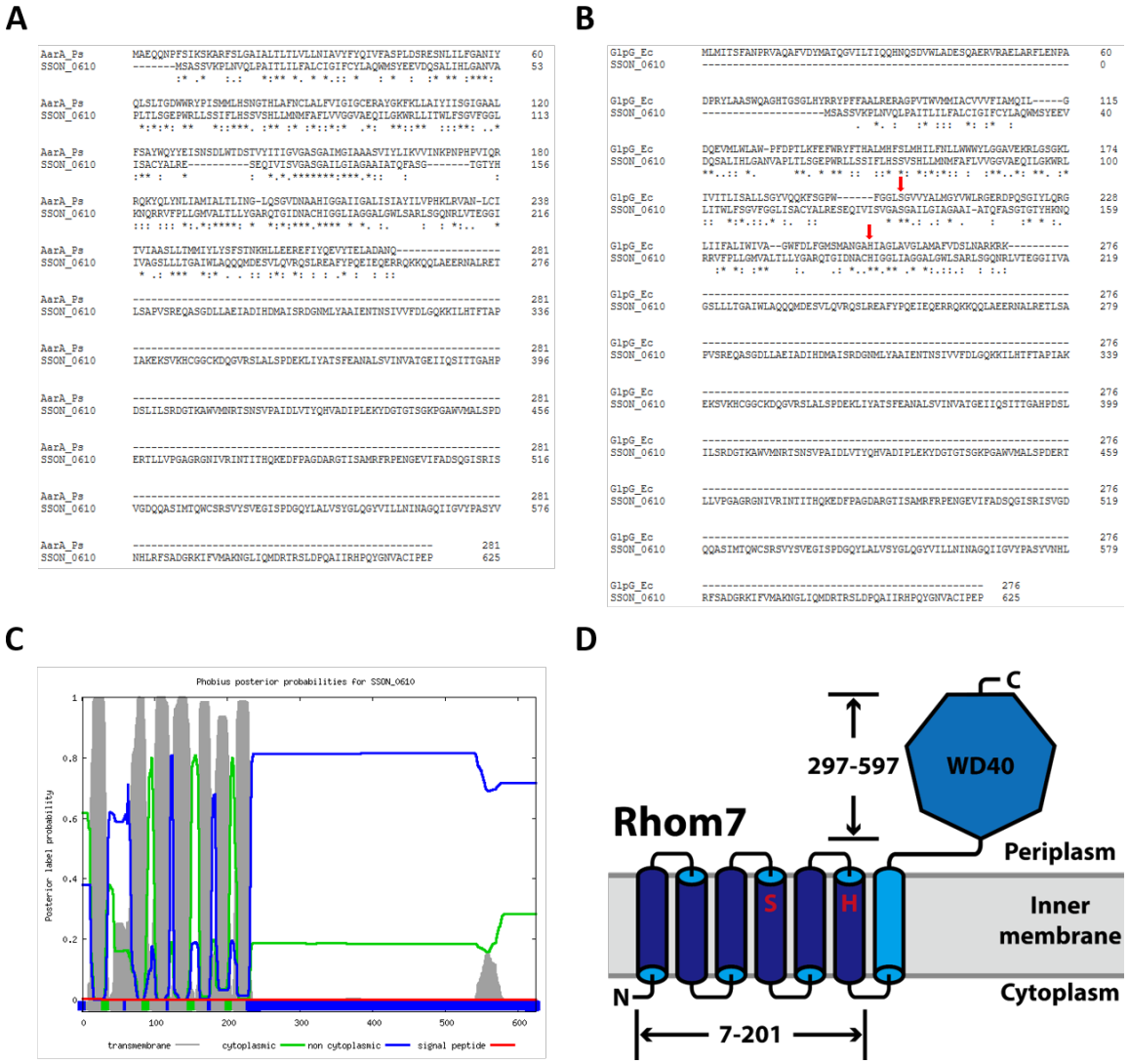


Figure 3.2. Sequence, topology and domain analyses of *S. sonnei* CS14 Rhom7 (SSON_0610). ***(A)*** *Amino acid sequence alignment of S. sonnei CS14 SSON_0610 with AarA.* ***(B)*** *Amino acid sequence alignment of S. sonnei CS14 SSON_0610 with E. coli GlpG. Potential active site residues of SSON_0610, Ser*^{133} *and His*^{187}*, are highlighted by red arrows.* ***(C)*** *Predicted topology of CS14 SSON_0610.* ***(D)*** *Schematic representation of CS14 Rhom7 with regard to the bacterial inner membrane. Putative catalytic serine and histidine residues are shown in red. Amino acids predicted to form the rhomboid core domain and the C-terminal WD40 domain are also shown. Alignments were performed by Clustal Omega multiple sequence alignment program with default parameters. Domain predictions*

are based on BLASTp analyses, while TMD and topology predictions were performed with Phobius.

Having identified putative rhomboid proteases in *S. sonnei* CS14, I next examined their genetic loci. Similar to *E. coli*, *S. sonnei* CS14 *glpG* is part of the *glpEGR* operon (Yang and Larson, 1998). In contrast, *rhom7* is located between the *gltI-sroC-gltJKL* operon (which encode a predicted glutamate/aspartate transporter and a small regulatory RNA (*sroC*)) and the *ybeZYX-lnt* operon (encoding a putative PhoE-like ATPase, (*ybeZ*), a putative endonuclease (*ybeY*), a putative Mg^{2+}/Ca^{2+} efflux protein (*ybeX*), and an apolipoprotein N-acyltransferase (*lnt*)) (**Figure 3.3**). Based on the orientation of adjacent open reading frames, *rhom7* in *S. sonnei* CS14 is predicted to form a monocistronic transcription unit. Interestingly *rhom7* is replaced by a transposable sequence in *E. coli* strain MG1655 (**Figure 3.3**). However, further bioinformatic searches showed that *rhom7* is also present in some non-K12 *E. coli* strains such as *E. coli* BL21.

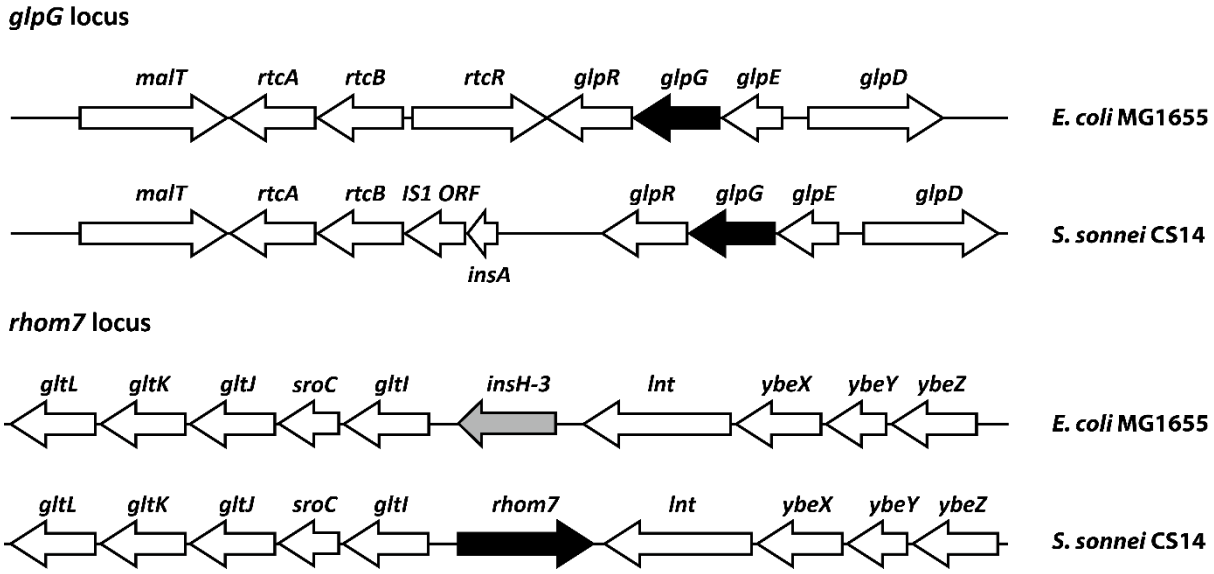


Figure 3.3. Genetic organisation of *S. sonnei* CS14 *glpG* and *rhom7* and comparison with their counterparts in *E. coli* MG1655. *Genes in S. sonnei CS14 were manually annotated. Briefly, open reading frames (ORFs) were first identified using SnapGene Viewer and the whole genome sequence of CS14, and the identity of the ORFs was then confirmed based on homology to the E. coli genome. Genes encoding putative rhomboids are shown in black while the insH-3 gene, which encodes an IS5 family transposase, is shown in grey. Size of the arrows is drawn to scale to the length of corresponding genes.*

3.2. Development of an assay to measure rhomboid activity using a TatA derived substrate in *S. sonnei* CS14

Rhom7 and GlpG are both predicted to be active rhomboids based on their homology to $AarA_{Ps}$ and $GlpG_{Ec}$ and, in particular they both have conserved putative catalytic residues: Ser^{133} and His^{187} for Rhom7, and Ser^{201} and His^{254} for Rhom7. I therefore sought to demonstrate their proteolytic activity by testing their ability to cleave known rhomboid substrates in *S. sonnei* CS14. To attribute any observed cleavage specifically to Rhom7 or GlpG activity, it was necessary to perform assays in a background lacking each or both

rhomboids. Therefore, I constructed CS14Δ*glpG*, CS14Δ*rhom7* and CS14Δ*glpG*Δ*rhom7* without introducing a chromosomal antibiotic cassette using a suicide vector (pCONJ4s) harbouring a construct designed for Rhom7/GlpG deletion as described in section 2.3.1. through bacterial mating (Kaniga *et al.,* 1991). pCONJ4s was delivered from *E. coli* SM10 λ*pir* to *S. sonnei* CS14 by conjugation. Homologous recombination between pCONJ4s and the chromosome integrated pCONJ4s to the chromosomal locus of either *rhom7* or *glpG*. A second recombination event occurring at low frequency and stochastically, led to excision of pCONJ4s. If this second recombination event occurs at the same flanking sequence as the first recombination, this process will 'reverse' the plasmid integration and leave the 'wild-type' *glpG* or *rhom7* locus intact. Alternatively, if the second recombination event occurs at the other flanking sequence as the first recombination, the mutant of interest will be generated as a result. As *glpG* is flanked by *glpE* and *glpR* in the *glpEGR* operon (**Figure 3.3**), only the first 506 bp of the *glpG* ORF was deleted, to preserve internal promoters in the *glpEGR* operon. Similarly, although *rhom7* does not appear to be part of any operon, it is not known whether the putative promoters of *rhom7* can influence the expression of downstream genes. Therefore, 1358 bp of the *rhom7* ORF starting from codon 15 was deleted to preserve putative promoters. All the mutants were confirmed by Sanger sequencing of the chromosomal regions subject to mutagenesis.

To measure the proteolytic activity of both rhomboids, an *in vivo* assay using a previously described artificial substrate was developed (Strisovsky *et al.,* 2009). This artificial substrate comprises the TMD (residues 1-50) of TatA from *P. stuartii*, the substrate of AarA, but which can also be cleaved by *E. coli* GlpG and *B. subtilis* YqgP between Ala^{8} and Ala^{9} (Stevenson *et al.,* 2007; Strisovsky *et al.,* 2009). In addition, the substrate has a perplasmic, N-terminal maltose-binding protein (MBP) with a triple-FLAG tag (3xFLAG), and a C-

terminal thioredoxin domain (Trx) with a His tag (**Figure 3.4A**) allowing detection of the N- or C-terminal products of cleavage by western blot analysis (Strisovsky *et al.,* 2009). Plasmid pKS508, constitutively expressing the artificial substrate, was introduced into *S. sonnei* Δ*glpG*Δ*rhom7*, together with pBAD33 *glpG* or pBAD33 *rhom7*, allowing arabinose inducible expression of GlpG or Rhom7, respectively; pBAD33 containing $glpG_{S201A}$ which lacks activity (Maegawa *et al.,* 2005) or an empty vector served as negative controls (**Figure 3.4B**).

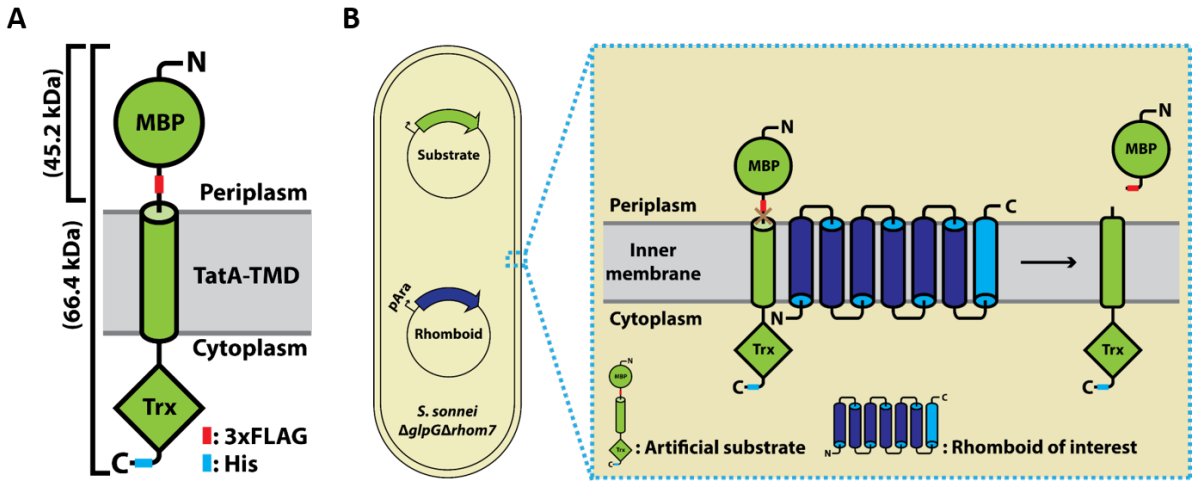


Figure 3.4. Representations of the artificial substrate and the *in vivo* cleavage assay to detect rhomboid activity in *S. sonnei*. *(A)* *The artificial substrate, which possesses an N-terminal maltose-binding protein (MBP) linked to the TMD (residues 1-50) of P. stuartii TatA by a triple-FLAG tag (3xFLAG, in red), and a thioredoxin domain (Trx) with a C-terminal His tag (in blue). The size of the full-length artificial substrate and products of rhomboid protease cleavage detected by anti-FLAG antibodies are indicated.* ***(B)*** *In vivo cleavage assay. The assay was performed in S. sonnei CS14ΔglpGΔrhom7, which does not have endogenous rhomboids. pKS508 constitutively expresses the artificial substrate from a tac promoter (**A**), while pBAD33 allows the expression of rhomboid/rhomboid mutants (blue) under the control of L-arabinose. Empty pBAD33 was used as a negative control.*

Bacteria were grown for 75 minutes in liquid LB before the addition of arabinose (final concentration, 0.2%) to induce rhomboid expression, then incubated for a further 120 minutes to allow cleavage of the substrate. After this, whole cell lysates were prepared for western blot analysis. The full-length, uncleaved substrate (predicted mol. mass, 66.4 kDa) and the N-terminal cleavage product (predicted mol. mass, 45.2 kDa) were detected with anti-FLAG antibodies (**Figure 3.5**). The empty vector control demonstrated that there is a degree of rhomboid-independent cleavage of the substrate, which is more pronounced in the absence of arabinose induction compared to arabinose induction (**Figure 3.5 lane 5 & 6**). Induced expression of *S. sonnei* $GlpG_{WT}$ increased the amount of cleaved substrate compared to when the expression of $GlpG_{S201A}$ is induced (**Figure 3.5 lane 2 & 4**), confirming that *S. sonnei* GlpG is an active rhomboid, and cleaves the substrate containing the TMD from TatA. Alanine substitution of the catalytic residue Ser^{201} abolished GlpG cleavage of the TatA TMD as expected (Maegawa *et al.,* 2005) (**Figure 3.5**).

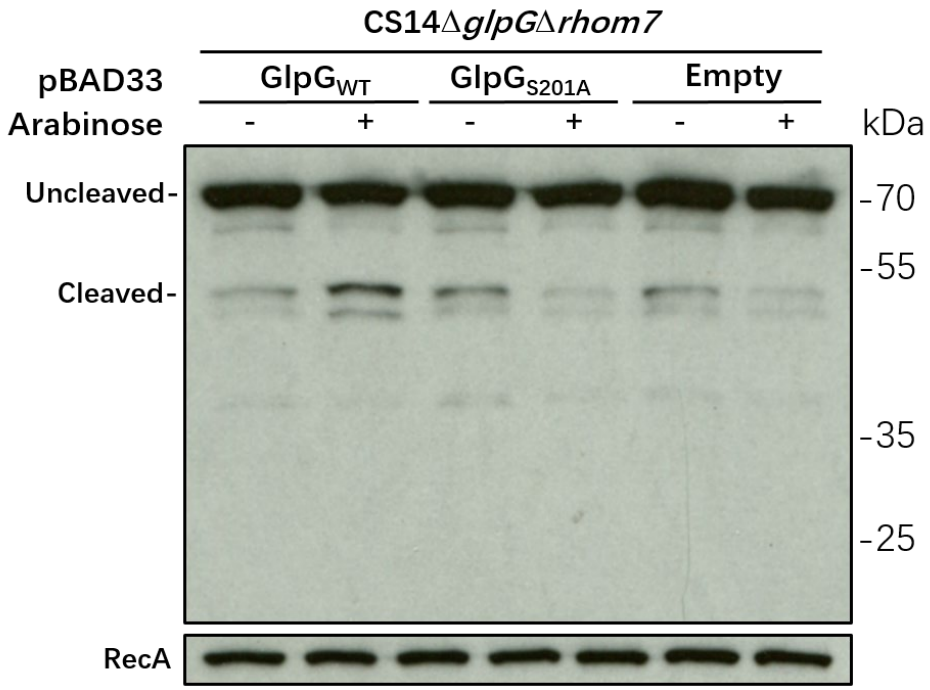


Figure 3.5. *S. sonnei* CS14 GlpG is an active rhomboid. *Western blot analysis of cleavage of the artificial substrate (**Figure 3.4A**) by S. sonnei CS14 GlpG. GlpG expressed from plasmid pBAD33 cleaves the full-length substrate. Mutation of the active site Ser^{201} of GlpG to alanine abolishes cleavage. Whole cell lysates were prepared from bacteria grown without or with arabinose induction of protease expression. CS14ΔglpGΔrhom7 transformed with pKS508 and pBAD33 empty vector was included as a control (n = 1). Cleavage of the artificial substrate was detected with anti-FLAG antibodies. RecA was used as a loading control.*

3.3. Serine 133 and Histidine 187 constitute the catalytic dyad of Rhom7

The proteolytic activity of Rhom7 was also tested using the *in vivo* cleavage assay. Similar to GlpG, expression of Rhom7 from pBAD33 resulted in cleavage of the artificial substrate (**Figure 3.6 lane 1 & 2**). Interestingly, upon expression of wild-type Rhom7, the uncleaved

artificial substrate was no longer detected, in contrast to GlpG whose cleavage was observed as the presence of additional cleaved substrate, but no observable changes in the uncleaved substrate (**Figure 3.5 lane 2**). Ala substitution of Ser^{133} or His^{187} in Rhom7, which form the predicted catalytic dyad, abolished Rhom7 cleavage of the artificial substrate (**Figure 3.6 lane 3-6**), consistent with Rhom7 being a functional rhomboid protease. In addition, an arabinose-dependent but rhomboid-independent decrease in the amount of uncleaved substrate was also observed (**Figure 3.6 lane 7-8**), suggesting the expression of the artificial substrate is also affected by arabinose.

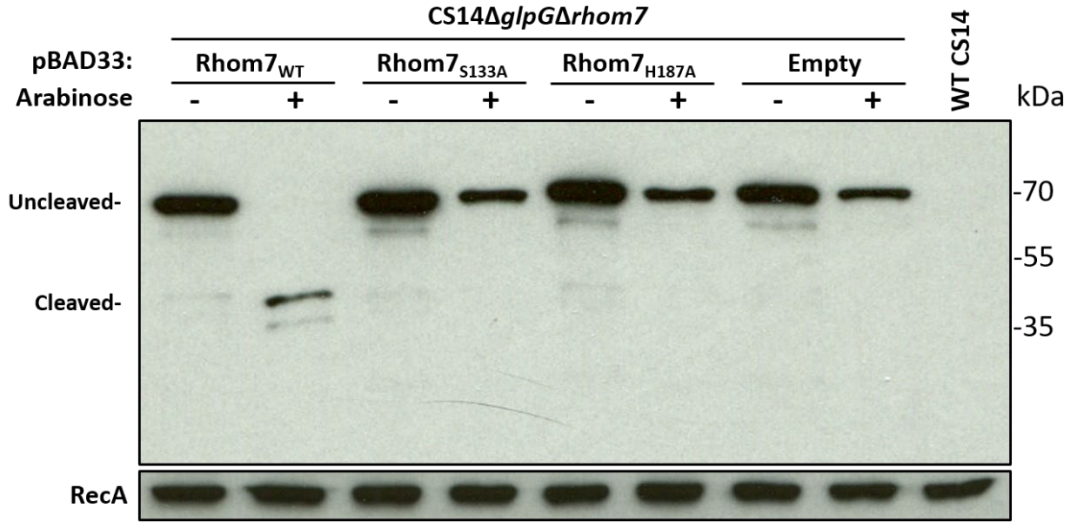


Figure 3.6. *S. sonnei* CS14 Rhom7 is an active rhomboid protease, with Ser^{133} and His^{187} required for its activity. *S. sonnei Rhom7 cleaves the artificial substrate. CS14ΔglpGΔrhom7 was co-transformed with pKS508 encoding the artificial substrate, and pBAD33 vectors encoding wild-type Rhom7, Rhom7 with mutations to putative active site residues (S133A and H187A). Whole cell lysates were prepared from bacteria grown without or with arabinose induction of protease expression. CS14ΔglpGΔrhom7 transformed with pKS508 and pBAD33 empty vector, and a whole cell lysate from WT S. sonnei CS14 were included as controls (n = 1). Cleavage of the artificial substrate was detected with anti-FLAG antibodies. RecA was used as a loading control.*

One hypothesis that could also explain lack of artificial substrate cleavage upon mutation of Ser^{133} or His^{187} is that both mutations abolished/reduced the expression of Rhom7. I therefore repeated the assay but with Rhom7 harbouring an N-terminal triple-HA tag (**Figure 3.7**). Interestingly, multiple bands were detected by western blot analysis with anti-HA antibodies, consistent with degradation of Rhom7. Of note, alanine substitution of Ser^{133} or His^{187} did not reduce, if not increased, cellular levels of Rhom7 compared with the wild-type protein (**Figure 3.7 lane 2 & 4 & 6**). In addition, in contrast to the absence of uncleaved substrate upon expression of wild-type Rhom7 (**Figure 3.6 lane 2**), uncleaved substrate could still be detected upon expression of 3xHA-$Rhom7_{WT}$ in the same way (**Figure 3.7 lane 2**), suggesting the triple-HA tagging might reduce the expression and/or the activity of wild-type Rhom7.

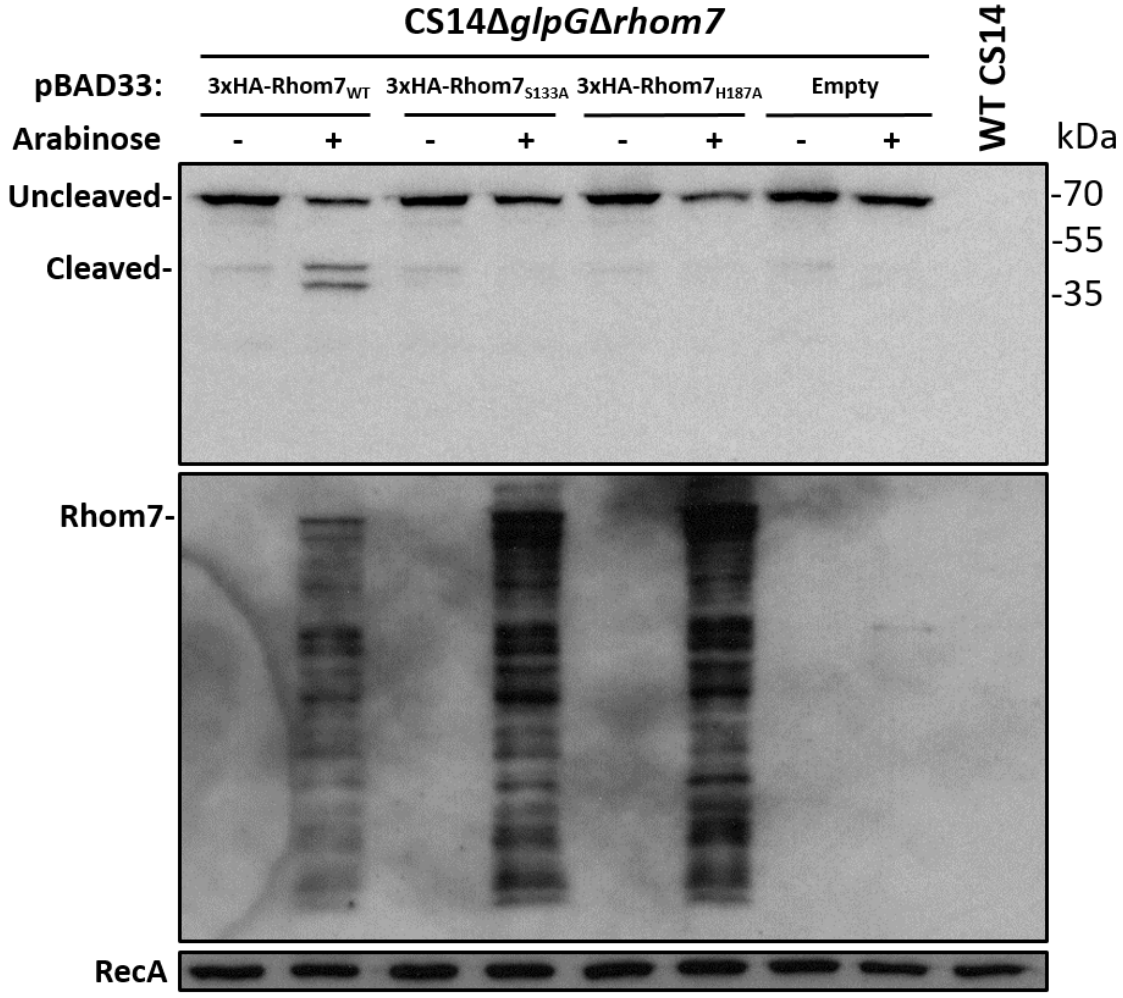


Figure 3.7. Neither S133A or H187A substitution of Rhom7 affects its expression. *CS14ΔglpGΔrhom7 was co-transformed with pKS508 encoding the artificial substrate, and pBAD33 vectors encoding wild-type Rhom7, Rhom7$_{S133A}$ or Rhom7$_{H187A}$ each with an N-terminal triple HA (3xHA) tag. Whole cell lysates were prepared from bacteria grown without or with arabinose induction of protease expression. CS14ΔglpGΔrhom7 transformed with pKS508 and pBAD33 empty vector (shown as -), and a whole cell lysate from WT S. sonnei CS14 were included as controls (n = 1). Cleavage of the artificial substrate was detected with anti-FLAG antibodies. Anti-HA antibodies were used to detect Rhom7 expression. RecA was used as a loading control.*

3.4. The C-terminal domain and 7th transmembrane domain of Rhom7 do not affect Rhom7-mediated cleavage of the TatA derived artificial substrate

As Rhom7 has distinct topology compared with GlpG (**Figure 3.2**), I examined the effect of removing the 7th TMD and the C-terminal domain ($Rhom7_{\Delta CTD\Delta TM7}$), or the CTD alone ($Rhom7_{\Delta CTD}$) on its activity. pBAD33 derivatives were generated expressing $Rhom7_{\Delta CTD}$ (amino acid 1-236) and $Rhom7_{\Delta CTD\Delta TM7}$ (amino acid 1-236). Results demonstrated that both the CTD and 7th TMD are dispensable for Rhom7 activity, as Rhom7-mediated cleavage of the artificial substrate was still detectable even in the absence of these regions (**Figure 3.8**).

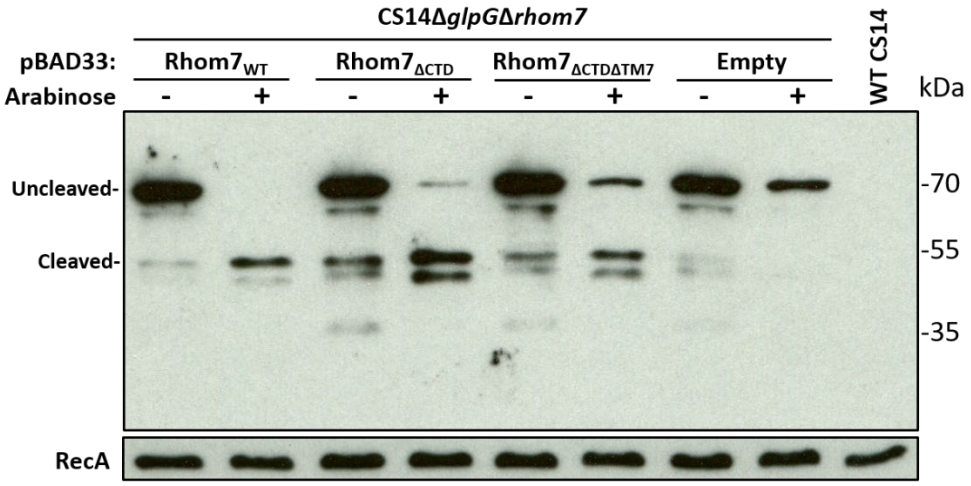


Figure 3.8. The 7th TMD and the CTD of Rhom7 are dispensable for Rhom7 activity. *CS14ΔglpGΔrhom7 was co-transformed with pKS508 encoding the artificial substrate, and pBAD33 vectors encoding wild-type Rhom7, $Rhom7_{\Delta CTD}$ (a.a. 1-236) and $Rhom7_{\Delta CTD\Delta TM7}$ (a.a. 1-213). Whole cell lysates were prepared from bacteria grown without or with arabinose induction of protease expression. CS14ΔglpGΔrhom7 transformed with pKS508 and pBAD33 empty vector, and a whole cell lysate from WT S. sonnei CS14 were included as controls (n = 1). Cleavage of the artificial substrate was detected with anti-FLAG antibodies. RecA was used as a loading control.*

3.5. Conclusion and discussion

In this chapter, two putative rhomboids, GlpG and Rhom7, were identified bioinformatically based on amino acid sequence homology to *E. coli* GlpG and *P. stuartii* AarA, respectively. GlpG possesses a cytosolic Rhomboid-N domain (amino acids 1-84) and a rhomboid core domain consisting of six TMDs (amino acids 95-276), whilst Rhom7 has a six-TMD rhomboid core domain (amino acids 7-201), plus a 7th TMD (amino acids 214-236), and a large C-terminal part (amino acids 237-625) containing a putative WD40 domain (amino acids 297-597). In addition, Alignment of Rhom7 amino acid sequence with that of *P. stuartii AarA* predicted that Ser^{133} or His^{187} constitutes the catalytic dyad of Rhom7. This was supported by the fact that mutation of either Ser^{133} or His^{187} into alanine completely abolished Rhom7 cleavage of an artificial substrate possessing the TMD of *P. stuartii* TatA. Furthermore, both the 7th TMD and the C-terminal part of Rhom7 are dispensable for Rhom7 to cleave the artificial substrate.

A previously described artificial substrate possessing the TMD of *P. stuartii* TatA was used to probe the activity of *S. sonnei* GlpG and Rhom7 as *P. stuartii* TatA is the only known bacterial rhomboid substrate to date (Stevenson *et al.,* 2007; Strisovsky *et al.,* 2009). The main advantage of using this artificial substrate for probing rhomboid proteolytic activity is the ease of detecting cleavage products by western blot analysis using anti-FLAG antibodies. However, this approach is limited by the potential of generating false negative results. For example, the absence of detectable cleavage of the substrate could be due to 1) the rhomboid is catalytically inactive or, 2) the rhomboid is active but does not cleave the TatA TMD due to specificity. Fortunately, our results were not confounded by this issue as both GlpG and Rhom7 demonstrated proteolytic activity against the artificial substrate. Also noticed was that the amount of uncleaved artificial substrate dropped upon arabinose induction (**Figure**

3.5 & 3.6). This was ruled out as a rhomboid-mediated effect as this decrease was observed even in the presence of the empty vector (**Figure 3.5 & 3.6**). Therefore, this difference in uncleaved artificial substrate amount was more likely to stem from metabolic changes induced by arabinose in *S. sonnei*.

S. sonnei GlpG differs from *E. soli* GlpG only by one residue (Ala^{130} in *S. sonnei* is Thr^{130}in *E. coli*). Based on the crystal structure of *E. coli* GlpG, Thr^{130} resides in the α-helix 2 (H2) of the L1 loop that connects TM1 and TM2 (**Figure 1.1**) (Wu *et al.,* 2006). So far, the role of Thr^{130} on GlpG activity has not been characterised, but the sequence alignment of evolutionarily distant rhomboids showed that threonine in this position is not well conserved, suggesting limited impact on GlpG activity (Wu *et al.,* 2006). Here, we confirmed that GlpG bearing an alanine in position 130 still maintains its activity *in vivo*.

Although both GlpG and Rhom7 exhibit proteolytic activity against the artificial substrate, the extent of substrate cleavage was markedly higher by Rhom7 than that of GlpG (**Figure 3.6 & 3.5**). This could be explained by two hypotheses. First, although GlpG and Rhom7 were transcriptionally driven by the same arabinose-inducible promoter on the same plasmid, the final levels of protein upon arabinose induction *in vivo* can be different. This is due to the fact that the native ribosomal binding sites of GlpG and Rhom7, which could vary in strength, were used and the intrinsic protein stability of each rhomboid could also be different. Alternatively, this difference in cleavage efficiency could result from difference in the intrinsic proteolytic activity of GlpG and Rhom7 towards the artificial substrate. This hypothesis is supported by the fact that Rhom7 is more homologous to *P. stuartii* AarA, the

rhomboid protease that cleaves TatA in physiological conditions, than GlpG (Stevenson *et al.,* 2007).

All active rhomboids known so far contain a rhomboid core domain consisting of six TMDs and may contain an additional 7th TMD either N-terminal (*e.g.,* human mitochondrial rhomboid PARL) or C-terminal (*e.g., P. stuartii* AarA) to the core domain (Koonin *et al.,* 2003; Saita *et al.,* 2017; Stevenson *et al.,* 2007). However, the role of these 7th TMDs has not been investigated. The results of this work demonstrate that the 7th TMD is not necessary for Rhom7 to cleave, at least, the artificial substrate. Thus, one role this TMD might be playing is to act as a topological determinant to place the C-terminal domain of Rhom7 in the periplasm.

Rhom7 is predicted to contain a C-terminal WD40 domain (cd00200, amino acids 297-597), a β-propeller domain composed of multiple repeats each with the length of around 40 amino acids containing the tryptophan-aspartic acid (WD) motif (Hu *et al.,* 2017). WD40 domain-containing proteins constitute one of the most abundant protein families in eukaryotes, however, they are far less abundant in prokaryotes (Hu *et al.,* 2017). WD40 domains act as protein interaction scaffolds coordinating multiprotein complex assemblies (Schapira *et al.,* 2017; Xu and Min, 2011). Therefore, it is conceivable that the predicted WD40 of Rhom7 may influence Rhom7 substrate specificity by mediating Rhom7-substrate interactions. In this work, the predicted WD40 domain was demonstrated to be dispensable for Rhom7 cleavage of the artificial substrate. However, since the substrate used in this assay is artificial and expressed on a multi-copy plasmid, the WD40 domain could still participate in Rhom7 recognition and cleavage of its physiological substrate(s). Thus, identification of

physiological substrate(s) of Rhom7 would be required for further investigation of the role of this domain in regulating Rhom7-mediated intramembrane proteolysis.

Chapter 4. Phenotypic characterisation of *S. sonnei* lacking GlpG and Rhom7

4.1. Biolog high-throughput phenotypic screen of *S. sonnei* CS14 Δ*glpG*Δ*rhom7*

After demonstrating that *S. sonnei* GlpG and Rhom7 are active rhomboids *in vivo*, I sought to determine their biological functions. I began by investigating the impact of loss of rhomboids on *S. sonnei* growth. GlpG and Rhom7 could have overlapping functions, and complement for the absence of each other. Therefore, the growth of CS14Δ*glpG*Δ*rhom7* (which lacks both rhomboids) was compared with the wild-type strain. There was no significant difference in the OD_{600} or CFU during aerobic growth of wild-type and CS14Δ*glpG*Δ*rhom7* in LB (**Figure 4.1**), consistent with the lack of a growth defect for *E. coli* Δ*glpG* (Clemmer *et al.,* 2006).

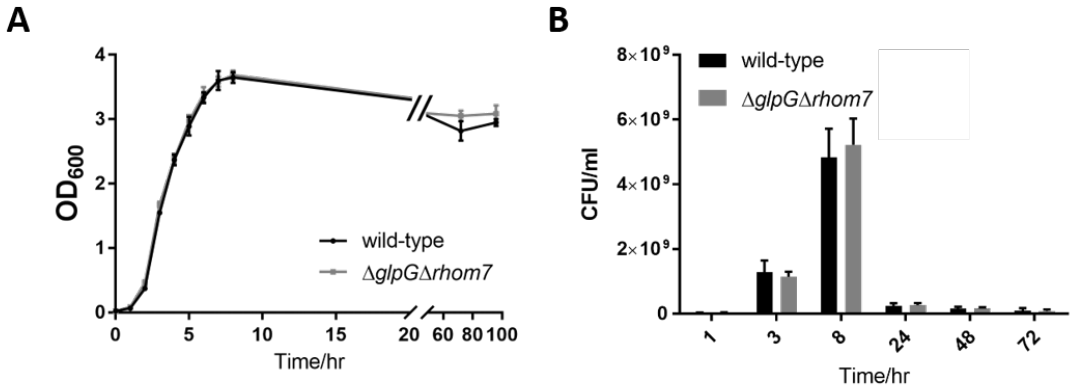


Figure 4.1. GlpG and Rhom7 are not required for aerobic growth of *S. sonnei* in LB. ***(A)*** *Wild-type S. sonnei CS14 and ΔglpGΔrhom7 were grown aerobically at 37°C from an initial OD_{600} of 0.02 in LB. Growth was measured by optical density at 600 nm.* ***(B)*** *Colony forming units of wild-type S. sonnei CS14 and ΔglpGΔrhom7 were enumerated from the same cultures used in* ***(A)****. Data shown in* ***(A)*** *and* ***(B)*** *are mean ± standard deviation from three independent experiments.*

To gain insights into the roles of rhomboids in *Shigella*, Biolog phenotype microarrays (PM1-PM10) were performed. The microarrays measure the aerobic growth of bacteria in more than 750 nutritional conditions including a variety of carbon, nitrogen, phosphorus, and sulphur sources; and more than 190 stress conditions covering various osmotic and pH stresses (Bochner *et al.,* 2001). The assays were performed by the company using wild-type CS14, CS14Δ*glpG*Δ*rhom7,* and CS14Δ*glpG*Δ*rhom7*+*glpG*, in which *glpG* was complemented on the chromosome under the control of a constitutive promoter. There were no statistically significant differences in growth of either mutant compared to the wild-type in any of the conditions tested (**Figure 4.2**, Appendices).

4.2. Growth of *S. sonnei* CS14Δ*glpG*Δ*rhom7* in stress conditions

The lack of phenotype being revealed by Biolog microarrays suggests that GlpG and/or Rhom7 may not have a function in standard laboratory conditions. In addition, Biolog microarrays do not cover oxidative stress, for which the defence mechanisms are critical for *Shigella* pathogenesis (Franzon *et al.,* 1990). Furthermore, *glpG* is in an operon with *glpE*, which encodes a thiosulphate:cyanide sulphurtransferase (Ray *et al.,* 2000). GlpE serves as a sulphur donor for the antioxidant thioredoxin 1 which mediates the assembly of the [2Fe-2S] cluster in SoxR, a major transcriptional regulator of the oxidative stress response in *E. coli* (Ding and Demple, 1998; Ray *et al.,* 2000). The *glpEGR* operonic structure suggests that GlpG might also be involved in oxidative stress. Therefore, I next examined whether GlpG and/or Rhom7 affect survival in the presence of oxidative stress.

Bacterial survival was evaluated following stress induced by the ROS-generating agents, hydrogen peroxide (H_2O_2) and paraquat (PQ) (Hassan and Fridovich, 1979; Imlay, 2013). PQ is a redox-cycling drug that extracts single electrons from the reduced flavins or metal

centres of redox enzymes. Reduced PQ can then transfer the electron to O_2, generating superoxide (O_2^-), which can be further reduced to form H_2O_2 (Hassan and Fridovich, 1979). Both O_2^- and H_2O_2 cause bacterial stress by oxidising proteins, lipids, carbohydrates and nucleic acids (Imlay, 2013).

As transcriptional responses to oxidative stress is a gradual adaptation process (Daugherty *et al.,* 2012), bacteria were initially grown to log phase, exposed to a sub-lethal dose of H_2O_2 (200 μM) for 20 minutes, then incubated in presence of 2 mM H_2O_2 for 15 minutes, after which the number of viable bacteria was measured by plating to solid media. Exposure of bacteria to 2 mM H_2O_2 without preincubation led to an average of 79% and 77% reduction (n = 3) in survival of wild-type CS14 (± 8.5% standard deviation) and CS14Δ*glpG*Δ*rhom7* (± 17% standard deviation), respectively, compared to bacteria not exposed to H_2O_2. Pre-incubation in 200 μM H_2O_2 increased survival from ≈20% to ≈50% for both strains in 2 mM H_2O_2, however, there was no statistically significant difference in survival of CS14 wild-type compared to CS14Δ*glpG*Δ*rhom7* (**Figure 4.3A**).

We next assessed the effect of PQ on strains with or without rhomboids. As *E. coli* is more resistant to PQ when grown in LB due to the protection conferred by polyamines in the medium (Chattopadhyay *et al.,* 2003; Minton *et al.,* 1990), CS14 wild-type and CS14Δ*glpG*Δ*rhom7* were grown in M9 minimal medium, incubated in low dose PQ (200 μM for 15 minutes), then exposed to 10 mM PQ for 45 minutes (Tamarit *et al.,* 1998). Again there was no significant difference in survival of wild-type *S. sonnei* and CS14Δ*glpG*Δ*rhom7* (**Figure 4.3B**). Therefore, GlpG and Rhom7 are not required for survival in oxidative stress induced by H_2O_2 or PQ.

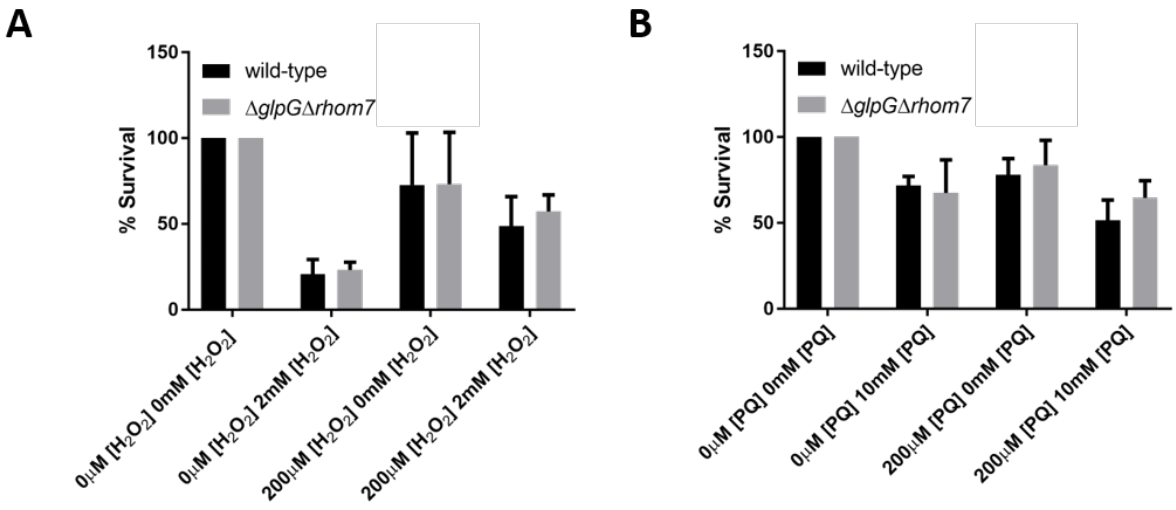


Figure 4.3. GlpG and Rhom7 do not impact *S. sonnei* survival in oxidative stress induced by H_2O_2 or PQ. ***(A)*** *Resistance of CS14 wild-type and CS14ΔglpGΔrhom7 to H_2O_2. S. sonnei was grown in LB to an OD_{600} of ≈ 0.5 then pre-incubated in 0 or 200 μM H_2O_2 for 20 mins before the incubation in 0 or 2 mM H_2O_2 for 15 minutes (pre-incubation/challenge shown below lanes). Bacterial survival was measured by plating. The percentage survival was calculated as the ratio of CFU recovered from H_2O_2-exposed bacteria to those that were not exposed to H_2O_2.* ***(B)*** *Resistance of WT S. sonnei and CS14ΔglpGΔrhom7 to paraquat (PQ). S. sonnei was grown as above, then pre-incubated in 0 or 200 μM PQ for 15 minutes before the incubation in 0 or 10 mM PQ for 45 minutes before plating to measure surviving bacteria. The percentage survival was calculated as the ratio of CFU recovered from PQ-treated bacteria to those that were not treated by PQ. Assays performed in* ***(A)*** *and* ***(B)*** *were performed by both Guangyu Liu and Miranda Rogers, an FHS project student from Sommerville college. Data shown in* ***(A)*** *and* ***(B)*** *are mean ± standard deviation from three independent experiments.*

4.3. Deletion of *glpG* and *rhom7* does not impair T3SS secretion in *S. sonnei* CS14

Secretion of effector proteins through the T3SS is critical for the virulence of *S. sonnei* (Buchrieser *et al.,* 2000). Several components of the T3SS are integral inner membrane proteins with one or multiple TMDs *e.g.* MxiA (eight TMDs), Spa40 (four putative TMDs), MxiG (one TMD), and MxiJ (one TMD) (Deane *et al.,* 2008; Galan *et al.,* 2014; Kuhlen *et al.,* 2018; McDowell *et al.,* 2011). As both rhomboids are also expected to reside in the inner membrane, we hypothesised that they could process inner membrane components of the T3SS, and affect T3SS assembly and function. Therefore, the effect of GlpG and Rhom7 on the secretion of effectors through the T3SS was examined. T3SS secretion was induced by the addtion of Congo Red (Bahrani *et al.,* 1997) to bacteria grown to exponential phase. The profile of secreted proteins assayed by SDS-PAGE and silver staining was annotated based on comparison with *S. flexneri* M90T which has a defined secretion profile (Roehrich *et al.,* 2010). There was no detectable difference in the effectors secreted by CS14 wild-type and CS14Δ*glpG*Δ*rhom7* (**Figure 4.4**). Thus, loss of GlpG and Rhom7 does not have a detectable impact on secretion through the T3SS in the presence of Congo Red.

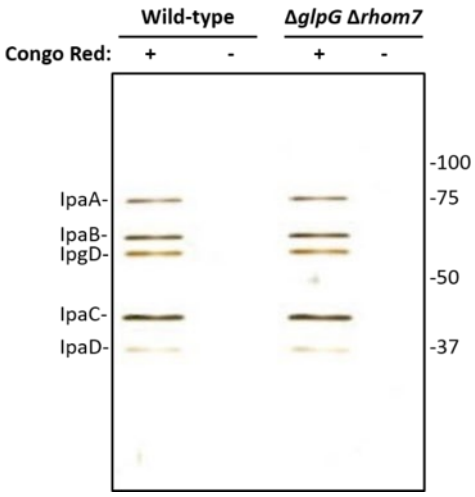


Figure 4.4. T3SS-mediated secretion of Ipa proteins of *S. sonnei* is independent of rhomboid proteases. *T3SS-mediated secretion by CS14 wild-type and CS14ΔglpGΔrhom7. Bacteria were grown to exponential phase and Congo red was added to a final concentration of 200 µg/ml to induce secretion through the T3SS. Bacterial supernatants were analysed by SDS-PAGE and silver staining. Effectors were identified based on size and previously described secretion profiles (McDowell et al., 2011).*

4.4. Conclusion and discussion

In this chapter, efforts were made to identify the biological function(s) of *S. sonnei* GlpG and Rhom7 by characterising the phenotypes of a strain lacking rhomboids (CS14Δ*glpG*Δ*rhom7*) compared to wild-type. The rationale behind deleting both GlpG and Rhom7 was to eliminate the possibility that putative functional redundancy between GlpG and Rhom7 could mask phenotypic effects of loss of a single rhomboid. The lack of a discernable impact on growth from loss of GlpG and Rhom7 in rich medium led us to invest in the high-throughput Biolog phenotypic microarrays, which measure the aerobic cellular respiration of bacteria in more than 950 nutritional and stress conditions. CS14Δ*glpG*Δ*rhom7*+*glpG* was included as the third strain for the Biolog screen, as chromosomal complementation of *glpG* under the control of a constititute promoter ensured *glpG* expression. The reason for including this strain is two-fold: first, if a phenotypic difference is observed between CS14 wild-type and CS14Δ*glpG*Δ*rhom7*, this strain would help identify which rhomboid is mediating the difference. Secondly, under circumstances when the endogenous *glpG* expression is switched off in CS14 wild-type, this strain might reveal certain non-physiological but potentially informative phenotypes due to GlpG over-expression. Although this screen still revealed no significant differences between CS14 wild-type and CS14Δ*glpG*Δ*rhom7*, the comprehensive list of nutritional conditions covered by the screen gave us enough confidence to cease further exploring the role of GlpG and Rhom7 in aerobic metabolism of *S. sonnei*.

GlpG was recently implicated in facilitating intestinal colonisation of extraintestinal pathogenic *E. coli* (ExPEC) in a murine model, however, the underlying mechanism remains unknown (Russell *et al.,* 2017). The pH stresses covered by Biolog microarrays are relevant for *S. sonnei* in infection settings, for example in stomach (Waterman and Small, 1996), and

the results suggest that, given the high genetic similarity between ExPEC and *Shigella*, the colonisation defect exhibited by the Δ*glpG* ExPEC was probably not due to compromised acid resistance.

The potential role of GlpG and Rhom7 in mediating oxidative stress response was investigated next as oxidative stress response is critical for *Shigella* to estabilish infection (Franzon *et al.,* 1990) and was not covered by Biolog microarrays. Internal ROS production initiates from molecular oxygen becoming partially reduced by redox enzymes, many of which are involved in respiration in the inner membrane (Unden *et al.,* 2014). Moreover, regulated intramembrane proteolysis (RIP) is widely involved in initiating signal transduction cascades (Brown *et al.,* 2000). It is therefore conceivable that rhomboids, which also reside in the inner membrane, might be able to sense local ROS concentration and transduce this information to initiate downstream stress responses. However, the results showed neither GlpG nor Rhom7 plays any significant roles in oxidative stress response in *S. sonnei*, suggesting that the colonisation defect exhibited by the Δ*glpG* ExPEC was probably not caused by attenuated oxidative stress response.

During infection, effector proteins secreted through T3SS play key roles in multiple stages of *Shigella*-host interaction and a T3SS-defective *Shigella* is completely avirulent (Buchrieser *et al.,* 2000; Sansonetti *et al.,* 1981). Rhomboids have been demonstrated to regulate secretion in both *Drosophila*, where rhomboid-1 regulates EGF ligand secretion, and in bacteria, where AarA-mediated N-terminal processing of TatA is essential for Tat transport in *P. stuartii* (Stevenson *et al.,* 2007; Urban *et al.,* 2001). Therefore, I speculated that GlpG and/or Rhom7 might modulate T3SS-mediated secretion through cleaving the integral membrane components of the T3SS in the inner membrane of *S. sonnei*. There are

four T3SS components known to be integral inner membrane proteins with one or more TMDs. MxiA (eight TMDs) is likely to form a nonameric ring in the T3SS apparatus and plays a key role in substrate selection (Abrusci *et al.,* 2013; Shen and Blocker, 2016). Spa40 (four putative TMDs) is part of the T3SS basal body and has been implicated in regulating the switching of substrate specificity (Deane *et al.,* 2008). MxiG is a type II memrane protein (single TMD, N-terminus in the cytoplasm) proposed to form a putative 24-membered ring, together with MxiJ (one TMD), constituting the inner membrane ring of the T3SS apparatus (Bergeron *et al.,* 2015; Kimbrough and Miller, 2000; McDowell *et al.,* 2011; Schraidt *et al.,* 2010). The T3SS-mediated secretion profile was assessed to infer T3SS assembly, which was not altered by the absence of both rhomboid proteases. These results suggest that GlpG and Rhom7 do not play any significant roles in T3SS-mediated secretion. This is result is not too surprising as among the four inner membrane-residing T3SS subunits, MxiA and Spa40 possess multiple TMDs and MxiG is a type II membrane protein, which are not typical rhomboid substrates found in other organisms. MxiJ, which possesses a type I C-terminal TMD, is confined in the inner part of the T3SS appratus and hence probably not accessible by GlpG and Rhom7.

Chapter 5. Identification of GlpG and Rhom7 substrates

5.1. Design of a genome-wide screen to search for substrates of GlpG and Rhom7

As the phenotypic characterisation of *S. sonnei* lacking rhomboid proteases in chapter four failed to identify any biological functions to either GlpG or Rhom7 in *S. sonnei*, I decided to take a substrate-oriented approach. The rationale behind this is that if substrates can be identified, this could shed light on the biological roles of rhomboids and provide avenues for more detailed analysis of their contribution to bacterial physiology and pathogenesis.

Rhomboids have propensity to cleave single-pass membrane proteins with a periplasmic N-terminus and a cytosolic C-terminus, *i.e.* type I (with a signal peptide) and type III (without a signal peptide) membrane proteins (Strisovsky *et al.,* 2009). Therefore, to identify the potential substrates of GlpG and/or Rhom7 in *S. sonnei*, we undertook a bioinformatic screen to identify type I and type III proteins, by predicting the membrane location and topology of all proteins in *S. sonnei*. To do this, TMHMM analysis was performed on amino acid sequences of the predicted proteome of *S. sonnei* SS046, the most closely related *S. sonnei* strain to CS14 with entire genome sequenced and bioinformatically annotated. TMHMM identifies TMDs in proteins based on a hidden Markov model method (HMM), which incorporates hydrophobicity, charge bias, and lengths of α-helices into one model (Krogh *et al.,* 2001). The output of TMHMM analysis, *i.e.* graphical representations of the topology of all proteins (n = 4911) were used to identify type I and type III proteins (**Figure 5.1**). By only retaining hits that are predicted to possess one TMD and a cytoplasmic C-terminus, the search yielded 16 initial potential rhomboid substrates. However, TMHMM also has several limitations. For instance, many proteins were predicted to possess TMD-like regions but were below the threshold to be classified as a TMD. In addition, the N-terminal signal

peptide of type I proteins resemble genuine TMDs, and hence these proteins can be classified as having two TMDs. Both cases would under-estimate the number of type I and type III proteins. Therefore, the predicted TMDs of the 16 candidate proteins were then used to identify further Type I and III proteins by performing searches using HHpred, a server with high sensitivity for detecting protein homologues based on their structural similarity (Soding *et al.,* 2005). One limitation of HHpred is that its database does not include a proteome of *S. sonnei*. Therefore, *E. coli* MG1655 proteome was chosen for protein homology searches. Additional candidates identified in this way were analysed with Phobius 1.01 and TOPCONS 2.0, two TMD prediction servers with improved ability to predict signal peptide, and to refine topology predictions (Kall *et al.,* 2007; Tsirigos *et al.,* 2015). Finally, the presence of candidates in *S. sonnei* CS14 was verified using the whole genome sequence of CS14. HHpred searches were then performed on all newly identified candidates. The process was run iteratively until no further candidates with predicted type I or type III topology were identified. This approach yielded 44 potential rhomboid substrates (**Figure 5.1**).

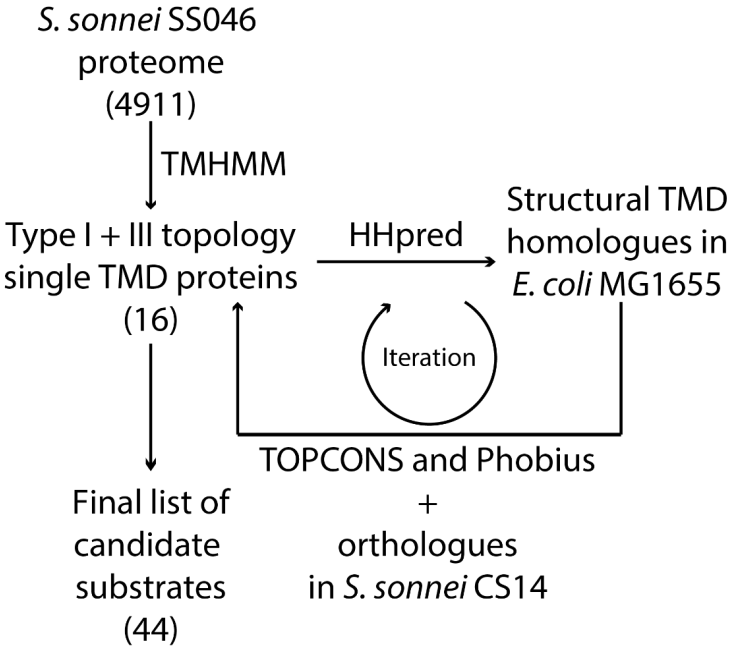


Figure 5.1. The workflow for the bioinformatic identification of *S. sonnei* candidate rhomboid substrates. *The number of candidate proteins at each stage is shown in brackets.*

5.2. Construction of artificial substrates containing candidate substrate transmembrane domains

To examine whether the proteins are cleaved by the *S. sonnei* rhomboids GlpG and Rhom7, I used the same approach to detect rhomboid-mediated proteolysis as described in section 3.2, except the TMD of TatA in the artificial substrate was replaced with the predicted TMD and 16 residues on the periplasmic side (if the periplasmic region is shorter than 16 residues, then the whole periplasmic region was included) of each of the 44 candidates (**Figure 5.2 and Table 5.1**). For some candidates such as HybA, a few residues C-terminal of the predicted TMD were also included due to their positve charge, which can be important for TMDs to adopt type I topology due to the 'positive-inside' rule (Elazar *et al.,* 2016).

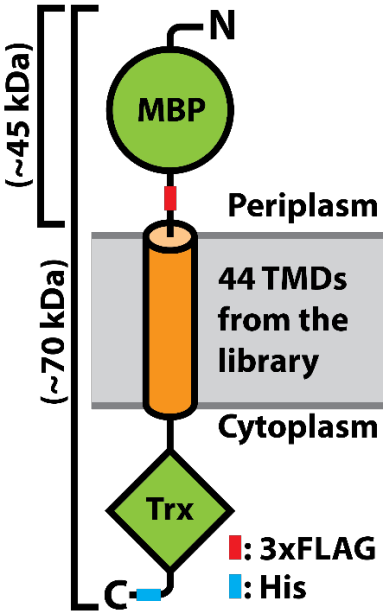


Figure 5.2. Schematic representation of the artificial substrate used for screening. *The substrate possesses an N-terminal MBP linked to TMDs from each of the 44 candidate substrates by a triple-FLAG tag (3xFLAG, in red), and a thioredoxin domain (Trx) with a C-terminal His tag (in blue).*

Table 5.1. The sequence of the predicted TMDs and additional amino acids of putative rhomboid substrates from *S. sonnei* CS14 used as substrates for *in vivo* screening.

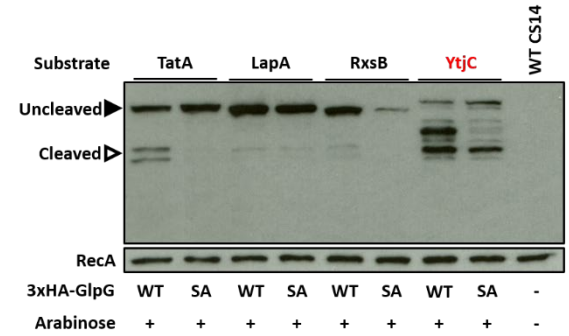
Putative substrate	**Sequence inserted into artificial substrate. Predicted TMD is underlined.**
CjrB	MNILHFSQSVKWSSWFICSLLLHGLIFLAFIWRFS
BcsB	GDVYYVGHLPWFERLWYALANHPILLAVLAAISVILLAWVLWRLL
CcmD	TPAFASWNEFFAMGGYAFFVWLAVVMTVIPLVVLVVHSVM
DgcJ	YVHDLPGGIRVSLSIDILYFITSSWKSVLFWILTALILLNMVRM
DjlA	QYWGKIIGVAVALLMGGGFWGVVLGLLIGHM
DjlB	ETVEPEPQPQPMPVDDGGLGCLGVIKIIFYIFIFAGLIGKILHLF
ElaB	YYYRAKQAVYRADDYVHEKPWQGIGVGAAVGLVLGLLLA
FdnH	LPKDPKIDTSVSLWKGALKPLAAAGFIATFAGLIFHYIGIG
FdoH	LPENPEISETVKFWKGIWKPLAAVGFAATFAASIFHYVGVGP
FliO	NNHATVQSSTPVSAAPLLQVSGALIAIIALILAAAWLV
Flk	NPFPPMMDTLQNMATRPALWILLVAIILMLVWLV
FtsH	LTKNVKVVGEPPEEPSLLASIFISWFPMLLLIGVWIFFMRQMQGG
HflK	GTGSGGGSSSQGPRPQLGGRVVTIAAAAIVIIWAASGFYTI
HyaA	SRVVDIPQMGTHSTADTVGLTALGVVAAAVGVHAVASAV
HybA	KLDDLSTGARSENIQHTLYKGMMLPLAVLAGLTVLVRR
HybO	TPRSQKPDVNAKEGGNVSAGAIGLLGGVVGLVAGVSVMAV
LapA	QQVTFNYLLAQGEYRISTLLAVLFAAGVAIGWLICGLFWLRV

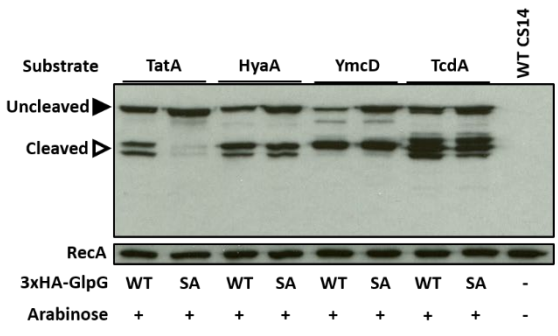
MxiJ	YVYTNVQPVKEIKSEFLTNEVIYLFLGMAVLVVILLVWAFKT GWF
NrfF	TERYGDFVRYNPPLTGQTLVLWALPVVLLLLMALILWRV
PpdC	SASLKNQQGFSLPEVMVAMVLMVMIVTALSGIQ
RS17255	SAQFWNRASIKQKIPLTILYILLFSVFLRLFFRHT
RsxB	NAIWIAVAAVSLLGLAFGAILGYAS
SecG	GASFGAGASATLFGSSGSGNFMTRMTALLATLFFIISLVLGNI
SohB	ELLSEYGLFLAKIVTVVLAIAAIAAIIVNVA
TcdA	VCAMKATAEGPKRMDCASGFGAATMVTATFGFVAVSHAL
TorS	NQHGLAHQEKASARGQYSLLLLGMVSLCALILILWRVVY
YajC	FFISDAVAATGAPAQGSPMSLILMLVVFGLIFYFMILR
YbdJ	RKHPGLFIGMYVAFFATLAVMLQSETLSGSVWLLVVLFILL
YbjT	SLEDKGDYRTIDVRAFWHPHGMPGLFYWLLMIPAHLFIFRG
YebO	NEVVNSGVMNIASLVVSVVVLLIGLILWFFI
YgaM	AARDAVGCADSFVRERPWCSVGTAAAVGIFIGALLSMR
YgiM	AQKKVDAASVQLDDKQRTIIMQWFMYGGGVLGLGLLLGLVL PHLI
YhcB	TWEYALIGLVVGIIIGAVAMRFGN
YhdP	VRRDLNMEAVVAPEISATVGVAAAFAVNPIVGAAVFAASKV L
YhhM	SKPPLFFIVIIGLIVVAASFRFM
YibN	QEIMQFVGRHPILSIAWIALLVAVLVTTFKSL
YjeT	MLYPKAWKKMISAMTNLPDNILRRFGGGLVVAGVVVYYML RKTI

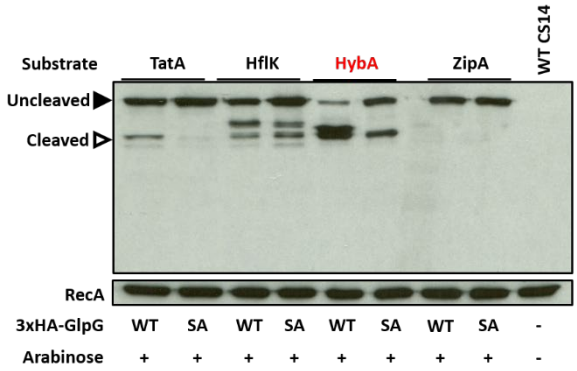
YmcD	TTTTGATTTATTTGGVAAGAVGTATVVGVATAVGVATLAVVA
YnaJ	EQYNIPLSEWTTSMYVIQSSMIFVYSLVFTVLLAIPLGIYFLGG
YqiK	DDIVNSVPSWMFTAIIAVCILFIIGIIFARLY
YqjD	TRVAAARADEYVRENPWTGVGIGAAIGVVLGVLLS
YtjB	DTHTLATEAQQVDNTTNILRLMLLLSLAIGVVLTRTLL
YtjC	NAALESCRDLPQGSRPLLVSHGIALGCLVSTILGLPAWA
ZipA	MQDLRLILIIVGAIAIIALLVHGFWTS

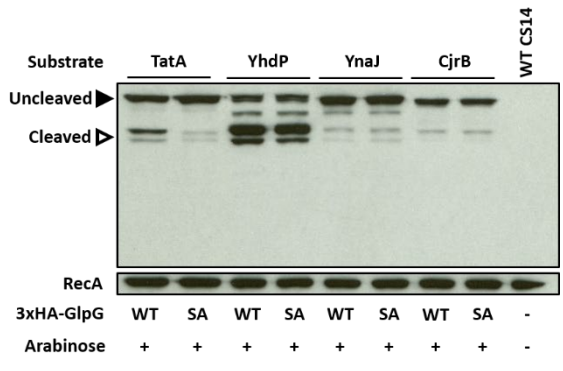
5.3. Results of genome-wide substrate screen and preliminary validation

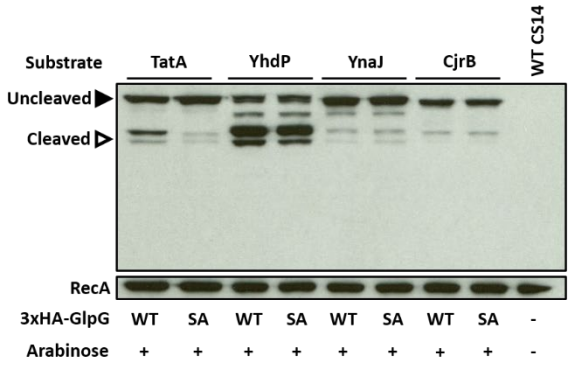
Each of the 44 artificial substrates was then expressed in CS14Δ*glpG*Δ*rhom7* along with pBAD33 harbouring active (wild-type, WT) or inactive (S201A for GlpG, S133A for Rhom7, SA) versions of GlpG or Rhom7. Cleavage of the substrates was detected by western blot analysis using anti-FLAG antibodies after arabinose induction of rhomboid expression. The artificial substrate with the TatA TMD was used as a positive control. The screen was performed once for all 44 candidates. A positive hit was defined when two criteria are met: 1) more cleavage product and 2) less uncleaved substrate were observed when expressed with active rhomboid (WT) compared to the inactive rhomboid (SA) **(Figure 5.3 and 5.4)**.

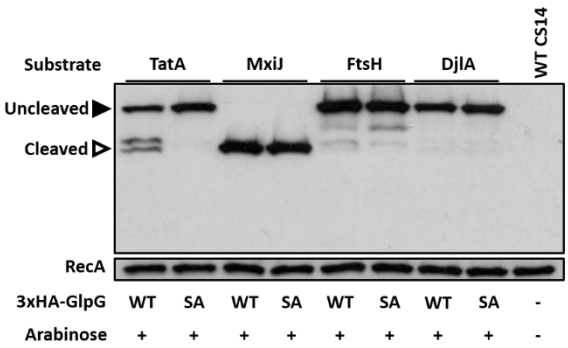

Substrate
TatA
LapA
RxsB
YtjC
WT CS14
Uncleaved
Cleaved
RecA
3xHA-GlpG WT SA WT SA WT SA WT SA -
Arabinose + + + + + + + + -

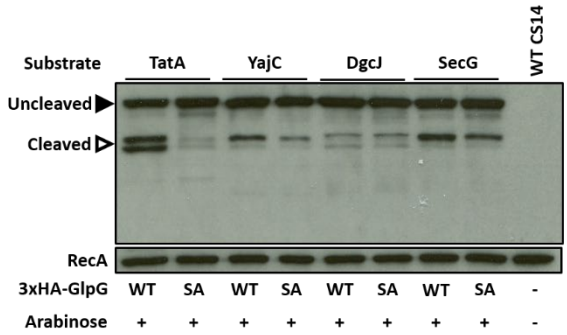

Substrate
TatA
HyaA
YmcD
TcdA
WT CS14
Uncleaved
Cleaved
RecA
3xHA-GlpG WT SA WT SA WT SA WT SA -
Arabinose + + + + + + + + -

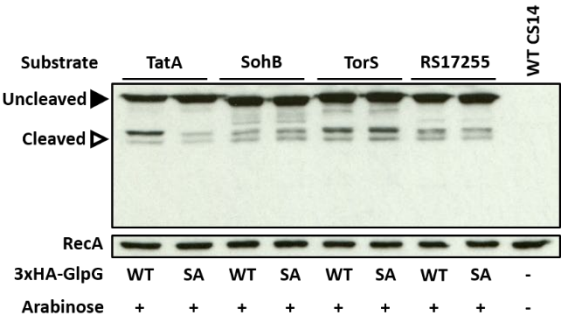

Substrate
TatA
HflK
HybA
ZipA
WT CS14
Uncleaved
Cleaved
RecA
3xHA-GlpG WT SA WT SA WT SA WT SA -
Arabinose + + + + + + + + -


Substrate
TatA
YhdP
YnaJ
CjrB
WT CS14
Uncleaved
Cleaved
RecA
3xHA-GlpG WT SA WT SA WT SA WT SA -
Arabinose + + + + + + + + -


Substrate
TatA
YhdP
YnaJ
CjrB
WT CS14
Uncleaved
Cleaved
RecA
3xHA-GlpG WT SA WT SA WT SA WT SA -
Arabinose + + + + + + + + -


Substrate
TatA
MxiJ
FtsH
DjlA
WT CS14
Uncleaved
Cleaved
RecA
3xHA-GlpG WT SA WT SA WT SA WT SA -
Arabinose + + + + + + + + -


Substrate
TatA
YajC
DgcJ
SecG
WT CS14
Uncleaved
Cleaved
RecA
3xHA-GlpG WT SA WT SA WT SA WT SA -
Arabinose + + + + + + + + -


Substrate
TatA
SohB
TorS
RS17255
WT CS14
Uncleaved
Cleaved
RecA
3xHA-GlpG WT SA WT SA WT SA WT SA -
Arabinose + + + + + + + + -

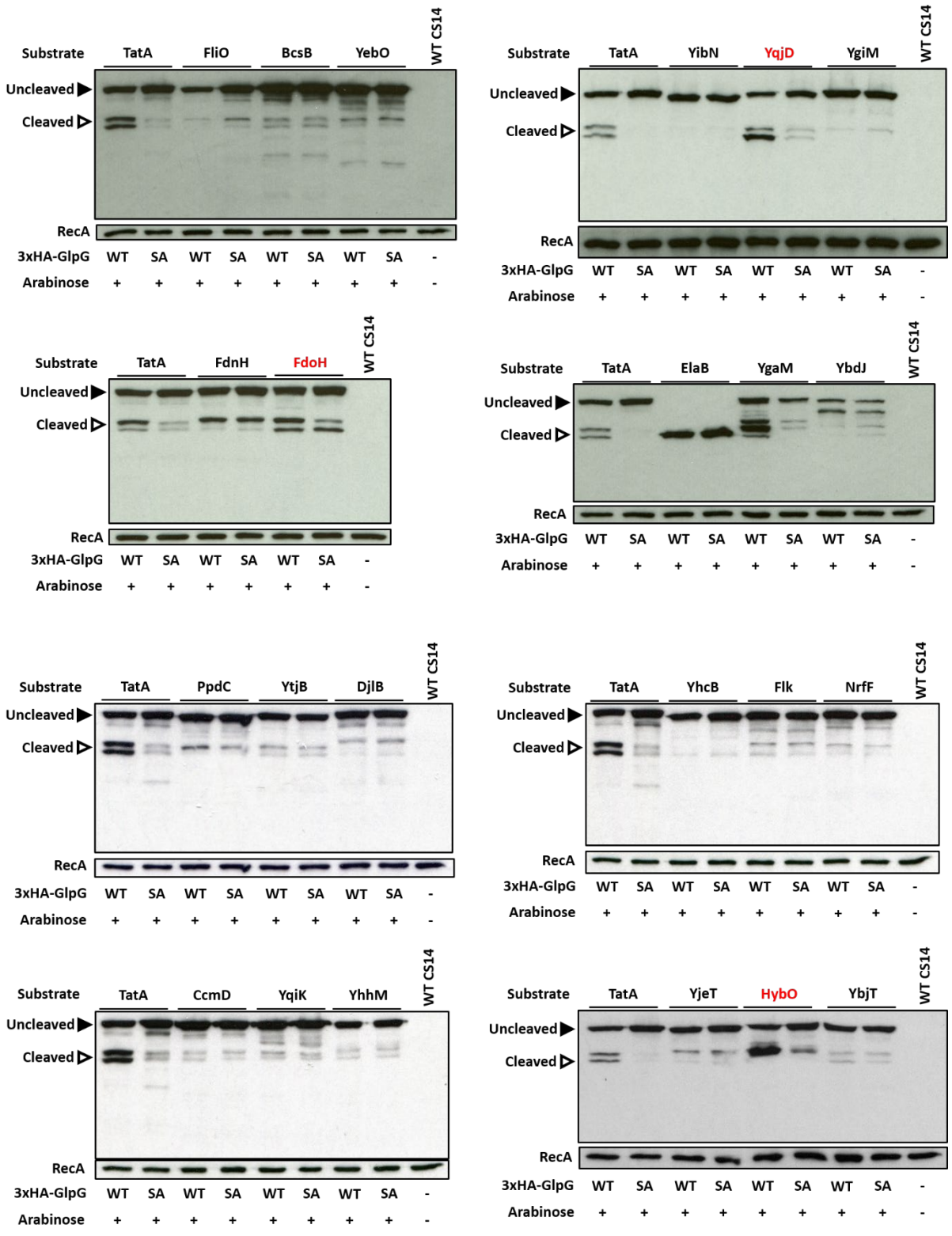
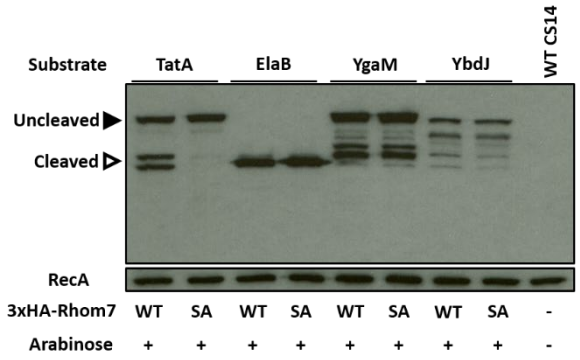
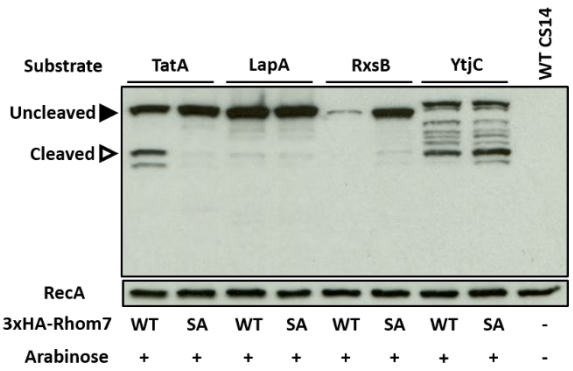
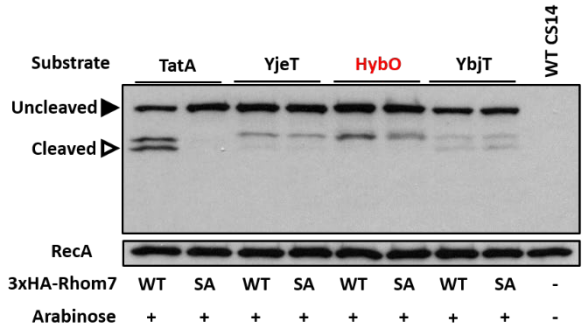


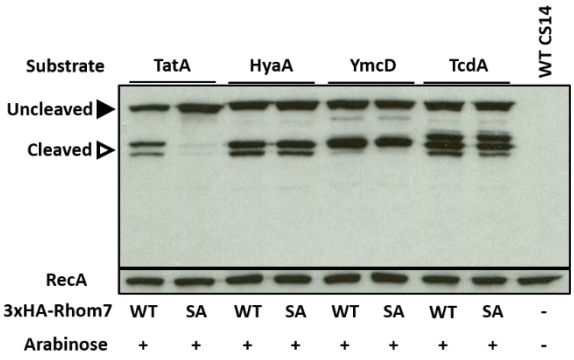
Figure 5.3. Cleavage of candidate substrates by GlpG. *The screen was performed with S. sonnei CS14ΔglpGΔrhom7 containing pKS508 encoding a putative substrate TMD, and pBAD33 encoding wild-type (WT) or inactive (SA) forms of GlpG. Cleavage was ascertained*

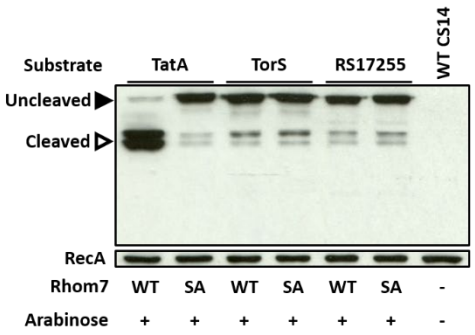
by western blot analysis using anti-FLAG antibodies. Positive hits (in red) were those for which there was an increase in the cleavage product (approximately 35 kDa) in the presence of the wild-type enzyme and a decrease in the amount of uncleaved substrate. RecA was used as a loading control. Differences in background intensity are due to different scanners used.

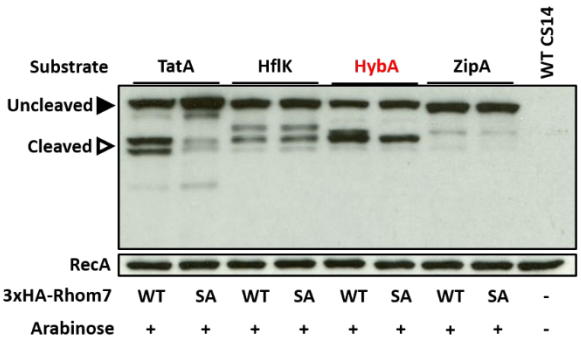

Substrate
TatA
ElaB
YgaM
YbdJ
WT CS14
Uncleaved
Cleaved
RecA
3xHA-Rhom7 WT SA WT SA WT SA WT SA -
Arabinose + + + + + + + + -

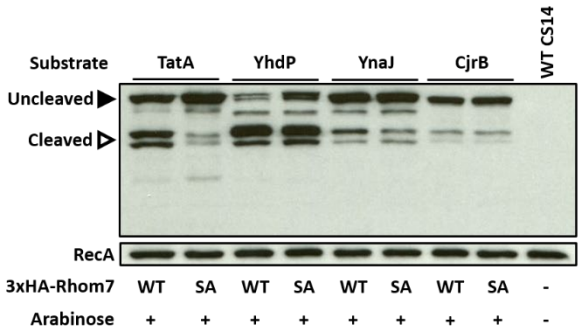

Substrate
TatA
LapA
RxsB
YtjC
WT CS14
Uncleaved
Cleaved
RecA
3xHA-Rhom7 WT SA WT SA WT SA WT SA -
Arabinose + + + + + + + + -

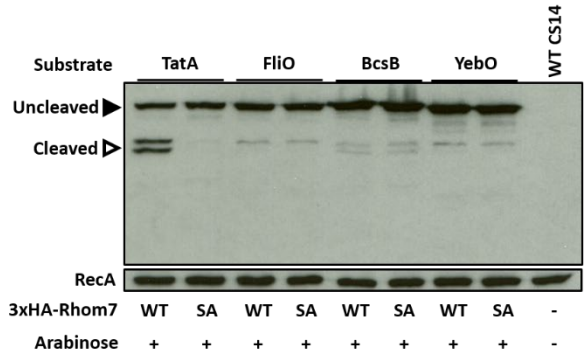

Substrate
TatA
YjeT
HybO
YbjT
WT CS14
Uncleaved
Cleaved
RecA
3xHA-Rhom7 WT SA WT SA WT SA WT SA -
Arabinose + + + + + + + + -


Substrate
TatA
HyaA
YmcD
TcdA
WT CS14
Uncleaved
Cleaved
RecA
3xHA-Rhom7 WT SA WT SA WT SA WT SA -
Arabinose + + + + + + + + -


Substrate
TatA
TorS
RS17255
WT CS14
Uncleaved
Cleaved
RecA
Rhom7 WT SA WT SA WT SA -
Arabinose + + + + + + -


Substrate
TatA
HflK
HybA
ZipA
WT CS14
Uncleaved
Cleaved
RecA
3xHA-Rhom7 WT SA WT SA WT SA WT SA -
Arabinose + + + + + + + + -


Substrate
TatA
YhdP
YnaJ
CjrB
WT CS14
Uncleaved
Cleaved
RecA
3xHA-Rhom7 WT SA WT SA WT SA WT SA -
Arabinose + + + + + + + + -


Substrate
TatA
FliO
BcsB
YebO
WT CS14
Uncleaved
Cleaved
RecA
3xHA-Rhom7 WT SA WT SA WT SA WT SA -
Arabinose + + + + + + + + -

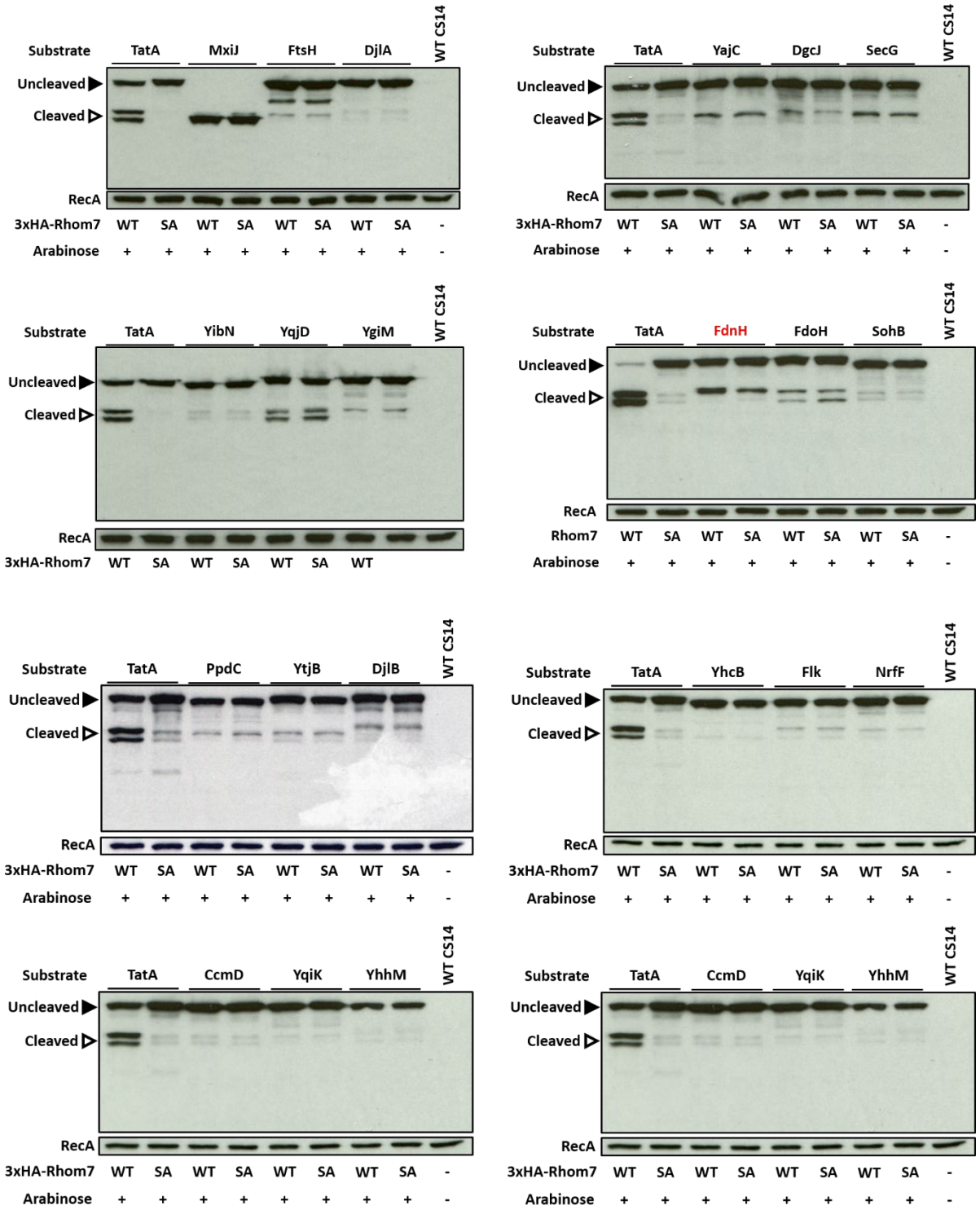


Figure 5.4. Cleavage of candidate substrates by Rhom7. *The screen was performed with S. sonnei CS14ΔglpGΔrhom7 containing pKS508 encoding a putative substrate TMD, and pBAD33 encoding wild-type (WT) or inactive (SA) forms of Rhom7. Assays were performed using eithet untagged Rhom7(Rhom7) or triple HA tagged Rhom7 (3xHA-Rhom7). Cleavage*

was ascertained by western blot analysis using anti-FLAG antibodies. Positive hits (in red) were those for which there was an increase in the cleavage product (approximately 35 kDa) in the presence of the wild-type enzyme and a decrease in the amount of uncleaved substrate. RecA was used as a loading control. Differences in background intensity are due to different scanners used.

Out of 44 tested candidate substrates, there was no evidence of cleavage for 29 TMDs. Nine proteins (RxsB, HyaA, SecG, YmcD, FliO, YhdP, YajC, DgcJ, and YgaM) only met one defining criterion *i.e.* only more cleavage product or less uncleaved substrate was observed when expressed with active rhomboid (WT) compared to the inactive rhomboid (SA). These 38 hits were excluded from further analysis. Six candidate TMDs (YqjD, HybO, HybA, FdoH, YtjC, and FdnH) showed susceptibility to cleavage by one or both of the active rhomboids. Therefore, cleavage assays were repeated with these substrates to verify the findings (**Figure 5.5**).

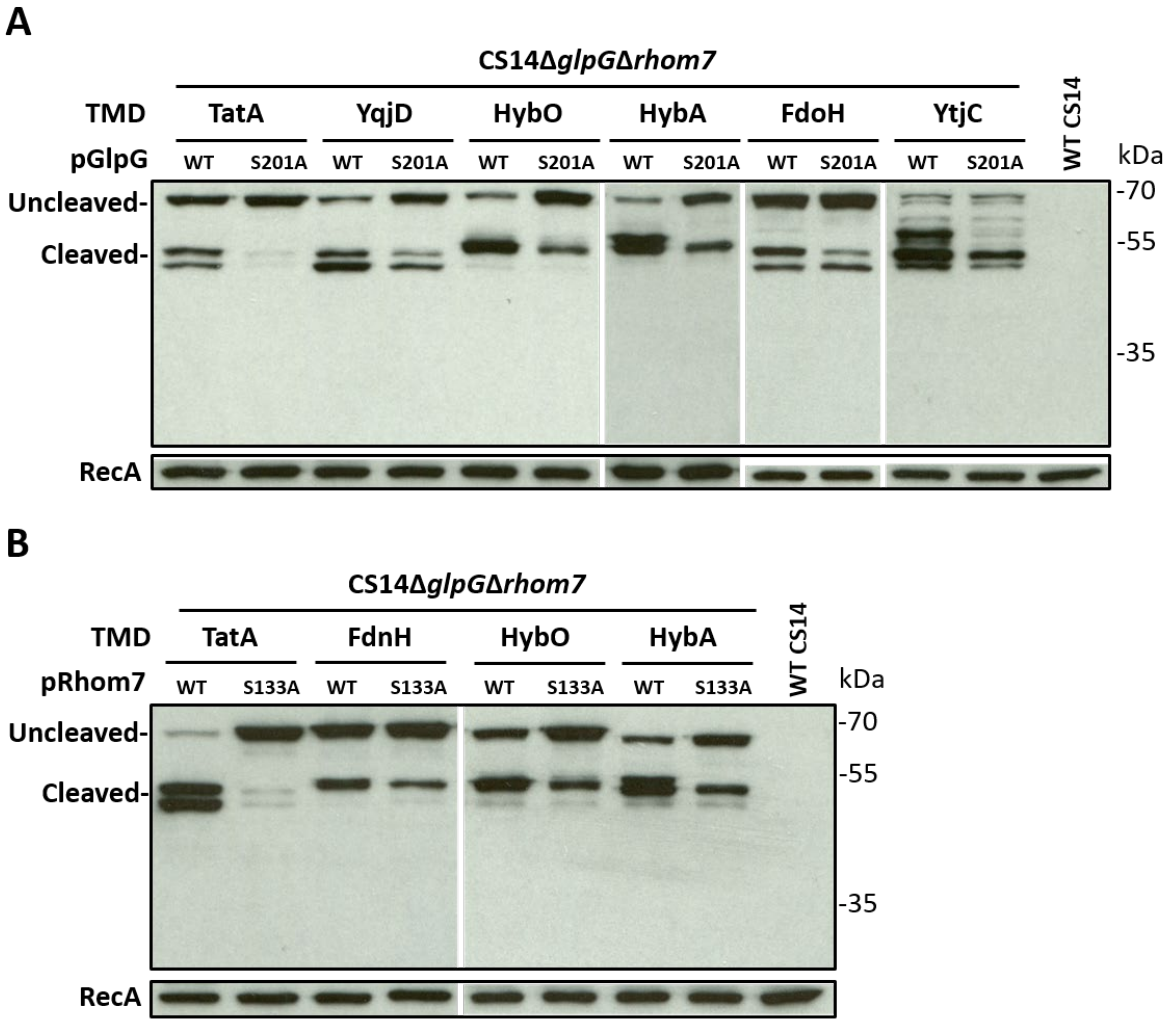


Figure 5.5. Confirmation of cleavage of six candidate substrates by GlpG or Rhom7. *The cleavage assays were performed with S. sonnei CS14ΔglpGΔrhom7 containing pKS508 encoding TMDs of the six candidates deriving from the screen, and pBAD33 encoding wild-type GlpG (**A**) or Rhom7 (**B**), or inactive versions (S201A for GlpG, S133A for Rhom7). Cleavage was detected by western blot analysis using anti-FLAG antibodies. The sizes of cleaved and uncleaved proteins are shown. RecA was used as a loading control.*

Therefore, from 44 bioinformatically identified rhomboid candidates, five TMDs were found to be cleaved by GlpG and three TMDs by Rhom7 (**Figure 5.6**): YqjD, an inner membrane ribosome-associated protein localised to cell poles during stationary phase

(Yoshida *et al.,* 2012); HybO, the small subunit of the hydrogenase 2 (Hyd-2) complex; HybA, an Iron-sulphur protein essential for Hyd-2 activity (Pinske *et al.,* 2015); FdoH, the β-subunit of formate dehydrogenase O (Abaibou *et al.,* 1995); and YtjC, a putative phosphatase. TMDs from HybO and HybA were also cleaved by Rhom7, along with FdnH, the β-subunit of formate dehydrogenase N (Jormakka *et al.,* 2002) (**Figure 5.6)**. Among these hits, YtjC was not futher characterised as it is known to be a soluble protein and its predicted 'TMD' is likely to be a hydrophobic helix packed in the protein core (Yip and Matsumura, 2013). Interestingly, except YqjD, all these proteins *i.e.* HybO/A, FdoH, and FdnH, share several common features: 1) they are iron-sulphur cluster containing proteins, 2) they are part of multi-subunit enzyme complexes involved in respiration, and 3) they are involved in the transfer of electrons in the complex. In addition, HybO and HybA are essential components of the same complex *i.e.* the Hyd-2 complex and are both substrates of the Tat system (Palmer and Berks, 2012; Richard *et al.,* 1999; Sargent *et al.,* 1998). Therefore, HybO and HybA were prioritised for further investigation.

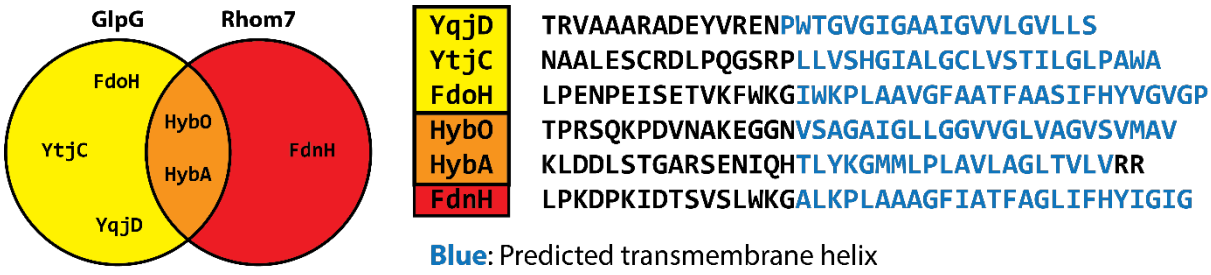


Figure 5.6. Venn diagram showing the putative substrates of GlpG and/or Rhom7 identified and alignment of amino acid sequence of TMD regions cloned into the artificial substrates.

5.4. Further investigation of HybA and HybO as substrates of GlpG and Rhom7

Next, to validate the results of the screen for rhomboid substrates, I examined rhomboid-mediated cleavage of the substrates in more physiological conditions by analysing the ability of endogenous rhomboids to cleave full-length substrates. To allow detection of substrate cleavage by western blot, I constructed strains in which HybO and HybA are expressed from their native chromosomal loci but carry a C-terminal sfCherry-3xFLAG tag. The HybA-sfCherry-3xFLAG fusion has a predicted molecular weight of 64 kDa, and the HybO fusion protein has a predicted molecular mass of 67.8 kDa. Cleavage of tagged HybO or HybA by GlpG or Rhom7 should therefore generate a product of approximately 32 kDa detectable with anti-FLAG antibodies (**Figure 5.7)**.

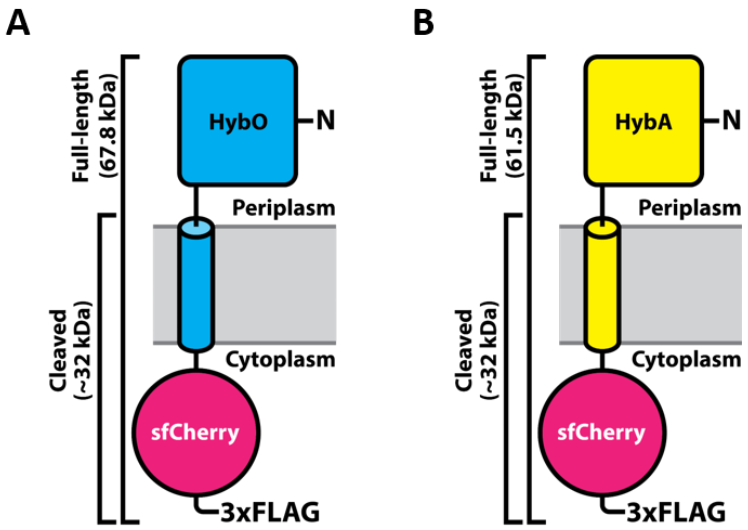


Figure 5.7. Schematic representation of full-length and cleaved tagged HybO and HybA. *The sequence encoding an sfCherry-3xFLAG tag was fused to the 3' end of the chromosomal hybO ORF or hybA ORF* ***(B)*** *to produce a protein detectable by western blot analysis using anti-FLAG antibodies. The sizes of the full-length HybO* ***(A)****/HybA* ***(B)****-sfCherry-3xFLAG and product of cleavage by rhomboid proteases detected by anti-FLAG antibodies are indicated.*

Importantly, as the Hyd-2 complex is only expressed in the absence of oxygen (Richard *et al.,* 1999), to monitor cleavage of HybO and HybA, *S. sonnei* was grown anaerobically in LB with 0.5% glycerol and 0.5% fumarate (Dubini *et al.,* 2002). Cleavage of the substrates was then assessed using strains in which GlpG or Rhom7 were expressed either from their native chromosomal site, or from a plasmid following induction by the addition of arabinose (final concentration 0.2%). As shown in **Figure 5.8**, cleavage of HybO-sfCherry-3xFLAG was detected in all strains, but was independent of the activity of GlpG or Rhom7. The apparent size of the rhomboid-independent band with the highest molecular mass is above 35 kDa, suggesting HybO is cleaved by a periplasmic protease and was ruled out as a rhomboid substrate. The discrepancy between results with full-length HybO and its TMD in an artificial substrate could be caused by access of the TMD in the artificial substrate, compared to when it is part of the full-length protein. Furthermore, this result highlights the importance of performing cleavage assays using full-length substrates to validate hits from the initial screen.

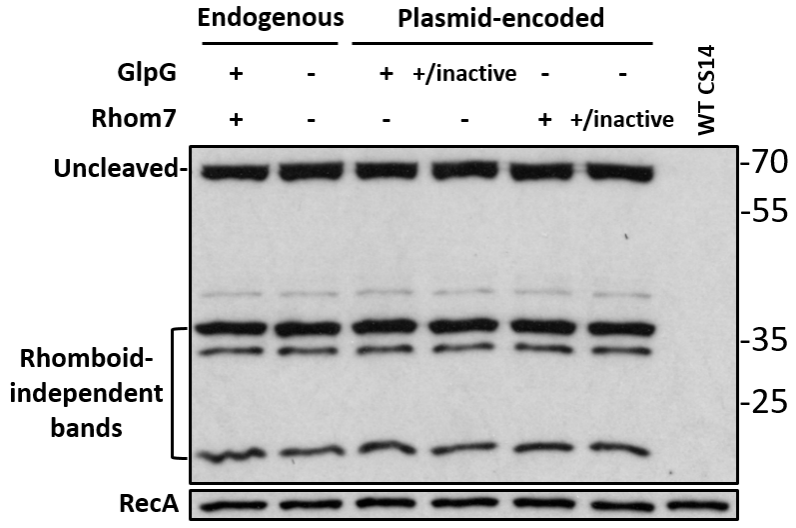


Figure 5.8. Full-length HybO is not amenable to cleavage by endogenous or plasmid-encoded GlpG or Rhom7. *Bacteria were grown anaerobically to exponential phase and rhomboids were expressed (endogenous, +) or deleted (endogenous, -) from their*

endogenous chromosomal locus. Alternatively, wild-type (plasmid-encoded, +) or inactive forms (plasmid-encoded, +/inactive) of rhomboids were expressed from pBAD33 by arabinose induction in a genetic background lacking endogenous GlpG or Rhom7 (n = 1). Inactive forms of rhomboids used were GlpG$_{S201A\ H254A}$ or Rhom7$_{S133A\ H187A}$. Cleavage was detected by western blot analysis using anti-FLAG antibodies. The sizes of cleaved and uncleaved proteins are shown. RecA was used as a loading control.

In contrast, HybA-sfCherry-3xFLAG cleavage was detected, mediated by both endogenous and plasmid encoded GlpG (**Figure 5.9**). Furthermore, no HybA cleavage was detected following expression of the plasmid-encoded inactive GlpG$_{S201A\ H254A}$; the ratio of cleaved HybA to uncleaved HybA decreased from 0.4 in the presence of wild-type GlpG to below 0.05 with inactive GlpG (**Figure 5.9, lanes 5 & 6)**. Interestingly, HybA was not cleaved by Rhom7 when the protease was expressed from its native chromosomal locus, but was cleaved when active Rhom7 was expressed from the plasmid (**Figure 5.9, lanes 2, 7 & 8).** This suggests that Rhom7 can cleave HybA, but that the endogenous levels of the rhomboid are insufficient to result in detectable cleavage. To rule out the possiblity that HybA cleavage by GlpG is an artefact of the presence of the C-terminal sfCherry-3xFLAG domain, this fusion domain was replaced by a C-terminal Halo-3xFLAG tag (**Figure 5.10**). Consistent with the previous findings, although a faint band correspoding to cleaved HybA Halo-3xFLAG could be seen with GlpG$_{S201A}$, GlpG$_{WT}$ produced more cleaved HybA Halo-3xFLAG, further supporting the conclusion that HybA is cleaved by GlpG (**Figure 5.10**). Taken together, our results confirm HybA as a substrate of GlpG in *S. sonnei*. To our knowledge HybA is the first substrate identified for GlpG.

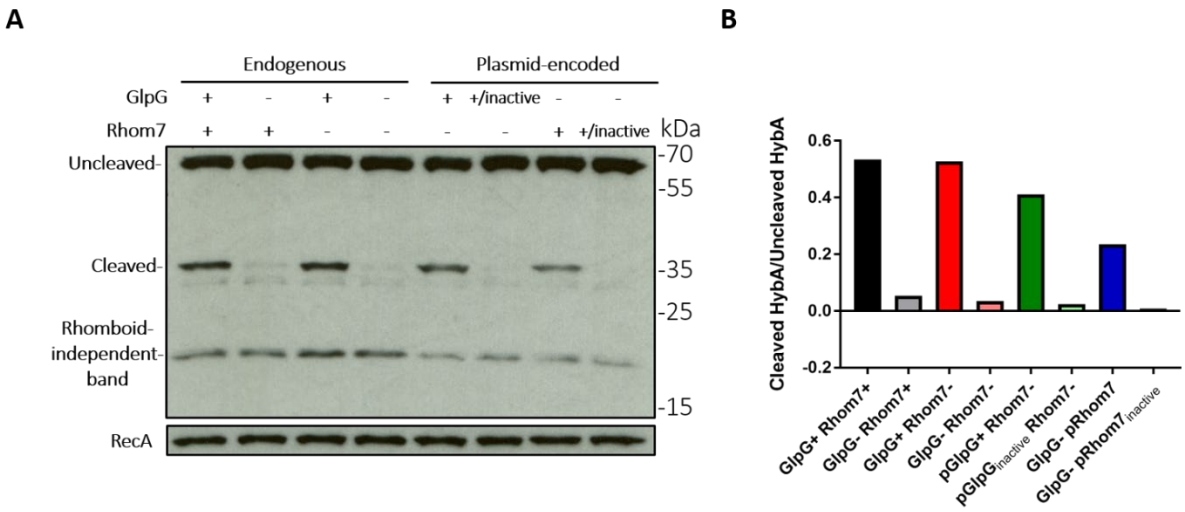


Figure 5.9. Endogenous GlpG cleaves full-length HybA but Rhom7 cleavage of HybA is dependent upon expression from a plasmid. *The sequence encoding an sfCherry-3xFLAG tag was fused to the 3' end of the hybA ORF on the chromosome to produce a fusion protein used to test HybA cleavage. Bacteria were grown anaerobically to exponential phase and rhomboids were expressed (endogenous, +) or deleted (endogenous, -) from their endogenous chromosomal locus. Plasmid-encoded wild-type (plasmid-encoded, +) or inactive forms (plasmid-encoded, +/inactive) of rhomboids were expressed from pBAD33 by arabinose induction in a genetic background lacking endogenous GlpG or Rhom7 (n = 1). Inactive forms of rhomboids used were* $GlpG_{S201A\ H254A}$ *or* $Rhom7_{S133A\ H187A}$*. Cleavage of HybA was detected with anti-FLAG antibodies (n = 1). RecA was used as a loading control. The ratio of cleaved HybA to uncleaved HybA from* ***(A)*** *was quantified and shown in* ***(B)***.

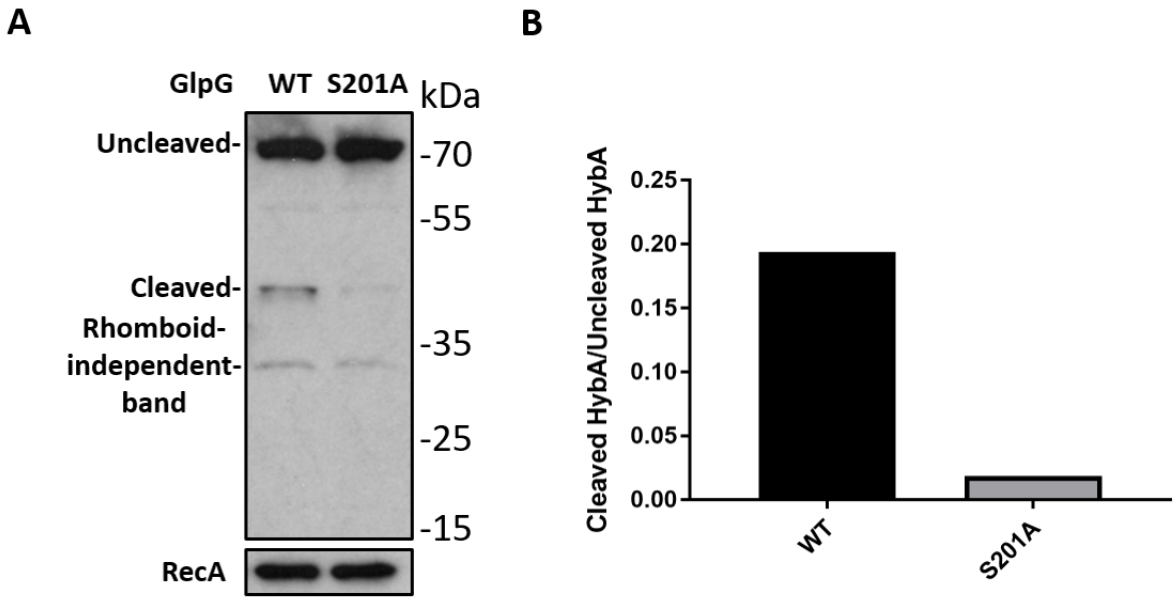


Figure 5.10. Endogenous GlpG cleaves HybA when HybA is tagged with a C-terminal Halo-tag. *The sequence encoding a Halo-3xFLAG tag was fused to the 3' end of the hybA ORF on the chromosome to produce a fusion protein used to detect HybA cleavage by endogenous wild-type GlpG (WT) or inactive* $GlpG_{S201A}$ *(S201A). Bacteria were grown anaerobically and the cleavage of HybA was detected with anti-FLAG antibodies (n = 1). RecA was used as a loading control. The ratio of cleaved HybA to uncleaved HybA from* ***(A)*** *was quantified and shown in* ***(B)****.*

5.5. Identification of HybA residues key to GlpG cleavage

The P1 residue (defined as the C-terminal residue generated by cleavage) of rhomboids substrates is usually a small aliphatic residue within or close to the periplasmic side of the TMD, even those from distant evolutionary backgrounds, *e.g.,* TatA in *P. stuartii*, TVAG_166850 in the eukaryote *Trichomonas vaginalis*, Spitz in *Drosophila*, and Smac in

human mitochondria (Riestra *et al.,* 2015; Saita *et al.,* 2017; Strisovsky *et al.,* 2009; Urban and Freeman, 2003). Therefore, identifying the P1 residue that has these properties and is essential for cleavage would provide further evidence that HybA is a *bona fide* GlpG substrate. In addition, testing whether the HybA-GlpG substrate-rhomboid pair also conforms to this rule became particularly important, as so far all enzymological studies of GlpG have been using artificial substrates. I identified Gly^{296} of HybA as a putative P1 residue by scanning the TOPCONS-predicted TMD sequence, as it is the only residue with small side-chain close to the periplasm. Using site directed mutagenesis, this residue was changed to a Phe (G296F) in chromosomally encoded HybA-sfCherry-3xFLAG, and cleavage of this modified substrate was assessed in presence of wild-type or inactive GlpG or Rhom7. Substitution of HybA Gly^{296} with Phe abolished cleavage by both GlpG and Rhom7, whether the proteases were plasmid or chromosomally expressed (**Figure 5.11A**); quantification of the ratio of cleaved HybA to uncleaved HybA decreased from 0.5 with wild-type HybA to below 0.1 with $HybA_{G296F}$. Furthermore, $HybA_{G296F}$ was expressed at similar levels as the wild-type protein (**Figure 5.11B**). These data are consistent with Gly^{296} being the P1 site in HybA for GlpG and Rhom7, suggesting that both rhomboids recognise conserved features of HybA even though they share low sequence identity (26% in the rhomboid core domain). However, to confirm that Gly^{296} is the P1 residue, N-terminal sequencing of cleaved HybA is required.

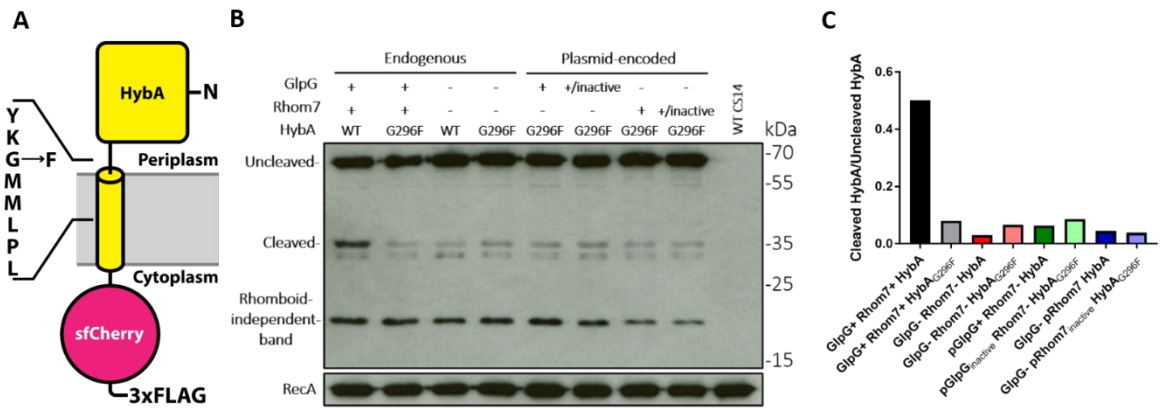


Figure 5.11. HybA Gly^{296} is critical for cleavage by the *Shigella* rhomboids, GlpG and Rhom7. *Introduction of a residue with a bulky side chain (Phe) in the putative P1 site of chromosomally-encoded HybA-sfCherry-3xFLAG (**A**) abolished HybA cleavage by either plasmid or chromosomally encoded GlpG and Rhom7. Bacteria were grown anaerobically and the cleavage of HybA was detected with anti-FLAG antibodies (n = 1). RecA was used as a loading control. The ratio of cleaved HybA to uncleaved HybA from **(B)** was quantified and shown in **(C)**.*

The TMDs of rhomboid substrates usually possess one or more residues that have helix-destabilising properties, *e.g.,* proline, which has the highest ability of all amino acids to increase the intrinsic TM dynamics (Cordes *et al.,* 2002). These helix-destabilising residues are proposed to facilitate substrate entry to the rhomboid active site (Moin and Urban, 2012). This was demonstrated by incorporating a proline into the TMDs of APP (amyloid precursor protein) and TGFα (neither of which is a rhomboid substrate), which converted them into GlpG substrates *in vivo* and *in vitro* (Moin and Urban, 2012). Inspection of the putative HybA TMD revealed a proline (Pro^{300}). Therefore, I examined whether substitution of this proline had any impact on GlpG or Rhom7-mediated cleavage of HybA. Replacement of Pro^{300} with Ala abrogated cleavage of HybA by both GlpG and Rhom7, whether the proteases were plasmid or chromosomally expressed (**Figure 5.12A**), shown by the ratio of

cleaved to uncleaved HybA which decreased from around 0.6 with wild-type HybA to below 0.1 with HybA$_{P300A}$, while not significantly affecting substrate expression (**Figure 5.12B**). The result demonstrates that Pro^{300} is critical for HybA cleavage and we propose that this could be due to its helix-destabilising property in the TMD.

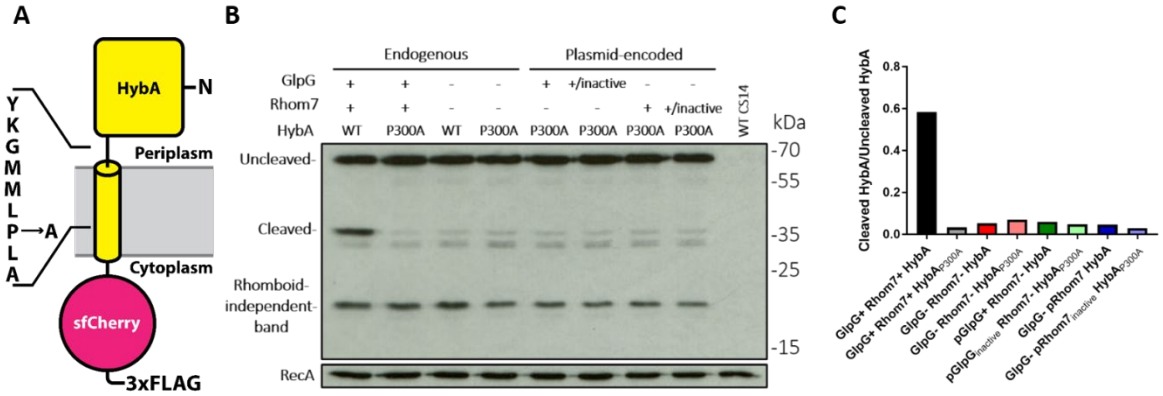


Figure 5.12. Pro^{300} of HybA is critical for cleavage by the *Shigella* rhomboids, GlpG and Rhom7. *Replacement of Pro^{300} (a putative helix-destabilising residue) with Ala of chromosomally-encoded HybA-sfCherry-3xFLAG (**A**) abolished HybA cleavage by either plasmid or chromosomally encoded GlpG and Rhom7. Bacteria were grown anaerobically and the cleavage of HybA was detected with anti-FLAG antibodies (n = 1). RecA was used as a loading control. The ratio of cleaved HybA to uncleaved HybA from **(A)** was quantified and shown in **(B)**.*

5.6. Conclusion and discussion

The work presented in this chapter aimed to identify substrates of GlpG and Rhom7 to provide clues into the biological functions of these rhomboids in *S. sonnei*. Due to the propensity of rhomboids to cleave type I and type III membrane proteins (Strisovsky *et al.,* 2009), a genome-wide bioinformatic screen aiming for identifying all potential type I and type III membrane proteins in *S. sonnei* CS14 was performed and resulted in identification of 44 candidate substrates. However, this bioinformatic approach has several limitations. Firstly, although rarely, rhomboids have been shown to cleave multiple-pass integral membrane proteins, *e.g.,* TSAP6 by human rhomboid RHBDD-1 (Wan *et al.,* 2012) and MdfA (only when truncated) by *E. coli* GlpG (Erez and Bibi, 2009). However, all potential multi-pass substrates were excluded from the screen. Secondly, although HHpred is capable of detecting distant structural homologs of the 16 initial hits generated by TMHMM (Soding *et al.,* 2005), it cannot pick up type I & III proteins bearing no homology to these 16 proteins. Therefore, the initial TMHMM step effectively formed a bottleneck, meaning 44 could be an under-estimation for total number of type I & III proteins in *S. sonnei*.

Overall, from 44 candidate substrates, six were identified as potential substrates, using western blot analysis of rhomboid-mediated cleavage. They were YqjD, YtjC, FdoH, HybO, and HybA for GlpG; FdnH, HybO, and HybA for Rhom7.

HybO and HybA were prioritised for further investigation as they are both components of the hydrogenase-2 (Hyd-2) complex (Beaton *et al.,* 2018; Sargent *et al.,* 1998). Full-length HybO was shown to be resistant to cleavage mediated by GlpG and Rhom7, hence was ruled out as a substrate. However, rhomboid-independent cleavage of full-length HybO was

observed. Moreover, the size of the largest rhomboid-independent proteolytic product suggests that HybO might be subject to proteolysis by other periplasmic protease(s) *in vivo*. This hypothesis is consistent with a previous publication reporting the presence of a putative flexible linker region connecting the periplasmic globular domain of HybO to its C-terminal TMD (Beaton *et al.,* 2018). To confirm whether this linker region is indeed where rhomboid-independent proteolysis occurs, mutagenesis of this region would be required. Moreover, if a proteolysis-resistant HybO mutant can be identified, Hyd-2 activity can be measured in this mutant to shed light on the physiological significance of HybO proteolysis.

Full-length HybA, in contrast, was amenable to cleavage by plasmid-encoded GlpG and Rhom7 and by endogenously expressed GlpG but not Rhom7, suggesting the levels of Rhom7 from the endogenous locus was not high enough to allow detectable HybA cleavage. Thus, HybA was established as the first natural substrate of GlpG. This was a critical step in the progress of this project as it fulfilled our aim to use the properties of rhomboid substrates to guide rational phenotypic characterisation of GlpG and Rhom7. Knowing the importance of HybA in Hyd-2, the role of GlpG in Hyd-2 activity was investigated and discussed in the next chapter.

Similar to the TMDs of other known rhomboid substrates, the predicted TMD of HybA shares several common features. For instance, the P1 residue of a rhomboid substrate is likely to be a small aliphatic residue (Riestra *et al.,* 2015; Saita *et al.,* 2017; Strisovsky *et al.,* 2009). Gly^{296} of HybA fullfils this criterion and was demonstrated to br critical for GlpG/Rhom7-mediated HybA cleavage. Another feature commonly possessed by known rhomboid substrates is the presence of a helix-destabilising residue, most likely proline, in

the TMD a few residues C-terminal to the P1 residue (Moin and Urban, 2012). Proline is unique among the 20 amino acids as its side chain cycles back to its backbone amide, causing the loss of an amide proton key to hydrogen bond formation with the carbonyl oxygen in position i-4 (where i is the prolin locus) and a consequent kink in an α-helix (Deber and Therien, 2002; Sal-Man *et al.,* 2014). In water-soluble proteins and peptides, proline is generally considered a helix breaker. In the TMDs of membrane proteins, however, prolines tend to induce the formation of molecular hinges located on average about four residues N-terminal to the proline residue (Cordes *et al.,* 2002; Senes *et al.,* 2004). Consistent with that, in this work, Pro^{300} was identified four residues C-terminal of the putative P1 residue Gly^{296} and was shown to be critical for mediating GlpG/Rhom7-mediated HybA cleavage. The significance of demonstrating the essentiality of the P1 and helix-destabilising residue in mediating GlpG/Rhom7 cleavage is beyond simply strengthening HybA as a *bona fide* rhomboid substrate, as it markedly narrows down the type of TMD cleavable by GlpG/Rhom7. This would be critical for future identification of multi-pass rhomboid substrates using western blot analysis, which is currently limited by the large number of TMDs required to screen.

HybO and HybA are both Tat-dependent membrane proteins located in the periplasmic face of the inner membrane by a C-terminal tail-anchor (Sargent *et al.,* 2002). However, as the Tat machinery is known to only transport folded proteins, how HybO and HybA are integrated to the membrane is not well defined. One mechanism proposed that the C-terminal tal-anchors behave as 'stop-transfer' signals. In this 'stop-transfer' model, the periplasmic domains of HybO/A would be sent to periplasm through the Tat channel, but the C-terminal tail-anchors will laterally escape into the inner membrane from the Tat channel (Sargent *et al.,* 2002). The alternative 'periplasmic re-entry' model states that the entire HybO/A

including the C-terminal tail-anchor is exported to the periplasm through the Tat channel, which is followed by re-insertion of the tail-anchor to the inner membrane spontaneously or through interacting with HybB, the third membrane-bound component of Hyd-2 (Pinske *et al.,* 2015; Sargent *et al.,* 2002). Our results shed light on this question by demonstrating that a C-terminal sfCherry-3xFLAG fusion does not affect the topology/localisation of HybO/A. This observation argues against the 'periplasmic re-entry' model and lends support to the 'stop-transfer' model as it would be hard to perceive, if the entire fusion protein is exported to the periplasm first, how the bulky C-terminal sfCherry-3xFLAG tag re-crosses the inner membrane to allow tail-anchor insertion.

Chapter 6. Impact of GlpG on Hyd-2 activity in *S. sonnei*

6.1. Impact of GlpG cleavage on Hyd-2-mediated hydrogen-dependent growth of *S. sonnei* CS14

As discussed in the introduction and the previous chapter, HybA is a key component of Hyd-2, a bi-functional complex capable of oxidising H_2 during anaerobic respiration and also generating H_2 using protons in the absence of other terminal electron acceptors (Pinske *et al.,* 2015). This inner membrane-bound complex faces the periplasm and consists of HybB, HybC and HybO in addition to HybA (Sargent *et al.,* 1998). HybC constitutes the catalytic subunit possessing the [NiFe] centre and requires HybO, the small subunit of the complex, for co-translocation to the periplasm through the Tat pathway (Rodrigue *et al.,* 1999). Both HybA and HybB are essential for Hyd-2-dependent electron transfer to the quinone pool but not required for the correct targeting of HybOC to the periplasm (Dubini *et al.,* 2002; Pinske *et al.,* 2015). However, although HybOC, the catalytic part of Hyd-2, has been structurally characterised (Beaton *et al.,* 2018), an intact functional Hyd-2 complex has never been isolated. Thus, the interactions between HybA and HybB with HybOC remain unknown. Moreover, Hyd-2 is unique for being the only protein complex in *Shigella* or *E. coli* with two integral membrane components (HybO and HybA) being Tat substrates. The C-terminal TMDs of HybO and HybA are integrated into the inner membrane *via* an unknown mechanism, which is fundamentally distinct from that employed by other integral membrane proteins (Hatzixanthis *et al.,* 2003). These knowledge gaps are part of the larger question of: 'How is Hyd-2 assembled *in vivo*?' Having established HybA as a physiological substrate of GlpG, our findings potentially open a novel pathway of regulating Hyd-2 assembly and activity. We predicted that GlpG cleavage of HybA could impact the function of the Hyd-2 complex in two opposite ways. First, if HybA TMD is essential for HybA incorporation into a functional Hyd-2, cleaved HybA will be liberated from the inner membrane due to loss of

its TMD. This will then disrupt the integrity of the Hyd-2 complex and hence the electron-transferring, leading to lower total Hyd-2 activity inside the cell. Alternatively, if cleaved HybA is in fact the mature HybA incorporated into Hyd-2, GlpG would then be acting as an activator, and inhibition of HybA cleavage would result in loss of Hyd-2 activity. Therefore, I sought to determine the impact of cleavage of HybA by GlpG on Hyd-2 activity *in vivo*.

First, I performed BLASTn searches of the *S. sonnei* SS046 genome using genes encoding Hyd-2 in *E. coli* MG1655 as queries, aligned the search results with the *S. sonnei* CS14 genome sequence, and confirmed the presence of the operon encoding Hyd-2 in CS14 (**Figure 6.1**). The entire operon has 99% DNA sequence identity to the *E. coli* Hyd-2 operon.

In addition, I examined the conservation of the loci encoding Hyd-1 and Hyd-3 in CS14. This revealed that *hyaC*, the gene encoding the inner membrane γ-subunit of Hyd-1 complex, has a nonsense mutation at codon 123. In addition, an IS1 transposon is inserted in the promoter region of the *hyaA-F* operon. Both features suggest that the Hyd-1 complex of *S. sonnei* CS14 is not functional (**Figure 6.1**). All genes encoding components of Hyd-3 are present in CS14 with high homology (97% DNA sequence identity across the Hyd-3 operon) to their counterparts in *E. coli*, suggesting *S. sonnei* CS14 possesses a functional Hyd-3 (**Figure 6.1**).

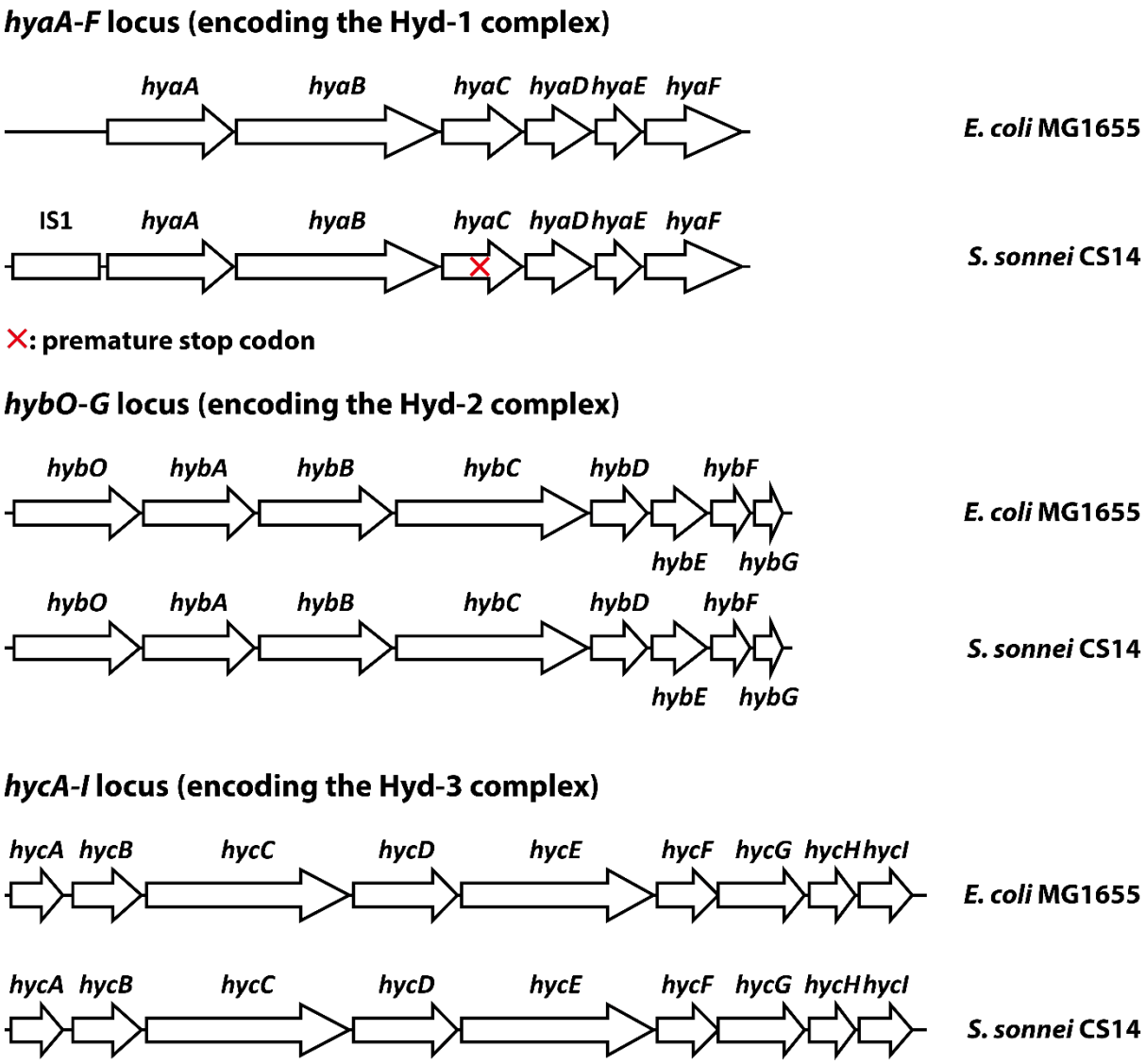


Figure 6.1. Schematic representations of the loci encoding Hyd-1, Hyd-2, and Hyd-3 in *E. coli* MG1655 and *S. sonnei* CS14. *Size of the arrows is drawn to scale to the length of corresponding genes.*

The above analyses show CS14 likely possesses functional Hyd-2, Hyd-3, and a potentially defective Hyd-1. Therefore, to allow the investigation of the impact of GlpG on Hyd-2 function in isolation, we generated a *S. sonnei* strain with a *hyaA-F* deletion to ensure removal of Hyd-1 activity, and also lacking *hycE* which encodes the catalytic subunit of Hyd-3. In addition, *rhom7* was also deleted to allow the impact of GlpG on Hyd-2 function

to be assessed in the absence of any other rhomboid. In this genetic background, we introduced further mutations to produce a panel of strains to enable us to determine the impact of GlpG and HybA cleavage on Hyd-2 activity; CS14Δ*hybO-G* (negative control), CS14Δ*glpG,* and CS14*hybA*$_{G296F}$, which expresses GlpG but has a mutation in HybA that renders it resistant to GlpG-mediated cleavage.

Initially, Hyd-2 activity was analysed by assessing growth in anaerobic conditions with H_2, fumarate and casamino acids (Dubini *et al.,* 2002). Specifically, after aerobic overnight growth, each strain was inoculated into M9 minimal medium supplemented with fumarate (0.5% w/v), 12.5 µg/ml nicotinic acid, and casamino acids (0.2% w/v) in the presence of 5% H_2. Fumarate was used as the major carbon source and an alternative electron acceptor in the absence of oxygen. Nicotinic acid was provided as all *Shigella* spp. are auxotrophic for nicotinic acid (Prunier *et al.,* 2007). This medium was designed to be limited in nutrients so any impact of Hyd-2 dependent H_2 metabolism on bacterial growth can be amplified; casamino acids was still added to stimulate the growth to control the duration of experiment within a practical level. Bacterial growth at 37°C was monitored for 24 hours (**Figure 6.2**). As expected, the strain lacking all three hydrogenase complexes was significantly impaired for growth compared to strains expressing the Hyd-2 complex, demonstrating that CS14 Hyd-2 is functional and enhances the growth of *S. sonnei* in H_2. However, loss of GlpG, or expression of HybA$_{G296F}$ had no statistically significant impact on growth, demonstrating that GlpG cleavage of HybA is not required for Hyd-2-dependent growth.

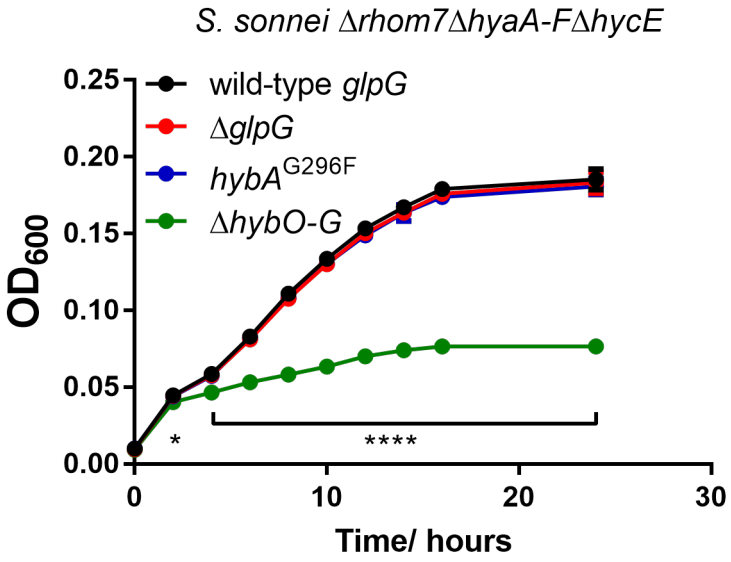


Figure 6.2. Hyd-2 contributes to *S. sonnei* growth in the presence of 5% H_2 but GlpG and cleavage of HybA have no effect. *Bacteria were grown in M9 medium supplemented with 0.5% fumarate, 12.5 µg/ml nicotinic acid and 0.2% casamino acids in the presence of 5% H_2, 10% CO_2 and 85% N_2 at 37°C with agitation. Growth of each strain was measured and compared every two hours by optical density at 600 nm. Data shown are mean ± standard deviation from three independent experiments. *, p<0.05; ****, p<0.0001 (Two-way ANOVA). All strains used in this experiment lack Rhom7, Hyd-1 and Hyd-3 activity.*

6.2. Impact of GlpG cleavage on Hyd-2-dependent hydrogen uptake in *S. sonnei* CS14

Although H_2-mediated growth of CS14 was influenced by Hyd-2 activity in these experiments, other factors could influence bacterial growth, and compensate for the absence of Hyd-2. In addition, Hyd-2 is a bi-functional enzyme with capacity for both H_2 uptake and generation (Pinske *et al.,* 2015). Therefore, to gain further insights into the impact of GlpG cleavage of HybA on Hyd-2 activity, we also measured Hyd-2-mediated H_2 generation of the same strains as above using gas chromatography. Each strain was grown in LB

supplemented with glycerol (0.5% v/v) and fumarate (0.5% w/v) anaerobically to stationary phase, washed and concentrated in PBS supplemented with glycerol (170 mM) to OD_{600} = 35 in a sealed vial. Glycerol was supplemented as it was shown to promote Hyd-2-mediated H_2 production in the absence of external electron acceptors (Pinske *et al.,* 2015). The amount of H_2 in the headspace was monitored over a period of 72 hours by gas chromatography (performed with Stephen Beaton from Department of Inorganic Chemistry, University of Oxford). However, we observed more than 99% cells died in the first 24 hours as determined by cfu measurements, during which little H_2 could be detected (data not shown).

This prompted us to instead measure Hyd-2-mediated H_2 uptake using the same strains as the ones used in H_2-dependent growth assays (**Figure 6.2**) by gas chromotography. After aerobic overnight growth, each strain was inoculated into M9 minimal medium supplemented with fumarate (0.5% w/v), 12.5 µg/ml nicotinic acid and casamino acids (0.2% w/v) in a sealed vial, and the headspace of cultures was purged with 10% H_2 in argon. The amount of H_2 in the headspace was monitored over a period of nine hours by gas chromatography (performed with Stephen Beaton). The decrease in headspace H_2 was used to assess H_2 uptake by the *S. sonnei* strains (**Figure 6.3A**). Viability of the mutants in these conditions was also assessed by measuring CFU of the initial and final time points (**Figure 6.3B, 6.3C**). Again, the strain lacking all three hydrogenase complexes did not consume H_2, consistent with the absence of all functional hydrogenase complexes. The remaining mutants each depleted H_2 in the headspace over time, providing evidence of Hyd-2 mediated H_2 uptake. Importantly, there was no significant difference in Hyd-2 dependent consumption of H_2 by Δ*rhom7*, Δ*rhom7*Δ*glpG,* and Δ*rhom7hybA*$_{G296F}$ (**Figure 6.3A**), providing further evidence that neither GlpG nor HybA cleavage modulates Hyd-2 activity. The number of bacteria in the input was the same for all four strains (**Figure 6.3B**). However, the number

of viable Δ*rhom7*Δ*hybO-G* at the end of experiments was significantly lower than the number of Δ*rhom7*, Δ*rhom7*Δ*glpG,* and Δ*rhom7hybA*$_{G296F}$ and no difference was observed among Δ*rhom7*, Δ*rhom7*Δ*glpG,* and Δ*rhom7hybA*$_{G296F}$, which followed the same trend as the H_2 uptake results, suggesting the H_2 consumed by bacteria was used to as an energy source for their growth.

To exclude the possibility that GlpG-mediated proteolytic cleavage of HybA was inhibited in these growth conditions, we repeated the assay using strains in which HybA carries a C-terminal sfCherry-3xFLAG tag and monitored cleavage by western blot analysis. The expected cleavage product was detected in the presence of GlpG at both 4.5 and 9 hours of growth in 10% H_2 (**Figure 6.3D, lane 1 and 4**), and cleavage was greatly reduced either in the absence of GlpG or in the presence of *hybA*$_{G296F}$ mutation (**Figure 6.3D, lane 2, 3, 5, and 6**). This demonstrates that GlpG cleavage of HybA still occurs in conditions of the H_2 uptake experiment, but still had no detectable impact on Hyd-2 activity. Of note, we observed that the pool of uncleaved HybA remained relatively stable regardless of the presence of active GlpG. If full-length HybA is the functional form of HybA in the Hyd-2 complex, this could account for the H_2 uptake results as the amount of HybA cleaved by GlpG was too small to allow us to detect changes in Hyd-2 activity.

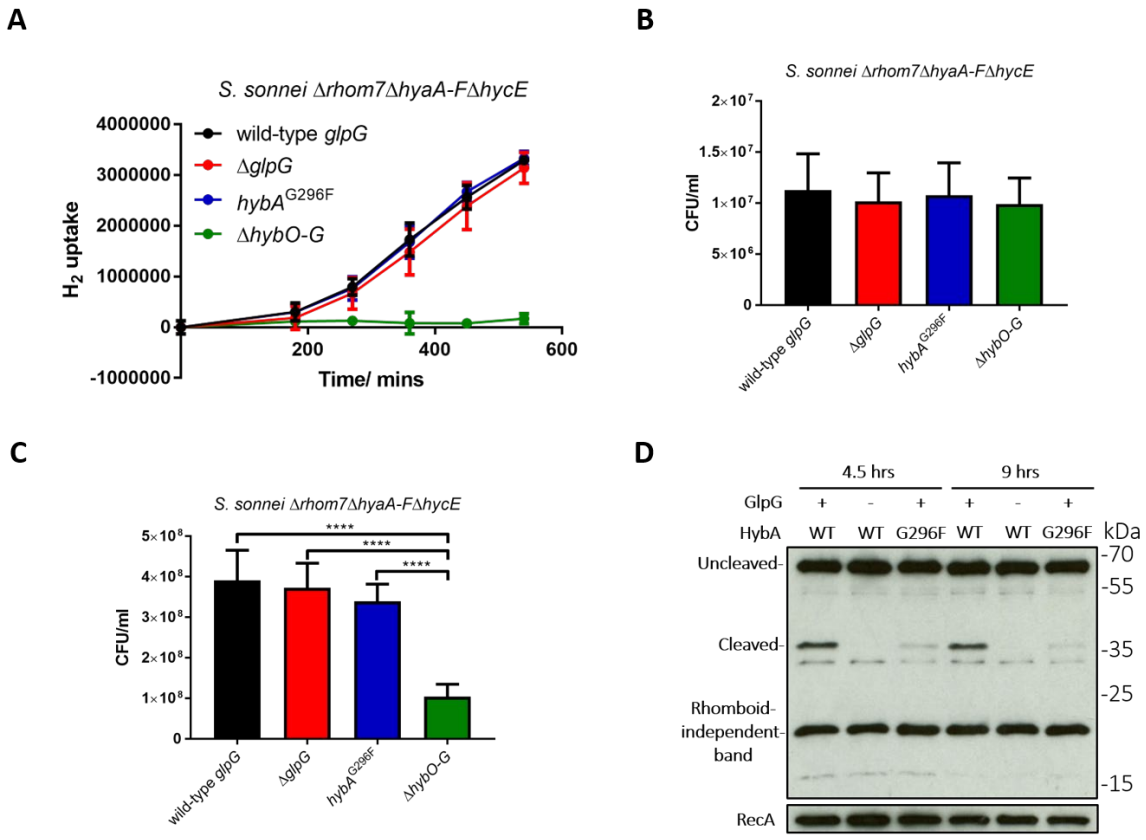


Figure 6.3. Neither GlpG nor HybA cleavage impacts Hyd-2-dependent H_2 uptake. ***(A)*** *Bacteria were grown aerobically in LB overnight then diluted into 5 ml M9 minimal medium supplemented with 0.5% fumarate, 12.5 µg/ml nicotinic acid and 0.2% casamino acids in a sealed glass chromatography vial. The headspace was purged with 10% H_2, 90% argon and cultures were incubated at 37°C with agitation for 9 hours (n = 3). The number of bacteria in the input (**B**) and at the end of experiments (**C**) were counted by plating (n = 3, ****, p<0.0001 (One-way ANOVA)). The amount of H_2 in the headspace was measured by gas chromatography. H_2 uptake at a given time point (T_x) is calculated by the formula: H_2 uptake T_x = amount of H_2 in the headspace at T_0 - amount of H_2 in the headspace at T_x.* ***(D)*** *Endogenous GlpG cleaves HybA during the H_2 uptake assay. Bacteria were grown in the same way as in **(A)**. Whole cell lysates were prepared at 4.5 hours and 9 hours post-purging for western blot analysis (n = 1). Cleavage of the full-length HybA was detected with anti-FLAG antibodies. RecA was used as a loading control. All strains*

*used in (**A**-**D**) lack Hyd-1 and Hyd-3 activity. Strains used in (**A**-**C**) harbour untagged HybA while strains used in **D*** possess sfCherry-3xFLAG tagged HybA for western blot analyses.

6.3. Conclusion and discussion

Based on the finding that HybA is a natural substrate of GlpG, we tested the hypothesis that HybA cleavage by GlpG modulates Hyd-2 activity in *S. sonnei* CS14. By assessing the effect of *glpG* deletion and $hybA_{G296F}$ mutation (which renders HybA refractory to GlpG cleavage), we conclude that neither GlpG or the propensity of HybA to be cleaved by GlpG has any significant effect on Hyd-2 activity *in vivo*. This is not consistent with either of our initial hypotheses, which predicted that HybA cleavage by GlpG will either decrease or increase Hyd-2 activity. Our findings therefore allow us to rule out GlpG as a HybA activator/maturase, as this hypothesis would predict Hyd-2 activity will almost be completely lost in CS14 lacking GlpG. However, it is less straight-forward to conclude that GlpG is also not a negative regulator of Hyd-2 activity. Although in our experiments, GlpG appeared not to down-regulate Hyd-2 activity, it was noted that there was only cleavage of a sub-population of HybA in the presence of endogenous levels of GlpG. Therefore, GlpG could still be a negative regulator in conditions not covered by our experiments. Alternatively, GlpG could have evolved not to influence Hyd-2 activity, but to only target HybA molecules that are not part of the Hyd-2 complex. This possibility is tested in the next series of experiments outlined in chapter 7.

Chapter 7. GlpG acts as a quality control enzyme cleaving uncomplexed substrates

7.1. HybA cleavage by GlpG is enhanced by deletion of HybB.

Hyd-2 *in vivo* is predicted to be a dimer of the HybOABC hetero-tetramer, which has a stoichiometry of 1:1:1:1 (Beaton *et al.,* 2018). Our observation that i) only a sub-population of full-length HybA is cleaved in the presence of endogenous levels of GlpG, and ii) that this cleavage of only a sub-population of HybA had no detectable impact on Hyd-2 function, prompted us to hypothesise that HybA may exist in two distinct populations: a 'cleavage-refractory' and a 'cleavage-susceptible' population. The latter population could be HybA that is not incorporated into the Hyd-2 complex. This could result from an imbalance in synthesis or transport of the components of Hyd-2, or damage/loss of parts of the complex. Under those circumstances, GlpG cleavage of 'uncomplexed' HybA would contribute to the quality control of Hyd-2 complex and removal of HybA that is not incorporated into Hyd-2 from the inner membrane. If GlpG does indeed only cleave uncomplexed HybA, it would predict that an increase in the amount of uncomplexed HybA would lead to an increase in the ratio of cleaved to uncleaved HybA. Therefore, I tested this hypothesis using two approaches. Firstly, HybA was over-expressed from a plasmid in addition to from its endogenous locus, which increased the relative abundance of this protein with respect to its partners HybO and HybC, and hence increased uncomplexed HybA. Under these conditions, the ratio of cleaved to uncleaved HybA was increased by >3-fold (**Figure 7.1A & B**) compared to when HybA was only expressed from its endogenous locus (**Figure 7.1A lane 5 to lane 1**).

Second, HybB, which is predicted to encode the 10-TMD subunit of Hyd-2, has been proposed to interact with the TMD of HybA (Beaton *et al.,* 2018). Hence, HybA cleavage was also assessed following *hybB* deletion, which was expected to yield 100% uncomplexed HybA. Consistent with our hypothesis, the ratio of cleaved to uncleaved HybA ratio increased by > 4-fold (**Figure 7.1A & B**) compared to when the Hyd-2 complex was intact (**Figure 7.1A lane 3 to lane 1**). However, uncleaved HybA could still be observed after *hybB* deletion, suggesting a population of uncomplexed HybA was not cleaved. This could a combined effect of continuous *de novo* synthesis of full-length HybA and the low catalytic efficiency of GlpG (Dickey *et al.,* 2013).

Surprisingly, the protein levels of GlpG were also increased when 1) GlpG possessed the S201A mutation, 2) *hybB* was deleted, and 3) when HybA was over-expressed. To rule out that increased HybA cleavage was due to increased GlpG expression rather than the increased proportion of uncomplexed HybA, I also assessed HybA cleavage upon over-expression of GlpG from the multi-copy plasmid pUC19 in the presence of HybB. Results demonstrated that, although GlpG was up-regulated by more than 5-fold (**Figure 7.1C**), in the presence of HybB, the extent of HybA cleavage remained lower than that when HybA cleavage was mediated by endogenous GlpG in the absence of HybB (**Figure 7.1A lane 7 to lane 3**). This finding suggests that perturbation of the relative abundance of components of the Hyd-2 complex, and consequently the increased amount of uncomplexed HybA, is the major determinant for increased HybA cleavage.

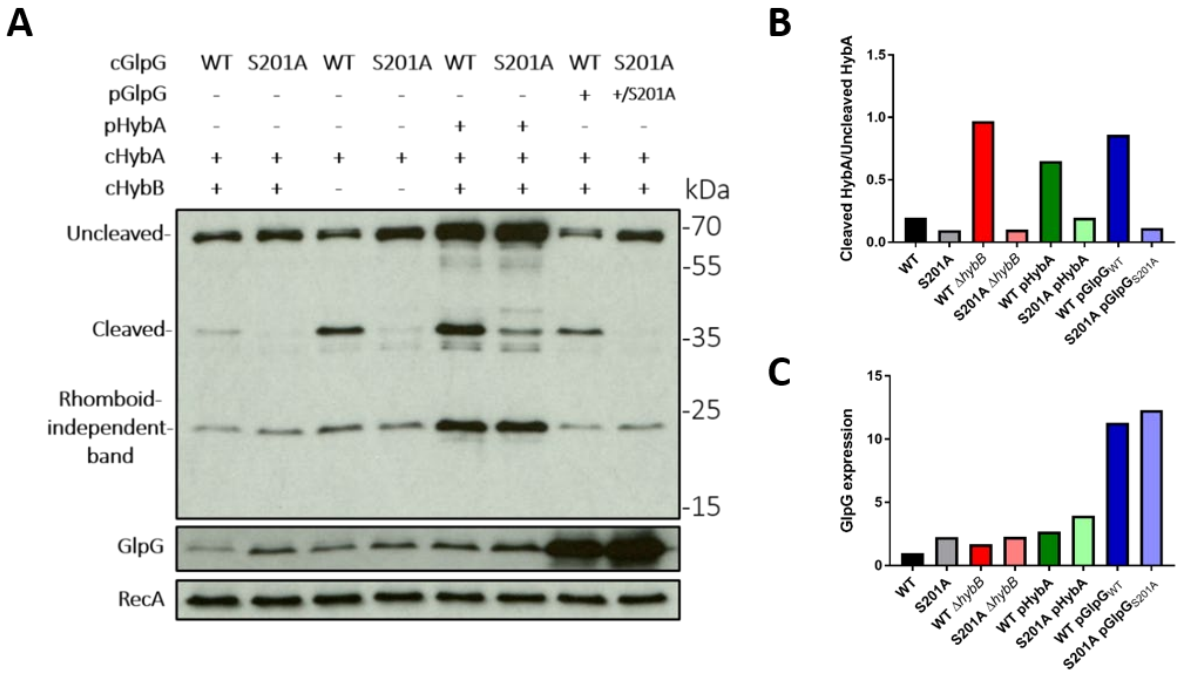


Figure 7.1. HybA cleavage by GlpG is increased upon deletion of *hybB*, or HybA over-expression. ***(A)*** *The sequence encoding an sfCherry-3xFLAG tag was fused to the 3′ end of the hybA ORF to produce a fusion protein used to test cleavage by endogenous GlpG (all strains in experiments shown in* ***Figure 7.1*** *lack rhom7). The Ser201Ala (S201A) substitution was introduced into GlpG to inactivate the enzyme. The sequence encoding a 3xHA tag was fused to the 5′ end of glpG to detect GlpG expression. Both plasmid-encoded HybA and GlpG (pHybA and pGlpG) were tagged in the same way as their chromosomal counterparts and expressed constitutively. Bacteria were grown anaerobically to exponential phase (OD_{600} = 0.5) before being lysed for western blot analysis (n = 1). Cleavage of the full-length HybA was detected with anti-FLAG antibodies. GlpG expressed from its endogenous locus was detected with anti-HA antibodies. RecA was used as a loading control. The ratio of cleaved to uncleaved HybA and GlpG expression from* ***(A)*** *were quantified and shown in* ***(B)*** *and* ***(C)****.*

In conclusion, these results are consistent with our hypothesis that uncomplexed HybA is more prone to cleavage by GlpG than HybA in the intact Hyd-2 complex. This leads to a 'complex protection' model in which uncomplexed HybA is preferentially cleaved by GlpG compared with when HybA is incorporated into the Hyd-2 complex (**Figure 7.2**).

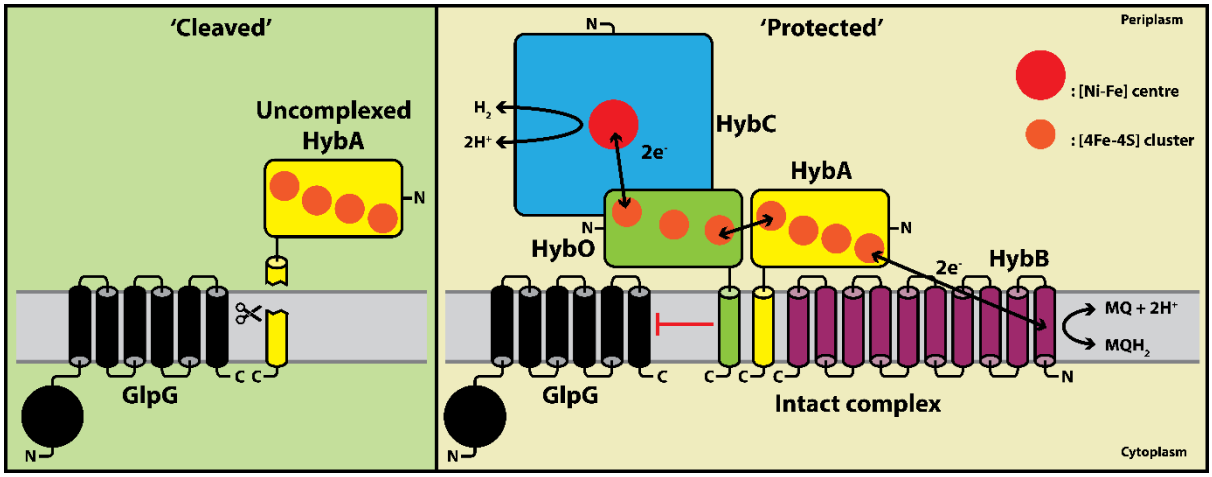


Figure 7.2. Schematic representation of the 'complex protection' model.

These findings prompted me to consider whether any of the other potential GlpG and Rhom7 substrates identified in our TMD screen (**Figure 5.5**), which are also components of multiprotein complexes might exhibit increased proteolysis in the absence of interacting partners. In addition to HybA, HybO, FdoH, and FdoH were identified as potential substrates for GlpG while HybO and FdnH were identified as potential substrates for Rhom7. According to published literature, YqjD has a C-terminal TMD anchoring it to the periplasmic face of the inner membrane and is expressed during stationary phase in *E. coli* (Yoshida *et al.,* 2012), however it is not known whether YqjD is part of a membrane protein complex. Therefore, YqjD was excluded from further analysis. HybO is also part of the Hyd-2 complex. The remaining substrates, FdoH for GlpG, and FdnH for Rhom7 are both known components of inner membrane-residing, periplasm-facing multiprotein complexes, and of

particular interest, are both involved in respiration like the Hyd-2 (Jormakka *et al.,* 2002; Unden *et al.,* 2014). Furthermore, FdoH, FdnH and HybA all belong to a superfamily of ferredoxin domain that mediate electron transfer (Berks *et al.,* 1995; Jormakka *et al.,* 2002).

Although HybO was shown to be cleaved in a rhomboid-independent manner (**Figure 5.7**), this was in context of wild-type levels of HybABC. Therefore, I tested whether HybO cleavage by endogenous GlpG alters upon deletion of *hybB* or over-expression of HybO. In contrast to HybA, neither genetic manipulation resulted in GlpG-mediated cleavage of HybO, confirming previous result that HybO is not a substrate of GlpG (**Figure 7.3**). Again, GlpG expression increased when GlpG underwent the S201A mutation or when *hybB* was deleted. However, in contrast to HybA, over-expression of HybO did not increase GlpG expression.

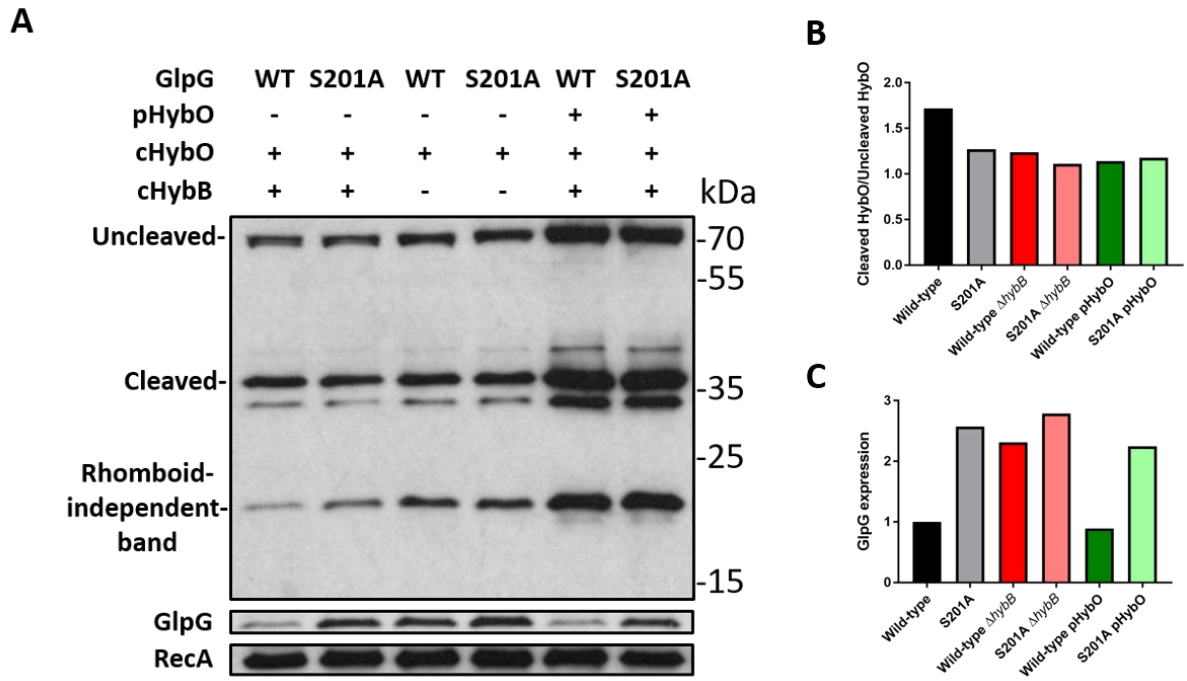


Figure 7.3. GlpG does not cleave HybO even upon deletion of *hybB*, or HybO over-expression. ***(A)*** *The sequence encoding an sfCherry-3xFLAG tag was fused to the 3′ end of the hybO ORF on the chromosome to test full-length HybO (cHybO) cleavage by endogenous GlpG (all strains in* ***Figure 7.3*** *underwent rhom7 deletion). Ser201Ala (S201A) mutation was introduced to the glpG ORF on the chromosome to inactivate the enzyme without perturbing glpG expression. The sequence encoding a 3xHA tag was fused to the 5′ end of glpG ORF on the chromosome to detect endogenous GlpG. Plasmid-encoded HybO (pHybO) was tagged in the same way as cHybO and was expressed constitutively. Bacteria were grown anaerobically to exponential phase (OD_{600} = 0.5) before lysed for western blot analysis (n = 1). Cleavage of the full-length HybO was detected with anti-FLAG antibodies. GlpG expressed from its endogenous locus was detected with anti-HA antibodies. RecA was used as a loading control. The ratio of cleaved to uncleaved HybO and GlpG expression from* ***(A)*** *were quantified and shown in* ***(B)*** *and* ***(C)****.*

7.2. FdoH cleavage by GlpG is only observed upon deletion of FdoI.

As discussed previously in section 1.5.2 of the Introduction, FdoH is part of the aerobic formate dehydrogenase-O (FDH-O) complex. The FDH-O complex is predicted to consist of three subunits: FdoG, FdoH, and FdoI (**Figure 7.4B**) based on its homologue: the anaerobic formate dehydrogenase-N (FDH-N) in *E. coli*, which consists of FdnG, FdnG, and FdnI as illustrated by its crystal structure (**Figure 7.4A**) (Jormakka *et al.,* 2002). Components of FDH-O share high sequence homology (77%, 78%, and 48% amino acid sequence identity with FdnG, FdnH, and FdnI, respectively) to their counterparts in FDH-O. FdoH is predicted to contain four [4Fe-4S] clusters implicated in electron transfer (Abaibou *et al.,* 1995; Jormakka *et al.,* 2002; Pinske *et al.,* 2015). In *S. sonnei* CS14, FDH-O is also encoded by the *fdoGHI-fdhE* operon with 99% DNA sequence identity to the locus in *E. coli*.

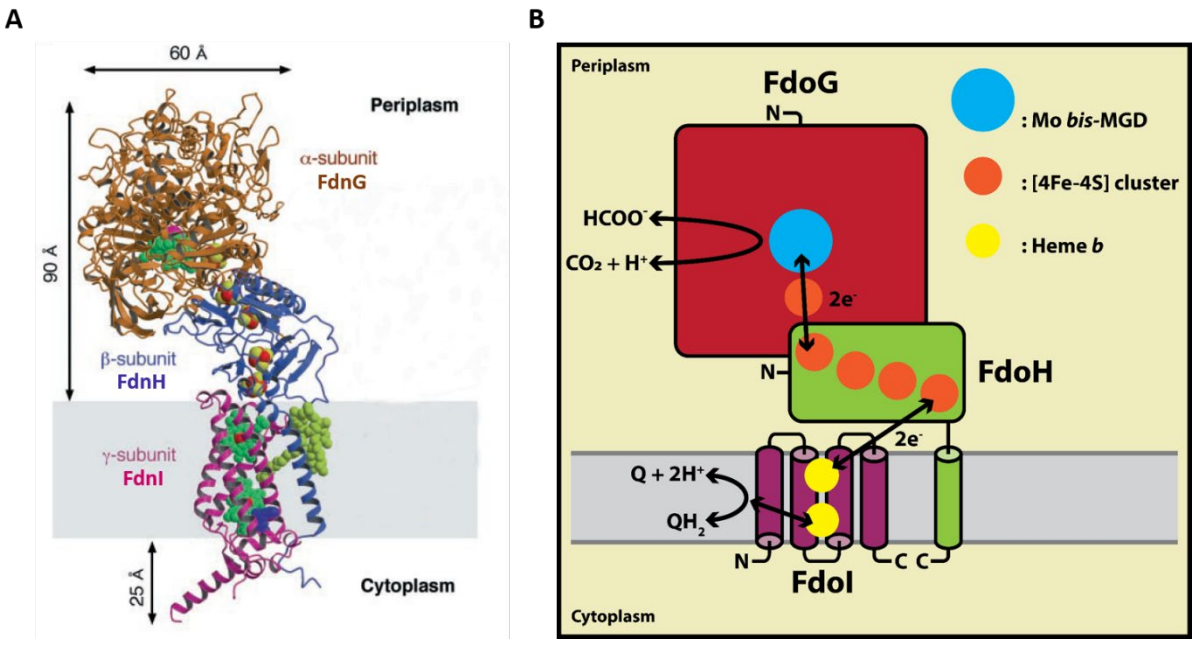


Figure 7.4. Structure of the formate dehydrogenase-N (FDH-N) complex and a schematic representation of the predicted organisation of the formate dehydrogenase-O (FDH-O) complex. *Q and QH_2 stand for oxidised and reduced quinones. Panel A is adapted from (Jormakka et al., 2002).*

To analyse GlpG-mediated cleavage of FdoH, a sequence encoding a C-terminal sfCherry-3xFLAG was added to the endogenous chromosomal copy of the *fdoH* (**Figure 7.5A**) to enable analysis of expression and proteolytic cleavage during aerobic, exponential growth by western blot analysis. To test whether the 'complex protection' model also holds true for FdoH cleavage by GlpG, FdoI, the multi-pass membrane subunit of the FdoGHI complex (**Figure 7.4B**), was deleted to artificially generate 'uncomplexed' FdoH. Unlike for HybA/O, generation of uncomplexed substrate through over-expression was not performed as FdoH itself does not contain a Tat-signal peptide and is dependent on FdoG binding for its localisation. Therefore, it is unlikely that over-expressed FdoH would be translocated across the inner membrane. Strikingly, FdoH cleavage by GlpG was only observed in the strain lacking FdoI (**Figure 7.5B**), consistent with the hypothesis that GlpG only cleaves uncomplexed FdoH. Of note, the increase in GlpG expression upon S201A mutation observed before was no longer detected.

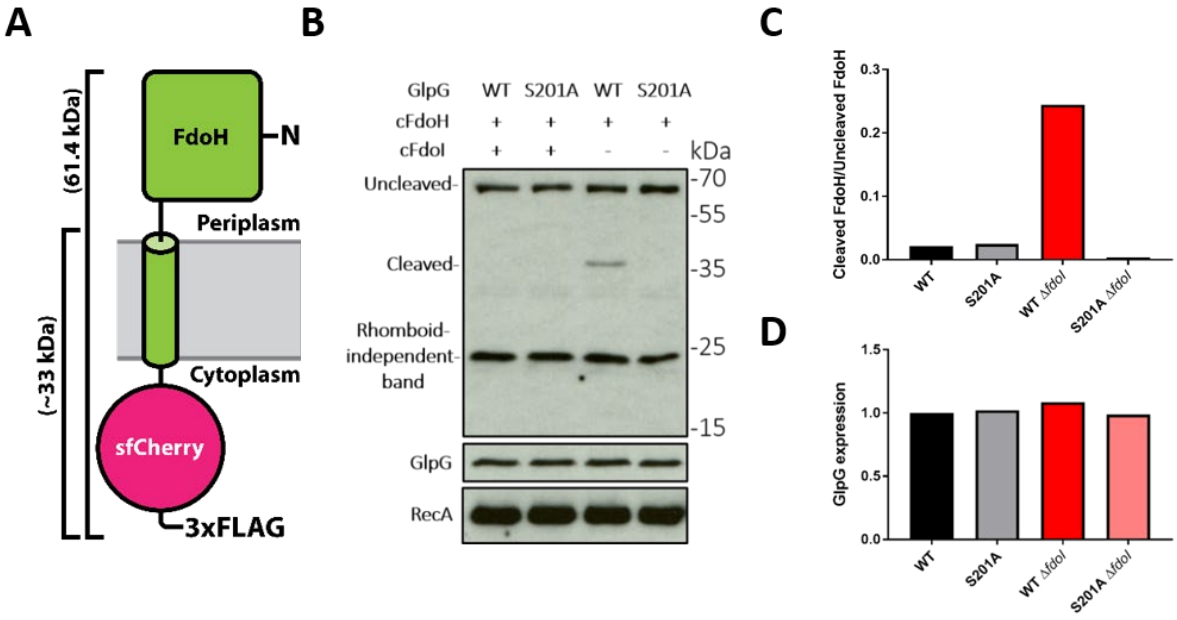


Figure 7.5. Endogenous GlpG cleaves full-length FdoH only in the absence of FdoI. ***(A)*** *Schematic representation of chromosomally tagged FdoH for cleavage analysis by GlpG. The sequence encoding an sfCherry-3xFLAG tag was fused to the 3' end of the fdoH ORF*

on the chromosome to test full-length FdoH cleavage in the presence of endogenous GlpG (all strains in ***Figure 7.4*** *underwent rhom7 deletion).* ***(B)*** *The sequence encoding a 3xHA tag was fused to the 5′ end of glpG to detect endogenous GlpG. Bacteria were grown aerobically to exponential phase (OD_{600} = 0.8) before being lysed for western blot analysis (n = 1). Cleavage of the full-length FdoH was detected with anti-FLAG antibodies. GlpG expressed from its endogenous locus was detected with anti-HA antibodies. RecA was used as a loading control. The ratio of cleaved to uncleaved FdoH and GlpG expression from* ***(B)*** *were quantified and shown in* ***(C)*** *and* ***(D)****.*

7.3. FdnH cleavage by Rhom7 is only observed upon deletion of *fdnI*.

FDH-N is encoded by the *fdnGHI* operon in *E. coli* (Berg and Stewart, 1991). Genome comparisons located a highly-homologous *S. sonnei* CS14 *fdnGHI* operon with 99% DNA sequence identity to the *E. coli* operon. To analyse Rhom7-mediated cleavage of FdnH, *fdnH* was tagged with sfCherry-3xFLAG at its C-terminus on the chromosome. FdnI, the multi-pass membrane subunit of the FdnGHI complex (**Figure 7.4A**), was deleted to produce uncomplexed FdnH. Again, FdnH over-expression was not performed as FdnH also does not contain the Tat-signal peptide is dependent on FdnG binding for its correct localisation. The expression and proteolytic cleavage of FdnH during anaerobic, exponential growth was assessed by western blot analysis. Notably, similar to its aerobic counterpart FdoH, FdnH cleavage was only observed in the absence of FdnI by plasmid-encoded Rhom7 (**Figure 7.6**), again supporting the 'complex protection' model. In addition, to test whether the loss of 'complex protection' could vary rhomboid substrate specificity observed in the initial screen (**Figure 5.5**), I also assessed the proteolytic cleavage of FdnH by GlpG during anaerobic, exponential growth by western blot analysis. The results demonstrated that FdnH was not

cleaved by GlpG regardless of the presence of FdnI, again supporting FdnH is not a substrate of GlpG (**Figure 7.7**).

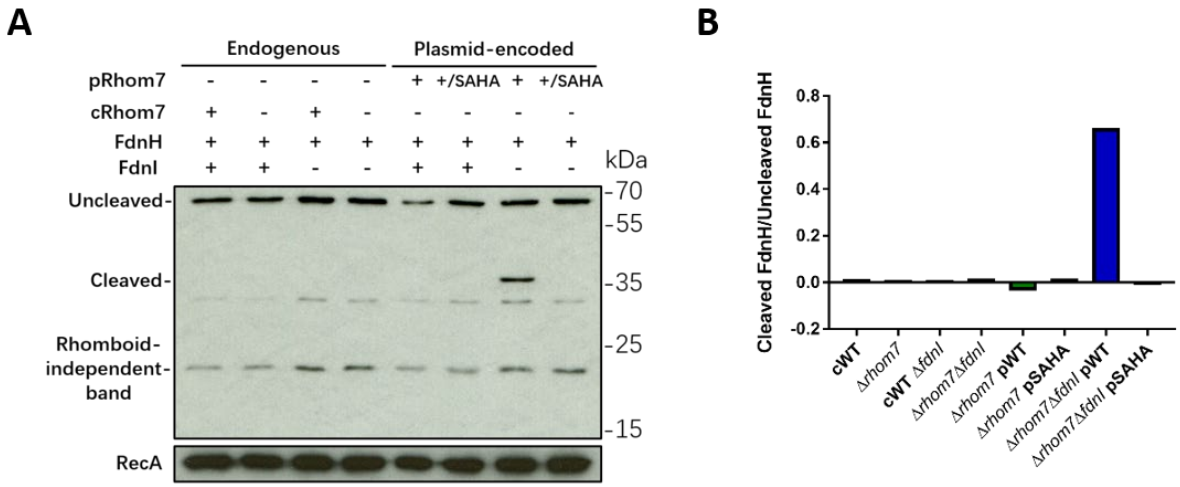


Figure 7.6. Plasmid-encoded Rhom7 cleaves full-length FdnH only in the absence of FdnI. ***(A)*** *The sequence encoding an sfCherry-3xFLAG tag was fused to the 3' end of the fdnH ORF on the chromosome to test full-length FdnH cleavage in the presence of endogenous and plasmid-encoded Rhom7 (all strains in* ***Figure 7.6*** *underwent glpG deletion). Bacteria were grown anaerobically to exponential phase (OD_{600} = 0.5) before being lysed for western blot analysis (n = 1). Cleavage of the full-length FdnH was detected with anti-FLAG antibodies. RecA was used as a loading control. The ratio of cleaved to uncleaved FdnH was quantified and shown in* ***(B)****.*

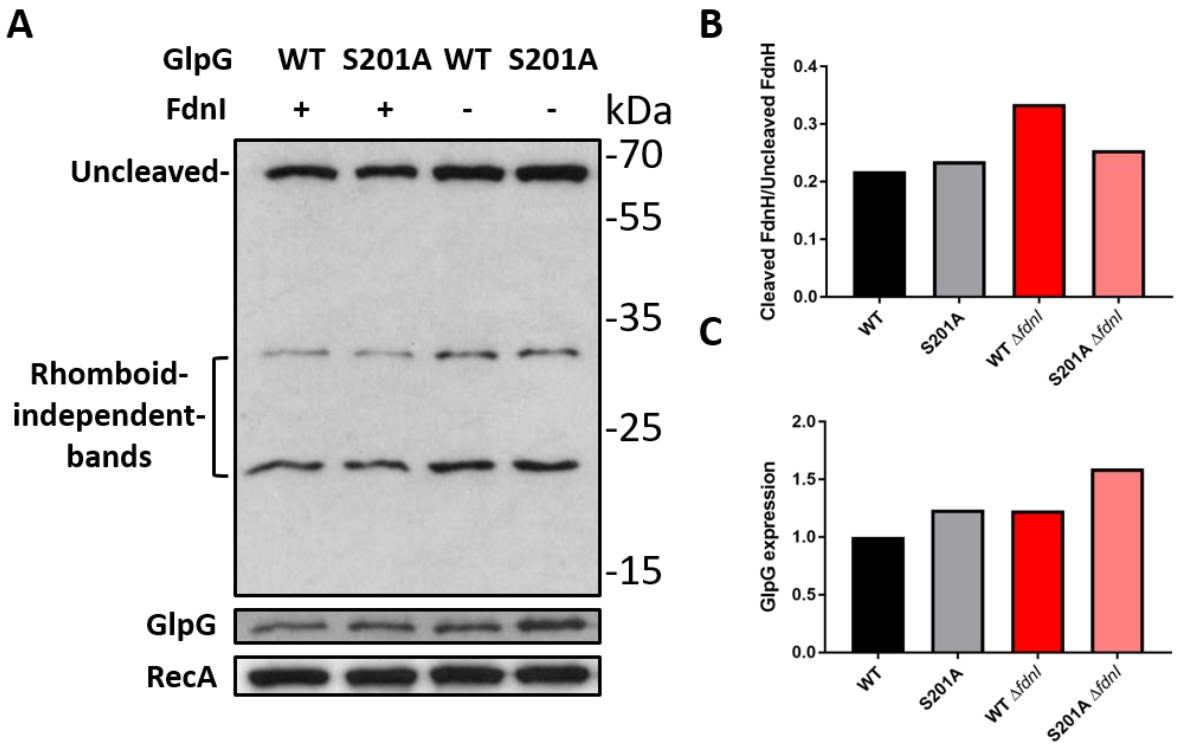


Figure 7.7. FdnH is not a substrate of GlpG. ***(A)*** *The fdnH ORF was tagged in the same way as in* ***Figure 7.3.1*** *to test full-length FdnH cleavage by endogenous Rhom7. The sequence encoding a 3xHA tag was fused to the 5' end of glpG to detect endogenous GlpG (all strains in* ***Figure 7.7*** *underwent rhom7 deletion). Bacteria were grown anaerobically to exponential phase (OD_{600} = 0.5) before being lysed for western blot analysis (n = 1). Cleavage of the full-length FdnH was detected with anti-FLAG antibodies. GlpG expressed from its endogenous locus was detected with anti-HA antibodies. RecA was used as a loading control. The ratio of cleaved to uncleaved FdoH and GlpG expression from* ***(A)*** *were quantified and shown in* ***(B)*** *and* ***(C)****.*

In summary, cleavage of three substrates: HybA, FdoH, and FdnH by GlpG and/or Rhom7, are consistent with the 'complex protection' model. Thus, we propose that rhomboids may

act as a quality control mechanism to initiate removal of uncomplexed components of membrane protein complexes from the inner membrane.

7.4. Proline-to-alanine mutations in rhomboid substrates abolish cleavage regardless of complex protection.

The mechanistic basis of the 'complex protection' model could theoretically arise from two non-mutually exclusive hypotheses (**Figure 7.8**). First, rhomboid substrates could be protected due to steric hindrance conferred by the membrane-residing interacting partners. For example, HybB could prevent GlpG from interacting with HybA. One analogy for this hypothesis would be that the substrate resembles the central filling in a sushi roll, with lateral contact to the filling prevented in the intact roll. Alternatively, the TMD of rhomboid substrates could still be accessible to rhomboids in their cognate complex, but the helix-destablisation process, which has been shown to be critical for GlpG cleavage (Moin and Urban, 2012), could be prevented due to helix stabilisation through TMD-TMD interaction/packing with TMDs from the interacting partners. The corresponding analogy for this hypothesis would be that the substrate is similar to ivy growing on a wall, in which the wall resembles the binding partner of the substrate. Without the wall, the ivy would lose its support, and become prone to cleavage (**Figure 7.8**).

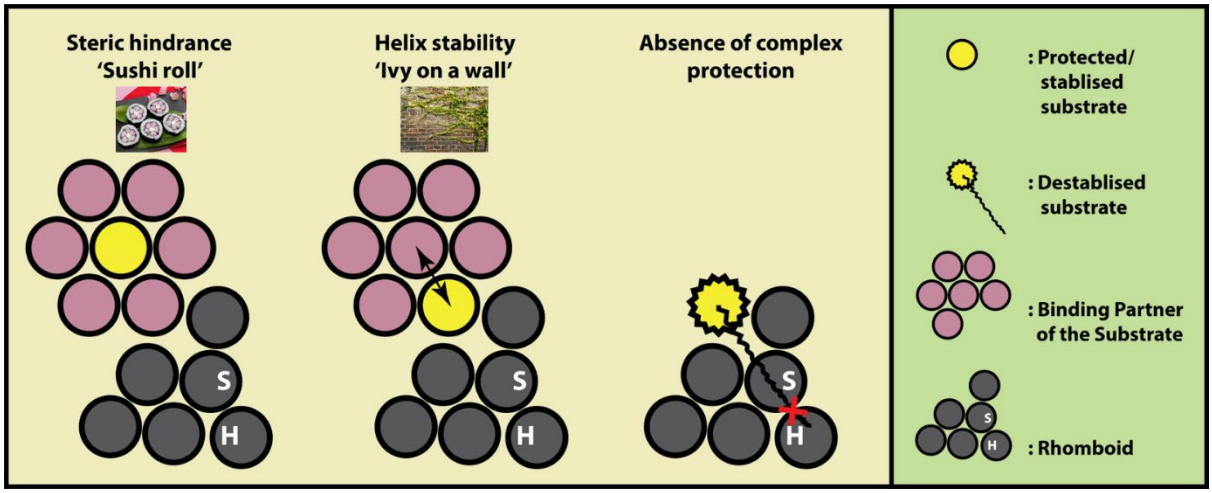


Figure 7.8. Schematic representation of the potential hypotheses accounting for the 'complex protection' model. *Protein TMDs are shown as top-down view from the periplasm. Serine and histidine residues in the catalytic dyad of the rhomboid are shown as S and H in white. Substrate stabilisation through TMD-TMD interaction is shown as the black double-headed arrow. Cleavage of the destabilised substrate is illustrated by the red cross.*

Current evidence shows that, in the crystal structure of the FDH-N complex (**Figure 7.4A**), the TMD of FdnH is not completely surrounded by FdnI, suggesting the TMD might still be available for cleavage, consistent with the 'Ivy on a wall' hypothesis. This structural feature is also likely to be conserved in the FDH-O complex due to their high sequence homology. However, due to the lack of knowledge of interactions of membrane-bound components of both FDH-O and Hyd-2 complexes at an atomic level, the 'Sushi roll' hypothesis cannot be ruled out. If the steric hindrance *i.e.,* the 'Sushi roll' hypothesis constitutes the sole mechanism mediating complex protection, then the hypothesis would predict that if the substrate binding partner is removed, the substrate will still be cleaved by their cognate rhomboids even when helix-destabilisation is abolished. Proline is well-known to confer

helix instability and induce hinges in TMDs (Cordes *et al.,* 2002), and Pro^{300} is critical for HybA cleavage by GlpG when the Hyd-2 complex is intact (**Figure 5.11**). Scanning the TMDs of FdoH and FdnH (**Figure 5.5**) identified proline residues (Pro^{259} in both proteins), which may be important for helix destabilisation. Therefore, I tested cleavage of all three substrates when the proline in the TMD of each was changed to an alanine, in the presence and absence of their membrane-bound putative binding partners. Of note, proline-to-alanine mutations abolished rhomboid-mediated cleavage of HybA, FdoH, and FdnH even in the absence of the membrane-bound interaction partners (**Figure 7.9-7.11**), consistent with the 'ivy on a wall' hypothesis.

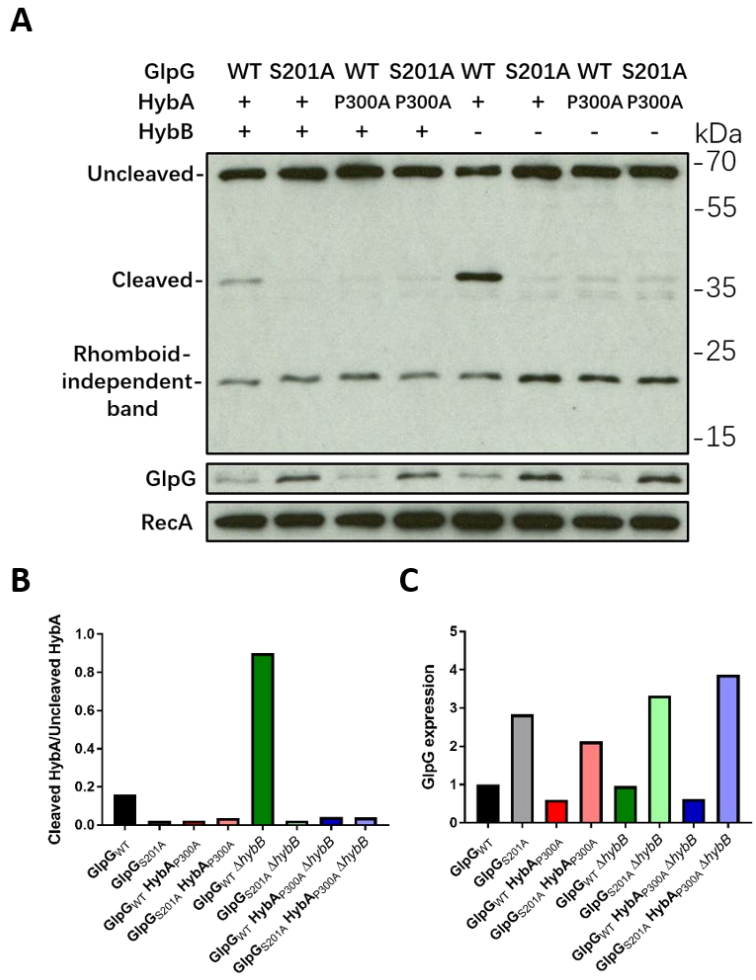


Figure 7.9. Mutation of the TMD Pro^{300} to alanine in HybA abolishes cleavage of uncomplexed HybA by GlpG. ***(A)*** *HybA cleavage by endogenous GlpG is abolished upon*

Pro300Ala (P300A) mutation of HybA. The sequence encoding an sfCherry-3xFLAG tag was fused to the 3′ end of hybA ORF on the chromosome to test full-length HybA cleavage by endogenous GlpG (all strains in ***Figure 7.9*** *lack Rhom7). Ser201Ala (S201A) mutation was introduced to the glpG ORF on the chromosome to inactivate the enzyme without perturbing glpG expression. The sequence encoding a 3xHA tag was fused to the 5′ end of glpG ORF on the chromosome to detect endogenous GlpG. Bacteria were grown anaerobically to exponential phase (OD_{600} = 0.5) before being lysed for western blot analysis (n = 1). Cleavage of the full-length HybA was detected with anti-FLAG antibodies. GlpG expressed from its endogenous locus was detected with anti-HA antibodies. RecA was used as a loading control. The ratio of cleaved to uncleaved HybA and GlpG expression from* ***(A)*** *were quantified and shown in* ***(B)*** *and* ***(C)****.*

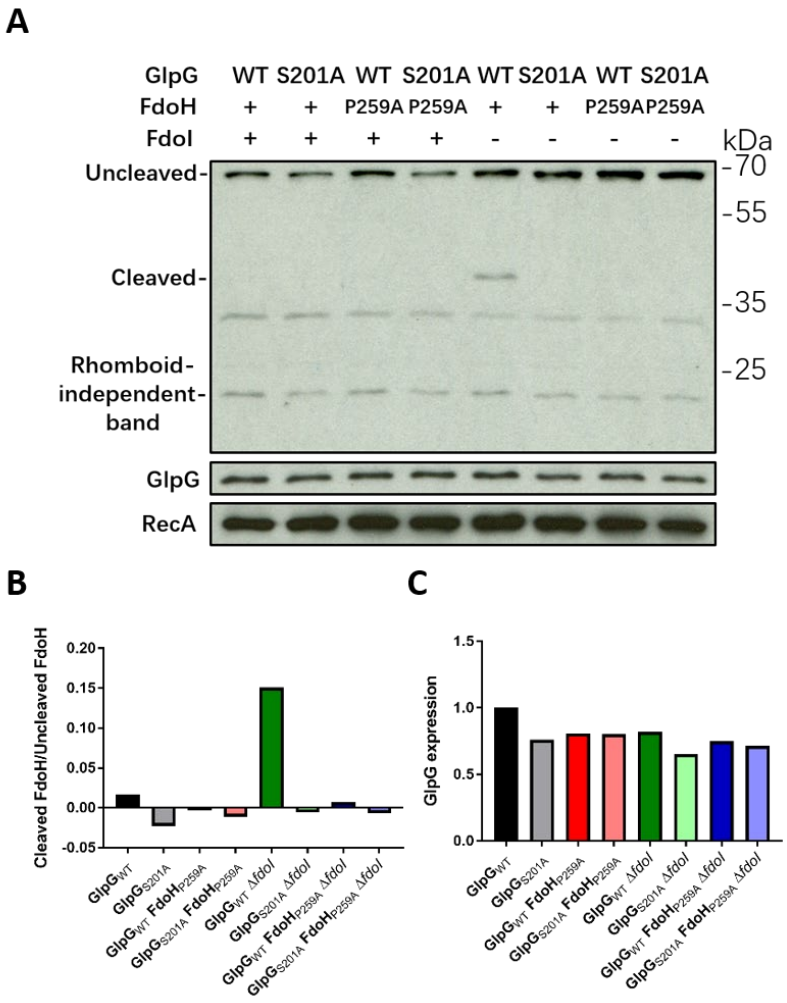


Figure 7.10. Mutation of TMD-embedded Pro^{259} to alanine in FdoH is sufficient to abolish cleavage of uncomplexed FdoH by GlpG. ***(A)*** *FdoH cleavage by endogenous GlpG is abolished upon Pro259Ala (P259A) mutation of FdoH. The sequence encoding an sfCherry-3xFLAG tag was fused to the 3' end of fdoH ORF on the chromosome to test full-length FdoH cleavage in the presence of endogenous GlpG (all strains in* ***Figure 7.10*** *lack Rhom7). GlpG was tagged in the same way as in* ***Figure 7.9****. Bacteria were grown aerobically to exponential phase (OD_{600} = 0.8) before being lysed for western blot analysis (n = 1). Cleavage of the full-length FdoH was detected with anti-FLAG antibodies. GlpG expressed from its endogenous locus was detected with anti-HA antibodies. RecA was used as a loading control. The ratio of cleaved to uncleaved FdoH and GlpG expression from* ***(A)*** *were quantified and shown in* ***(B)*** *and* ***(C)****.*

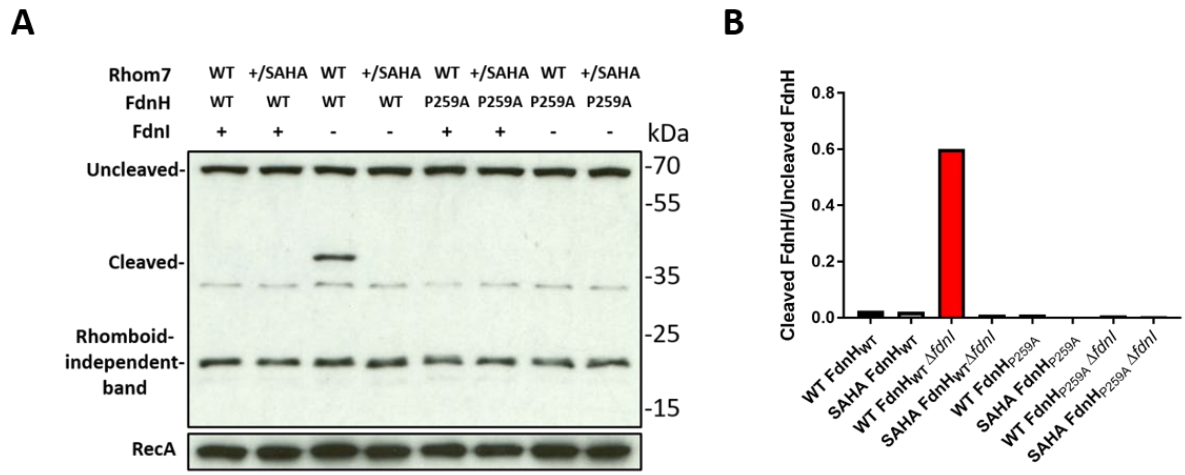


Figure 7.11. Mutation of TMD-embedded Pro^{259} to alanine in FdnH is sufficient to abolish cleavage of uncomplexed FdnH by plasmid-encoded Rhom7. ***(A)*** *FdnH cleavage by plasmid-encoded Rhom7 is abolished upon Pro259Ala (P259A) mutation of FdnH. The sequence encoding an sfCherry-3xFLAG tag was fused to the 3' end of fdnH ORF on the chromosome to test full-length FdnH cleavage by plasmid-encoded Rhom7 (all strains in* ***Figure 7.11*** *lack GlpG). Bacteria were grown anaerobically to exponential phase (OD_{600} = 0.5) before being lysed for western blot analysis (n = 1). Cleavage of the full-length FdnH was detected with anti-FLAG antibodies. RecA was used as a loading control. The ratio of cleaved to uncleaved FdnH was quantified and shown in* ***(B)****.*

7.5. Conclusion and discussion

In this chapter, I began by observing that GlpG-mediated cleavage of HybA was enhanced when HybA was not associated with the Hyd-2 complex, either through deletion of its interacting partner HybB, or through HybA over-expression. These observations led me to propose the 'complex protection' model, which states that GlpG preferentially cleaves uncomplexed substrates. This model was further tested with FdoH cleavage by GlpG and FdnH cleavage by Rhom7, and in both cases, only uncomplexed FdoH and FdnH were cleaved by their cognate rhomboid, consistent with the model.

Although increased cleavage by cognate rhomboid upon removal of known (in the case of FdnH) or proposed (in the case of HybA and FdoH) membrane-bound interacting partner was observed for all three substrates, the basal levels of cleavage of these substrates differed when their cognate complexes were intact. While HybA displayed detectable cleavage when the Hyd-2 complex is intact, cleavage of FdoH and FdnH was only detectable upon removal of FdoI and FdnI. This difference in basal cleavage susceptibility could result from two steps. First, Hyd-2 is a tetramer with two Tat-dependent components whereas FDH-O and FDH-N are both trimers with one Tat-dependent component (Dubini and Sargent, 2003; Jormakka *et al.,* 2002; Unden *et al.,* 2014). Therefore, it is conceivable that the error rate of HybA incorporation into the Hyd-2 complex is higher than its counterparts to FDH-O and FDH-N, resulting in more uncomplexed HybA to be targerted by GlpG. Another hypothesis would be that there is a mechanism by which rhomboid substrates dissociate from their respective complexes under certain circumstances, *e.g.,* when they are damaged. In that case, our observations could be interpreted as HybA has a higher chance of complex dissociation than FdoH and FdnH, possibly due to HybA being more prone to damage.

I also investigated the mechanistic basis of the 'complex protection' model. Two non-mutually exclusive hypotheses were raised: 1) the 'Sushi roll' hypothesis states that rhomboid substrates are protected from its binding partner through steric hindrance, *i.e.,* substrate access by rhomboid is physically blocked by the binding partner of the substrate; 2) the 'Ivy on a wall' hypothesis states that the TMD of rhomboid substrates, is intrinsically unstable, and prone to rhomboid cleavage due to the presence of helix-destabilising residues, *e.g.,* proline. However, the TMD can be stabilised by its interaction with the TMDs of another complex subunit, which leads to it being refractory to rhomboid cleavage (Moin and Urban, 2012). I demonstrated that proline-to-alanine mutation alone can confer HybA, FdoH, and FdnH resistance to cleavage by their cognate rhomboid in the absence of their interacting partners. These data argue in favour of the 'Ivy on a wall' hypothesis as the molecular mechanism for the 'complex protection' model.

Notably, endogenous levels of GlpG were also observed to vary according to several factors. First, wild-type GlpG always seemed to be less abundant than $GlpG_{S201A}$, but this difference was only observed under anaerobic conditions (**Figure 7.1 & 7.3**). Secondly, both HybA over-expression and HybB removal increased wild-type GlpG levels this effect was less clear for $GlpG_{S201A}$ (**Figure 7.1 & 7.3**). In contrast, over-expression of HybO did not alter wild-type GlpG levels (**Figure 7.3**). Difference in the amount of GlpG could stem from differential regulation at three major levels: transcription, mRNA stability and translation, and protein stability/degradation. First, it is hard to explain how transcriptional regulation mediates this difference as the S201A point mutation on the chromosome did not alter any DNA sequence in the promoters regulating *glpG* (Yang and Larson, 1998). Secondly, not only the promoters but also the 5' and 3' UTRs of *glpG* were not altered at the sequence level by the S201A point mutation. Therefore, this point mutation might either stabilise the *glpG*

transcript through making it less prone to degradation or alternatively, the point mutation may allow formation of additional secondary structures with the UTRs to make translation more efficient. But, for either hypothesis to be true, an RNase (for the mRNA degradation hypothesis) or an RNA chaperone (for the mRNA stability hypothesis) needs to respond in an oxygen- and HybA/HybB-dependent manner. Finally, wild-type GlpG might be more susceptible for degradation than $GlpG_{S201A}$ in anaerobic conditions. A GlpG self-degradation model seems to be consistent with that but it is hard to explain how the self-degradation is inhibited by oxygen as GlpG is clearly active in aerobic conditions (**Figure 3.5**). Thus, if GlpG degradation is occurring, it is likely to be executed by a different protease. Following this model, the higher susceptibility to degradation of wild-type GlpG to degradation than $GlpG_{S201A}$ could be explained by that $GlpG_{S201A}$ could trap HybA in its active site due to the inactivation mutation, forming a $GlpG_{S201A}$-HybA complex that locks $GlpG_{S201A}$ in a 'degradation-refractory' state. This hypothesis is also consistent with the observations that HybA but not HybO over-expression and HybB removal increased wild-type GlpG levels. As what HybA over-expression and HybB removal had in common was to expand the pool of uncomplexed HybA to form complex with GlpG, locking GlpG in the 'degradation-refractory' state. Although wild-type GlpG is able to exit this state by cleaving HybA, the cleavage capacity may still be saturated by excessive amount of HybA as the catalytic rate of GlpG has an extremely low k_{cat} ($\approx 0.006\ s^{-1}$) (Dickey *et al.,* 2013).

Chapter 8. Discussion and future directions

In summary, this work identified two rhomboids: GlpG and Rhom7, in *S. sonnei*. In addition, this work established HybA and FdoH as physiological substrates for GlpG, and FdnH as physiological substrate Rhom7 (although a condition in which the endogenous rhom7 locus is robustly expressed is yet to be found). These three rhomboid substrates share many common features: 1) structurally, they possess an N-terminal ferredoxin domain containing four [4Fe-4S] clusters and a C-terminal TMD; 2) they are part of a membrane-bound multi-subunit respiratory complex using alternative electron donors and acceptors (Abaibou *et al.,* 1995; Dubini and Sargent, 2003; Jormakka *et al.,* 2002); and 3) they are substrates of the Tat protein-transport system, although FdoH and FdnH do not possess a Tat signal peptide themselves (Sargent *et al.,* 2002). The identification of HybA, FdoH, and FdnH as substrates for GlpG and/or Rhom7 provides some explanation as to why the Biolog phenotype microarrays did not reveal any phenotypes of CS14Δ*glpG*Δ*rhom7*. The Biolog screens are performed in aerobic conditions, in which both Hyd-2 and FDH-N are not expressed. FDH-O, although expressed aerobically, requires formate as its substrate. However, formate was not included as one of the carbon sources for the screen, but instead used at non-physiological concentrations as an osmolyte stress (Bochner *et al.,* 2001). Biolog analysis has previously been employed to characterise *E. coli* GlpG (Clemmer *et al.,* 2006). The results showed that *E. coli* Δ*glpG* exhibited was a slight increase in resistance to cefotaxime, a β–lactam antibiotic (Clemmer *et al.,* 2006). Although FDH-N-mediated formate metabolism has been shown to protect stationary-phase *E. coli* against a human antimicrobial peptide BPI (bacteriocidal/permeability-increasing protein) (Barker *et al.,* 2000), whether FDH-N is applicable to cefotaxime resistance and whether rhomboids can modulate it require further testing.

Rhomboid proteases are a family of master regulators involved in modulating key biological processes across different kingdoms of life. They were first implicated in initiating epidermal growth factor signalling during *Drosophila* development (Urban *et al.,* 2002), and later in mediating bacterial quorum sensing (Stevenson *et al.,* 2007), they have roles in controlling invasion and adherence of eukaryotic parasites to host cells (Baker *et al.,* 2006; Riestra *et al.,* 2015); and regulate mitochondrial quality control and host cell apoptosis (Jin *et al.,* 2010; Saita *et al.,* 2017). In each of these key processes, the precise regulation of rhomboid activity is paramount. One well-known mechanism of regulating rhomboid activity is compartmentalisation. In *Drosophila*, rhomboid-1 is localised in the Golgi apparatus and the cleavage of its substrate Spitz is critically dependent on Star, a chaperone that trafficks Spitz from the endoplasmic reticulum (ER) to the Golgi apparatus (Urban *et al.,* 2001). Another method of regulation of rhomboid activity is stimulation by specific environmental cues. Again in *Drosophila*, rhomboid-4 (DmRho4) contains an N-terminal EF-hand domain capable of sensing cytosolic Ca^{2+} concentrations (Baker and Urban, 2015). Ca^{2+} binding to the EF-hand domain was shown both *in vitro* and *in vivo* to potently stimulate DmRho4 intramembrane proteolysis of Spitz by facilitating substrate access to the active site (Baker and Urban, 2015).

In this work, a potentially novel mechanism of regulation of rhomboid protease activity has been unveiled. This regulation is exerted at the level of the integrity of the complex in which the rhomboid substrate resides so that only uncomplexed substrates are amenable to cleavage by their cognate substrates, namely, the 'complex protection' model. This model is in some way reminiscent of regulation through compartmentalisation in the case of *Drosophila* rhomboid-1 (Urban *et al.,* 2001), but as bacteria do not possess complex organellar structures

such as ER and Golgi apparatus, ‘compartmentalisation’ is achieved at a more microscopic level.

Multi-component membrane protein complexes play key roles in a wide range of biological processes, for instance, T-cell receptor (TCR) in adaptive immunity, mitochondrial respiratory complexes in energy metabolism, and T3SS in bacterial pathogenesis (Chaban *et al.,* 2014; Coburn *et al.,* 2007; Wucherpfennig *et al.,* 2010). Correct assembly of individual components is crucial for the function of these intricate molecular machines, as demonstrated by multiple diseases caused by complex assembly defects (Ghezzi and Zeviani, 2018; Tjon *et al.,* 2008). However, our knowledge and understanding of quality control mechanisms at the level of protein complex assembly is surprisingly scarce. One of the best-characterised examples illustrating assembly-dependent quality control is the biogenesis of αβTCR (**Figure 8.1**) (Feige and Hendershot, 2013).

αβTCR biogenesis is rigorously controlled, as only properly assembled receptors are capable of traversing the secretory pathway and reaching the cell surface (Klausner *et al.,* 1990). A mature αβTCR consists of the α and β chains and the coreceptor chains, CD3γ, CD3δ, CD3ε, and the ξ chain (**Figure 8.1A**). The α chain contains two positively charged residues: one lysine and one arginine, in its TMD, making the TMD energetically unfavourable when integrated in the membrane alone. Therefore, during TCR biogenesis, if the α chain is properly assembled with CD3ε and CD3δ, each of which has a negatively charged aspartic acid in their TMD, the α chain is stabilised in the ER membrane through TMD-TMD interactions and hence able to assemble with two ξ chains and proceed in the secretory pathway to reach the cell surface (**Figure 8.1B**) (Feige and Hendershot, 2013). In contrast, misassembled α chain inserted in the ER membrane will exit the membrane and becomes

captured by an ER luminal chaperone Bip, which targets this α chain for degradation through ER-associated degradation (ERAD) (**Figure 8.1B**) (Feige and Hendershot, 2013). Although the αβTCR is very distant from the respiratory enzyme complexes studied in this work, both phylogenetically and functionally, they illustrate one common challenge faced by the cell: 'How to ensure only correctly assembled complexes are present in the correct place?'. In the case of αβTCR, the built-in check points are the charged residues in the TMD of its subunits, for which incorrect assembly leads to membrane exit and degradation (Feige and Hendershot, 2013). In the case of Hyd-2, FDH-O, and FDH-N, we propose that the built-in check points are the helix-destabilising proline residues in the TMD of HybA, FdoH, and FdnH, respectively. These proline residues could serve as assembly sensors, and instead of promoting membrane exit, work in concert with GlpG and/or Rhom7 to trigger the intramembrane proteolysis of uncomplexed HybA, FdoH, and FdnH. Interestingly, in *Wolinella succinogenes*, a bacterium isolated from bovine rumen fluid (Wolin *et al.,* 1961), the formate dehydrogenase adopts a different architecture. The FdnH orthologue lacks the C-terminal TMD, instead, this TMD resides in the FdnI homologue, making FdnI a five-TMD protein (Stanley *et al.,* 2002). In this case, the 'assembly' of these five TMDs is achieved through genetic fusion instead of protein-protein interactions. Consistent with our proposed model, BLASTp searches failed to identify any rhomboids in *W. succinogenes*.

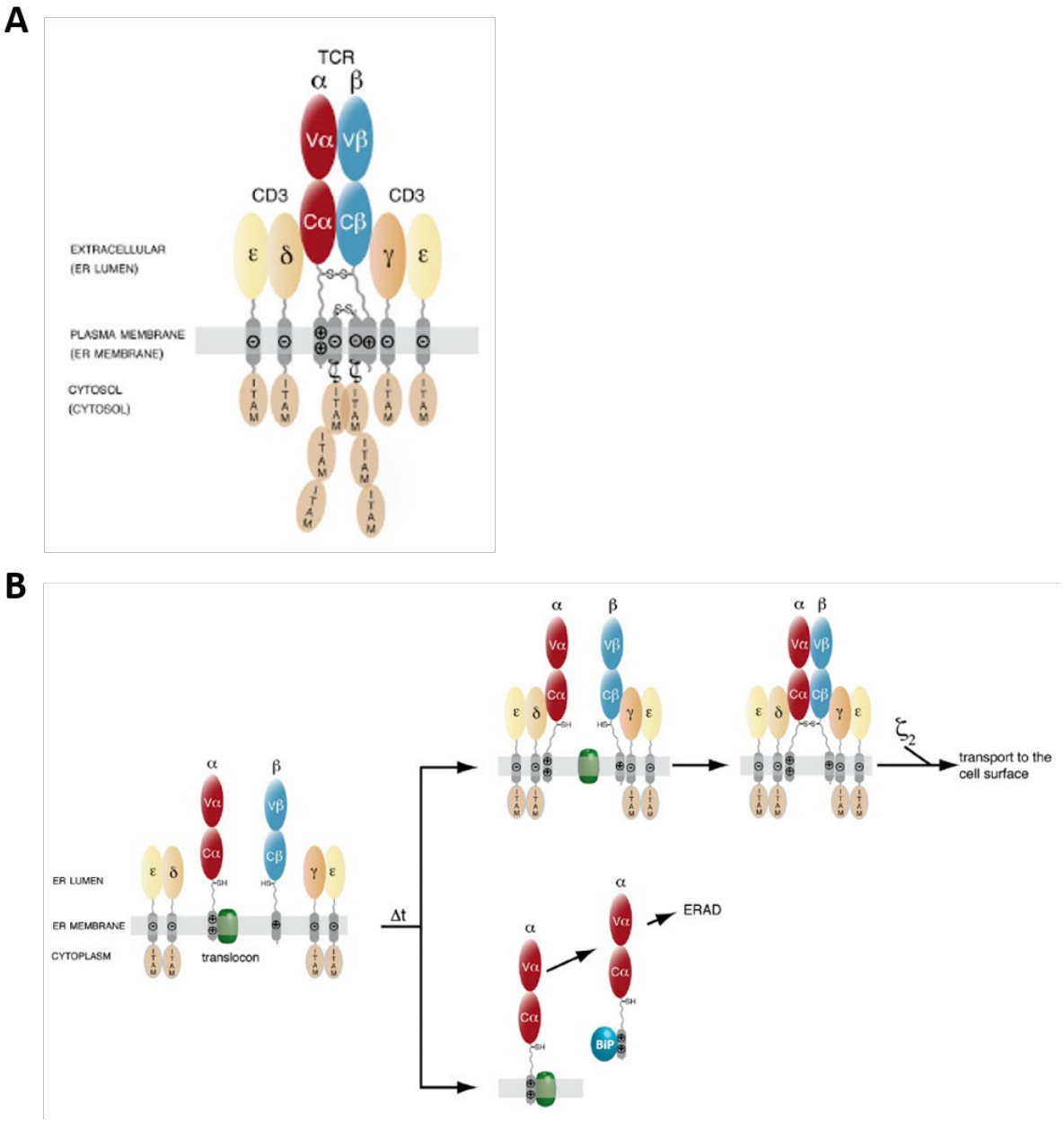


Figure 8.1. Quality control of the αβTCR by assembly-dependent membrane integration. ***(A)*** *Schematic representation of the TCR complex. Topology of TCR components with respect to different cellular compartments are indicated. Positively- (+) and negatively- (-) charged residues embedded in TMDs (shown as grey rods) and interchain disulphide bridges are shown. Cytosolic Immunoreceptor Tyrosine-based Activation Motifs (ITAMs) are also shown.* ***(B)*** *Model for the assembly-dependent quality control mechanism in αβTCR biosynthesis. The Sec translocon for initial membrane integration of individual components of TCR is shown in green, reduced disulphide bridge*

is shown as two thiols (-SH). BiP (in blue), an ER-resident chaperone, targets uncomplexed α-chains that exit the membrane for ERAD. This figure is adapted from (Feige and Hendershot, 2013).

GlpG has been shown to be enzymatically inefficient ($k_{cat} \approx 0.006\ s^{-1}$) due to its catalytically weak dyad instead of a classical Ser-His-Asp triad (Dickey *et al.,* 2013). GlpG was proposed to have evolved to be kinetically slow, which would protect non-substrates by ensuring they have sufficient time to dissociate from GlpG before cleavage commences (Dickey *et al.,* 2013). Following this logic, GlpG was proposed to act as a membrane-patrolling enzyme searching for unfolded membrane proteins as a repair mechanism analogous to DNA glycosylases, enzymes that patrol the chromosome looking for damaged bases (Dickey *et al.,* 2013; Jacobs and Schar, 2012). Our proposal for GlpG as a quality control enzyme for multi-protein membrane complexes strongly echoes this previous work. Moreover, a successful quality control mechanism should in theory not interfere with the system when the system is operating normally, which is consistent with the observations that GlpG did not modulate Hyd-2 activity.

A striking feature shared by all rhomboid substrates identified in this work is the presence of a UXKGUUXP motif in their TMD, where U stands for a hydrophobic residue, X for a residue that can be polar or non-polar, G and P are the predicted P1 and helix-destabilising residues, respectively (**Figure 8.2**).

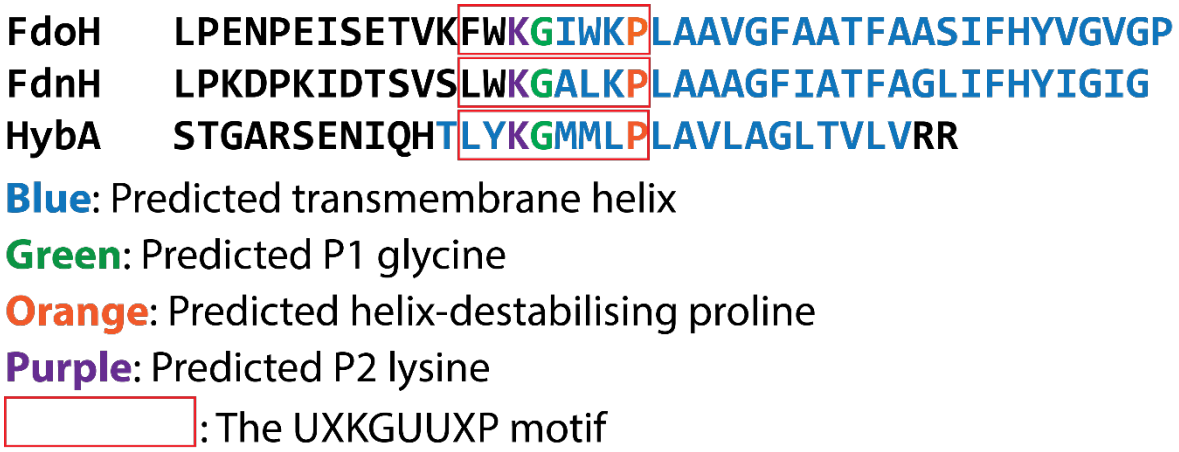


Figure 8.2. Alignment of amino acid sequences of TMD regions of FdoH, FdnH, and HybA. *Predicted TMD residues are shown in blue, and the UXKGUUXP motifs are circled with red boxes.*

In addition to the G and P, which are critical for HybA cleavage by GlpG and Rhom7 as disscussed in chapter 5, a lysine residue occupies the P2 position in all three substrates. This lysine may also contribute to rhomboid cleavage as its positively charged side chain is predicted to form a hydrogen bond with the nitrogen atom from the side chain of His^{150} of GlpG (or His^{76} of Rhom7 based on alignment), further stabilising the oxyanion hole (Cho *et al.,* 2016). If this hypothesis is true, it would predict that mutating this lysine to aspartic/glutamic acid or non-polar residues will reduce GlpG/Rhom7-mediated cleavage. Interestingly, although FdoH and FdnH are specifically cleaved by GlpG and Rhom7, respectively, the motifs are highly similar in these two proteins, with the putative P1 Gly^{255} and the helix-destabilising Pro^{259} completely conserved (**Figure 8.2**). Differences between FdoH and FdnH lie in the putative P4, P1', and P2' residues. For the putative P4 residues, both phe^{252} in FdoH and leu^{252} and FdnH are hydrophobic residues, which are well-tolerated by *E. coli* GlpG (Cho *et al.,* 2016) and hence not likely to be responsible for specificity. The rhomboid domain of Rhom7 shares higher homology to that of *P. stuartii* AarA than to GlpG

(**Figure 3.2**), and when the P1' and P2' residues were mutated to I and W (as in FdoH) in TatA, the physiological substrate of AarA, this resulted in > 50% reduction of TatA cleavage by AarA *in vitro* (Strisovsky *et al.,* 2009). In constrast, when the P1' and P2' of TatA were mutated to A and L (as in FdnH) or M and M (as in HybA), more than 75% cleavage by AarA was retained (Strisovsky *et al.,* 2009). If we assume that higher sequence homology between rhomboids leads to higher similarity in substrate specifity, this potentially offers an explanation why FdnH and HybA are cleaved by Rhom7 *in vivo* while FdoH is not. To futher test these hypotheses about substrate specificity possessed by GlpG and Rhom7, firstly I need to confirm that the glycine residues are indeed the P1 residue in each substrate, using mass spectroscopy or N-terminal sequencing. Subsequently, mutagenesis of the confirmed P4, P1', and P2' residues will reveal the determinants that govern substrate specificity of both GlpG and Rhom7.

Current efforts to further decipher the molecular basis of rhomboid catalysis is hindered by the lack of a physiological rhomboid-substrate pair that is amenable to X-ray crystallography. GlpG seems to be the best candidate for structural characterisation as demonstrated by the fruitful insights provided by the previous structural and enzymological studies of this rhomboid. Our work has identified two physiological substrates of GlpG: HybA and FdoH. Future work could centre around obtaining the structure of a scission complex formed by GlpG with the TMD of HybA/FdoH (**Figure 1.3**) (Cho *et al.,* 2016). In addition, the structure of GlpG covalently linked to a HybA/FdoH-mimicking peptide aldehyde inhibitor would shed more light on the interactions of P1-P4 residues with GlpG (Cho *et al.,* 2016).

Regarding the 'complex protection' model, this work has provided several lines of evidence that are consistent with the model (**Chapter 7**). If the model is correct, it would predict that GlpG only cleaves HybA/FdoH that is not complexed. This is challenging to demonstrate experimentally as the cleavage product of complexed HybA/FdoH and uncomplexed HybA/FdoH would be indistinguishable. However, as inactive rhomboids have been shown to trap substrates in a complex stable enough for co-immunoprecipitation (Saita *et al.,* 2017), the formation of a stable $GlpG_{S201A}$-HybA/FdoH complex could be exploited to investigate the model further. For example, the 'complex protection' model would predict that $GlpG_{S201A}$ will co-immunoprecipitate with HybA but not with HybB. The model would also predict that HybB will co-immunoprecipitate with HybA but not with $GlpG_{S201A}$.

Finally, both hydrogenase-2 and FDH-N have been shown to be important for mediating pathogenesis of enteric pathogens *in vivo* (Maier *et al.,* 2013; Winter *et al.,* 2013). The fact that all three rhomboid substrates identified in this work possess iron-sulphur clusters, which are prone to damage under many circumstances such as oxidative stresses and metal stress, point to the possibility that GlpG/Rhom7 could exert their quality control role beyond scavenging uncomplexed substrates resulting from basal levels of assembly errors during lab culture, but also including damaged substrates. This could have profound implications in infection, during which *S. sonnei* would experience stresses that can damage these respiratory enzymes from both host innate immune response and antibiotic treatments (Goldblatt, 2014; Slauch, 2011; Belenky *et al.,* 2015; Dwyer *et al.,* 2014). Under those circumstances, the physiological significance of rhomboids would amplify and potentially account for the colonisation defect exhibited by the Δ*glpG* ExPEC (Russell *et al.,* 2017). To test this hypothesis, it would be interesting to test whether the presence of stresses that are

known to damage iron-sulphur clusters enhances rhomboid-dependent cleavage of HybA, FdoH, and FdnH.

Overall, my work identified two rhomboids: GlpG and Rhom7, in *S. sonnei*. In addition, by identifying three substrates: HybA, FdoH, and FdnH for GlpG and/or Rhom7 and their characterisation, I propose that rhomboids act as quality control mechanisms scavenging uncomplexed subunits of multi-protein respiratory complexes important for pathogenesis of enteric pathogens, which could have implications on *S. sonnei* infection. Moreover, the identification of natural substrates for GlpG will pave the way for future kinetic and structural analyses on this fascinating family of enzymes.

Appendices

Table 2.2. Primers used in this work

Primer name	Sequence (5' to 3')
PGL97	AAACGCAAAAGAAAATGCCGATATCCTATTG
PGL98	AATAGGATATCGGCATTTTCTTTTGCGTTTTCATTTGTAGAGCTCATCC
PGL154	CATTATACCTAGGACTGAGCTAGCTGTCAATTCACTAAATAATAGTGAACGGCAGGTATATGTG
PGL155	TTGACAGCTAGCTCAGTCCTAGGTATAATGCTAGCCCAACATGTTACACAATAATGGAGTAATGAACATATGAGCAAAGGAGAAGAAC
PGL185	GGAAGTGTATCGTACAGTAGACGGAGTATAGTAAAACGACGGCCAGTGTTTAAACGTCATAGCTGTTTCCTG
PGL186	AATAATTCCACGGACTATAGACTATACTAGCAGGAAACAGCTATGACGTTTAAACACTGGCCGTCGTTTTAC
PGL222	AATTGCTAACGCAGTCAGGCACCGTGTATGCGCGGAACCCCTATTTGTT
PGL223	CGCCGGACGCATCGGGCCGGATCTAGATATTTACCAATGCTTAATCAGTGAGGC
PGL224	ATATCTAGATCCGGCCCGATGCG
PGL225	CATACACGGTGCCTGACTGCG
PGL228	CGGAGTATAGTAAAACGACGGCCAGTGTTTCGCCATACGAGTAAACGACG

PGL231	CTATACTAGCAGGAAACAGCTATGACGTTTTATAGAATCAAG TAGCCTACAGGGCGGCGATTACCAGGCT
PGL242	AGTCGCAAAACTCACCTTTTTTGTTGTTATCCCTCAATTACCA CAAGTAAAAGGAG
PGL243	CACTATACCTAGGACTGAGCTAGCCGTCAAATAACAACAAAA AAGGTGAGTTTTGCG
PGL244	TTGACGGCTAGCTCAGTCCTAGGTATAGTGCTAGCCCAACATG TACAATAATAATGGAGTCATGAACATATGTTGATGATTACCTC TTTTGCTA
PGL245	ATAACAACAAAAAAGGTGAGTTTTGCGACTCACCTTTTTTATT TGTTTATTTTCGTTTTCGCGCATTGAGCGAATC
PGL251	GGAGTATAGTAAAACGACGGCCAGTGTTTCTTCAGAGGTAGT GCCGGAAG
PGL252	GGGTAATTGAACATTTAACGGC
PGL253	TCGGTTAAGCCGTTAAATGTTCAATTACCCAGGTAACATCGTG CGGATCA
PGL254	ATACTAGCAGGAAACAGCTATGACGTTTCGAATCAGCCGATC GCTATAAC
PGL255	TCTAGAAAGGAAGAAACACCTTATGTCTGC
PGL256	GCATGCTTACGGTTCAGGGATACAAGCC
PGL257	GCATGCTTAATCCATCTGCTGCTGCGC
PGL263	CGGTAGCGGTAAGCTAATTGTTATTACTC
PGL264	GAGTAATAACAATTAGCTTACCGCTACCGGCGTTACGCGCCG TACG

PGL271	TCTAGAGGAATAAGCGACAGCAACGATGTT
PGL272	GCATGCTTATTTTCGTTTTCGCGCATTGAG
PGL305	CAACGCGCCACCCGCAATCAGGCCGCCAATAGCGCAAGCGTTATCTATTCCTG
PGL306	ATTGGCGGCCTGATTGCGGGTGGCGCGTTG
PGL323	CTATACTAGCAGGAAACAGCTATGACGTTTATGCCGCTTATCCCCAGAATGCCGAAATCA
PGL328	CGGAGTATAGTAAAACGACGGCCAGTGTTTATATTAGCAAAGAATCGCGC
PGL353	AAGCCCGCCAAACCACGGCCCGCTGAATTT
PGL354	AAATTCAGCGGGCCGTGGTTTGGCGGGCTTGCGGGCGTGGTGTATG
PGL360	TGCCCCAACGCTGATGA
PGL361	AGTGAGCAGATAGTCATCAGCGTTGGGGCAGCGGGGGCGATTCTGGGA
PGL362	ATACCACGACGATTTCCGG
PGL363	TGTAGAAACTGCCGGAAATCGTCGTGGTATTCACTCCAGAGCGATGAAAACGTTTCAGTT
PGL377	TTGCATGCTTATTCCGTAACGAGTCGGT
PGL378	CTCGTTACGGAATAAGCATGCAAGCTTGGC
PGL414	TATCCATATGATGTTCCAGATTATGCTTATCCCTACGATGTACCAGATTATGCATATCCATACGATGTGCCCGATTATGCATCTGCATCTTCGGTTAAGCCG

PGL415	ATAAGCATAATCTGGAACATCATATGGATACATAAGGTGTTTCTTCCTTT
PGL416	CGGAGTATAGTAAAACGACGGCCAGTGTTTAAGGGAAAAAATAAATATCGGCG
PGL417	AATCTGTAGTTTTTCGTTAGCG
PGL418	CATTTTGACGCTAACGAAAAACTACAGATTGTTGGGCATGGGGAACTTGTTTCTATTGAG
PGL419	CTATACTAGCAGGAAACAGCTATGACGTTTGCAAGTAAAATGCTTGAAATAACAAGATCA
PGL430	CCAGAATCAATGGCCATTAAAG
PGL431	CAAGCGCAACAAATAATACAAG
PGL446	CTTGCGCAGCAGCAGATGGATTAAGCATGCAAGCTTGGC
PGL447	ATCCGTGGACCGGCGTGGGAATTGGCGCTGCAATCGGTGTAGTGCTCGGCGTTCTGCTGTCGTCTAGAATGAGCGATAAAATTATTCACCTG
PGL448	TCCCACGCCGGTCCACGGATTTTCACGCACATACTCATCAGCACGCGCCGCCGCGACACGGGTCGGTACCTTGTCATCGTCATCCC
PGL449	GAGTATCGCCTGGATCGCGTTACTGGTGGCGGTTCTTGTGACTACGTTTAAGAGCCTGTCTAGAATGAGCGATAAAATTATTCACCTG
PGL450	ACGCGATCCAGGCGATACTCAGTATGGGATGACGGCCAACAAATTGCATAATTTCTTGCGGTACCTTGTCATCGTCATCCCCGAGACC

PGL451	CATCATGCAATGGTTTATGTATGGCGGTGGCGTGCTGGGGCTTGGCTTGCTGCTCGGTCTGGTACTGCCGCACCTGATCTCTAGAATGAGCGATAAAATTATTCACCTG
PGL452	TACATAAACCATTGCATGATGATGGTGCGCTGTTTGTCATCCAGCTGTACGCTGGCGGCATCGACCTTTTTCTGCGCCGGTACCTTGTCATCGTCATCCCCGAGACC
PGL453	ATCCGCACGGTATGCCGGGGCTGTTTTACTGGTTATTGATGATCCCCGCGCATCTGTTTATTTTTCGCGGATCTAGAATGAGCGATAAAATTATTCACCTG
PGL454	CCCCGGCATACCGTGCGGATGCCAGAAAGCGCGGACATCGATAGTACGATAGTCGCCTTTATCTTCCAGGCTCGGTACCTTGTCATCGTCATCCC
PGL458	TCAGGTGAATAATTTTATCGCTCATTCTAGAAATCGTTTTCCTCAACATGTAGTAGACCACAACGCCCGCAACCACAAGTCCACCGCCAAAACGACGTAAAATATTATCGGGC
PGL459	CGGGTCTCGGGGATGACGATGACAAGGTACCGACTCCGCGTTCACAGAAACCGGATGTTAACGCTAAAGAGGGCGGCAACGTCTCTGCAGGCGCTATTGGTT
PGL460	TCAGGTGAATAATTTTATCGCTCATTCTAGACACCGCCATCACGCTGACACCGGCAACCAGCCCAACCACACCGCCGAGCAAACCAATAGCGCCTGCAGAGA
PGL461	GGTCTCGGGGATGACGATGACAAGGTACCGAATAACCACGCTACTGTGCAATCTTCCACGCCGGTTTCTGCTGCGCCACTGCTGCAGGTCAGCGGCG

PGL462	CAGGTGAATAATTTTATCGCTCATTCTAGATACCAGCCAGGCAGCAGCGAGGATCAGGGCAATAATGGCGATGAGTGCGCCGCTGACCTGCAGCAGT
PGL463	GGTCTCGGGGATGACGATGACAAGGTACCGTTGACCAAGAACGTCAAGGTTGTCGGTGAACCGCCTGAAGAACCAAGCCTGCTGGCTTCTATCTTCATCTCCTGGTTCCC
PGL464	CAGGTGAATAATTTTATCGCTCATTCTAGAGCCGCCCTGCATTTGACGCATGAAGAAGATCCACACACCAATCAGCAACAGCATCGGGAACCAGGAGATGAAGATAGAAG
PGL465	GGTCTCGGGGATGACGATGACAAGGTACCGCAACAGGTGACGTTTAATTATCTGTTAGCGCAAGGGGAGTACCGTATTTCCACATTGCTGGCGGTATTGTTTGC
PGL466	CAGGTGAATAATTTTATCGCTCATTCTAGAAACTCGCAGCCAGAACAGGCCACAAATCAACCAACCGATAGCAACCCCCGCAGCAAACAATACCGCCAGCAATG
PGL467	GGTCTCGGGGATGACGATGACAAGGTACCGCAGTATTGGGGAAAAATCATTGGCGTGGCCGTGGCCTTACTGATGGGCGGCGGCTTT
PGL468	CAGGTGAATAATTTTATCGCTCATTCTAGACATATGGCCAATTAACAGGCCTAACACTACGCCCCAAAAGCCGCCGCCCATCAGTA
PGL469	GGTCTCGGGGATGACGATGACAAGGTACCGTATGTTTATACAAATGTACAACCTGTTAAGGAAATTAAATCGGAATTTTTAACAAATGAAGTAATATATTTATTTCTCGGG

PGL470	CAGGTGAATAATTTTATCGCTCATTCTAGAGAACCACCCTGTTTTGAATGCCCATACCAAAAGAATGACAACTAGTACAGCCATCCCGAGAAATAAATATATTACTTCATTTG
PGL471	GATGACGATGACAAGGTACCGGGCGACGTTTATTACGTAGGTCATCTGCCGTGGTTCGAGCGCTTGTGGTATGCGCTGGCAAACCATCCGATTCTGCT
PGL472	CAGGTGAATAATTTTATCGCTCATTCTAGACAGCAGACGCCACAGTACCCATGCCAGCAATATCACACTGATTGCCGCCAGCACCGCCAGCAGAATCGGATGGTTTGCC
PGL473	GGTCTCGGGGATGACGATGACAAGGTACCGAACGAAGTTGTAAATTCAGGCGTGATGAACATTGCGTCTTTGGTTGTATCGGTGGTGG
PGL474	CAGGTGAATAATTTTATCGCTCATTCTAGAGATAAAAAACCACAAGATGAGCCCGATAAGAAGAACCACCACCGATACAACCAAAGAC
PGL475	GGGTCTCGGGGATGACGATGACAAGGTACCGTATTACTATCGGGCGAAGCAGGCAGTTTATCGTGCTGATGACTACGTCCACGAAAAACCCTGGCAAGGA
PGL476	CAGGTGAATAATTTTATCGCTCATTCTAGATGCCAGCAACAGTCCTAGTACCAGCCCAACGGCCGCGCCCACACCAATTCCTTGCCAGGGTTTTTCGT
PGL477	GGTCTCGGGGATGACGATGACAAGGTACCGCGTAAACATCCGGGGCTGTTTATCGGTATGTACGTTGCTTTTTTTGCCACCCTGGCGGTGATGTTGCAGTCC

PGL478	CAGGTGAATAATTTTATCGCTCATTCTAGAAAGCAGGATAAA TAATACAACCAGTAGCCAGACAGAGCCTGACAGCGTTTCGGA CTGCAACATCACCGCCA
PGL479	GGGATGACGATGACAAGGTACCGGCCGCACGCGATGCCGTTG GCTGCGCGGATTCTTTTGTTCGTGAAAGACCCTGGTGTAGCGT GGGTA
PGL480	CGTCAGTCAGGTGAATAATTTTATCGCTCATTCTAGAGCGCAT GCTTAACAGTGCGCCGATAAAAATACCTACCGCAGCTGCTGT ACCCACGCTACACCAGGGT
PGL481	CGGGTCTCGGGGATGACGATGACAAGGTACCGTCAAATATCA CCTTTTCTCATGCTTTATTCTATCAAAAATCGCTCATTTTTACT TTTTTTACCATAATGTTACTCAG
PGL482	CAGGTGAATAATTTTATCGCTCATTCTAGACGAGATATAAAA AACAACGTGTAAGTTATGTAAATTGCCGATGCTACTGAGTAA CATTATGGTAAAAAAAG
PGL483	GGTCTCGGGGATGACGATGACAAGGTACCGGTGTGTGCGATG AAAGCCACGGCAGAAGGGCCGAAGCGGATGGATTGTGCATCT GGATTTGGCGCGGCA
PGL484	CAGGTGAATAATTTTATCGCTCATTCTAGACAGCGCATGAGA AACCGCAACAAACCCAAAGGTGGCGGTCACCATCGTTGCCGC GCCAAATCCAGATG
PGL485	GGGATGACGATGACAAGGTACCGAATGCGGCACTGGAATCTT GCCGGGACTTACCGCAGGGAAGCCGACCGCTGCTGGTGAGCC ATGGTATT

PGL486	CAGGTGAATAATTTTATCGCTCATTCTAGATGCCCATGCTGGTAATCCGAGAATCGTACTCACCAGACATCCCAGTGCAATACCATGGCTCACCAGCAG
PGL487	GGTCTCGGGGATGACGATGACAAGGTACCGTTTTTTATTTCTGATGCGGTAGCGGCAACGGGTGCACCGGCGCAAGGTAGCCCGATGTCTTTGATTT
PGL488	CAGGTGAATAATTTTATCGCTCATTCTAGAACGCAGGATCATGAAATAGAAAATCAGACCGAATACCACCAGCATCAAAATCAAAGACATCGGGCTACC
PGL489	GGGATGACGATGACAAGGTACCGAATGCTATCTGGATTGCCGTTGCCGCCGTGAGCCTGCTGGGCCTGGCG
PGL490	CAGTCAGGTGAATAATTTTATCGCTCATTCTAGAGGAGGCATAACCCAGAATGGCGCCAAACGCCAGGCCCAGCAGGCTCA
PGL491	GGTCTCGGGGATGACGATGACAAGGTACCGTACGTCCATGACTTACCGGGCGGCATTCGTGTCTCGTTATCCATTGATATTCTTTACTTTATCACGTCTTCGTGGAA
PGL492	CAGGTGAATAATTTTATCGCTCATTCTAGACATTCGGACCATATTCAGCAAAATTAACGCCGTCAAAATCCAGAACAGAACGCTTTTCCACGAAGACGTGATAAAGT
PGL493	GATGACGATGACAAGGTACCGTCAGCTTCCCTGAAGAATCAACAAGGCTTTAGCCTGCCGGAGGTAATGGTTGCGATGGTG
PGL494	CAGGTGAATAATTTTATCGCTCATTCTAGACTGGATACCCGATAACGCAGTGACAATCATCACCATCAGCACCATCGCAACCATTACCT

PGL495	GGATGACGATGACAAGGTACCGAGCCGCGTGGTCGATATTCCGCAAATGGGTACTCATTCCACCGCCGATACCGTCGGTTTAACCGCGCTT
PGL496	CAGGTGAATAATTTTATCGCTCATTCTAGAAACGGCGCTGGCGACTGCGTGCACACCAACAGCCGCAGCCACCACACCAAGCGCGGTTAAACCGACGG
PGL497	GGTCTCGGGGATGACGATGACAAGGTACCGACGACCACTACCGGCGCAACCACGACTGCTACTACCACTGGTGGCGTGGCTGCTGGCGCTGTCGGCACGGCGA
PGL498	CAGGTGAATAATTTTATCGCTCATTCTAGATGCAACAACTGCCAACGTCGCCACACCAACAGCCGTTGCCACACCGACAACTGTCGCCGTGCCGACAGCGCCA
PGL499	GGATGACGATGACAAGGTACCGGGAGCCTCCTTCGGAGCAGGCGCTTCCGCTACGCTGTTTGGTTCAAGTGGTTCTGGTAACTTCATGACCCGCATG
PGL500	CAGGTGAATAATTTTATCGCTCATTCTAGAGATGTTACCCAGCACCAGACTGATGATAAAGAATAACGTTGCCAGCAGCGCCGTCATGCGGGTCATGAAGTTAC
PGL501	GGGATGACGATGACAAGGTACCGGAAACCGTCGAGCCGGAGCCGCAACCGCAGCCAATGCCCGTTGACGATGGCGGGCTGGGCTGTCTGGGCGTCATCAAA
PGL502	CAGGTGAATAATTTTATCGCTCATTCTAGAGAAGAGATGGAGTATTTTTCCTATCAAACCGGCAAAGATAAAAATGTAGAAAATAATTTTGATGACGCCCAGACAGC

PGL508	GGTCTCGGGGATGACGATGACAAGGTACCGGACACCCATACG CTCGCCACCGAAGCCCAACAGGTGGATAACACCACTAACATT TTACGCCTGATGTTG
PGL509	TCAGGTGAATAATTTTATCGCTCATTCTAGATAGCAGCGTGCG GGTCAGCACTACGCCGATTGCCAGTGAGAGCAGCAACATCAG GCGTAAAATGTTAG
PGL516	CGGGTCTCGGGGATGACGATGACAAGGTACCGGATGATATTG TTAATTCTGTGCCCTCCTGGATGTTTACCGCGATTATTGCCGT ATGCA
PGL517	CAGGTGAATAATTTTATCGCTCATTCTAGAATAGAGCCTGGCG AAAATAATTCCAATAATAAACAGAATGCATACGGCAATAATC GCG
PGL518	GGTCTCGGGGATGACGATGACAAGGTACCGAATCCTTTTCCA CCGATGATGGACACGTTGCAAAATATGGCAACGCGTCCCGCG CTGTGGA
PGL519	CAGGTGAATAATTTTATCGCTCATTCTAGAAACCAGCCAGACC AGCATCAGGATAATCGCGACTAACAGTATCCACAGCGCGGGA CGCGTT
PGL520	GGATGACGATGACAAGGTACCGACCGAACGCTACGGAGATTT TGTTCGCTATAACCCACCGTTAACGGGTCAGACGCTGGTGTTG TGGGCG
PGL521	CAGGTGAATAATTTTATCGCTCATTCTAGACACTCGCCAGAGG ATCAGTGCCATCAGCAATAACAACACCACTGGCAGCGCCCAC AACACCAGCGTCT

PGL522	CGGGTCTCGGGGATGACGATGACAAGGTACCGACCTGGGAATATGCGCTAATTGGGTTAGTCGTCGGCATCATTATTG
PGL523	CAGGTGAATAATTTTATCGCTCATTCTAGAATTACCAAAACGCATGGCCACAGCACCAATAATGATGCCGACGACT
PGL524	TCTCGGGGATGACGATGACAAGGTACCGACCCCTGCATTTGCTTCCTGGAATGAATTTTTCGCAATGGGCGGTTACGCCTTTTTTGTCTGGCTGGCGG
PGL525	CAGGTGAATAATTTTATCGCTCATTCTAGACATCACCGAGTGCACGACCAAAACCACCAGCGGAATAACGGTCATCACCACCGCCAGCCAGACAAAAAAG
PGL526	CGGGTCTCGGGGATGACGATGACAAGGTACCGATGCAGGATTTGCGTCTGATATTAATCATTGTTGGCGCGATCGCCATAATC
PGL527	CAGGTGAATAATTTTATCGCTCATTCTAGAGCTGGTCCAGAAACCATGTACCAGTAAAGCGATTATGGCGATCGCGCCAAC
PGL528	GGATGACGATGACAAGGTACCGAGCAAACCACCTCTTTTCTTTATTGTTATCATTGGCTTAATTGTCGTC
PGL529	CAGGTGAATAATTTTATCGCTCATTCTAGACATAAAACGAAACGATGCGGCGACGACAATTAAGCCAATGATAAC
PGL530	GGGCGGGTCTCGGGGATGACGATGACAAGGTACCGTCTGCCCAGTTCTGGAATCGCGCTTCCATAAAACAAAAAATCCCTCTCACCATCCTCTACATTC
PGL531	CAGGTGAATAATTTTATCGCTCATTCTAGATGTGTGACGGAAGAATAAACGTAGAAAAACGCTAAACAAGAGAATGTAGAGGATGGTGAGAGG

PGL538	GGTCTCGGGGATGACGATGACAAGGTACCGAAACTGGACGATCTTTCTACCGGTGCGCGTTCCGAAAATATTCAACACACCCTGTATAAAGGCATGA
PGL539	CAGTCAGGTGAATAATTTTATCGCTCATTCTAGAGCGACGAACCAGCACGGTTAAGCCCGCCAGCACAGCCAGTGGTAGCATCATGCCTTTATACAGGGTGTG
PGL540	GGTCTCGGGGATGACGATGACAAGGTACCGGTACGTCGCGACCTGAATATGGAAGCGGTTGTCGCACCAGAGATTTCTGCGACGGTGGGCGTGGCTGCGGCTT
PGL541	CAGGTGAATAATTTTATCGCTCATTCTAGACAGCACTTTACTAGCGGCAAAAACTGCCGCGCCAACAATGGGGTTAACCGCAAAAGCCGCAGCCACGCCCACC
PGL542	GGATGACGATGACAAGGTACCGGGCACCGGATCTGGCGGTGGCAGTTCATCGCAAGGCCCGCGCCCGCAGCTTGGCGGTCGTGTCGTTACCATC
PGL543	CAGGTGAATAATTTTATCGCTCATTCTAGAAATGGTATAGAAACCACTGGCCGCCCAGATAATGACAATCGCTGCCGCTGCGATGGTAACGACACGACCGC
PGL544	GGTCTCGGGGATGACGATGACAAGGTACCGATGAATATTCTCCACTTCTCACAGTCGGTAAAATGGAGCAGCTGGTTTATCTGTTCTCTACTGC
PGL545	CAGGTGAATAATTTTATCGCTCATTCTAGAGGAAAAACGCCAGATAAACGCAAGGAAAATCAGTCCATGAAGCAGTAGAGAACAGATAAACCAGC

PGL546	GGTCTCGGGGATGACGATGACAAGGTACCGGAACAGTACAATATTCCGCTGTCTGAGTGGACGACATCAATGTATGTGATTCAGTCATCGATGATTTTTGTTTATAGCC
PGL547	AGTCAGGTGAATAATTTTATCGCTCATTCTAGAGCCGCCAAGGAAATAAATTCCCAACGGGATTGCCAGCAACACAGTAAAGACCAGGCTATAAACAAAAATCATCGATGACTG
PGL548	GGTCTCGGGGATGACGATGACAAGGTACCGGAATTGTTGTCTGAATATGGTTTGTTTTTGGCGAAAATCGTTACCGTTGTGCTAGCG
PGL549	CAGGTGAATAATTTTATCGCTCATTCTAGAAGCAACATTGACAATAATGGCGGCAATCGCCGCAATCGCTAGCACAACGGTAACGA
PGL550	GGTCTCGGGGATGACGATGACAAGGTACCGAATCAGCACGGACTGGCGCATCAGGAAAAAGCCAGTGCACGCGGGCAATACAGCCTGTTATTGCTGGGG
PGL551	CAGGTGAATAATTTTATCGCTCATTCTAGAATAAACCACGCGCCAGAGGATCAGAATCAGTGCGCAAAGTGAAACCATCCCCAGCAATAACAGGCTGT
PGL552	GGTCTCGGGGATGACGATGACAAGGTACCGCTGCCGAAAGATCCGAAGATCGACACCTCGGTAAGCCTGTGGAAAGGCGCGTTGAAACCGCTGGCAGCGGCT
PGL553	CAGGTGAATAATTTTATCGCTCATTCTAGAGCCAATACCGATGTAGTGGAAAATCAACCCGGCAAAAGTGGCAATAAAGCCAGCCGCTGCCAGCGGTTTCA

PGL554	GGGATGACGATGACAAGGTACCGTTGCCGGAGAACCCGGAA ATCAGCGAAACCGTGAAATTCTGGAAAGGCATCTGGAAACCG CTCGCGGCTGTTG
PGL555	CAGGTGAATAATTTTATCGCTCATTCTAGACGGACCGACACCC ACGTAGTGGAAGATACTGGCCGCAAAGGTAGCCGCAAAGCCA ACAGCCGCGAGCGGTTTC
PGL564	CGGAGTATAGTAAAACGACGGCCAGTGTTTCATCTCACGTGA AGCGATGG
PGL565	ATAAGACTGAAAGCGATACTTACCGTCTTTCAGCGGCGGATT ACCTTCT
PGL566	AAAGACGGTAAGTATCGCTTTCAG
PGL567	CTATACTAGCAGGAAACAGCTATGACGTTTATCCACAATGCA GAGAGATTTG
PGL568	GCGCCAGTGCAGAAGGATGAGC
PGL569	CGATATCAAAATGCGCACCGG
PGL570	CGGAGTATAGTAAAACGACGGCCAGTGTTTGATATATAGCGA ATGAGGTGAACGATG
PGL571	TTATAGTTATTCCGTTGCGAAGACC
PGL572	TGCCAGGTCTTCGCAACGGAATAACTATAATGATTAGCTTTCG TAGGCCTG
PGL573	CTATACTAGCAGGAAACAGCTATGACGTTTTTTTACCATTACG CGCCGATG
PGL574	CTGGCTGTTCTGGTTACAGGG
PGL575	GCGCAAAAGTCGTTCTTTGCC

PGL612	CGGAGTATAGTAAAACGACGGCCAGTGTTTGGATCACATCCG CAAAGCG
PGL615	CTATACTAGCAGGAAACAGCTATGACGTTTACTTCGGCACGTT GTACGG
PGL617	CTGGTTGCACAGCTCGCAC
PGL618	CCATCCCATTAAAAGATAACGG
PGL619	CGGAGTATAGTAAAACGACGGCCAGTGTTTGTCAACAAAGAC CAGGAGGAG
PGL620	CAGCCAGTGGTAGCATCATAAATTTATACAGGGTGTGTTGAA TATTTTCG
PGL621	TGTATAAATTTATGATGCTACCACTGGCTGTG
PGL622	CTATACTAGCAGGAAACAGCTATGACGTTTAACATAACCTTGT TCAGACGC
PGL623	CAGCCAGTGGTAGCATCATAAA
PGL624	CCGAGATCATCAGTGACCCC
PGL625	CCTGGTCGAACAACGACAAAC
PGL626	TTCTTTGATGATGGCCATGTTATCTTCTTCTTCTCCCCGTGAGT CAGCG
PGL627	GAAGAAGATAACATGGCCATCATC
PGL628	TCATCCTTGTAATCGATATCATGATCTTTATAATCACCGTCAT GGTCTTTGTAGTCACCACCGGTACTATGACG
PGL629	TAAAGATCATGATATCGATTACAAGGATGACGATGACAAGTA AACGGGGAGAATAACCGTG

PGL630	TTCTTTGATGATGGCCATGTTATCTTCTTCTGACTCATGATCGTCTCCTCC
PGL631	TAAAGATCATGATATCGATTACAAGGATGACGATGACAAGTAAACGGAGGAGACGATCATGAG
PGL632	TTCTTTGATGATGGCCATGTTATCTTCTTCTGACTCATGATCGTCTCCTCCGTCGTGATGGTCGTTTTTGGTGTTGCGACGAACCAGCACGG
PGL633	TAAAGATCATGATATCGATTACAAGGATGACGATGACAAGTAATAAAAGCTTGGCACTGGCCG
PGL640	CGAGCACTTCACCAACAAGGACCATAGCATATGACTGGAGATAACACCCTC
PGL641	TTCTTTGATGATGGCCATGTTATCTTCTTCTTCTCCCCGTGAGTCAGCG
PGL653	CGGAGTATAGTAAAACGACGGCCAGTGTTTTTGGCTGCGGTTATTGCATTG
PGL715	CACTAAATAATAGTGAACGGCAGG
PGL716	CACTGATGAGAATATCGTCGG
PGL717	CATACACCACGCCAGAAAGC
PGL718	CAGATCTTTGGTTTCCATGCTGC
PGL746	CGGAGTATAGTAAAACGACGGCCAGTGTTTATTTCTGGTCTGGATACCGG
PGL747	GCTCAAACTCTCTTTAATCACGCC
PGL748	GCTGGCGGCGTGATTAAAGAGAGTTTGAGCGAGCGTTACAGCATTGAGCG

PGL749	CTATACTAGCAGGAAACAGCTATGACGTTTTAATAAATTGCTCATCAGATGGCCTC
PGL751	GCTGGCGATGGATTCATCC
PGL752	GCCGCAACTGGGTGATTTAATC
PGL809	ATCCTCTAGAGTCGACCTGCAGGCATGCATTATTTTCGTTTTCGCGCATTG
PGL810	CAGGAAACAGCTATGACCATGATTACGCCAGGAATAAGCGACAGCAACGATG
PGL880	TAAAGATCATGATATCGATTACAAGGATGACGATGACAAGTAAGGTACCGTGAGGTCCGC
PGL881	CTATACTAGCAGGAAACAGCTATGACGTTTCTCCTGAGTCGATGCCTTC
PGL882	CGGAGTATAGTAAAACGACGGCCAGTGTTTATTTCCAGTCGGAAAACTGC
PGL883	TTCTTTGATGATGGCCATGTTATCTTCTTCCTCATGATGATCCTCCTCGTCATC
PGL884	TAAAGATCATGATATCGATTACAAGGATGACGATGACAAGTAACGATGACGAGGAGGATCATC
PGL885	CTATACTAGCAGGAAACAGCTATGACGTTTAGCGAACAACCTGCATCG
PGL886	TAAAGATCATGATATCGATTACAAGGATGACGATGACAAGTAATCTCTTTTGAACTTTAAGCTGAAAATGGC
PGL887	CTATACTAGCAGGAAACAGCTATGACGTTTATTATTCAGGAATCACTGGACG
PGL888	GGCGCTTTATCGCTGCTG

PGL889	CTGTGAAGATCCCGGCTG
PGL890	CCCAGAATGCCATATACATATGG
PGL891	GAAGGAACGCTACTGAGAG
PGL749	CTATACTAGCAGGAAACAGCTATGACGTTTTAATAAATTGCTCATCAGATGGCCTC
PGL751	GCTGGCGATGGATTCATCC
PGL752	GCCGCAACTGGGTGATTTAATC
PGL904	CTATACTAGCAGGAAACAGCTATGACGTTTTAATAAATTGCTCATCAGATGGCCTC
PGL905	CGGAGTATAGTAAAACGACGGCCAGTGTTTATCATTACTTACATGGTGACCTTTG
PGL906	TTCTTTGATGATGGCCATGTTATCTTCTTCTGCCTGTTTATCCTCATTGCC
PGL907	TAAAGATCATGATATCGATTACAAGGATGACGATGACAAGTAATGAGGATAAACAGGCATGAGC
PGL908	CTATACTAGCAGGAAACAGCTATGACGTTTAAGAAATAGCCAGGGGATG
PGL909	CACCGACAAACCCATTATC
PGL927	GGGGTCGAATGGAAAGCCAGTACCGATTTCTGACTCATGATCGTCTCCTCC
PGL928	GAAATCGGTACTGGCTTTCC
PGL929	TCATCCTTGTAATCGATATCATGATCTTTATAATCACCGTCATGGTCTTTGTAGTCACCGGAAATCTCCAGAGTAG
PGL963	CTGGAAAGCTCTCGCGGCTGTTGGCTTTG

PGL964	GCAAAGCCAACAGCCGCGAGAGCTTTCCAGATGCCTTTCCAGAATTTCACG
PGL965	CTATACTAGCAGGAAACAGCTATGACGTTTTTTTCTGCATAACCGGACC
PGL966	GAAAGGCATCTGGAAAGCT
PGL970	CCAGAGCTTTCAACGCGCCTTTCC
PGL971	TGTGGAAAGGCGCGTTGAAAGCTCTGGCAGCGGCTGGCTTT
PGL972	GAAAGGCGCGTTGAAAGCT
PGL1060	TCATTTGCGCTCCTCGTC
PGL1061	CACGAAGAGAAAGACGAGGAGCGCAAATGAGGTACCGTGAGGTCCGC

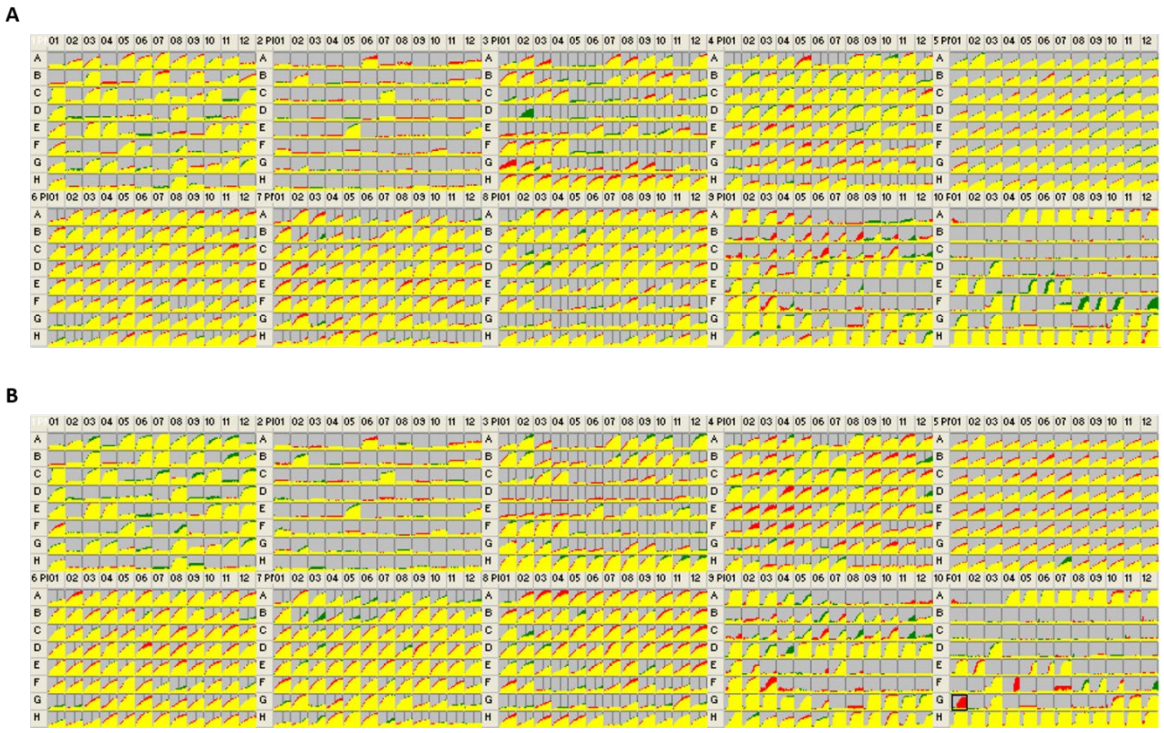


Figure 4.2. Characterisation of *S. sonnei* rhomboid mutants by Biolog phenotypic microarrays. ***(A)*** *Comparisons of cellular respiration (NADH production) of S. sonnei ΔglpGΔrhom7 (green) to wild-type S. sonnei (red) in conditions provided by PM1-10.* ***(B)*** *Comparisons of cellular respiration (NADH production) of S. sonnei ΔglpGΔrhom7 + glpG (green) to wild-type S. sonnei (red) in conditions provided by PM1-10. Areas of overlap are coloured yellow, while differences are shown as patches of red or green. Wild-type S. sonnei seemed to respire better than S. sonnei ΔglpGΔrhom7 + glpG in well G01 of PM10 (pH 9.5 + anthranilic acid). However, this result was not reproducible.*

References

Abaibou, H., Pommier, J., Benoit, S., Giordano, G., and Mandrand-Berthelot, M.-A. (1995). Expression and characterization of the *Escherichia coli fdo* locus and a possible physiological role for aerobic formate dehydrogenase. Journal of bacteriology *177*, 7141-7149.

Abrusci, P., Vergara-Irigaray, M., Johnson, S., Beeby, M.D., Hendrixson, D.R., Roversi, P., Friede, M.E., Deane, J.E., Jensen, G.J., Tang, C.M.*, et al.* (2013). Architecture of the major component of the type III secretion system export apparatus. Nat Struct Mol Biol *20*, 99-104.

Akira, S., Uematsu, S., and Takeuchi, O. (2006). Pathogen recognition and innate immunity. Cell *124*, 783-801.

Al-Hasani, K., Rajakumar, K., Bulach, D., Robins-Browne, R., Adler, B., and Sakellaris, H. (2001). Genetic organization of the she pathogenicity island in *Shigella flexneri* 2a. Microbial pathogenesis *30*, 1-8.

Anderson, M.C., Vonaesch, P., Saffarian, A., Marteyn, B.S., and Sansonetti, P.J. (2017). *Shigella sonnei* Encodes a Functional T6SS Used for Interbacterial Competition and Niche Occupancy. Cell Host Microbe *21*, 769-776 e763.

Andrews, S.C., Berks, B.C., McClay, J., Ambler, A., Quail, M.A., Golby, P., and Guest, J.R. (1997). A 12-cistron *Escherichia coli* operon (*hyf*) encoding a putative proton-translocating formate hydrogenlyase system. Microbiology *143*, 3633-3647.

Arbibe, L., Kim, D.W., Batsche, E., Pedron, T., Mateescu, B., Muchardt, C., Parsot, C., and Sansonetti, P.J. (2007). An injected bacterial effector targets chromatin access for transcription factor NF-kappaB to alter transcription of host genes involved in immune responses. Nature immunology *8*, 47-56.

Ashida, H., Kim, M., Schmidt-Supprian, M., Ma, A., Ogawa, M., and Sasakawa, C. (2010). A bacterial E3 ubiquitin ligase IpaH9.8 targets NEMO/IKKgamma to dampen the host NF-kappaB-mediated inflammatory response. Nature cell biology *12*, 66-73; sup pp 61-69.

Atlung T, and L., B. (1994). Role of the transcriptional activator AppY in regulation of the cyx *appA* operon of *Escherichia coli* by anaerobiosis, phosphate starvation, and growth phase. Journal of bacteriology *176*, 5414-5422.

Avital, A., Maayan, C., and Goitein, K.J. (1982). Incidence of Convulsions and Encephalopathy in Childhood *Shigella* Infections. Clinical Pediatrics *21*, 645-648.

Bagramyan, K., Mnatsakanyan, N., Poladian, A., Vassilian, A., and Trchounian, A. (2002). The roles of hydrogenases 3 and 4, and the F_0F_1- ATPase, in H_2 production by *Escherichia coli* at alkaline and acidic pH. FEBS letters *516*, 172-178.

Bahrani, F.K., Sansonetti, P.J., and Parsot, C. (1997). Secretion of Ipa Proteins by *Shigella flexneri*. Infection and Immunity *65*, 4005-4010.

Baker, K.S., Dallman, T.J., Ashton, P.M., Day, M., Hughes, G., Crook, P.D., Gilbart, V.L., Zittermann, S., Allen, V.G., Howden, B.P.*, et al.* (2015). Intercontinental dissemination of azithromycin-resistant shigellosis through sexual transmission: a cross-sectional study. The Lancet Infectious Diseases *15*, 913-921.

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

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

Ballantine, S.P., and Boxer, D.H. (1985). Nickel-Containing Hydrogenase Isoenzymes from Anaerobically Grown *Escherichia coli* K-12. Journal of bacteriology *163*, 454-459.

Barker, H.C., Kinsella, N., Jaspe, A., Friedrich, T., and O'Connor, C.D. (2000). Formate protects stationary-phase *Escherichia coli* and *Salmonella* cells from killing by a cationic antimicrobial peptide. Molecular Microbiology *35*, 1518-1529.

Barta, M.L., Guragain, M., Adam, P., Dickenson, N.E., Patil, M., Geisbrecht, B.V., Picking, W.L., and Picking, W.D. (2012). Identification of the bile salt binding site on ipad from *Shigella flexneri* and the influence of ligand binding on IpaD structure. Proteins: Structure, Function, and Bioinformatics *80*, 935-945.

Bassis, C.M., Theriot, C.M., and Young, V.B. (2014). Alteration of the murine gastrointestinal microbiota by tigecycline leads to increased susceptibility to *Clostridium difficile* infection. Antimicrob Agents Chemother *58*, 2767-2774.

Baxt, L.A., and Goldberg, M.B. (2014). Host and bacterial proteins that repress recruitment of LC3 to *Shigella* early during infection. PLoS One *9*, e94653.

Beaton, S.E., Evans, R.M., Finney, A.J., Lamont, C.M., Armstrong, F.A., Sargent, F., and Carr, S.B. (2018). The structure of hydrogenase-2 from *Escherichia coli*: implications for H_2-driven proton pumping. Biochem J *475*, 1353-1370.

Belenky, P., Ye, J.D., Porter, C.B., Cohen, N.R., Lobritz, M.A., Ferrante, T., Jain, S., Korry, B.J., Schwarz, E.G., Walker, G.C.*, et al.* (2015). Bactericidal Antibiotics Induce Toxic Metabolic Perturbations that Lead to Cellular Damage. Cell Rep *13*, 968-980.

Benoit, S., Abaibou, H., and Mandrand-Berthelot, M.-A. (1998). Topological Analysis of the Aerobic Membrane-Bound Formate Dehydrogenase of *Escherichia coli*. Journal of bacteriology *180*, 6625-6634.

Berg, B.L., and Stewart, V. (1990). Structural genes for nitrate-inducible formate dehydrogenase in *Escherichia coli* K-12. Genetics *125*, 691-702.

Berg, B.L., and Stewart, V. (1991). Nitrate-inducible formate dehydrogenase in *Escherichia coli* K-12. Genetics *125*, 691-702.

Bergeron, J.R.C., Worrall, L.J., De, S., Sgourakis, N.G., Cheung, A.H., Lameignere, E., Okon, M., Wasney, G.A., Baker, D., McIntosh, L.P., *et al.* (2015). The modular structure of the inner-membrane ring component PrgK facilitates assembly of the type III secretion system basal body. Structure *23*, 161-172.

Berks, B.C., Page, M.D., Richardson, D.J., Reilly, A., Cavill, A., Outen, F., and Ferguson, S.J. (1995). Sequence analysis of subunits of the membrane-bound nitrate reductase from a denitrifying bacterium: the integral membrane subunit provides a prototype for the dihaem electron-carrying arm of a redox loop. Molecular Microbiology *15*, 319-331.

Bernardini, M.L., Mounier, J., d'Hauteville, H., Coquis-Rondon, M., and Sansonetti, P.J. (1989). Identification of icsA, a plasmid locus of *Shigella flexneri* that governs bacterial intra- and intercellular spread through interaction with F-actin. Proceedings of the National Academy of Sciences of the United States of America *86*, 3867-3871.

Beyer, L., Doberenz, C., Falke, D., Hunger, D., Suppmann, B., and Sawers, R.G. (2013). Coordination of FocA and pyruvate formate-lyase synthesis in *Escherichia coli* demonstrates preferential translocation of formate over other mixed-acid fermentation products. J Bacteriol *195*, 1428-1435.

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

Blokesch, M., Magalon, A., and Bock, A. (2001). Interplay between the specific chaperone-like proteins HybG and HypC in maturation of hydrogenases 1, 2, and 3 from *Escherichia coli*. J Bacteriol *183*, 2817-2822.

Blokesch, M., Rohrmoser, M., Rode, S., and Böck, A. (2004). HybF, a Zinc-Containing Protein Involved in NiFe Hydrogenase Maturation. Journal of bacteriology *186*, 2603-2611.

Bochner, B.R., Gadzinski, P., and Panomitros, E. (2001). Phenotype microarrays for high-throughput phenotypic testing and assay of gene function. Genome Res *11*, 1246-1255.

Böhm, R., Sauter, M., and Böck, A. (1990). Nucleotide sequence and expression of an operon in *Escherichia coli* coding for formate hydrogenylase components. Molecular microbiology *4*, 231-243.

Bohnhoff, M., Drake, B.L., and Miller, C.P. (1954). Effect of streptomycin on susceptibility of intestinal tract to experimental *Salmonella* infection. Experimental biology and medicine *86*, 132-137.

Bourdet-Sicard, R., Rudiger, M., Jockusch, B.M., Gounon, P., Sansonetti, P.J., and Nhieu, G.T. (1999). Binding of the *Shigella* protein IpaA to vinculin induces F-actin depolymerization. The EMBO journal *18*, 5853-5862.

Brinkmann, V., and Zychlinsky, A. (2012). Neutrophil extracellular traps: is immunity the second function of chromatin? The Journal of cell biology *198*, 773-783.

Brown, M.S., Ye, J., Rawson, R.B., and Goldstein, J.L. (2000). Regulated Intramembrane Proteolysis. Cell *100*, 391-398.

Buchrieser, C., Glaser, P., Rusniok, C., Nedjari, H., D'Hauteville, H., Kunst, F., Sansonetti, P., and Parsot, C. (2000). The virulence plasmid pWR100 and the repertoire of proteins secreted by the type III secretion apparatus of *Shigella flexneri*. Molecular Microbiology *38*, 760-771.

Caboni, M., Pedron, T., Rossi, O., Goulding, D., Pickard, D., Citiulo, F., MacLennan, C.A., Dougan, G., Thomson, N.R., Saul, A.*, et al.* (2015). An O antigen capsule modulates bacterial pathogenesis in *Shigella sonnei*. PLoS Pathog *11*, e1004749.

Capitani, G., De Biase, D., Aurizi, C., Gut, H., Bossa, F., and Grütter, M.G. (2003). Crystal structure and functional analysis of *Escherichia coli* glutamate decarboxylase. EMBO Journal *22*, 4027-4037.

Carbonero, F., Benefiel, A.C., and Gaskins, H.R. (2012). Contributions of the microbial hydrogen economy to colonic homeostasis. Nat Rev Gastroenterol Hepatol *9*, 504-518.

Chaban, Y., Boekema, E.J., and Dudkina, N.V. (2014). Structures of mitochondrial oxidative phosphorylation supercomplexes and mechanisms for their stabilisation. Biochim Biophys Acta *1837*, 418-426.

Chattopadhyay, M.K., Tabor, C.W., and Tabor, H. (2003). Polyamines protect *Escherichia coli* cells from the toxic effect of oxygen. Proceedings of the National Academy of Sciences of the United States of America *100*, 2261-2265.

Chatzidaki-Livanis, M., Geva-Zatorsky, N., and Comstock, L.E. (2016). *Bacteroides fragilis* type VI secretion systems use novel effector and immunity proteins to antagonize human gut Bacteroidales species. Proc Natl Acad Sci U S A *113*, 3627-3632.

Cho, S., Dickey, S.W., and Urban, S. (2016). Crystal Structures and Inhibition Kinetics Reveal a Two-Stage Catalytic Mechanism with Drug Design Implications for Rhomboid Proteolysis. Mol Cell *61*, 329-340.

Chun, C.D., Liu, O.W., and Madhani, H.D. (2007). A link between virulence and homeostatic responses to hypoxia during infection by the human fungal pathogen *Cryptococcus neoformans*. PLoS Pathog *3*, e22.

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

Coburn, B., Sekirov, I., and Finlay, B.B. (2007). Type III secretion systems and disease. Clin Microbiol Rev *20*, 535-549.

Connor, T.R., Barker, C.R., Baker, K.S., Weill, F.X., Talukder, K.A., Smith, A.M., Baker, S., Gouali, M., Pham Thanh, D., Jahan Azmi, I.*, et al.* (2015). Species-wide whole genome

sequencing reveals historical global spread and recent local persistence in *Shigella flexneri*. Elife *4*, e07335.

Cordes, F.S., Bright, J.N., and Sansom, M.S.P. (2002). Proline-induced Distortions of Transmembrane Helices. Journal of Molecular Biology *323*, 951-960.

Cossart, P., and Sansonetti, P.J. (2004). Bacterial invasion: the paradigms of enteroinvasive pathogens. Science (New York, N.Y.) *304*, 242-248.

Costa, S.C., and Lesser, C.F. (2014). A multifunctional region of the *Shigella* type 3 effector IpgB1 is important for secretion from bacteria and membrane targeting in eukaryotic cells. PLoS One *9*, e93461.

Darwin, A., Tormay, P., Page, L., Griffiths, L., and Cole, J. (1993). Identification of the formate dehydrogenases and genetic determinants of formate-dependent nitrite reduction by *Escherichia coli* K12. Journal of General Microbiology *139*, 1829-1840.

Daugherty, A., Suvarnapunya, A.E., and Runyen-Janecky, L. (2012). The role of oxyR and soxRS in oxidative stress survival in *Shigella flexneri*. Microbiol Res *167*, 238-245.

De Strooper, B., Annaert, W., Saftig, P., Craessaerts, K., Mumm, J., Schroeter, E., Schrijvers, V., Wolfe, M., Ray, W., Cupers, P*., et al.* (1999). A presenilin-1-dependent big gamma-secretase-like protease mediates release of Notch intracellular domain. Nature *398*, 518 -522

Deane, J.E., Graham, S.C., Mitchell, E.P., Flot, D., Johnson, S., and Lea, S.M. (2008). Crystal structure of Spa40, the specificity switch for the *Shigella flexneri* type III secretion system. Mol Microbiol *69*, 267-276.

Deber, C.M., and Therien, A.G. (2002). Putting the β-breaks on membrane protein misfolding. Nature structural biology *9*, 318-319.

Deretic, V., and Levine, B. (2009). Autophagy, immunity, and microbial adaptations. Cell host & microbe *5*, 527-549.

Dickenson, N.E., Zhang, L., Epler, C.R., Adam, P.R., Picking, W.L., and Picking, W.D. (2011). Conformational changes in IpaD from *Shigella flexneri* upon binding bile salts provide insight into the second step of type III secretion. Biochemistry *50*, 172-180.

Dickey, S.W., Baker, R.P., Cho, S., and Urban, S. (2013). Proteolysis inside the membrane is a rate-governed reaction not driven by substrate affinity. Cell *155*, 1270-1281.

Ding, H., and Demple, B. (1998). Thiol-Mediated Disassembly and Reassembly of [2Fe-2S] Clusters in the redox-regulated transcription factor SoxR. Biochemistry *37*, 17280 -17286.

Dolla, A., Pohorelic, B.K.J., Voordouw, J.K., and Voordouw, G. (2000). Deletion of the *hmc* operon of *Desulfovibrio vulgaris* subsp. *vulgaris* Hildenborough hampers hydrogen metabolism and low-redox-potential niche establishment. Archives of Microbiology *174*, 143-151.

Dubini, A., Pye, R.L., Jack, R.L., Palmer, T., and Sargent, F. (2002). How bacteria get energy from hydrogen: A genetic analysis of periplasmic hydrogen oxidation in *Escherichia coli*. International journal of hydrogen energy *27*, 1413-1420.

Dubini, A., and Sargent, F. (2003). Assembly of Tat-dependent [NiFe] hydrogenases: identification of precursor-binding accessory proteins. FEBS Letters *549*, 141-146.

Dupont, H.L. (1989). Inoculum Size in Shigellosis and Implications for Expected Mode of Transmission. The Journal of infectious diseases *159*, 1126-1128.

Dwyer, D.J., Belenky, P.A., Yang, J.H., MacDonald, I.C., Martell, J.D., Takahashi, N., Chan, C.T., Lobritz, M.A., Braff, D., Schwarz, E.G.*, et al.* (2014). Antibiotics induce redox-related physiological alterations as part of their lethality. Proc Natl Acad Sci U S A *111*, E2100-2109.

Egile, C., Loisel, T.P., Laurent, V., Li, R., Pantaloni, D., Sansonetti, P.J., and Carlier, M.F. (1999). Activation of the CDC42 effector N-WASP by the *Shigella flexneri* IcsA protein

promotes actin nucleation by Arp2/3 complex and bacterial actin-based motility. The Journal of cell biology *146*, 1319-1332.

Eilers, B., Mayer-Scholl, A., Walker, T., Tang, C., Weinrauch, Y., and Zychlinsky, A. (2010). Neutrophil antimicrobial proteins enhance *Shigella flexneri* adhesion and invasion. Cellular microbiology *12*, 1134-1143.

Elazar, A., Weinstein, J.J., Prilusky, J., and Fleishman, S.J. (2016). Interplay between hydrophobicity and the positive-inside rule in determining membrane-protein topology. Proc Natl Acad Sci U S A *113*, 10340-10345.

Enoch, H.G., and Lester, R.L. (1975). The purification and properties of formate dehydrogenase and nitrate reductase from *Escherichia coli*. Journal of biological chemistry *250*, 6693-6705.

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

Faherty, C.S., Redman, J.C., Rasko, D.A., Barry, E.M., and Nataro, J.P. (2012). *Shigella flexneri* effectors OspE1 and OspE2 mediate induced adherence to the colonic epithelium following bile salts exposure. Mol Microbiol *85*, 107-121.

Feige, M.J., and Hendershot, L.M. (2013). Quality control of integral membrane proteins by assembly-dependent membrane integration. Mol Cell *51*, 297-309.

Forzi, L., and Sawers, R.G. (2007). Maturation of [NiFe]-hydrogenases in *Escherichia coli*. Biometals *20*, 565-578.

Franzon, V.L., Arondel, J., and Sansonetti, P.J. (1990). Contribution of superoxide dismutase and catalase activities to *Shigella flexneri* pathogenesis. Infection and Immunity *58*, 529-535.

Freeman, M. (2014). The rhomboid-like superfamily: molecular mechanisms and biological roles. Annu Rev Cell Dev Biol *30*, 235-254.

Freter, R. (1955). The fatal enteric cholera infection in the guinea pig, achieved by inhibition of normal enteric flora. The Journal of infectious diseases *97*, 57-65.

Fritsch, M.J., Krehenbrink, M., Tarry, M.J., Berks, B.C., and Palmer, T. (2012). Processing by rhomboid protease is required for *Providencia stuartii* TatA to interact with TatC and to form functional homo-oligomeric complexes. Mol Microbiol *84*, 1108-1123.

Galan, J.E., Lara-Tejero, M., Marlovits, T.C., and Wagner, S. (2014). Bacterial type III secretion systems: specialized nanomachines for protein delivery into target cells. Annu Rev Microbiol *68*, 415-438.

Gallio, M., Sturgill, G., Rather, P., and Kylsten, P. (2002). A conserved mechanism for extracellular signaling in eukaryotes and prokaryotes. Proceedings of the National Academy of Sciences of the United States of America *99*, 12208-12213.

Ghezzi, D., and Zeviani, M. (2018). Human diseases associated with defects in assembly of OXPHOS complexes. Essays Biochem *62*, 271-286.

Giannella, R.A., Broitman, S.A., and Zamcheck, N. (1972). Gastric acid barrier to ingested microorganisms in man: studies *in vivo* and *in vitro*. Gut *13*, 251-256.

Goldblatt, D. (2014). Recent advances in chronic granulomatous disease. J Infect *69 Suppl 1*, S32-35.

Guzman, L.-M., Belin, D., Carson, M.J., and Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. Journal of Bacteriology *177*, 4121-4130.

Handa, Y., Suzuki, M., Ohya, K., Iwai, H., Ishijima, N., Koleske, A.J., Fukui, Y., and Sasakawa, C. (2007). *Shigella* IpgB1 promotes bacterial entry through the ELMO-Dock180 machinery. Nature cell biology *9*, 121-128.

Hassan, H.M., and Fridovich, I. (1979). Intracellular production of superoxide radical and of hydrogen peroxide by redox active compounds. Archives of biochemistry and biophysics *196*, 385-395.

Hatzixanthis, K., Palmer, T., and Sargent, F. (2003). A subset of bacterial inner membrane proteins integrated by the twin-arginine translocase. Molecular Microbiology *49*, 1377-1390.

Havelaar, A.H., Kirk, M.D., Torgerson, P.R., Gibb, H.J., Hald, T., Lake, R.J., Praet, N., Bellinger, D.C., de Silva, N.R., Gargouri, N.*, et al.* (2015). World Health Organization Global Estimates and Regional Comparisons of the Burden of Foodborne Disease in 2010. PLoS Med *12*, e1001923.

High, N., Mounier, J., Prevost, M.C., and Sansonetti, P.J. (1992). IpaB of *Shigella flexneri* causes entry into epithelial cells and escape from the phagocytic vacuole. EMBO J *11*, 1991-1999.

Holt, K.E., Baker, S., Weill, F.X., Holmes, E.C., Kitchen, A., Yu, J., Sangal, V., Brown, D.J., Coia, J.E., Kim, D.W.*, et al.* (2012). *Shigella sonnei* genome sequencing and phylogenetic analysis indicate recent global dissemination from Europe. Nat Genet *44*, 1056-1059.

Hood, R.D., Singh, P., Hsu, F., Guvener, T., Carl, M.A., Trinidad, R.R., Silverman, J.M., Ohlson, B.B., Hicks, K.G., Plemel, R.L.*, et al.* (2010). A type VI secretion system of *Pseudomonas aeruginosa* targets a toxin to bacteria. Cell Host Microbe *7*, 25-37.

Hoshino, K., Takeuchi, O., Kawai, T., Sanjo, H., Ogawa, T., Takeda, Y., Takeda, K., and Akira, S. (1999). Cutting edge: Toll-like receptor 4 (TLR4)-deficient mice are hyporesponsive to lipopolysaccharide: evidence for TLR4 as the Lps gene product. Journal of immunology (Baltimore, Md. : 1950) *162*, 3749-3752.

Hu, X.J., Li, T., Wang, Y., Xiong, Y., Wu, X.H., Zhang, D.L., Ye, Z.Q., and Wu, Y.D. (2017). Prokaryotic and Highly-Repetitive WD40 Proteins: A Systematic Study. Sci Rep *7*, 10585.

Huang, J., and Brumell, J.H. (2014). Bacteria-autophagy interplay: a battle for survival. Nature reviews. Microbiology *12*, 101-114.

Huang, Z., Sutton, S.E., Wallenfang, A.J., Orchard, R.C., Wu, X., Feng, Y., Chai, J., and Alto, N.M. (2009). Structural insights into host GTPase isoform selection by a family of bacterial GEF mimics. Nat Struct Mol Biol *16*, 853-860.

Hughes, E.R., Winter, M.G., Duerkop, B.A., Spiga, L., Furtado de Carvalho, T., Zhu, W., Gillis, C.C., Buttner, L., Smoot, M.P., Behrendt, C.L.*, et al.* (2017). Microbial Respiration and Formate Oxidation as Metabolic Signatures of Inflammation-Associated Dysbiosis. Cell Host Microbe *21*, 208-219.

Imlay, J.A. (2008). Cellular defenses against superoxide and hydrogen peroxide. Annu Rev Biochem *77*, 755-776.

Imlay, J.A. (2013). The molecular mechanisms and physiological consequences of oxidative stress: lessons from a model bacterium. Nat Rev Microbiol *11*, 443-454.

Inoue, H., Nojima, H., and Okayama, H. (1990). High efficiency transformation of *Escherichia coli* with plasmids. Gene (Amsterdam) *96*, 23-28.

Jacobs, A.L., and Schar, P. (2012). DNA glycosylases: in DNA repair and beyond. Chromosoma *121*, 1-20.

Jehl, S.P., Nogueira, C.V., Zhang, X., and Starnbach, M.N. (2012). IFNgamma inhibits the cytosolic replication of *Shigella flexneri* via the cytoplasmic RNA sensor RIG-I. PLoS Pathog *8*, e1002809.

Jeong, H.J., Jang, E.S., Han, B.I., Lee, K.H., Ock, M.S., Kong, H.H., Chung, D.I., Seol, S.Y., Cho, D.T., and Yu, H.S. (2007). Acanthamoeba: could it be an environmental host of *Shigella*? Exp Parasitol *115*, 181-186.

Jin, S.M., Lazarou, M., Wang, C., Kane, L.A., Narendra, D.P., and Youle, R.J. (2010). Mitochondrial membrane potential regulates PINK1 import and proteolytic destabilization by PARL. J Cell Biol *191*, 933-942.

Jones, R.W. (1980). Proton translocation by the membrane-bound formate dehydrogenase of *Escherichia coli*. FEMS microbiology letters *8*, 167-171.

Jormakka, M., Byrne, B., and Iwata, S. (2003). Protonmotive force generation by a redox loop mechanism. FEBS Letters *545*, 25-30.

Jormakka, M., Törnroth, S., Byrne, B., and Iwata, S. (2002). Molecular Basis of Proton Motive Force Generation Structure of Formate Dehydrogenase-N. Science *295*, 1863-1868.

Kall, L., Krogh, A., and Sonnhammer, E.L. (2007). Advantages of combined transmembrane topology and signal peptide prediction--the Phobius web server. Nucleic Acids Res *35*, W429-432.

Kamada, N., Chen, G.Y., Inohara, N., and Nunez, G. (2013). Control of pathogens and pathobionts by the gut microbiota. Nat Immunol *14*, 685-690.

Kaniga, K., Delor, I., and Cornelis, G.R. (1991). A wide-host-range suicide vector for improving reverse genetics in Gram-negative bacteria inactivation of the *blaA* gene of *Yersinia enterocolitica*. Gene *109*, 137-141.

Kanjee, U., and Houry, W.A. (2013). Mechanisms of acid resistance in *Escherichia coli*. Annu Rev Microbiol *67*, 65-81.

Kateete, D.P., Katabazi, F.A., Okeng, A., Okee, M., Musinguzi, C., Asiimwe, B.B., Kyobe, S., Asiimwe, J., Boom, W.H., and Joloba, M.L. (2012). Rhomboids of Mycobacteria:

characterization using an *aarA* mutant of *Providencia stuartii* and gene deletion in *Mycobacterium smegmatis*. PLoS One *7*, e45741.

Kateete, D.P., Okee, M., Katabazi, F.A., Okeng, A., Asiimwe, J., Boom, H.W., Eisenach, K.D., and Joloba, M.L. (2010). Rhomboid homologs in mycobacteria insights from phylogeny and genomic analysis. BMC Microbiology *10*.

Kayath, C.A., Hussey, S., El hajjami, N., Nagra, K., Philpott, D., and Allaoui, A. (2010). Escape of intracellular *Shigella* from autophagy requires binding to cholesterol through the type III effector, IcsB. Microbes and infection / Institut Pasteur *12*, 956-966.

Kenne, L., Lindberg, B., Petersson, K., Katzenellenbogen, E., and Romanowska, E. (1978). Structural Studies of *Shigella flexneri* O-Antigens. European Journal of Biochemistry *91*, 279-284.

Kentner, D., Martano, G., Callon, M., Chiquet, P., Brodmann, M., Burton, O., Wahlander, A., Nanni, P., Delmotte, N., Grossmann, J.*, et al.* (2014). *Shigella* reroutes host cell central metabolism to obtain high-flux nutrient supply for vigorous intracellular growth. Proc Natl Acad Sci U S A *111*, 9929-9934.

Kern, R., Malki, A., Abdallah, J., Tagourti, J., and Richarme, G. (2007). *Escherichia coli* HdeB is an acid stress chaperone. J Bacteriol *189*, 603-610.

Kim, D.W., Lenzen, G., Page, A.L., Legrain, P., Sansonetti, P.J., and Parsot, C. (2005). The *Shigella flexneri* effector OspG interferes with innate immune responses by targeting ubiquitin-conjugating enzymes. Proceedings of the National Academy of Sciences of the United States of America *102*, 14046-14051.

Kimbrough, T.G., and Miller, S.I. (2000). Contribution of *Salmonella typhimurium* type III secretion components to needle complex formation. Proceedings of the National Academy of Sciences of the United States of America *97*, 11008-11013.

Klausner, R.D., Lippincott-Schwartz, J., and Bonifacino, J.S. (1990). The T cell antigen receptor insights into organelle biology. Annual Review of Cell Biology *6*, 403-431.

Knopf, R.R., and Adam, Z. (2012). Rhomboid proteases in plants - still in square one? Physiologia plantarum *145*, 41-51.

Konovalova, A., Sogaard-Andersen, L., and Kroos, L. (2014). Regulated proteolysis in bacterial development. FEMS microbiology reviews *38*, 493-522.

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

Kotloff, K.L., Riddle, M.S., Platts-Mills, J.A., Pavlinac, P., and Zaidi, A.K.M. (2017). Shigellosis. The Lancet.

Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J Mol Biol *305*, 567-580.

Kuhlen, L., Abrusci, P., Johnson, S., Gault, J., Deme, J., Caesar, J., Dietsche, T., Mebrhatu, M.T., Ganief, T., Macek, B.*, et al.* (2018). Structure of the core of the type III secretion system export apparatus. Nat Struct Mol Biol *25*, 583-590.

Kuniyoshi, T.M., Balan, A., Schenberg, A.C., Severino, D., and Hallenbeck, P.C. (2015). Heterologous expression of proteorhodopsin enhances H_2 production in *Escherichia coli* when endogenous Hyd-4 is overexpressed. J Biotechnol *206*, 52-57.

Laurinavichene, T.V., and Tsygankov, A.A. (2001). H_2 consumption by *Escherichia coli* coupled via hydrogenase 1 or hydrogenase 2 to different terminal electron acceptors. FEMS microbiology letters *202*, 121-124.

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

Lehtio, L., and Goldman, A. (2004). The pyruvate formate lyase family: sequences, structures and activation. Protein Eng Des Sel *17*, 545-552.

Lemieux, M.J., Fischer, S.J., Cherney, M.M., Bateman, K.S., and James, M.N.G. (2007). The crystal structure of the rhomboid peptidase from *Haemophilus influenzae* provides insight into intramembrane proteolysis. Proceedings of the National Academy of Sciences of the United States of America *104*, 750-754.

Levine, M.M., Dupont, H.L., Formal, S.B., Hornick, R.B., Takeuchi, A., Gangarosa, E.J., Snyder, M.J., and Libonati, J.P. (1973). Pathogenesis of *Shigella dysenteriae* 1 (Shiga) dysentery. The Journal of infectious diseases *127*, 261-270.

Li, H., Xu, H., Zhou, Y., Zhang, J., Long, C., Li, S., Chen, S., Zhou, J.M., and Shao, F. (2007). The phosphothreonine lyase activity of a bacterial type III effector family. Science (New York, N.Y.) *315*, 1000-1003.

Li, J., and Stewart, V. (1992). Localization of upstream sequence elements required for nitrate and anaerobic induction of *fdn* (formate dehydrogenase-N) operon expression in *Escherichia coli* K-12. Journal of bacteriology *174*, 4935-4942.

Lin, J., In Soo Lee, Frey, J., Slonczewski, J.L., and Foster, J.W. (1995). Comparative analysis of extreme acid survival in *Salmonella typhimurium*, *Shigella flexneri*, and *Escherichia coli*. Journal of bacteriology *177*, 4097-4104.

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

Lu, P., Bai, X.C., Ma, D., Xie, T., Yan, C., Sun, L., Yang, G., Zhao, Y., Zhou, R., Scheres, S.H.*, et al.* (2014). Three-dimensional structure of human gamma-secretase. Nature *512*, 166-170.

Lubitz, W., Ogata, H., Rudiger, O., and Reijerse, E. (2014). Hydrogenases. Chem Rev *114*, 4081-4148.

Luke, I., Butland, G., Moore, K., Buchanan, G., Lyall, V., Fairhurst, S.A., Greenblatt, J.F., Emili, A., Palmer, T., and Sargent, F. (2008). Biosynthesis of the respiratory formate dehydrogenases from *Escherichia coli*: characterization of the FdhE protein. Arch Microbiol *190*, 685-696.

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

Magdalena, J., Hachani, A., Chamekh, M., Jouihri, N., Gounon, P., Blocker, A., and Allaoui, A. (2002). Spa32 Regulates a Switch in Substrate Specificity of the Type III Secreton of *Shigella flexneri* from Needle Components to Ipa Proteins. Journal of Bacteriology *184*, 3433-3441.

Maier, L., Vyas, R., Cordova, C.D., Lindsay, H., Schmidt, T.S., Brugiroux, S., Periaswamy, B., Bauer, R., Sturm, A., Schreiber, F.*, et al.* (2013). Microbiota-derived hydrogen fuels *Salmonella* typhimurium invasion of the gut ecosystem. Cell Host Microbe *14*, 641-651.

Manolaridis, I., Kulkarni, K., Dodd, R.B., Ogasawara, S., Zhang, Z., Bineva, G., O'Reilly, N., Hanrahan, S.J., Thompson, A.J., Cronin, N.*, et al.* (2013). Mechanism of farnesylated CAAX protein processing by the intramembrane protease Rce1. Nature *504*, 301-305.

Marteyn, B., West, N.P., Browning, D.F., Cole, J.A., Shaw, J.G., Palm, F., Mounier, J., Prevost, M.C., Sansonetti, P., and Tang, C.M. (2010). Modulation of *Shigella* virulence in response to available oxygen *in vivo*. Nature *465*, 355-358.

McDowall, J.S., Murphy, B.J., Haumann, M., Palmer, T., Armstrong, F.A., and Sargent, F. (2014). Bacterial formate hydrogenlyase complex. Proc Natl Acad Sci U S A *111*, E3948-3956.

McDowell, M.A., Johnson, S., Deane, J.E., Cheung, M., Roehrich, A.D., Blocker, A.J., McDonnell, J.M., and Lea, S.M. (2011). Structural and functional studies on the N-terminal domain of the *Shigella* type III secretion protein MxiG. J Biol Chem *286*, 30606-30614.

McNorton, M.M., and Maier, R.J. (2012). Roles of H_2 uptake hydrogenases in *Shigella flexneri* acid tolerance. Microbiology *158*, 2204-2212.

Menon, N.K., Chatelus, C.Y., Dervartanian, M., Wendt, J.C., Shanmugam, K.T., Peck Jr., H.D., and Przybyla, A.E. (1994). Cloning, sequencing, and mutational analysis of the *hyb* operon encoding *Escherichia coli* hydrogenase 2. Journal of bacteriology *176*.

Menon, N.K., Robbins, J., Peck Jr., H.D., Chatelus, C.Y., Choi, E.-S., and Przybyla, A.E. (1990). Cloning and sequencing of a putative *Escherichia coli* [NiFe] hydrogenase-1 operon containing six open reading frames. Journal of bacteriology *172*, 1969-1977.

Merritt, M.E., and Donaldson, J.R. (2009). Effect of bile salts on the DNA and membrane integrity of enteric bacteria. J Med Microbiol *58*, 1533-1541.

Mesak, L.R., Mesak, F.M., and Dahl, M.K. (2004). Expression of a novel gene, *gluP*, is essential for normal *Bacillus subtilis* cell division and contributes to glucose export. BMC Microbiology *4*, 1-16.

Miller, M.B., and Bassler, B.L. (2001). Quorum sensing in bacteria. Annual Review of Microbiology *55*, 165-199.

Minton, K.W., Tabor, H., and Tabor, C.W. (1990). Paraquat toxicity is increased in *Escherichia coli* defective in the synthesis of polyamines. Proceedings of the National Academy of Sciences of the United States of America *87*, 2851-2855.

Mishra, S., and Imlay, J. (2012). Why do bacteria use so many enzymes to scavenge hydrogen peroxide? Arch Biochem Biophys *525*, 145-160.

Mizushima, N., and Komatsu, M. (2011). Autophagy: renovation of cells and tissues. Cell *147*, 728-741.

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

Mostowy, S., Bonazzi, M., Hamon, M.A., Tham, T.N., Mallet, A., Lelek, M., Gouin, E., Demangel, C., Brosch, R., Zimmer, C.*, et al.* (2010). Entrapment of intracytosolic bacteria by septin cage-like structures. Cell host & microbe *8*, 433-444.

Mullineaux-Sanders, C., Suez, J., Elinav, E., and Frankel, G. (2018). Sieving through gut models of colonization resistance. Nat Microbiol *3*, 132-140.

Nesbit, A.D., Fleischhacker, A.S., Teter, S.J., and Kiley, P.J. (2012). ArcA and AppY antagonize IscR repression of hydrogenase-1 expression under anaerobic conditions, revealing a novel mode of O_2 regulation of gene expression in *Escherichia coli*. J Bacteriol *194*, 6892-6899.

Neutra, M.R., Pringault, E., and Kraehenbuhl, J.P. (1996). Antigen sampling across epithelial barriers and induction of mucosal immune responses. Annual review of immunology *14*, 275-300.

Newton, H.J., Pearson, J.S., Badea, L., Kelly, M., Lucas, M., Holloway, G., Wagstaff, K.M., Dunstone, M.A., Sloan, J., Whisstock, J.C.*, et al.* (2010). The type III effectors NleE and NleB from enteropathogenic *E. coli* and OspZ from *Shigella* block nuclear translocation of NF-kappaB p65. PLoS pathogens *6*, e1000898.

Nickerson, K.P., Chanin, R.B., Sistrunk, J.R., Rasko, D.A., Fink, P.J., Barry, E.M., Nataro, J.P., and Faherty, C.S. (2017). Analysis of *Shigella flexneri* Resistance, Biofilm Formation, and Transcriptional Profile in Response to Bile Salts. Infect Immun *85*.

Nigro, G., Fazio, L.L., Martino, M.C., Rossi, G., Tattoli, I., Liparoti, V., De Castro, C., Molinaro, A., Philpott, D.J., and Bernardini, M.L. (2008). Muramylpeptide shedding modulates cell sensing of *Shigella flexneri*. Cell Microbiol *10*, 682-695.

Ogawa, M., Yoshimori, T., Suzuki, T., Sagara, H., Mizushima, N., and Sasakawa, C. (2005). Escape of intracellular Shigella from autophagy. Science *307*, 727-731.

Okuda, J., Toyotome, T., Kataoka, N., Ohno, M., Abe, H., Shimura, Y., Seyedarabi, A., Pickersgill, R., and Sasakawa, C. (2005). *Shigella* effector IpaH9.8 binds to a splicing factor U2AF(35) to modulate host immune responses. Biochemical and biophysical research communications *333*, 531-539.

Owen, R.L., and Jones, A.L. (1974). Epithelial cell specialization within human Peyer's patches: an ultrastructural study of intestinal lymphoid follicles. Gastroenterology *66*, 189-203.

Palmer, T., and Berks, B.C. (2012). The twin-arginine translocation (Tat) protein export pathway. Nat Rev Microbiol *10*, 483-496.

Parente, J., Casabuono, A., Ferrari, M.C., Paggi, R.A., De Castro, R.E., Couto, A.S., and Gimenez, M.I. (2014). A rhomboid protease gene deletion affects a novel oligosaccharide N-linked to the S-layer glycoprotein of *Haloferax volcanii*. J Biol Chem *289*, 11304-11317.

Pedelacq, J.D., Cabantous, S., Tran, T., Terwilliger, T.C., and Waldo, G.S. (2006). Engineering and characterization of a superfolder green fluorescent protein. Nat Biotechnol *24*, 79-88.

Pedron, T., Thibault, C., and Sansonetti, P.J. (2003). The invasive phenotype of Shigella flexneri directs a distinct gene expression pattern in the human intestinal epithelial cell line Caco-2. The Journal of biological chemistry *278*, 33878-33886.

Peters, J.W., Schut, G.J., Boyd, E.S., Mulder, D.W., Shepard, E.M., Broderick, J.B., King, P.W., and Adams, M.W. (2015). [FeFe]- and [NiFe]-hydrogenase diversity, mechanism, and maturation. Biochim Biophys Acta *1853*, 1350-1369.

Phalipon, A., and Sansonetti, P.J. (2007). *Shigella*'s ways of manipulating the host intestinal innate and adaptive immune system: a tool box for survival? Immunology and cell biology *85*, 119-129.

Philpott, D.J., Sorbara, M.T., Robertson, S.J., Croitoru, K., and Girardin, S.E. (2014). NOD proteins: regulators of inflammation in health and disease. Nat Rev Immunol *14*, 9-23.

Pinske, C., Jaroschinsky, M., Linek, S., Kelly, C.L., Sargent, F., and Sawers, R.G. (2015). Physiology and bioenergetics of [NiFe]-hydrogenase 2-catalyzed H_2-consuming and H_2-producing reactions in *Escherichia coli*. J Bacteriol *197*, 296-306.

Pinske, C., McDowall, J.S., Sargent, F., and Sawers, R.G. (2012). Analysis of hydrogenase 1 levels reveals an intimate link between carbon and hydrogen metabolism in *Escherichia coli* K-12. Microbiology *158*, 856-868.

Pinske, C., and Sawers, R.G. (2016). Anaerobic Formate and Hydrogen Metabolism. EcoSal Plus *7*.

Pollard, T.D., and Borisy, G.G. (2003). Cellular Motility Driven by Assembly and Disassembly of Actin Filaments. Cell *112*, 453-465.

Pommier, J., Mandrand, M.A., Holt, S.E., Boxer, D.H., and Giordano, G. (1992). A second phenazine methosulphate-linked formate dehydrogenase isoenzyme in *Escherichia coli*. Biochimica et biophysica acta. Biomembranes *1107*, 305-313.

Pope, L.M., Reed, K.E., and Payne, S.M. (1995). Increased protein secretion and adherence to HeLa cells by *Shigella* spp. following growth in the presence of bile salts. Infection and immunity *63*, 3642-3648.

Prunier, A.L., Schuch, R., Fernandez, R.E., Mumy, K.L., Kohler, H., McCormick, B.A., and Maurelli, A.T. (2007). *nadA* and *nadB* of *Shigella flexneri* 5a are antivirulence loci responsible for the synthesis of quinolinate, a small molecule inhibitor of *Shigella* pathogenicity. Microbiology *153*, 2363-2372.

Puhar, A., Tronchere, H., Payrastre, B., Nhieu, G.T., and Sansonetti, P.J. (2013). A *Shigella* effector dampens inflammation by regulating epithelial release of danger signal ATP through production of the lipid mediator PtdIns5P. Immunity *39*, 1121-1131.

Pupo, G.M., Lan, R., and Reeves, P.R. (2000). Multiple independent origins of *Shigella* clones of *Escherichia coli* and convergent evolution of many of their characteristics. Proceedings of the National Academy of Sciences of the United States of America *97*, 10567-10572.

Qiu, S., Xu, X., Yang, C., Wang, J., Liang, B., Li, P., Li, H., Yi, S., Liu, H., Cui, X.*, et al.* (2015). Shift in serotype distribution of *Shigella* species in China, 2003-2013. Clin Microbiol Infect *21*, 252 e255-258.

Rajakumar, K., Sasakawa, C., and Adler, B. (1997). Use of a novel approach, termed island probing, identifies the *Shigella flexneri* she pathogenicity island which encodes a homolog of the immunoglobulin A protease-like family of proteins. Infection and immunity *65*, 4606-4614.

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

Rather, P.N., Parojcic, M.M., and Paradise, M.R. (1997). An extracellular factor regulating expression of the chromosomal aminoglycoside 2'-*N*-acetyltransferase of *Providencia stuartii*. Antimicrobial Agents and Chemotherapy *41*, 1749-1754.

Rathinam, V.A., Vanaja, S.K., Waggoner, L., Sokolovska, A., Becker, C., Stuart, L.M., Leong, J.M., and Fitzgerald, K.A. (2012). TRIF licenses caspase-11-dependent NLRP3 inflammasome activation by gram-negative bacteria. Cell *150*, 606-619.

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

Ray, K., Bobard, A., Danckaert, A., Paz-Haftel, I., Clair, C., Ehsani, S., Tang, C., Sansonetti, P., Tran, G.V., and Enninga, J. (2010). Tracking the dynamic interplay between bacterial and host factors during pathogen-induced vacuole rupture in real time. Cellular microbiology *12*, 545-556.

Ray, W.K., Zeng, G., Potters, M.B., Mansuri, A.M., and Larson, T.J. (2000). Characterization of a 12-kilodalton rhodanese encoded by *glpE* of *Escherichia coli* and its interaction with thioredoxin. Journal of bacteriology *182*, 2277-2284.

Redwood, M.D., Mikheenko, I.P., Sargent, F., and Macaskie, L.E. (2008). Dissecting the roles of *Escherichia coli* hydrogenases in biohydrogen production. FEMS Microbiol Lett *278*, 48-55.

Richard, D.J., Sawers, G., Sargent, F., McWalter, L., and Boxer, D.H. (1999). Transcriptional regulation in response to oxygen and nitrate of the operons encoding the [NiFe] hydrogenases 1 and 2 of *Escherichia coli*. Microbiology *145*, 2903-2912.

Riestra, A.M., Gandhi, S., Sweredoski, M.J., Moradian, A., Hess, S., Urban, S., and Johnson, P.J. (2015). A *Trichomonas vaginalis* Rhomboid Protease and Its Substrate Modulate Parasite Attachment and Cytolysis of Host Cells. PLoS Pathog *11*, e1005294.

Rodrigue, A., Chanal, A., Beck, K., Müller, M., and Wu, L.-F. (1999). Co-translocation of a Periplasmic Enzyme Complex by a Hitchhiker Mechanism through the Bacterial Tat Pathway. Journal of Biological Chemistry *274*, 13223-13228.

Roehrich, A.D., Martinez-Argudo, I., Johnson, S., Blocker, A.J., and Veenendaal, A.K. (2010). The extreme C terminus of *Shigella flexneri* IpaB is required for regulation of type III secretion, needle tip composition, and binding. Infect Immun *78*, 1682-1691.

Russell, C.W., Richards, A.C., Chang, A.S., and Mulvey, M.A. (2017). The Rhomboid Protease GlpG Promotes the Persistence of Extraintestinal Pathogenic *Escherichia coli* within the Gut. Infect Immun *85*.

Sack, D.A., Hoque, S., Etheridge, M., and Huq, A. (1994). Is protection against shigellosis induced by natural infection with Plesiomonas shigelloides? The Lancet *343*, 1413-1415.

Saita, S., Nolte, H., Fiedler, K.U., Kashkar, H., Venne, A.S., Zahedi, R.P., Kruger, M., and Langer, T. (2017). PARL mediates Smac proteolytic maturation in mitochondria to promote apoptosis. Nat Cell Biol *19*, 318-328.

Sal-Man, N., Gerber, D., and Shai, Y. (2014). Proline localized to the interaction interface can mediate self-association of transmembrane domains. Biochim Biophys Acta *1838*, 2313-2318.

Sana, T.G., Flaugnatti, N., Lugo, K.A., Lam, L.H., Jacobson, A., Baylot, V., Durand, E., Journet, L., Cascales, E., and Monack, D.M. (2016). *Salmonella* Typhimurium utilizes a T6SS-mediated antibacterial weapon to establish in the host gut. Proc Natl Acad Sci U S A *113*, E5044-5051.

Sanada, T., Kim, M., Mimuro, H., Suzuki, M., Ogawa, M., Oyama, A., Ashida, H., Kobayashi, T., Koyama, T., Nagai, S.*, et al.* (2012). The *Shigella flexneri* effector OspI deamidates UBC13 to dampen the inflammatory response. Nature *483*, 623-626.

Sander, L.E., Davis, M.J., Boekschoten, M.V., Amsen, D., Dascher, C.C., Ryffel, B., Swanson, J.A., Muller, M., and Blander, J.M. (2011). Detection of prokaryotic mRNA signifies microbial viability and promotes immunity. Nature *474*, 385-389.

Sansonetti, P.J., Arondel, J., Huerre, M., Harada, A., and Matsushima, K. (1999). Interleukin-8 Controls Bacterial Transepithelial Translocation at the Cost of Epithelial Destruction in Experimental Shigellosis. Infection and Immunity *67*, 1471-1480.

Sansonetti, P.J., Kopecko, D.J., and Formal, S.B. (1981). *Shigella sonnei* plasmids: evidence that a large plasmid is necessary for virulence. Infection and immunity *34*, 75-83.

Sansonetti, P.J., Kopecko, D.J., and Formal, S.B. (1982). Involvement of a plasmid in the invasive ability of *Shigella flexneri*. Infection and immunity *35*, 852-860.

Sansonetti, P.J., and Phalipon, A. (1999). M cells as ports of entry for enteroinvasive pathogens: mechanisms of interaction, consequences for the disease process. Seminars in immunology *11*, 193-203.

Sansonetti, P.J., Phalipon, A., Arondel, J., Thirumalai, K., Banerjee, S., Akira, S., Takeda, K., and Zychlinsky, A. (2000). Caspase-1 activation of IL-1beta and IL-18 are essential for *Shigella flexneri*-induced inflammation. Immunity *12*, 581-590.

Sargent, F. (2016). The Model [NiFe]-Hydrogenases of *Escherichia coli*. Advances in Microbial Physiology *68*, 433-507.

Sargent, F., Ballantine, S.P., Rugman, P.A., Palmer, T., and Boxer, D.H. (1998). Reassignment of the gene encoding the *Escherichia coli* hydrogenase 2 small subunit. Identification of a soluble precursor of the small subunit in a *hypB* mutant. European journal of biochemistry *255*, 746-754.

Sargent, F., Berks, B.C., and Palmer, T. (2002). Assembly of membrane-bound respiratory complexes by the Tat protein-transport system. Arch Microbiol *178*, 77-84.

Sawers, G. (1994). The hydrogenases and formate dehydrogenases of *Escherichia coli*. Antonie van Leeuwenhoek *66*, 57-88.

Sawers, R.G., Ballantine, S.P., and Boxer, D.H. (1985). Differential expression of hydrogenase isoenzymes in *Escherichia coli* K-12 Evidence for a third isoenzyme. Journal of bacteriology *164*, 1324-1331.

Sawers, R.G., and Boxer, D.H. (1986). Purification and properties of membrane-bound hydrogenase isoenzyme 1 from anaerobically grown *Escherichia coli* K12. European journal of biochemistry *156*, 265-275.

Schapira, M., Tyers, M., Torrent, M., and Arrowsmith, C.H. (2017). WD40 repeat domain proteins: a novel target class? Nat Rev Drug Discov *16*, 773-786.

Schraidt, O., Lefebre, M.D., Brunner, M.J., Schmied, W.H., Schmidt, A., Radics, J., Mechtler, K., Galan, J.E., and Marlovits, T.C. (2010). Topology and organization of the *Salmonella typhimurium* type III secretion needle complex components. PLoS Pathog *6*, e1000824.

Schroeder, G.N., and Hilbi, H. (2008). Molecular pathogenesis of *Shigella* spp.: controlling host cell signaling, invasion, and death by type III secretion. Clin Microbiol Rev *21*, 134-156.

Self, W.T., Hasona, A., and Shanmugam, K.T. (2004). Expression and Regulation of a Silent Operon, *hyf*, Coding for Hydrogenase 4 Isoenzyme in *Escherichia coli*. Journal of Bacteriology *186*, 580-587.

Senerovic, L., Tsunoda, S.P., Goosmann, C., Brinkmann, V., Zychlinsky, A., Meissner, F., and Kolbe, M. (2012). Spontaneous formation of IpaB ion channels in host cell membranes reveals how Shigella induces pyroptosis in macrophages. Cell Death Dis *3*, e384.

Senes, A., Engel, D.E., and DeGrado, W.F. (2004). Folding of helical membrane proteins: the role of polar, GxxxG-like and proline motifs. Curr Opin Struct Biol *14*, 465-479.

Shah, N., Ldupont, H., and Ramsey, D.J. (2009). Global Etiology of Travelers' Diarrhea Systematic Review from 1973 to the Present. The American journal of tropical medicine and hygiene *80*, 609-614.

Shen, D.K., and Blocker, A.J. (2016). MxiA, MxiC and IpaD Regulate Substrate Selection and Secretion Mode in the T3SS of *Shigella flexneri*. PLoS One *11*, e0155141.

Shepherd, J.G., Wang, L., and Reeves, P.R. (2000). Comparison of O-antigen gene clusters of *Escherichia coli* (*Shigella*) sonnei and *Plesiomonas shigelloides* O17: Sonnei gained its current plasmid-borne O-antigen genes from *P. shigelloides* in a Recent event. Infection and immunity *68*, 6056-6061.

Simon, R., Priefer, U., and Pühler, A. (1983). A Broad Host Range Mobilization System for In Vivo Genetic Engineering Transposon Mutagenesis in Gram Negative Bacteria. Bio/technology *1*, 784-791.

Singer, M., and Sansonetti, P.J. (2004). IL-8 is a key chemokine regulating neutrophil recruitment in a new mouse model of *Shigella*-induced colitis. Journal of immunology (Baltimore, Md. : 1950) *173*, 4197-4206.

Sirianni, A., Krokowski, S., Lobato-Marquez, D., Buranyi, S., Pfanzelter, J., Galea, D., Willis, A., Culley, S., Henriques, R., Larrouy-Maumus, G.*, et al.* (2016). Mitochondria mediate septin cage assembly to promote autophagy of *Shigella*. EMBO Rep *17*, 1029-1043.

Sistrunk, J.R., Nickerson, K.P., Chanin, R.B., Rasko, D.A., and Faherty, C.S. (2016). Survival of the Fittest: How Bacterial Pathogens Utilize Bile To Enhance Infection. Clin Microbiol Rev *29*, 819-836.

Skarlatos, P., and Dahl, M.K. (1998). The Glucose Kinase of *Bacillus subtilis*. Journal of Bacteriology *180*, 3222-3226.

Skibinski, D.A.G., Golby, P., Chang, Y.-S., Sargent, F., Hoffman, R., Harper, R., Guest, J.R., Attwood, M.M., Berks, B.C., and Andrews, S.C. (2002). Regulation of the Hydrogenase-4 Operon of *Escherichia coli* by the σ^{54}-Dependent Transcriptional Activators FhlA and HyfR. Journal of Bacteriology *184*, 6642-6653.

Slauch, J.M. (2011). How does the oxidative burst of macrophages kill bacteria? Still an open question. Mol Microbiol *80*, 580-583.

Soding, J., Biegert, A., and Lupas, A.N. (2005). The HHpred interactive server for protein homology detection and structure prediction. Nucleic Acids Res *33*, W244-248.

Stanley, N.R., Sargent, F., Buchanan, G., Shi, J., Stewart, V., Palmer, T., and Berks, B.C. (2002). Behaviour of topological marker proteins targeted to the Tat protein transport pathway. Molecular microbiology *43*, 1005-1021.

Steggerda, F.R. (1968). Gastrointestinal gas following food consumption. Annals of the New York Academy of Sciences *150*, 57-66.

Steinberg, B.E., and Grinstein, S. (2007). Unconventional roles of the NADPH oxidase: signaling, ion homeostasis, and cell death. Sci STKE *2007*, pe11.

Stensrud, K.F., Adam, P.R., La Mar, C.D., Olive, A.J., Lushington, G.H., Sudharsan, R., Shelton, N.L., Givens, R.S., Picking, W.L., and Picking, W.D. (2008). Deoxycholate interacts with IpaD of *Shigella flexneri* in inducing the recruitment of IpaB to the type III secretion apparatus needle tip. J Biol Chem *283*, 18646-18654.

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

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

Suzuki, S., Franchi, L., He, Y., Munoz-Planillo, R., Mimuro, H., Suzuki, T., Sasakawa, C., and Nunez, G. (2014). *Shigella* type III secretion protein MxiI is recognized by Naip2 to induce Nlrc4 inflammasome activation independently of Pkcδ. PLoS Pathog *10*, e1003926.

Suzuki, T., Franchi, L., Toma, C., Ashida, H., Ogawa, M., Yoshikawa, Y., Mimuro, H., Inohara, N., Sasakawa, C., and Nunez, G. (2007). Differential regulation of caspase-1 activation, pyroptosis, and autophagy via Ipaf and ASC in *Shigella*-infected macrophages. PLoS pathogens *3*, e111.

Suzuki, T., Miki, H., Takenawa, T., and Sasakawa, C. (1998). Neural Wiskott-Aldrich syndrome protein is implicated in the actin-based motility of *Shigella flexneri*. The EMBO Journal *17*, 2767-2776.

Takeuchi, A., Sprinz, H., LaBrec, E.H., and Formal, S.B. (1965). Experimental bacillary dysentery. An electron microscopic study of the response of the intestinal mucosa to bacterial invasion. The American journal of pathology *47*, 1011-1044.

Tamarit, J., Cabiscol, E., and Ros, J. (1998). Identification of the Major Oxidatively Damaged Proteins in *Escherichia coli* Cells Exposed to Oxidative Stress. Journal of biological chemistry *273*, 3027-3032.

Tapley, T.L., Franzmann, T.M., Chakraborty, S., Jakob, U., and Bardwell, J.C. (2010). Protein refolding by pH-triggered chaperone binding and release. Proc Natl Acad Sci U S A *107*, 1071-1076.

Tattoli, I., Carneiro, L.A., Jehanno, M., Magalhaes, J.G., Shu, Y., Philpott, D.J., Arnoult, D., and Girardin, S.E. (2008). NLRX1 is a mitochondrial NOD-like receptor that amplifies NF-kappaB and JNK pathways by inducing reactive oxygen species production. EMBO reports *9*, 293-300.

Thompson, C.N., Duy, P.T., and Baker, S. (2015). The Rising Dominance of *Shigella sonnei*: An Intercontinental Shift in the Etiology of Bacillary Dysentery. PLoS Negl Trop Dis *9*, e0003708.

Thurston, T.L., Wandel, M.P., von Muhlinen, N., Foeglein, A., and Randow, F. (2012). Galectin 8 targets damaged vesicles for autophagy to defend cells against bacterial invasion. Nature *482*, 414-418.

Tjon, J.M., Verbeek, W.H., Kooy-Winkelaar, Y.M., Nguyen, B.H., van der Slik, A.R., Thompson, A., Heemskerk, M.H., Schreurs, M.W., Dekking, L.H., Mulder, C.J.*, et al.* (2008). Defective synthesis or association of T-cell receptor chains underlies loss of surface T-cell receptor-CD3 expression in enteropathy-associated T-cell lymphoma. Blood *112*, 5103-5110.

Tran Van Nhieu, G., Caron, E., Hall, A., and Sansonetti, P.J. (1999). IpaC induces actin polymerization and filopodia formation during *Shigella* entry into epithelial cells. The EMBO journal *18*, 3249-3262.

Trchounian, K., Soboh, B., Sawers, R.G., and Trchounian, A. (2013). Contribution of hydrogenase 2 to stationary phase H_2 production by *Escherichia coli* during fermentation of glycerol. Cell Biochem Biophys *66*, 103-108.

Troeger, C., Forouzanfar, M., Rao, P.C., Khalil, I., Brown, A., Reiner, R.C., Fullman, N., Thompson, R.L., Abajobir, A., Ahmed, M*., et al.* (2017). Estimates of global, regional, and national morbidity, mortality, and aetiologies of diarrhoeal diseases: a systematic analysis for the Global Burden of Disease Study 2015. The Lancet Infectious Diseases *17*, 909-948.

Tsirigos, K.D., Peters, C., Shu, N., Kall, L., and Elofsson, A. (2015). The TOPCONS web server for consensus prediction of membrane protein topology and signal peptides. Nucleic Acids Res *43*, W401-407.

Unden, G., Steinmetz, P.A., and Degreif-Dunnwald, P. (2014). The Aerobic and Anaerobic Respiratory Chain of *Escherichia coli* and *Salmonella enterica*: Enzymes and Energetics. EcoSal Plus *6*.

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

Urban, S., and Freeman, M. (2003). Substrate Specificity of Rhomboid Intramembrane Proteases Is Governed by Helix-Breaking Residues in the Substrate Transmembrane Domain. Molecular Cell *11*, 1425-1434.

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

Urban, S., Lee, J.R., and Freeman, M. (2002). A family of Rhomboid proteases activates all *Drosophila* intramembrane phila membrane-tethered EGF ligands. The EMBO Journal *21*, 4277-4286.

Vinothkumar, K.R., Pierrat, O.A., Large, J.M., and Freeman, M. (2013). Structure of rhomboid protease in complex with beta-lactam inhibitors defines the S2' cavity. Structure *21*, 1051-1058.

Vokes, S.A., Reeves, S.A., Torres, A.G., and Payne, S.M. (1999). The aerobactin iron transport system genes in *Shigella flexneri* are present within a pathogenicity island. Molecular microbiology *33*, 63-73.

Volbeda, A., Darnault, C., Parkin, A., Sargent, F., Armstrong, F.A., and Fontecilla-Camps, J.C. (2013). Crystal structure of the O(2)-tolerant membrane-bound hydrogenase 1 from *Escherichia coli* in complex with its cognate cytochrome *b*. Structure *21*, 184-190.

Vollaard, E.J., and Clasener, H.A.L. (1994). Colonization resistance. Antimicrobial agents and chemotherapy *38*, 409-414.

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

Wang, F., Jiang, Z., Li, Y., He, X., Zhao, J., Yang, X., Zhu, L., Yin, Z., Li, X., Wang, X.*, et al.* (2013). *Shigella flexneri* T3SS effector IpaH4.5 modulates the host inflammatory response via interaction with NF-kappaB p65 protein. Cellular microbiology *15*, 474-485.

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

Warren, J.W. (1986). *Providencia stuartii*: a common cause of antibiotic-resistant bacteriuria in patients with long-term indwelling catheters. Reviews of infectious diseases *8*, 61-67.

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

Waterman, S.R., and Small, P.L.C. (1996). Identification of σs-dependent genes associated with the stationary-phase acid-resistance phenotype of *Shigella flexneri*. Molecular microbiology *21*, 925-940.

Weinrauch, Y., Drujan, D., Shapiro, S.D., Weiss, J., and Zychlinsky, A. (2002). Neutrophil elastase targets virulence factors of enterobacteria. Nature *417*, 91-94.

West, N.P., Sansonetti, P., Mounier, J., Exley, R.M., Parsot, C., Guadagnini, S., Prévost, M.-C., Prochnicka-Chalufour, A., Delepierre, M., Tanguy, M.*, et al.* (2005). Optimization of virulence functions through glucosylation of *Shigella* LPS. Science *307*, 1313-1317.

Willingham, S.B., Bergstralh, D.T., O'Connor, W., Morrison, A.C., Taxman, D.J., Duncan, J.A., Barnoy, S., Venkatesan, M.M., Flavell, R.A., Deshmukh, M.*, et al.* (2007). Microbial pathogen-induced necrotic cell death mediated by the inflammasome components CIAS1/cryopyrin/NLRP3 and ASC. Cell host & microbe *2*, 147-159.

Winter, S.E., Winter, M.G., Xavier, M.N., Thiennimitr, P., Poon, V., Keestra, A.M., Laughlin, R.C., Gomez, G., Wu, J., Lawhon, S.D.*, et al.* (2013). Host-Derived Nitrate Boosts Growth of *E. coli* in the Inflamed Gut. Science *339*, 708-711.

Wolf, P.G., Biswas, A., Morales, S.E., Greening, C., and Gaskins, H.R. (2016). H_2 metabolism is widespread and diverse among human colonic microbes. Gut Microbes *7*, 235-245.

Wolin, M.J., Wolin, E.A., and Jacobs, N.J. (1961). Cytochrome-producing anaerobic Vibrio, *Vibrio succinogenes*, sp. n. Journal of Bacteriology *81*, 911-917.

Wormann, M.E., Horien, C.L., Johnson, E., Liu, G., Aho, E., Tang, C.M., and Exley, R.M. (2016). *Neisseria cinerea* isolates can adhere to human epithelial cells by type IV pilus-independent mechanisms. Microbiology *162*, 487-502.

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

Wucherpfennig, K.W., Gagnon, E., Call, M.J., Huseby, E.S., and Call, M.E. (2010). Structural biology of the T-cell receptor: insights into receptor assembly, ligand recognition, and initiation of signaling. Cold Spring Harb Perspect Biol *2*, a005140.

Xu, C., and Min, J. (2011). Structure and function of WD40 domain proteins. Protein Cell *2*, 202-214.

Yang, B., and Larson, T.J. (1998). Multiple promoters are responsible for transcription of the glpEGR operon of *Escherichia coli* K-12. Biochimica et biophysica acta. N, Gene structure and expression *1396*, 114-126.

Yanisch-Perron, C., Vieira, J., and Messing, J. (1985). Improved M13 phage cloning vectors and host strains nucleotide sequences of the M13mpl8 and pUC19 vectors. Gene (Amsterdam) *33*, 103-119.

Yip, S.H., and Matsumura, I. (2013). Substrate ambiguous enzymes within the *Escherichia coli* proteome offer different evolutionary solutions to the same problem. Mol Biol Evol *30*, 2001-2012.

Yoshida, H., Maki, Y., Furuike, S., Sakai, A., Ueta, M., and Wada, A. (2012). YqjD is an inner membrane protein associated with stationary-phase ribosomes in *Escherichia coli*. J Bacteriol *194*, 4178-4183.

Zhang, L., Ding, X., Cui, J., Xu, H., Chen, J., Gong, Y.N., Hu, L., Zhou, Y., Ge, J., Lu, Q., *et al.* (2012). Cysteine methylation disrupts ubiquitin-chain sensing in NF-kappaB activation. Nature *481*, 204-208.

Zhang, Y., and Lin, K. (2012). A phylogenomic analysis of *Escherichia coli* / *Shigella* group: implications of genomic features associated with pathogenicity and ecological adaptation. BMC evolutionary biology *12*, 174.

Zoll, S., Stanchev, S., Began, J., Skerle, J., Lepsik, M., Peclinovska, L., Majer, P., and Strisovsky, K. (2014). Substrate binding and specificity of rhomboid intramembrane protease revealed by substrate-peptide complex structures. EMBO J *33*, 2408-2421.