# Experimental characterization of a catalytically active flagellin variant in *Clostridium haemolyticum*

by

Hina Bandukwala

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Biology

Waterloo, Ontario, Canada, 2017

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

**Abstract**

The bacterial flagellum is made up of approximately 20,000 subunits of the monomeric protein, flagellin, and plays a role in cell motility and pathogenesis. The extreme sequence diversity within the hypervariable region of flagellin genes observed across phyla suggests hidden functional diversity. This thesis outlines the discovery of the first family of flagellin variants with proteolytic activity. A multi-faceted approach revealed a conserved HExxH motif within the hypervariable region (HVR) of these flagellin variants. The motif is characteristic of the Gluzincin family of thermolysin-like peptidases and was found to be conserved in 74 bacterial species spanning over 32 genera. Experimental validation began with the recombinant expression and purification of the HVR of the flagellin FliA(H) from the species *Clostridium haemolyticum*, an animal pathogen. An approach using mass spectrometry and proteomics revealed that the substrate specificity of this flagellin protease is similar to that of zinc-dependant matrix metallopeptidases (MMPs). Furthermore, peptide sequencing of harvested *C.haemolyticum* flagellar filaments revealed that the proteolytic flagellin was the second most dominant flagellin component and was also shown to have MMP-like protease activity. Considering the expanded functional repertoire of this organelle in the recent years, this flagellin-associated protease may play a role in chemotaxis, biofilm formation, adhesion and pathogenesis.

**Acknowledgements**

First and foremost, I would like to thank my supervisor, Dr. Andrew Doxey for giving me the opportunity to work on an invigorating, multi-faceted project and supervising me through it. I would also like to thank my co-supervisor Dr. Trevor Charles as well as my committee members, Dr. Todd Holyoak and Dr. Brendan McConkey for their insightful feedback throughout my graduate career enabling me to be a critical scientist.

I am also very grateful to my lab mates Briallen Lobb, Mike Mansfield and Jeremy Adams for their wonderful friendship, support and especially scientific feedback. I would also like to thank Matthew Mcleod for training me in protein purification and Dustin Sigurdson for guiding me through Immunohistochemistry as well as encouraging me to climb harder. I would like to extend my gratitude to the members of the Charles Lab, Jiujun Cheng, John Heil, Maya D'Alessio and Kathy Lam for their assistance in various aspects of molecular biology. I would also like to thank Mark Lubberts for being an inspiring colleague as well my iGEM pals.

Lastly, I would like to thank my family for their encouragement and love without whom I would not have had the opportunity to come to Canada. I would like to thank my cousin, Kumail Bandukwala for his immense amount of help. Lastly, I would like to thank Daniel Puzzuoli for his confidence in my abilities and his support that is beyond words.

## Dedication

I would like to dedicate this thesis to my father, Abbas Asghar Bandukwala, whose hard work allowed me the privilege to study in Canada, my mother, Fatema Abbas, my nana (grandfather), Hussainy Bandukwala and most importantly to my nani (grandmother), Bina Hussainy Bandukwala, who has been an exemplary role model in my life.

# Table of Contents

**List of Tables** ix

**List of Figures** x

**List of Abbreviations** xii

**1 Introduction** 1

1.1 The bacterial flagellum . . . 1

1.1.1 Structure . . . 1

1.1.2 Assembly . . . 5

1.1.3 Regulation . . . 7

1.2 Function of flagella in the environment . . . 8

1.2.1 Motility . . . 8

1.2.2 Pathogenesis . . . 9

1.3 Engineering and modularity of the bacterial flagellum . . . 14

1.4 Zinc-dependent metalloproteases . . . 14

1.4.1 Background . . . 14

1.4.2 Gluzincins . . . 17

1.4.3 Metzincins . . . 18

1.5 *Clostridium haemolyticum* . . . 19

1.6 Computational predication of a novel proteolytic flagellin family . . . 20

1.6.1 Computational methods . . . 20
1.7 Research objective . . . 24

**2 General Materials and Methods** **25**
2.1 Bacterial plasmids . . . 25
2.2 Bacterial strains . . . 26
2.3 Maintenance of bacterial cultures . . . 26
2.4 Standard molecular biology protocols for DNA manipulation . . . 27
2.4.1 Preparation of competent cells . . . 27
2.4.2 Bacterial transformation . . . 27
2.4.3 Plasmid DNA mini prep . . . 27
2.4.4 Restriction enzyme digestion . . . 28
2.4.5 Agarose gel electrophoresis . . . 28
2.5 Standard protocols for Protein Analysis . . . 28
2.5.1 SDS-PAGE . . . 28
2.5.2 BCA Protein Assay . . . 28

**3 Investigating protease activity from FliA(H) *in vitro*** **30**
3.1 Background, Hypotheses and Rationale . . . 30
3.2 Methodology . . . 32
3.2.1 Bacterial strains and construct design . . . 32
3.2.2 Expression and purification of recombinant proteins . . . 33
3.2.3 PICS Assay . . . 34
3.2.4 Quenched fluorometric assays . . . 36
3.3 Results . . . 37
3.3.1 Recombinant FliA(H) . . . 37
3.3.2 PICS peptidase specificity assays . . . 40
3.3.3 Quenched fluorometric peptide cleavage assays . . . 42
3.3.4 Peptide Docking Model . . . 43
3.4 Discussion and future directions . . . 45

**4 Investigating protease activity *in vivo* from *Clostridium haemolyticum* 47**
4.1 Hypotheses and Rationale . . . 47
4.2 Methodology . . . 48
4.2.1 Isolation of flagellar proteins from *C. haemolyticum* . . . 48
4.2.2 Peptide sequencing by LC-MS/MS . . . 49
4.2.3 Immunogold Transmission Electron Microscopy . . . 49
4.2.4 PICS Assay with purified flagellar filaments . . . 50
4.2.5 Quenched fluorometric assays . . . 50
4.3 Results . . . 50
4.3.1 LC-MS/MS . . . 50
4.3.2 Immunogold TEM . . . 52
4.3.3 PICS and QF peptide cleavage assays with purified flagellar filaments 54
4.4 Discussion and future directions . . . 57

**5 Additional insights into the biological roles of proteolytic flagellins 59**
5.1 Summary . . . 59
5.2 Insights into the molecular evolution of proteolytic flagellin HVR . . . 60
5.3 Insights into the biological function of proteolytic flagellins . . . 66
5.4 Conclusion . . . 67

**References 69**

**APPENDICES 79**

**A Recipes 80**
A.1 Molecular Biology . . . 80
A.2 Protein purification . . . 80
A.3 Western Blotting . . . 81
A.4 Immunogold TEM . . . 82

# List of Tables

5.1 **Flagellin associated domain architectures and their abundance in the NCBI database as detected using the Conserved Domain Database.** . . . . . 62

5.2 **Predicted "proteolytic flagellins" identified in the NCBI database.** 63

# List of Figures


1.1 Structure of the bacterial flagellum based on *Salmonella enterica serovar Typhimurium* . . . 2
1.2 Role of bacterial flagella in pathogenesis . . . 10
1.3 Overview of zincin family metalloprotease structure . . . 16
1.4 Structural model of the flagellar filament and identification of uncharacterized domains in the surface-exposed flagellin HVR. . . . 22
1.5 Sequence analysis and structural modeling of the proteolytic flagellin family. . . . 23

3.1 Design and expression of Trx-FliA(H)-HVR . . . 38
3.2 LC-MS/MS . . . 39
3.3 PICS assay with recombinant Trx-FliA(H)-HVR. . . . 41
3.4 Quenched Fluorometric peptide cleavage assay with Trx-FliA(H)-HVR. . . . 43
3.5 Peptide docking model . . . 44

4.1 Isolation of flagellar filaments from *Clostridium haemolyticum* . . . 51
4.2 Sequence coverage of the *Clostridium haemolyticum* FliA(H) obtained through mass spectrometry analysis . . . 52
4.3 ECL-based western blot for verification of anti-FliA(H) primary antibodies . . . 53
4.4 Immunogold TEM with flagellar filaments . . . 54
4.5 PICS assay with isolated flagellar filaments . . . 55

4.6 **QF assay with isolated flagellar filaments** . . . . . 56

5.1 **Phylogenomic distribution of proteolytic flagellins on the bacterial tree of life** . . . . . 61

5.2 **Genomic context of proteolytic flagellin gene** . . . . . 65

5.3 **Hypothesized function of the flagellin-HVR associated protease** . 67

# List of Abbreviations

| | |
|---|---|
| TEM | Transmission electron microscopy |
| HVR | Hypervariable region |
| LB | Luria-Bertani Broth |
| IPTG | Isopropyl $\beta$-D-1- thiogalactopyranoside |
| SUMO | Small Ubiquitin-like Modifier |
| Trx | Thioredoxin |
| HAPs | Hook-associated proteins |
| fTTSS | flagellar Type Three Secretion System |
| LPS | Lipopolysaccharides |
| BSA | Bovine serum albumin |
| PBS | Phosphate buffered saline |
| HBB | Hook-Basal body complex |
| PAMPs | Pathogen-associated molecular patterns |
| CCW | Counter-clockwise |
| CW | Clockwise |
| EHEC | Enterohemorrhagic *Escherichia coli* |
| EPS | Exopolysaccharides |
| EPEC | Enteropathogenic *Escherichia coli* |
| PRRs | Pattern recognition receptors |
| TLR | Toll-like receptor |
| MPs | Metalloproteases |
| ECM | Extracellular matrix |
| MMPs | Matrix Metalloproteases |
| CDART | Conserved domain architecture retrieval tool |
| PSI-BLAST | Position-specific iterative Basic Local Alignment Search Tool |
| PICS | Proteomic identification of substrate cleavage sites |
| QF | Quenched fluorometric |
| DMSO | Dimethyl sulfoxide |
| BCA | Bicinchoninic acid |
| ECL | Enhanced chemiluminescent |
| SDS-PAGE | Sodium dodecyl sulphate- polyacrylamide gel electrophoresis |
| EDTA | Ethylenediamine tetraacetic acid |
| HRP | Horseradish peroxidase |
| TBS-T | Tris buffered saline - Tween 20 |
| PVDF | Polyvinylidene difluoride |
| GFP | Green fluorescent protein |
| SDS | Sodium dodecyl sulphate |
| LC-MS/MS | Liquid Chromatography Mass Spectrometry |

# Chapter 1

# Introduction

## 1.1 The bacterial flagellum

The bacterial flagellum is a remarkable self-assembling molecular nanomachine. Historically, its function has primarily been associated with locomotion, however recent studies have expanded its functional repertoire [1, 2]. The flagellum is now known to be actively interacting with the environment. The flagellum is involved in pathogenesis, chemotaxis as well as has antigenic properties. The bacterial flagellum is a highly complex organelle. Approximately 40 genes are involved in the expression and regulation of the flagellum out of which approximately 24 are structural proteins [3].

### 1.1.1 Structure

The structure and arrangement of the flagellum varies greatly between flagellated species. Structural diversity is also mirrored at the sequence level. The overall structure of the bacterial flagellum has been elucidated based on structural studies on *Escherichia coli* and *Salmonella enterica serovar Typhimurium*. As shown in Figure 1.4, the bacterial flagellum consist of three parts which include the basal body, hook and the filament [4].

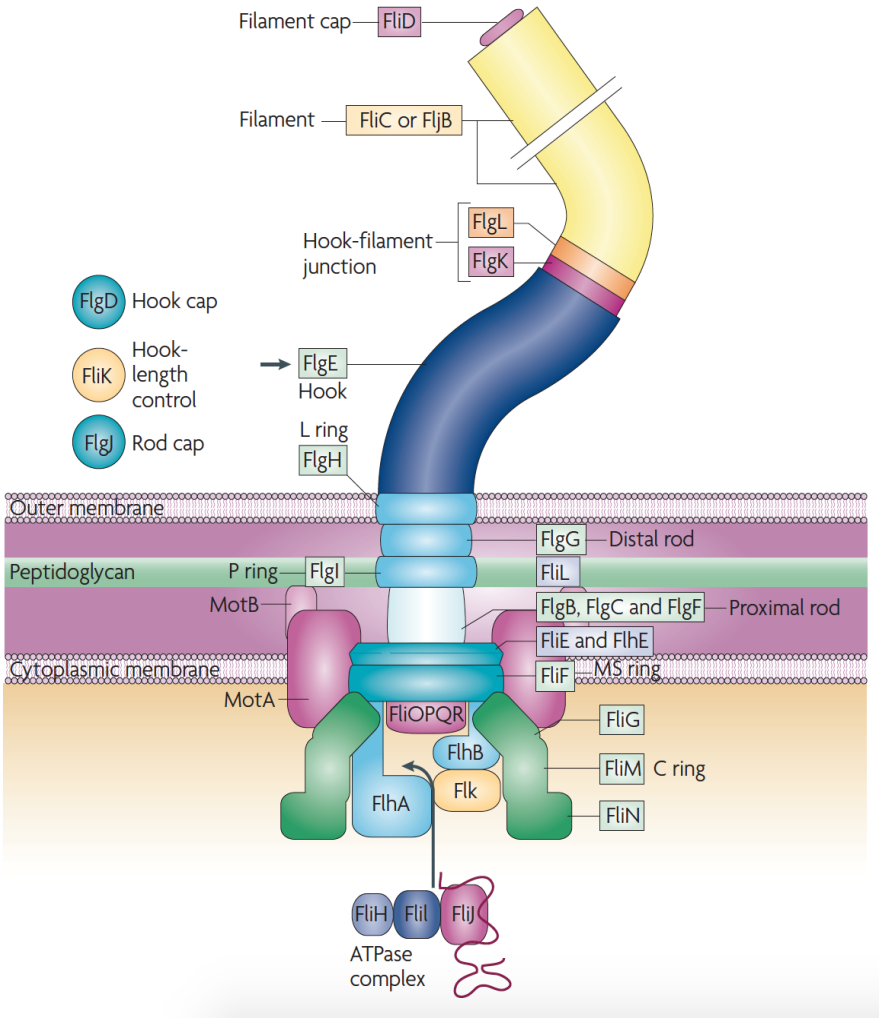


Figure 1.1: **Structure of the bacterial flagellum based on *Salmonella enterica serovar Typhimurium***

**The figure represents components of an assembled bacterial flagella based on *Salmonella enterica serovar Typhimurium*. The It consists of the basal body and hook which span the cell wall and membranes, and a flagellar filament that protrudes outwards. This figure was adapted from [4] with permission.**

### Basal Body

The basal body acts as the engine that allows for the generation of torque and thereby the rotation of the flagellar filament. The basal body is a multi-unit component and is located within the cytoplasmic membrane. The basal body consists of several rings mounted onto a central rod (further divided into proximal and distal rods) and a motor which can be further subdivided into rotor and stator complexes. The rings include an inner-membrane embedded MS ring (proximal end of the basal body), a P ring which is attached to the peptidoglycan layer and an outer-membrane embedded L ring [5].

The rotor or switch complex is composed of the proteins FliG, FliM and FliN [3] that assemble together to form the cytoplasmic-facing C ring. These rotor proteins interact with the phosphorylated subunit of the regulatory protein CheY (part of the two-component chemosensory system) to allow the motor to alternate between clockwise and counterclockwise directions [4]. This is how the organelle responds to chemical changes and thus allowing bacteria to move towards favourable environments (chemotaxis), as well as near-surface swimming and mechanosensing. The stator complexes, in contrast to the rotor, are stationary. Two cytoplasmic-membrane spanning proteins, MotA and MotB, which appear in a ratio of 4:2, form the stator complex [4] [6]. MotA and MotB proteins tend to be arranged in a circular array around the MS and C rings [6] with the periplasmic domain of MotB attached to the peptidoglycan layer [4].The stator is an ion-channel capable of harvesting electrochemical energy generated from the proton motive force (or sodium motive force in some species) and generate torque via the rotor complex.

### Hook

The hook is a short, curved structure which connects the filament to the basal body. The hook is primarily composed of 120 copies of the protein FlgE that polymerize to form a helical tube consisting of 11 protofilaments [7, 8]. The hook tends to act as a universal joint that transmits torque from the cell-membrane embedded motor and transfers it to the outwardly protruding filament [7]. The hook is a highly flexible structure and thus allows organisms to maneuver their rigid flagellar filaments without breaking them off the cell e.g. the flexibility conferred by the hook allows petrichious flagella in various organisms to function properly as a bundle [4]. The hook extends to 55 nm on average, but the length can vary depending on the assembled structure [9].The length of the hook is controlled by the protein FliK during flagellar assembly [7]. The hook has a 25 Å channel running through the centre[9].

## Filament

The filament is an appendage that protrudes from the cell body and acts as a helical propellor that allows for the propulsion of cells. The number and arrangement of flagella differ between bacterial species, for example they may be polar (*Pseudoalteromonas tunicata*), lateral (*Aeromonas spp.*), peritrichous (*Escherichia coli*, *Salmonella spp.*), amphitrichious or periplasmic (*Spirochaetes*). The flagellar filament is a protein complex composed of 11 protofilaments which are in turn composed of approximately 20,000 subunits of the monomeric structural protein, flagellin (FliC)[10]. The protofilaments are axially staggered relative to the central axis by approximately half a subunit [2]. These flagellin protein subunits self-assemble to form a hollow tube which is, on average, 20 nm wide [11] and extends up to 10 - 15 $\mu$m long [1].

Depending on the arrangement of subunits within the protofilaments, they may form a right or left-handed helix i.e. R or L-type[2]. The R-type refers to short protofilaments in which the subunits are closer together. The L-type, on the other hand, refers to long protofilaments in which the subunits are further apart. Depending on the ratio of the type of protofilaments present at any given time, can result in 12 different conformations of the filament i.e. if both R and L-type protofilaments are present, the filament will be helical which can occur in 10 different ways, if all protofilaments are L-type, the filament will be long and if they are all R-type, the filament will be short [6]. Bacteria exhibiting different forms of motility (e.g. run and tumble) can do so via the flagellar filament switching between right and left-handed supercoiled forms. These forms, in turn are produced by the alternation between the L and R type protofilaments explained above [12]. Like any proteins, flagellar filaments take different polymorphic forms based on the amino acids in their respective flagellins as well as the pH of the surrounding environment [8].

The sequences of these flagellin proteins display an interesting pattern. Each flagellin protein tends to have slowly evolving N- and C-terminal regions (corresponding to D0 and D1 domains in *Salmonella typhimurium*) that form the core of the assembled filament, however the central/intervening region, called the hypervariable region or HVR, is highly variable in sequence and length within and between species. The HVR forms the surface-exposed region in the assembled appendage [13]. Each flagellin molecule takes an inverted gamma conformation with the vertical arm being the N- and C-termini and the horizontal arm being the surface-exposed HVR [2]. The N- and C-termini tend to be rich in hydrophobic residues which allows the domains to be coiled-coil [2], [13]. Several hypotheses exist to explain the dual sequence diversity and conservation of flagellin sequences within and between various bacterial species. Since the N- and C-termini tend

to be important for the structural integrity of the assembled flagella, they remain conserved. The surface-exposed region, however, interacts with the environment e.g. it has antigenic properties, it tends to be under positive selection [2]. Another factor driving the variability of the HVR is due to lateral gene transfer of foreign genes as well as recombination [14].

### 1.1.2 Assembly

*Escherichia coli* and *Salmonella spp.* are primary models for studying flagellar structure and assembly. More than 40 genes are involved in the regulation and assembly of a flagellum with approximately 24 being structural proteins [3]. The assembly process tends to be linear/sequential and starts from the innermost (proximal) to the outermost (distal) structures. To recap, the flagellar apparatus consists of the multi-component basal body, the flexible hook and the flagellar filament. To understand the assembly, it is also important to remember that the cell envelope of bacteria include an inner cytoplasmic membrane, a thick or thin peptidoglycan layer (depending on whether they are Gram-negative or Gram-positive) and an LPS-rich outer-membrane in Gram-negative bacteria.

#### MS ring

The assembly of the basal body (and therefore, the entire flagellar complex) begins with the MS ring [4]. The assembly of the MS ring is dependent on the Sec export mechanism.The MS ring complex is composed of approximately 26 copies of a single monomeric protein, FliF. These monomers assemble into a single ring structure that appear as two adjacent loops consisting of various sub-structures.The entire MS ring complex spans the cytoplasmic membrane[9]. Once the assembly of the MS ring is completed, assembly of flagellar apparatus occurs inwards, outwards and laterally [15].

#### C-ring/Rotor/Switch complex

The inward assembly starts with the switch complex/C-ring. This complex is composed of three proteins: FliG, FliM and FliN. Once FliG binds to the cytoplasmic face of the MS-ring, the proteins FliM and FliN are recruited. FliM acts as an intermediate between FliG and FliN. The N-terminus of FliM interacts with the phosphorylated subunit of CheY, the middle domain interacts with FliG and the C-terminus interacts with FliN.

The N-terminus and the middle region are involved with rotational switching and the C-terminus is involved in flagellar assembly. The C-ring is the only structure that is completely embedded within the cytoplasmic membrane of bacteria [9].

### Flagellar Type Three Secretion System

The Flagellar Type Three Secretion System or fTTSS apparatus is homologous to the apparatus responsible for the export of virulence factors in pathogenic bacteria [15]. The fTTSS assembles within the C-ring.

### Stator

The lateral assembly of the flagellar apparatus involves the assembly of the stator. The Stator is composed of two proteins MotA and MotB that together form the ion channel through the plasma membrane. These proteins assemble in a ring structure consisting of 10 studs with the periplasmic domain of MotB attached to the peptidoglycan layer [15] [3]. The stator is assembled around the rotor after the hook-basal body (HBB) structure is completed [16]. As mentioned before, the stator is responsible for torque generation.

### Rod

The rod is a helical tube with a coiled-coil conformation. It can be divided into two distinct units i.e. proximal and distal. The distal component of the rod is composed of the protein FlgG and FliE. The proximal rod is formed by 6 subunits each of FlgB, FlgC and FlgF [4] [9]. These proteins are exported across the cytoplasmic membrane by the fTTSS and assemble in the periplasm [4].

### P and L rings

The P ring is composed of 26 FlgL monomers that assemble around the rod. The P-ring connects to the peptidoglycan cell wall. The L ring is composed of 28 monomers of FlgH and connects to the outer-membrane. It is known that the C-terminus of FliT is essential for the assembly of the L-ring. Together they form the homopolymeric P/L ring complex that act as the rod bushing [4].

#### Hook

The Hook is made of 120 copies of the monomer FlgE. Hook assembly is initiated when the rod cap protein FlgJ is displaced by the hook-capping protein FlgD at the distal end. Monomers of FlgE are exported across the cytoplasmic membrane by the fTTSS. These protein subunits pass through the central channel and self-assemble endogenously from base to tip. Once the hook reaches a certain length, the fTTSS undergoes a switch in substrate specificity from early flagellar substrates to late flagellar substrates. This occurs due to the fTTSS protein FlhB and mediated by the hook-length control protein - FliK. The hook cap is then discarded and replaced with hook-associated proteins or HAPs. This marks the end of the assembly of the HBB (Hook-Basal body complex).

#### Filament

HAPs are the monomeric proteins: FliD, FlgL and FlgK. Before the assembly of the filament begins, the HAPs momentarily form a complex at the distal end. These are the first filament-type substrates to be transported by the fTTSS. FliD is a pentagonal disc that forms the filament cap, whereas FlgL and FlgK form the hook-filament junction [9], [15]. The filament is composed of approximately 20,000 subunits of the monomeric protein flagellin (FliC). These protein subunits travel the entire length of the central channel until they assemble at the distal end of the growing filament.

### 1.1.3 Regulation

Flagellar gene networks are highly complex and are responsible for the expression of over 40 structural and regulatory genes [17]. Synthesis of structural genes for assembly of a bacterial flagella occurs sequentially with gene products being transcribed as they are needed during assembly. Temporal regulation of flagellar genes is highly complex and coordinated. It occurs via networks that employ checkpoints within the flagellar expression pathway. Overall, the process begins with the expression of class I genes which encode a master regulator that triggers the hierarchal expression of class II genes (i.e. structural and regulatory genes) [18].

## 1.2 Function of flagella in the environment

As mentioned above, the bacterial flagellum is an organelle responsible for locomotion. Flagellum-mediated motility is coupled with chemoreceptor systems, allowing bacterial species to sense and move towards favourable environments. For pathogenic bacteria, these environments may be sites of infection thus assisting flagellated microbes to move towards hosts and/or target sites during infection. The flagellum can secrete virulence factors, aid in biofilm formation and have adhesive properties [1]. Flagella also tend to be antigenic and can be detected by the human immune system.

### 1.2.1 Motility

The primary function of the bacterial flagellum is locomotion. The two types of flagellar-mediated motility adopted by microorganisms include swimming and swarming.

Swimming refers to the movement of individual bacterial cells within a liquid (though sometimes, it may also occur on top of a surface). Bacteria such as *Escherichia coli* swim by executing run and tumble processes. When *Escherichia coli* moves in a series of smooth straight paths, the movement is referred to as a run. A tumble on the other hand, is characterized as a sudden turn/reversal that results in a random re-orientation of the bacterial cell. After a tumble, the bacterial cell will resume a run in the new direction. Chemotactic bacteria alter their frequency of tumbles when moving towards attractants or moving away from repellents. For example, the frequency of tumbles will be reduced to extend each run when moving towards a favourable environment. This kind of movement is also referred to as a directed random walk [19]. The run or tumble is a result of changing rotational directions of the flagellar rotor as well as polymorphic transitions of protofilaments within the filament. For a run, the rotor rotates in a CCW direction and the protofilaments transition to a left-handed helix bundle. On the other hand, for a tumble to occur, the rotor rotates CW and the protofilaments adopt a right-handed conformation [1]. Details about the different protofilament conformations have been outlined under section (filament). It is also important to note that not all filaments need to be in a bundle to cause a run, in fact only one or two need to change rotation to CW and cause a tumble (tumble varies depending on ratio of CW-rotating flagella and duration of CW rotation) [20].

Swarming refers to the movement of a group of bacterial cells on a surface and/or through viscous liquids. Swarming motility can have different phenotypes and is also inducible. There is a correlation between bacteria that have peritrichous flagella and swarming

motility. This is because petrichious flagella bundle together when rotated by the motor which results in increased flagellar stiffness and therefore a greater propulsive force allowing bacteria to move within viscous fluids [21]. It is important to note that bacterial species such as *Rhodosprillum centenum* and *Aeromonas spp.* have a single polar flagellum and can selectively induce expression of petrichious flagella to swarm. Since polar and lateral flagella are controlled by different motors and also expressed through different genes, the pathogenic bacterium, *Pseudomonas aeruginosa*, unlike other species, retains its polar flagella but synthesizes an alternative motor to allow for swarming motility [21].

### 1.2.2 Pathogenesis

The bacterial flagellum plays an important role in various aspects of pathogenesis. The pathogenic cycle can be divided into four parts 1) reaching the host/target site, 2) colonization or invasion, 3) growth and maintenance and 4) dispersal to new host or target sites (Figure 1.2) [22]. How the flagellum plays a role in each of these parts will be discussed in detail.

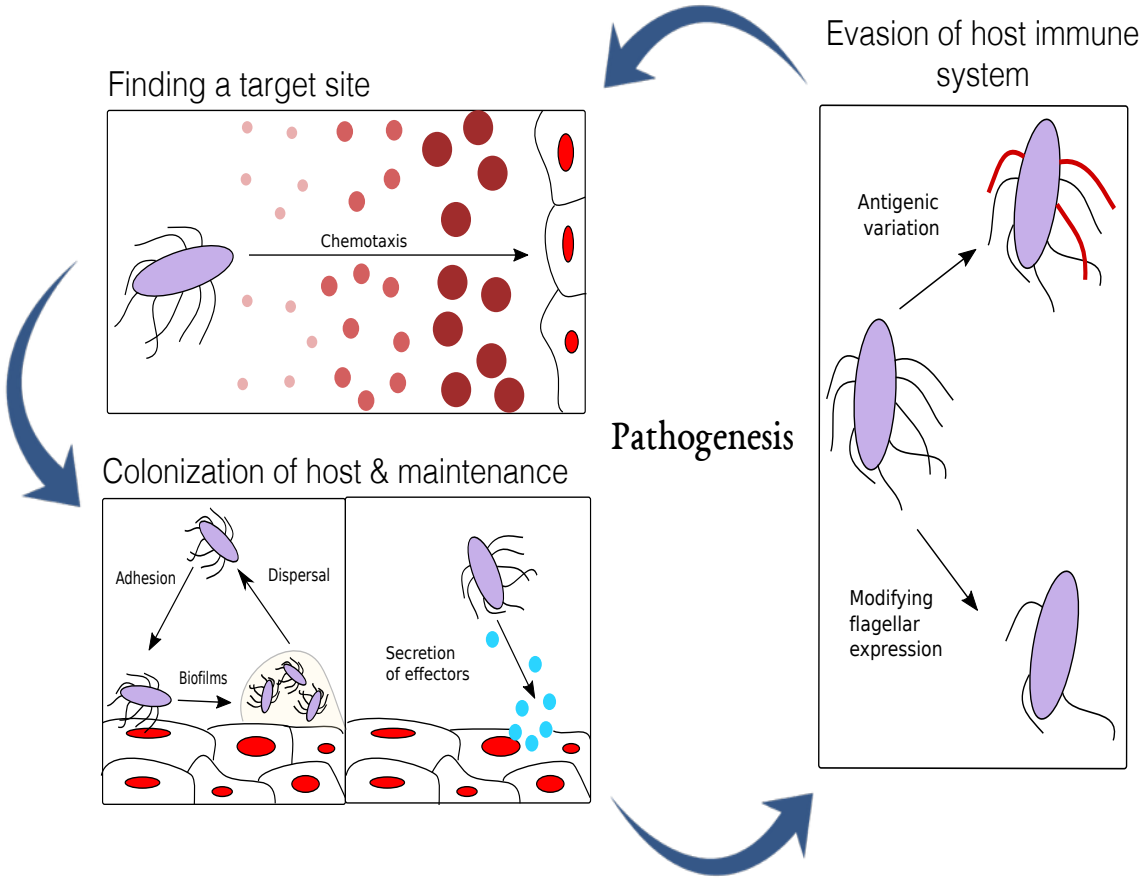


Figure 1.2: **Role of bacterial flagella in pathogenesis**

**The figure represents the role of bacterial flagella in 4 stages of pathogenesis. Flagellar mediated motility and chemotaxis is used to disperse to favourable environments or find host sites. Role of flagella in biofilm formation as well as the secretion of virulence factors by the flagellar type three secretion system allows bacterial cells to colonize hosts. Further evasion of host immune system is done by modifying expression of flagellar genes as well as antigenic variation and post-translational modification. This figure is based on [22].**

### Reaching the host/target site

The first task of a pathogenic bacteria is to find the target site and move towards it. Chemotaxis is the process by which bacteria actively sense their environment and move towards favourable sites. The interaction between chemotaxis and flagellar motility allows the bacteria to have a directed trajectory towards a host as opposed to depending on Brownian motion. As mentioned in section 1.1.1, the phosphorylated subunit of the signalling protein, CheY interacts with the flagellar rotor and thereby controls it's direction of rotation. In *Escherichia coli*, high amounts of phosphorylated CheY results in frequent tumbling and low amounts lead to more runs. Pathogenic bacteria tend to have chemoreceptor proteins in their flagellum which have evolved to control the amount of phosphorylated CheY proteins depending on their environment [22]. Examples of pathogenic organisms that utilize chemotaxis to move selectively towards target sites, include, *Salmonella spp.* which colonize the intestinal epithelium, *C.jejuni*, which colonize zones in intestines which are rich in mucin/glycoproteins, and many others. Stimuli other than chemicals are also known to be important for directing the movement of pathogenic bacteria [22].

Many pathogens can sense their environment via regulatory systems other than chemotaxis. Upon sensing the environment, these pathogenic bacteria can enhance their motility through flagellar regulation. For example, Enterohemorrhagic *Escherichia coli* needs to penetrate the polysaccharide and glycoprotein rich mammalian intestinal surface, before being able to colonize the epithelial cell surfaces. EHEC can detect short chain fatty acids in the large intestine upon which they can trigger their motility. The detection of fatty acids however, does not occur through chemotaxis, but instead through two transcriptional regulatory steps for flagellar biosynthesis. Another example of alternate flagellar-driven motility is *Vibrio cholera* which detects high bile levels through a two-component regulatory system and in turn increases motility [22].

Flagellated bacterial pathogens also use near-surface swimming to find entry sites for infection and colonize hosts more efficiently. Near-surface swimming is a phenomenon in which bacteria are entrapped at surfaces thus swimming due to physical and hydrodynamic forces. This allows flagellated organisms to search for receptors or entry sites on a 2D plane as opposed to a complex 3D plane [22]. *Salmonella Typhimurium* is a flagellated pathogen that adheres to a host intestinal surface and triggers formation of membrane ruffles. Multiple bacteria tend to invade by entering through the same ruffle. It has been shown *in vitro* that these bacteria reach the target site by non-chemotaxis flagellar motility and are subsequently trapped on the surface (near-surface swimming) that increases the pathogen density and makes the scanning of the host topology more effi-

cient. Together this allows the pathogens to efficiently enter through the same route [23].

Pathogenic bacteria need to recognize when they arrive at an optimal target site and then trigger colonization. The bacterial flagellum acts as a mechanical sensor that can allow bacteria to do this. For example, in members of the alpha and gamma proteobacteria, bacterial flagella can initiate differentiation into swarmer cells and therefore start pathogenesis. Finally, the bacterial flagellum, can also trigger biofilm formation at target site in bacteria such as *Pseudomonas aeruginosa* and *V. cholera*. Flagellar stators tend to be the primary organelles responsible for mechano-sensing for biofilm formation [22].

Flagellar-mediated motility is important for detection and subsequent colonization of hosts. This has been shown by various examples. For example, it has been shown that non-motile, flagella-less mutants of *Pseudomonas fluorescens* show reduced colonization of their host, potato roots [24]. *Campylobacter jejuni* is another organism where non-motile mutant strains show reduced colonization of the intestine [25].

### Colonization or invasion

After reaching the site of infection, the flagellum continues to assist the pathogen by being directly or indirectly involved in the subsequent progression of infection. There are multiple ways in which this may happen and these will be reviewed in this section.

Adherence to a surface is an essential step in the invasion of a host. Adhesive properties of flagellins assists in adhering to target surfaces via bacterial filament. Flagellin is known to act as an adherence molecule in various *Escherichia coli*strains. In enteropathogenic *Escherichia coli* (EPEC), flagellin's adhesive capabilities assist them in epithelial cell invasion and in *Escherichia coli* O157:H7, flagellin allows for interaction with the bovine intestinal epithelium. It has also been shown that *fliC* deletion mutants in EPEC and EHEC show reduced binding to host proteins such as mucin from bovine mucus [26]. *Pseudomonas aeruginosa* is another pathogen that uses both flagellin and flagellin cap protein (FliD) as adhesin molecules that allow it to adhere to mucin. *Pseudomonas aeruginosa* has also been known to use FliF (MS-ring) for adhering to mucin [22]. Adhesive properties of flagella have also been demonstrated in *Clostridium difficile* in which adherence of crude flagella, FliC and FliD to axenic mouse cecal mucus *in vitro* was shown [27].

Pathogenic bacteria also tend to establish biofilms after successfully identifying and reaching an infection site. Biofilms are formed when multiple bacteria adhere to a surface and differentiate into aggregates (single or multiple species) encased within an EPS-rich matrix [28]. For a flagellated bacterial cell to transition from free-swimming to a

biofilm lifestyle, it must slow down by inhibiting it's flagellar rotation. In *Bacillus subtilis*, cyclic di-GMP acts as the control protein which controls flagellar rotation by interfering with FliG(C-ring) and MotA(stator) interaction. Similarly, in gram-negative bacteria such as *Escherichia coli* and *Salmonella*, cyclic di-GMP acts as a messenger protein enabling control of the flagellar motor [22]. Overall, flagellar motility as well as the flagellar molecule itself are important for the initial adherence of the bacterium to the surface as well as the subsequent differentiation into biofilms for a successful invasion of the host. It is important to note that the flagella assist non-pathogenic bacteria in surface colonization by forming biofilms as well. An example is *Pseudoalteromonas tunicata*, a marine bacterium, which is known for its ability to form biofilms. This bacterium tends to compete with the epiphytic bacterial community associated with eukaryotic surfaces to form biofilms [29, 28, 30].

Another important role of the flagellum in colonization is through secretion of virulence factors via the fTTSS or the flagellar export apparatus. Mechanistic insights into the role of fTTSS in virulence have been gained from sequence and structural similarities shared between the the non-flagellar TTSS and the flagellar export apparatus [31]. Non-flagellar TTS is a structure used by Gram-negative bacteria to translocate virulence proteins directly into the cytosol of their respective eukaryotic hosts. TTSS containing pathogens tend to cause a diverse number of diseases. Some of these pathogens include *Salmonella*, *Escherichia coli* (EPEC and EHEC), and *Pseudomonas aeruginosa* [32]. Phylogenetic analysis has indicated that the bacterial TTSS has evolved from an ancestral bacterial flagellum thus suggesting parallels in the flagellum's involvement in pathogenesis [22].

### Growth, maintenance and dispersal

Once a pathogen has entered its host and established an infection, it enters a growth stage. To replicate within the host, it needs to evade the immune system in place to recognize and destroy these pathogens. Eukaryotic hosts such as mammals have pattern recognition receptors or PRRs that have evolved to recognize conserved components of microbial pathogens. These components are called pathogen-associated molecular patterns or PAMPs. Since flagellin proteins have conserved N- and C- termini imperative to assembly of the filament, they become ideal PAMPs. It has, however, also been shown, however, that deletions in the central region of flagellin in *Escherichia coli* K-12 strains also has a direct effect on H-antigenicity (a flagellum-specific bacterial antigen) [33]. Hosts other than mammals, use Toll-like receptors or TLRs to recognize PAMPs [34]. TLR5 primarily recognizes bacterial flagellins. Considering how the bacterial flagella has

antigenic properties, flagellated species evade the immune system by various methods. These include reducing or eliminating flagellar expression, phase variation which refers to the alternate expression of flagellar proteins and varying the proportion of bacteria that express flagella. Other evasion strategies include ways in which the bacterial flagellin is unrecognizable to TLR5; i.e. post-translational modifications such as glycosylation of flagellins and slightly altering the core sequence of flagellins, for example, mutations in the conserved N-terminus [22].

Lastly, at the end of the pathogenic cycle, bacteria need to escape their hosts and disperse. These bacteria use flagellar-mediated motility to leave the site of infection, and move towards new favourable environments [22].

## 1.3 Engineering and modularity of the bacterial flagellum

Self-polymerizing flagellin monomers are highly modular and are being utilized for genetic engineering. For example, the hypervariable domain (D3) of *Salmonella* FliC is a good target for the construction of a GFP (green fluorescent protein) fusion which subsequently results in a fluorescent filament[35]. As mentioned in 1.1.1, the conserved N- and C- termini of flagellins are responsible for polymerization and therefore extensive modification of the HVR does not affect self-assembly of the flagellar filament. Engineering HVRs of flagellins to construct protein fusions would result in up to 20,000 such HVRs simultaneously expressing the target proteins thus optimizing recombinant protein expression. Modularity of the bacterial flagellum is also utilized for bacterial surface display techniques in various facets of protein expression as well as screening and analysis of large protein libraries [36, 37].

## 1.4 Zinc-dependent metalloproteases

### 1.4.1 Background

Metalloproteases or MPs are a group of proteolytic enzymes that require a divalent metal ion for catalysis. They are responsible for the cleavage of peptide bonds and play a role in various physiological processes [38, 39].

Classification of MPs based on the MEROPS database is one of the most comprehensive and is also broadly accepted. The MEROPS database uses sequence similarity and evolutionary distance to classify proteases into families, which are then grouped into clans based on homology. The classification is made more robust with three-dimensional structural data of proteases when available [38, 40]. MPs can be mono or di-metalate. Dimetalate MPs are usually exopeptidases [38] and include proteases such as the amidohydrolase superfamily [41]. Monometalate or mononuclear MPs are divided into different tribes including, inverzincins, zincins, HybD, LAS MPs and $\alpha\beta\alpha$-exopeptidases, with the two main ones being Metzincins and Gluzincins. The names reflect their loop structures: i.e. Met-turn and Ser/Gly turn in the C-terminus subdomain respectively. Zinc-dependent MPs have a common domain structure. This consists of a signal peptide, a propeptide and a catalytic domain [42].

Most monometallic MPs belong to the zincin tribe that have more than 100 coding genes in humans [38]. Most of these MPs require a divalent metal ion $Zn^{2+}$, and sometimes cobalt for activity. The zincin tribe is characterized by the catalytic HExxH motif. This motif consists of two zinc-binding histidine residues and a glutamate that performs general base catalysis by providing the nucleophile for cleavage of peptide bonds. The general reaction catalyzed by these MPs includes a single step in which the glutamate acts as a general acid/base that is responsible for activating a zinc-bound $H_2O$ molecule. This provides the nucleophile for cleaving peptide bonds without a covalent intermediate. These reactions follow Michaelis-Menten kinetics at neutral pH [43, 38]. This mechanism was derived from research on the thermolysin from *Bacillus thermoproteolyticus* and bovine carbopeptidase A [44, 45, 46, 38].

A general mechanism for peptide-bond hydrolysis by MPs has been elucidated with structural data (X-ray crystallography) and structural modelling. Based on work by Schechter and Berger on the endopeptidase, papain and its substrate, alanine peptides, It has been shown that the substrate tends to bind the active site in an extended conformation [47]. The active site cleft itself is also in an "optimal" conformation prior to catalysis and is therefore accessible to the peptidic substrate [47]. The active site cleft has been divided into two parts with respect to the scissile peptide bond. The part upstream the active site cleft is termed as the N-terminal upstream region and the part below as the C-terminus downstream region. The upstream region is annotated to be "non-prime" and the downstream region as "prime" [47, 38].

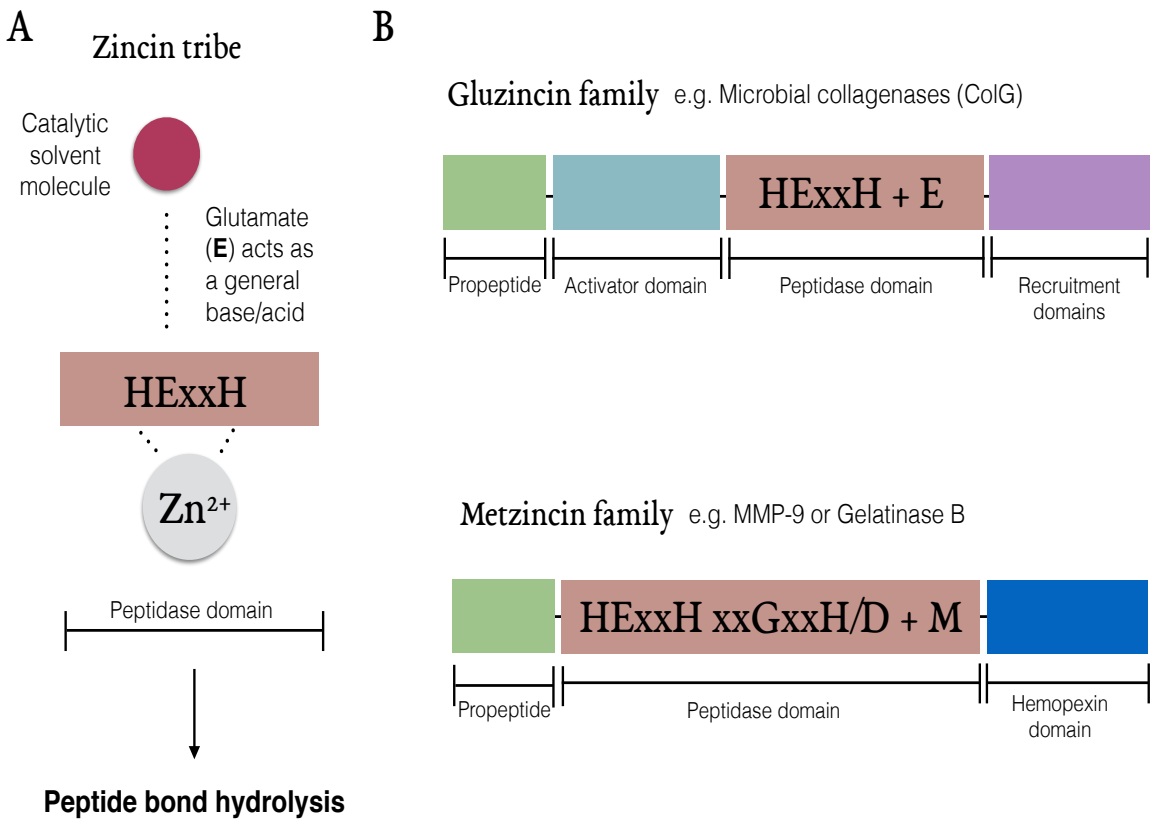


Figure 1.3: **Overview of zincin family metalloprotease structure**

**(A) Metalloproteases belonging to the zincin tribe have the consensus catalytic motif HExxH. The two histidines are responsible for binding zinc whereas, the glutamate acts as a general base and and activates a catalytic solvent molecule to provide nucleophile for the cleavage of peptide bonds. (B) The zincin tribe is divided into gluzincin and metzincin. The gluzincin family of proteases contain an additional downstream glutamate (33 to 35 aa downstream) within their peptidase domains. The domain architecture is shown for ColG, a microbial collagenase from *Clostridium histolyticum*.The metzincin family of proteases contain an additional methionine in the C-terminus peptidase subdomain. The domain architecture is shown for MMP-9 or gelatinase B.**

### 1.4.2 Gluzincins

The gluzincin superfamily of MPs forms one of the clans of the zincin tribe. Gluzincins are distinguished from other MPs due to a characteristic Ser/Gly turn as well as a glutamate which acts as the third metal binding residue. The Gluzincin clan has following consensus sequence - HExxH. Additionally another motif [E/H]XXZ is found 33 - 35 amino acids downstream where X is any amino acid and Z is one of the following: A/F/S/G/T [38] [48]. This clan is further divided into families including, bacterial collagenases, clostridial neurotoxins, thermolysins and others.

#### Microbial collagenases

Microbial collagenases are MPs that belong to the MEROPS peptidase family M9 and together are involved in the hydrolysis of different types of native collagen. Microbial collagenases act as extracellular virulent factors and are known to be utilized by bacteria for pathogenesis, and are implicated in human disease.

Collagen is the primary component of the ECM (extracellular matrix). Furthermore, various types of collagens are also the most abundant proteins in animal tissues and organs. Within the ECM, collagen fibres provide structural integrity and organization whereas, collagen fibrils act as anchors for the attachment of macromolecules and other cells. A collagen molecule is composed of three intertwined $\alpha$ chains that together form a right-handed helix. Each $\alpha$ chain is composed of repeating units of Gly-X-Y, where X and Y are any amino acids. These motifs are present in human collagens as well as some prokaryotic collagens. Based on the structural differences, collagens are divided into two major forms: fibrillar and non-fibrillar. Fibrillar collagen has a structure that is characterized by undisturbed helices. It is comprised of more classic collagens i.e., type I, II and III. Non-fibrillar collagens on the other hand contain non-helical regions interspersed within the largely helical molecule. This category consists of type IV and VI collagens. Type I collagen forms 95% of collagens within most animal tissue [49].

Clostridial collagenases were the first microbial collagenases to be biochemically characterized and understood. These are large multi-domain proteins of approximately 115 kDa. The domain organization of clostridial collagenases has been identified through structural studies on various existing isoforms of collagenases from different clostridia species (ColG, ColH and ColT) [50, 51]. The basic structure consists of a collagenase enzyme module and the recruitment domains. The collagenase module consists of a predomain of variable length that contains the export signal and an N terminal domain (activator and peptidase domains) that contains the catalytic zinc. The recruitment mod-

ule consists of polycystic kidney disease-like (PKD-like) and collagen-binding (CBD) domains. The number of each of these domains vary between collagenases [50, 51]. Bacterial collagenases in contrast to vertebrate collagenases, can cleave collagen at multiple sites and degrade it completely. Interestingly, clostridial collagenases have been shown to require calcium along with zinc for catalytic activity [51]. The substrate specificity of purified clostridial collagenases has confirmed preference for the characteristic collagen motifs (Gly-Pro-X and Gly-X-Hyp, where X is any amino acid and Hyp is hydroxyproline) [52].

### 1.4.3 Metzincins

Metzincins form a clan belonging to the zincin tribe. This family consists of multi-domain, zinc-dependent matrix metalloproteases (MMPs). This clan further subdivides into 12 distinct families out which of at least 7 have been structurally characterized. Each of these families have distinct structural components. The domain architecture of metzincins includes an N terminal pro-domain, a protease domain consisting the following consensus sequence, HExxHxxGxx[H/D] and a characteristic downstream Met-turn [53]. Additionally, all MMPs except for MMP-7 and MMP-26 have a carboxyterminal hemopexin domain [54].

#### MMP-9 or Gelatinase B

Matrix metalloproteases are principal mediators of extracellular matrix (ECM) maintenance via remodelling and degradation. MMPs are also known for being essential players in the host immune system. They are responsible for functions such as reduction of inflammation, killing of pathogens via effector cells and cleavage of cytokines. Irregular activity of MMPs within the host is implicated in immunopathology [55].

MMP-9 (gelatinase B), has one of the most complex structures amongst the MMPs. In addition to common structural features, MMP-9 contains an additional gelatin-binding fibronectin domain as well as a Ser/Thr/Pro-rich collagen type V domain. Gelatinase B is a heavily glycosylated protease and undergoes heterodimerization via disulfide bonding. These factors complicate crystallization efforts [54, 56]. Though, gelatinase B without the collagen and hemopexin domains has been crystallized, its catalytic domain has been studied through crystallization of gelatinase A, an MMP which is the closest in sequence to gelatinase B [54].

MMP-9 is an endoprotease whose substrate specificity depends on several factors. This includes, within the catalytic site of the three-dimensional protein, availability of the cleavage site on the substrate, primary sequence of the substrate and interestingly exosites on the protein which interact with distal sites on the substrate and are important for hydrolysis [54, 57]. Cleavage assays with synthetic peptides has revealed the primary sequence specificity of gelatinase B. A strong preference of hydrophobic residues at the P1' and P2 sites have been shown. This corresponds to the fact that the S1' site of MMPs contain a deep hydrophobic pocket. Preference for small amino acids such as Gly, Ala and Pro has been shown at the P1 site. Physiological substrates of MMP-9 include gelatin, which is a form of denatured collagens. Studies on gelatinase A have revealed that the fibronectin domain is responsible for interacting with collagenous substrates. Studies on type II gelatin, a primary substrate of gelatinase B has revealed that gelatinase B cleaves after a Gly residue at P1 position [54, 56].

## 1.5 *Clostridium haemolyticum*

*Clostridium haemolyticum* a.k.a *Clostridium novyi* type D is an anaerobic soil-borne bacterium capable of forming spores. It is an animal pathogen responsible for causing Bacillary hemoglobinuria or redwater disease in cattle [58]. It is ubiquitous in various livestock-producing countries [59]. *Clostridium novyi* species are divided into types A, B, C and D based on the type of toxin they produce. Types A and B are known to be pathogenic and secrete types $\alpha$ and $\beta$ toxins respectively, and type C is non-pathogenic. Originally *Clostridium haemolyticum* was thought to be a member the type B species due to biological similarities but phylogenetic analysis of several clostridia species by 16S rRNA sequencing has identified it as a separate species. The analysis revealed that although *Clostridium haemolyticum* clusters with *Clostridium botulinum* type C and D, it shares more than 99.9% sequence similarity with *Clostridium novyi* type B [60]. *Clostridium haemolyticum* was also considered to be a separate species due to its pathogenesis not being related to alpha toxin production as is the case for types A and B. Furthermore, the diseases caused by types A and B i.e., gas gangrene in animals/humans and necrotic hepatitis in sheep respectively, are also different from *Clostridium haemolyticum* which affects cattle [61].

## 1.6 Computational predication of a novel proteolytic flagellin family

As outlined in Section 1.1.1, bacterial flagellin sequences tend to have conserved N- and C- termini that form the core of the assembled flagellar filament, however, their central regions are highly variable and form the surface-exposed regions of the assembled filament (Figure 1.4). Given the sequence pattern displayed by the HVR as well as the increasingly important role of flagella in pathogenesis and biofilm formation, the HVRs of 26,587 flagellin sequences from the NCBI non-redundant database were analyzed for hidden functional diversity through discovery of novel domain architectures. Detection of new domain architectures is an effective bioinformatic strategy for novel function discovery [62]. 62 such conserved domain architectures were discovered through CDART (Figure 1.4). Even though no natural flagella has been known to have enzymatic activity, uncharacterized protein domains with predicted enzymatic activity were found within flagellin HVRs (Table 5.1).

The most frequent domain found was the M9 peptidase domain which is similar to Gluzincin family of zinc-dependent MPs, a flagellin architecture first identified by Doxey et al. [63]. This domain was the 6th most common flagellin architecture detected (Figure 1.4), and was found in the highest number of species. This was further verified by in-depth sequence analysis and structural modelling (Figure 1.5). PSI-BLAST searches revealed the presence of this domain in 74 different bacterial species spanning 32 genera Table 5.2 and were largely distributed amongst Proteobacteria and Firmicutes. Some of these include animal pathogens such as *Clostridium haemolyticum* and *Clostridium botulinum* C/D str. DC5 as well as avid biofilm former *Pseudoaltermonas tunicata*. The phylogenetic diversity of the species containing the peptidase domain indicates the mechanism to be lateral gene transfer (Figure 5.1).

In-depth sequence analysis (Figure 1.4), structural modelling (Figure 1.5) and reciprocal PSI-BLAST searches verified that the peptidase domain is closely related to clostridial collagenases, members of the gluzincin family. These identified flagellins contain a conserved HExxH catalytic motif, which as indicated in Section 1.4.2, is characteristic of gluzincin-like MPs.

### 1.6.1 Computational methods

The following work was a collaborative effort with **Mike Mansfield** at the University of Waterloo.

The CDART tool [64] was used to identify all proteins in the NCBI non- redundant protein database containing a predicted flagellin N-hypervariable region-C domain architecture (search performed on January 31, 2017). Identified sequences possessing a central gluzincin domain within the flagellin hypervariable region were retrieved and aligned using MUSCLE version v3.8.31[65] with default parameters, and visualized with Jalview version 2.80b1 [66]. A structural model of the FliA(H)-hypervariable region was generated using the modeling servers SWISS-MODEL [67] and iTASSER [68] with the crystal structures of clostridial collagenase G (PDB entry 2Y3U), H (PDB entry 4AR1), and T (PDB entry 4AR8) as templates [50, 51]. Phyre 2.0's intensive modeling procedure was also used and yielded highly similar results [69]. Energy minimization and secondary structure assignment were performed within the UCSF Chimera package [70].

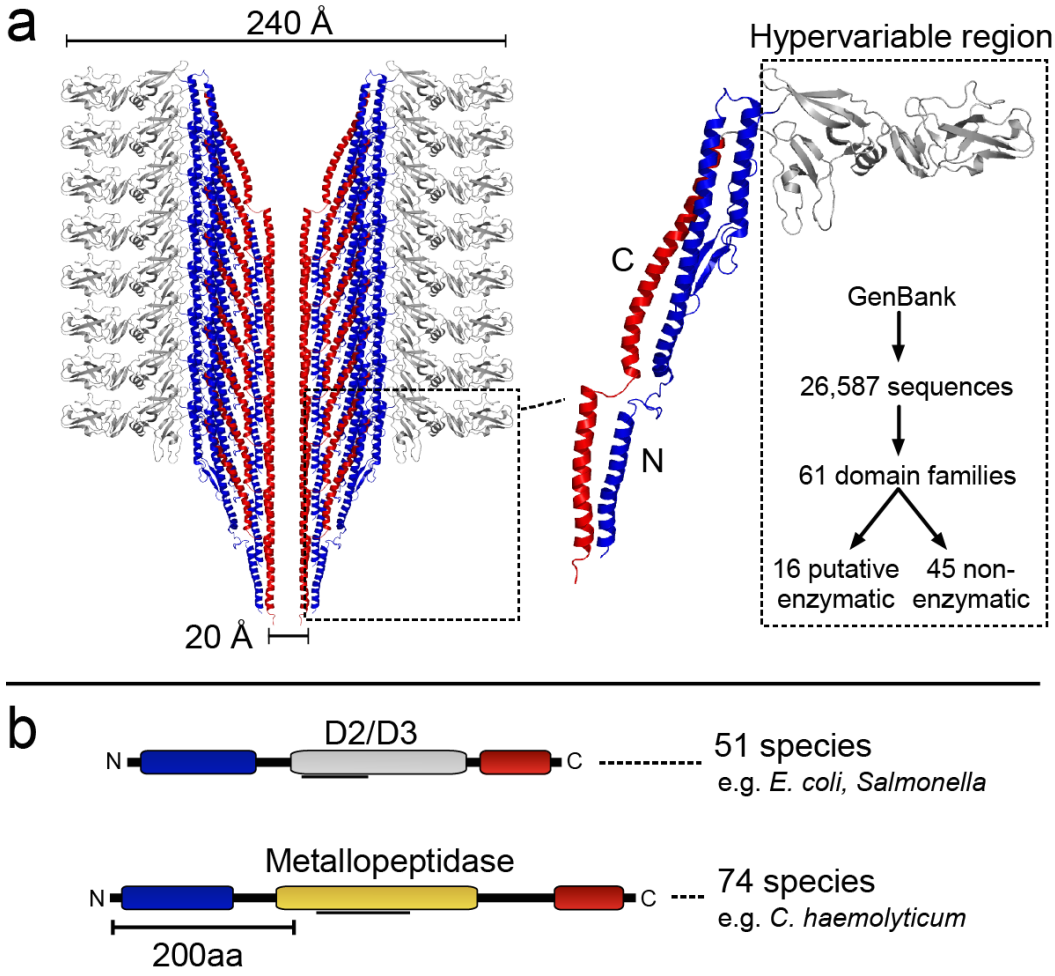


Figure 1.4: **Structural model of the flagellar filament and identification of uncharacterized domains in the surface-exposed flagellin HVR.**

**(A) Structural model of the flagellar filament and constituent flagellins (left), highlighting the interior flagellin N-terminal and C-terminal domain and the surface-exposed hypervariable region domain (right). The model is based on the structure of FliC from *Salmonella* (PDB 1UCU). Flagellin hypervariable regions from the NCBI GenBank database were analyzed, revealing 61 putative domain families including several with potential enzymatic function. (B) A schematic depicting the commonly studied D2/D3 flagellin domain variant, as well as a novel putative metallopeptidase domain predicted within 74 species. The lines below the domains indicate regions matched by the Conserved Domain Database, which were manually refined through subsequent analysis (Eckhard et al. 2017, in review).**

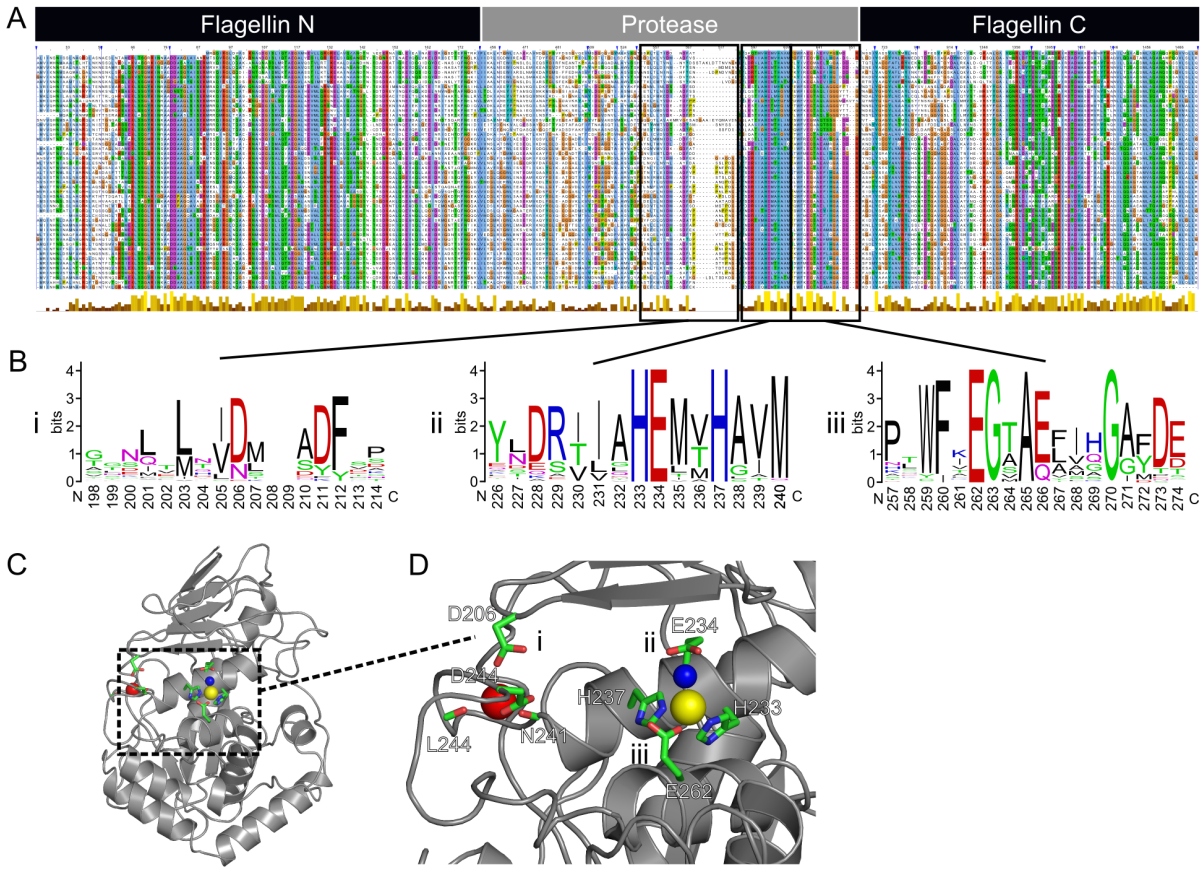


Figure 1.5: **Sequence analysis and structural modeling of the proteolytic flagellin family.**

**(A) 86-sequence multiple alignment of proteolytic flagellins, with key functional motifs depicted assequence logos (B) including a putative Ca-binding site (i), the zinc-binding HEXXH motif (includes the zinc-binding His233 and His237 and general base Glu234 for hydrolysis; sequencenumbering according to FliA(H), UniProt entry Q8RR94) (ii) and a motif containing the third zinc-binding residue, Glu262 (iii). In addition, the hydrophobic basement-forming residue(Ala265) located below the active site zinc is also conserved. (C) Structural model of proteolytic flagella (FliA(H)-hypervariable region) from *Clostridium haemolyticum* based on the peptidase domain of collagenase H from *Clostridium histolyticum* (PDB 4AR1). (D) The zinc-binding HEXXH motif and Ca-binding site with important residues labeled.(Eckhard et al. 2017, in review).**

## 1.7 Research objective

Computational prediction of a metallopeptidase domain in the HVR of various bacterial species prompted the investigation and experimental validation of the first family of HVR-associated proteases with predicted enzymatic activity. The investigation was carried out using FliA(H) from *Clostridium haemolyticum*. This species was chosen due to it's industrial importance as an animal pathogen. The experimental approach was two-fold and investigated FliA(H) from *Clostridium haemolyticum* both *in vitro* and *in vivo*. Both approaches are covered in detail in Chapters 2 and 3.

# Chapter 2

# General Materials and Methods

## 2.1 Bacterial plasmids

### pET32a(+)

pET32a(+) is part of the pET expression vector collection designed to optimize expression and purification of recombinant proteins in bacterial cells. The plasmid pET32a+ is derived from pBR322 and therefore maintains that naming convention. This pET vector allows for constitutive and high-level expression of cloned sequences via a T7 promoter. With multiple cloning sites available, the cloned sequences can have cleavable Trx Tag$^{TM}$ , His Tag® or S Tag$^{TM}$ for subsequent purification and detection of the expressed proteins [71]. Other features include an ampicillin resistance gene for maintenance of the plasmid in transformed bacterial cells as well as the pMB1 ori which results in a high copy number of cells.

### pET21a(+)

pET21a(+) is also a part of the pET expression vector collection designed to optimize expression and purification of recombinant proteins in bacterial cells. Details about the plasmid are the same as the vector described above with the only difference being that this plasmid allows for a C-terminal His Tag®.

## 2.2 Bacterial strains

***Escherichia coli* - DH5$\alpha$**

Genotype: $F^-$ *endA1 glnV44 thi-1 recA1 relA1 gyrA96 deoR nupG purB20* $\phi$80*dlacZ*$\delta$M15 $\delta$*(lacZYA-argF)*U169, hsdR17($r_K^-$ $m_K^-$), $\lambda^-$ [72].

This strain of *E.coli* is a common laboratory strain used in molecular biology. It is derived from the K-12 strain and also referred to as a Hoffman-Berling 1100 strain derivative. This strain was used to maintain frozen stocks for some constructs.

***Escherichia coli* - BL21(DE3)**

Genotype: $F^-$ *ompT gal dcm lon hsdS*$_B$($r_B^-$ $m_B^-$) $\lambda$(DE3 [*lacI lacUV5-T7p07 ind1 sam7 nin5*]) [$malB^+$]$_{K-12}$ $\lambda^S$

This strain of *E.coli* is compatible with the pET expression vectors used for the expression of recombinant proteins. This strain is derived from the *E.coli* B strain. DE3 refers to the $\lambda$ prophage inserted in the strain which carries the T7 RNA polymerase gene and $lacI^q$. Since the gene for T7 RNA polymerase in under the control of a lac promoter, the expression of the cloned sequence under the control of it's respective T7 promoter (as is the case with pET expression vectors) can be repressed until IPTG is used [73].

***Clostridium haemolyticum***

Strain ATCC® 9650$^{TM}$ ; NCIMB 10664

This strain of *Clostridium haemolyticum* contains the flagellin variant FliA(H) and was used for this project.

## 2.3 Maintenance of bacterial cultures

*Escherichia coli* strains were grown at 37°C in LB broth and LB agar under aerobic conditions with continuous shaking.

*Clostridium haemolyticum* was grown at 37°C under anaerobic conditions. The anaerobic gas mixture is defined as 80% $N_2$, 10% $CO_2$ and 10% $H_2$.

## 2.4 Standard molecular biology protocols for DNA manipulation

### 2.4.1 Preparation of competent cells

*Escherichia coli* cells were grown overnight and subcultured 1:100 into fresh LB media. The cells were allowed to grow until the $OD_{600}$ was between 0.4 to 0.6 (log phase). Please refer to section 2.3 for details on cell culturing protocols. The cells were subsequently pelleted via centrifugation at 13,000 rpm and resuspended gently in ice-cold 0.1M $CaCl_2$. The cells were centrifuged again at 6000 rpm at 4°C and resuspended in ice-cold 0.1M $CaCl_2$. The volume of 0.1M $CaCl_2$ used for the final resuspension was 40% of the original volume of LB that the cells were grown in. Competent cells were made freshly before every transformation.

### 2.4.2 Bacterial transformation

100 ng or 1$\mu$L of purified plasmid, mini-prepped plasmid (2.4.3) or ligated sample was added to 100 $\mu$L of chemically-competent cells. This mixture was allowed to sit for 30 minutes on ice. After this, heat-shock of the cells was carried out by incubating the mixture at 42°C for 30 seconds. The cells were incubated on ice again, for 2 minutes. 1 mL of fresh media was added to the mixture and allowed to incubate at 37°C with shaking. The cells were pelleted again at 13,000 rpm for 2 minutes and resuspended in 50$\mu$L of fresh media and subsequently plated on LB agar plates containing the appropriate antibiotic.

### 2.4.3 Plasmid DNA mini prep

Transformed bacterial cells were grown overnight and pelleted via centrifugation at 13,000 rpm for 2 minutes. The pellet was resuspended in 250 $\mu$L of resuspension solution. 250 $\mu$L of lysis buffer and subsequently 350 $\mu$L of neutralization buffer is added. The mixture was then centrifuged for 5 minutes at 13,000 rpm. The supernatant was added to a flow-through tube and centrifuged at 13,000 rpm for 2 minutes after which the flow-through is discarded. The pelleted plasmid DNA is washed three times with ethanol containing wash buffer and is then eluted in 30 $\mu$L of elution buffer.

### 2.4.4 Restriction enzyme digestion

The steps in the protocol for restriction enzyme digestion depends on several factors. Considering that a double-digest of 1000 ng of purified plasmid is being carried out, 1 $\mu$L of each restriction enzyme is added to the appropriate volume of purified plasmid DNA corresponding to an amount of 1000 ng. 10X restriction digest buffer (based on the kit and type of enzymes are being used) is then added to the reaction mixture. The mixture is topped up to 20 $\mu$L with milliQ water. The reaction mixture is incubated at the appropriate temperature and for the appropriate amount of time, depending on the enzyme being used.

### 2.4.5 Agarose gel electrophoresis

To cast a 1% agarose gel, 50g of agarose is added to 50 mL of TAE buffer and heated until all the agarose is dissolved. After the mixture is allowed to cool down for approximately 2 minutes, 1:10,000 of GelRed$^{TM}$ nucleic acid stain is added. The mixture is poured into the appropriate casting system, an appropriate comb is inserted and allowed to dry. Excess TAE buffer is adding to the system containing the solidified gel and DNA solution containing loading dye is added to the wells. The gel was run at 90V until the loading dye ran off.

## 2.5 Standard protocols for Protein Analysis

### 2.5.1 SDS-PAGE

Protein samples were prepared with loading dye containing SDS and beta-mercaptoethanol. These were loaded on a 4 % stacking / 12% resolving polyacrylamide gel and run at 100V for 1 hour [74]. The gel was stained using Coomassie blue.

### 2.5.2 BCA Protein Assay

Pierce$^{TM}$ BCA Protein Assay kit (Thermo Scientific$^{TM}$) was used for protein quantification. It uses a standard curve generated from a range of diluted BSA protein standards of known concentration. This was done by diluting a 2 mg/mL BSA standard to prepare diluted standards at the following concentrations: 2.0, 1.5, 1.0, 0.75, 0.50, 0.25, 0.125,

0.025 and 0 mg/mL. 25 $\mu$L of the standards along with the protein sample of unknown concentration were plated in duplicate on a 96 well plate. A BCA working solution was prepared by adding BCA reagent A and BCA reagent B at a ratio of 50:1. 200 $\mu$L was added to each well containing the standard or the protein sample. The plate was incubated at 37°C for 30 minutes after which the absorbance was measured using a plate reader at an absorbance of 562 nm. A standard curve of absorbance vs BSA concentration was generated and used to calculate the concentration of the protein sample.

# Chapter 3

# Investigating protease activity from FliA(H) *in vitro*

## 3.1 Background, Hypotheses and Rationale

As mentioned in Section 1.6, the 6th most common flagellin architecture, which was also found in the highest number of species, contained a conserved HExxH motif within the HVR of bacterial flagellins. Considering the multi-faceted role of flagella in pathogenesis explained in section 1.2.2, *Clostridium haemolyticum* FliA(H) was chosen as the representative flagellin variant for investigating the biological origins and function of this protein. *Clostridium haemolyticum* is an animal pathogen and is therefore of importance in industry. Considering that the peptidase domain was predicted to be the most closely related to clostridial collagenases, *Clostridium haemolyticum* was an ideal candidate for studying the flagellin variant. To begin characterization of the putative enzymatic flagellin variant, the first avenue of research was to study the HVR *in vitro*. The rationale for this chapter is divided into two main questions that were asked and investigated. The questions are as follows:

**I. Does the HVR containing the gluzincin-like domain from *Clostridium haemolyticum* FliA(H) produce a folded protein product?**

To investigate whether the obtained protease domain was in fact functional, the HVR of FliA(H) was cloned into an expression vector and transformed in to a compatible *Escherichia coli* expression strain. The HVR without the N and C termini was used for investigating recombinant expression.

## II. Does the folded protein exhibit catalytic activity similar to the predicted function?

To investigate potential protease activity in FliA(H), two methods were employed.

1. Proteomic identification of substrate cleavage sites or PICS assay

   PICS assay was developed in the Overall laboratory at the University of British Columbia. This assay involves incubation of the test protease with a peptide library. The peptide library is obtained by digesting the genome of *Escherichia coli* K-12 with the serine proteases, Trypsin and endoproteinase GluC to create two 'random' peptide libraries. Trypsin cleaves peptide chains selectively at the C-terminus of amino acids, lysine and arginine, except when followed by proline. Endoproteinase GluC, on the other hand, cleaves peptides at the C-terminus of the amino acid, glutamic acid. The two libraries are incubated separately with the test protease. Cleavage of peptides by the test protease results in an exposed N terminus. These prime-side cleavage products are labelled with biotin and purified by affinity chromatography with streptavidin and sequenced using LC-MS/MS. The corresponding sequences of the non-prime side are generated through bioinformatics. The sequences of all the cleaved peptides are aligned to get a consensus sequence which corresponds to the putative substrate cleavage sites.

   PICS was performed with the recombinant Trx-FliA(H)-HVR to investigate the putative protease activity and whether its substrate cleavage profile is similar to that of substrates cleaved by clostridial collagenases.

2. Quenched fluorometric (QF) assays

   To further validate the substrate specificity of the recombinant FliA(H), QF peptide cleavage assays with three synthetic quenched fluorometric peptides were carried out. The synthetic peptides were chosen to validate the preference for the hydrophobic amino acid leucine in the P1' position. The three peptides were as follows: ALG-L, which closely matches the cleavage specificity determined by PICS; the classical MMP substrate QF24 (PLG-L); and PLG-V to test the dependency for leucine in P1'.

## 3.2 Methodology

### 3.2.1 Bacterial strains and construct design

All constructs were synthesized by Biobasic.Inc and are as follows:

1. **Trx-FliA(H)-HVR**

   The hypervariable region of FliA(H) ($Lys^{149}$-$Ile^{448}$) WP_039229452.1) from *Clostridium haemolyticum* was cloned into the expression vector pET-32a(+) using the restriction enzymes KpnI and XhoI. N- and C- termini from FliA(H) were removed. These cloning sites allow the fusion protein to have a cleavable His.Tag®as well as a Trx.Tag™. The Trx tag results in a 109 amino acid long thioredoxin protein to be produced as a fusion with the cloned recombinant protein. It was used to ensure the production of soluble recombinant proteins [75, 76]. The polyhistidine affinity tag consists of several consecutive histidine residues and are appended to proteins to allow for purification by nickel affinity chromatography with an Ni-NTA resin [77].

2. **SUMO-FliA(H)-HVR**

   The hypervariable region (300 amino acids long) of FliA(H) from *C.haemolyticum* along with a N-terminal SUMO tag was cloned into the expression vector pET-21a(+) using the restriction enzymes NdeI and HindIII. These cloning sites allow the fusion protein to have an additional N-terminus His.Tag®.

3. **SUMO-FliA(H)-HVR, mutant**

   The SUMO-FliA(H)-HVR construct was modified by mutating the glutamic acid to alanine (E234A) in the catalytic site -HExxH- within the hypervariable region (300 amino acids long) of FliA(H) by site-directed mutagenesis. This mutation would knock-out HVR-associated protease activity and thus also remove possibility of self-proteolysis. This results in a more stable protein that can be used for the production of polyclonal antibodies as well as for X-ray crystallography studies.

All constructs were verified by protocols outlined in 2.4.

### 3.2.2 Expression and purification of recombinant proteins

1. **Expression and Purification of recombinant SUMO-FliA(H)-HVR protein**

   For expression of SUMO-FliA(H)-HVR recombinant proteins, the constructs were transformed into BL21(DE3) cells as highlighted in Section 2.4.2. The transformed cells were grown at 37 °C until $OD_{600}$ was 0.4 at which point they were induced with 0.1mM IPTG. The induced cells were allowed to grow at 30 °C for 3 hours before they were harvested by centrifugation. Constructs were verified by protocols described in Section 2.4.

   SUMO-FliA(H)-HVR proteins were purified using Immobilized metal ion affinity-flow chromatography. A nickel resin that chemically binds to His-tagged proteins was added to a gravity-flow column. Cells were incubated with lysis buffer supplemented with Protease Inhibitor Cocktail Set VII (Calbiochem®#539138 M), PMSF protease inhibitor and EDTA. The homogenate was then subjected to lysis using a french press. The cell lysate was centrifuged to remove pelleted cell debris. The supernatant was incubated with 60% ammonium sulphate overnight. The overnight suspension was subsequently centrifuged to obtain the pelleted recombinant protein. The pellet was resuspended in excess lysis buffer and incubated with the charged Ni-NTA agarose resin (Qiagen-Cat No. 30210) for 45 minutes. The lysate was collected as flow-through and discarded. The bound protein was eluted with a high imidazole elution buffer. Flow through was collected in multiple fractions until $OD_{280} < 0.1$ and subsequently concentrated to 10 mL using a nitrogen-based concentrator. The concentrated protein was subsequently incubated overnight (10 hrs) with SUMO protease. The SUMO protease digested sample was buffer exchanged into 25mM HEPES buffer, pH 7.5 and subsequently incubated with the Nickel resin again. The elutant was frozen in 10$\mu$L pellets in liquid nitrogen and stored at -80°C.

   Recipes for buffers used is provided in the Appendix.

2. **Expression and Purification of recombinant Trx-FliA(H)-HVR protein**

   Expression and purification of recombinant Trx-FliA(H)-HVR protein was done by Biobasic.Inc

   The Trx-FliA(H) construct was transformed into competent BL21(DE3) cells. Cells were grown in LB media supplemented with ampicillin at 37°C until the logarithmic phase was reached after which it was induced with 0.1 mM IPTG for 3 hours

at 30°C. The His®tagged proteins were purified using Nickel affinity chromatography and further purified by size-exclusion via gel-filtration. Purification was confirmed by SDS-PAGE as outlined in section 2.5.1.

### 3.2.3 PICS Assay

Proteomic identification of substrate cleavage sites or PICS assay was developed and performed by the Overall laboratory at the University of British Columbia [78]. The protocol was written and carried out by **Dr. Ulrich Eckhard**.

Proteomic identification of FliA(H) cleavage sites by PICS using whole proteome peptide libraries were performed [79, 80, 81, 78]. In brief, as a proteome source to generate peptide libraries, cell pellets collected from *Escherichia coli* K-12 cultures were lysed in presence of protease inhibitors and cell debris was removed by centrifugation. Soluble proteins were denatured with 4 M guanidine hydrochloride and cysteine side chains were reduced and protected with iodoacetamide. After chloroform/methanol precipitation, pellets were re-suspended and digested with trypsin (TPCK treated, Sigma-Aldrich) or GluC (Staphylococcus aureus protease V8, Worthington) at a protease to proteome ratio of 1:100 (w/w). Primary amines were then blocked by reductive dimethylation and samples were desalted by size exclusion chromatography and further purified by reversed-phase chromatography. Eluates were concentrated, re-suspended in water, and stored at -80°C as peptide libraries of 200 $\mu$g. PICS cleavage assays were performed by incubating one peptide library aliquot with active recombinant FliA(H). Cleaved peptides presenting neo N-termini were selectively biotinylated, affinity purified using streptavidin Sepharose and desalted using reversed-phase solid phase extraction. Eluates were vacuum dried to near dryness and stored at -80°C until LC-MS/MS analysis on a high resolution quadrupole Time-Of-Flight mass spectrometer (impact II, Bruker Daltonics).

#### Mass spectrometry

The protocol was written and carried out by **Dr. Ulrich Eckhard**.

LC-MS/MS analysis was performed on a nano-LC system (EASYnLC1000, Thermo Scientific, USA) coupled to a high resolution quadrupole Time-Of-Flight mass spectrometer (Impact II, Bruker Daltonics) using the CaptiveSpray ionizationsource (Bruker Daltonics), a 2-cm long, 75-$\mu$m inner diameter fused silica trap column, and a 20 cm long, 50-$\mu$m inner diameter fused silica fritted analytical column. The trap column was packed with 5-?m diameter Aqua C18 beads (Phenomenex, USA), while the analytical column

was packed with 1.9-$\mu$m diameter Reprosil-Pur C18-AQ beads (Dr.Maisch, Ammerbuch, Germany). Peptides were eluted using a 0 to 80% gradient of organic phase over 90 min. Buffer A was 0.1% formic acid and buffer B was 100% (v/v) acetonitrile with 0.1% formic acid. MS/MS data were acquired automatically using the otofControl software (Bruker Daltonics) for information-dependent acquisition. Error of mass measurement was usually within 5 ppm and was not allowed to exceed 10 ppm. Peak lists and mzXML-files of the acquired high resolution Time-Of-Flight mass spectrometer data were created using the Compass DataAnalysis software 4.2 (Bruker Daltonics).

### LC-MS/MS of recombinant FliA(H)-HVR

The protocol was written and carried out by **Dr. Ulrich Eckhard**.

Tandem mass spectrometry was performed essentially as described in detail by [82]. In brief, free sulfhydryl groups of 10 $\mu$g of FliA(H) hypervariable region protease in 1x phosphate buffer saline (PBS; 10 mM Na2HPO4, 1.8 mM KH2PO4, 137 mM NaCl, 2.7mM KCl, pH 7.4) were alkylated with iodoacetamide (20 mM iodoacetamide, 1 h, 20°C) after initial cysteine reduction with 10 mM dithiothreitol (DTT; 30 min, 20°C). Labeling was stopped by adding further DTT (10 mM, 15 min, 20°C). Trypsin Gold (Promega)was added in a 1:50 w/w ratio and the sample was digested over night at 37 C. The reaction was stopped by adding formic acid to a final concentration of 2.5% and desalted by C18-STAGE tip [83] prior mass spectrometry analysis. LC-MS/MS analysis was performed as described below. Peptides were identified at a 1% false discovery rate from a combined UniProt database, containing both human and *Escherichia coli* proteins and supplemented by our target protein, thioredoxin-tagged FliA(H)-hypervariable region. Two search engines, Mascot v2.4.1 (Matrix Science, London, UK) and X!Tandem [84] were used in conjunction with Peptide Prophet [85] as implemented in the Trans Proteomic Pipeline v.4.6. Search parameters included a mass tolerance of 15 ppm for the parental ions and 0.1 Da for fragment ions, and allowed for one missed cleavage. Carbamidomethylation of cysteine residues (+57.02 Da) was used as fixed modification, whereas methionine oxidation (+15.99 Da) was set as variable. Peptide lists from both search engines were combined within the Trans Proteomic Pipeline for further analysis.

### In-gel digestion and mass spectrometry

The protocol was written and carried out by **Dr. Ulrich Eckhard**.

In-gel digests were performed as described by [86]. Recombinant FliA(H) was resolved by 12% SDS-PAGE, stained with Coomassie G-250, and nine bands were excised, destained

with 60% acetonitrile, 20mM ammonium bicarbonate, and then washed in 100% acetonitrile before lyophilization. Gel bands were rehydrated with 10 $\mu$L Trypsin Gold (Promega) in 20 mM ammonium bicarbonate (12 ng/$\mu$L) by passive diffusion for 1 h at 4 °C. Excess solution was removed and 20 $\mu$L 20 mM ammonium bicarbonate was added to the gel plugs and digested overnight at 37 °C. Supernatants were removed and three rounds of active extraction were performed using 20 $\mu$L of 1% formic acid, and two times 20 $\mu$L of 30% acetonitrile, 1% formic acid. All supernatants were pooled and concentrated to 10 $\mu$L using a SpeedVac, and 1-2 $\mu$L were analyzed by LC-MS/MS using the same set-up as described in section 3.2.3, but 30 min gradients. Search parameters included 15 ppm tolerance for MS and 0.10 Da for MS/MS, variable methionine oxidation and propionamide cysteine, and a maximum of one missed cleavage.

#### Spectrum to peptide matching and identification

The protocol was written and carried out by **Dr. Ulrich Eckhard**.

Peptides were identified at a 1% false discovery rate from the UniProt *Escherichia coli* K12 database (November 2013) using two search engines, Mascot v2.4.1 (Matrix Science, London, UK) and X! Tandem [84], in conjunction with Peptide Prophet [85] as implemented in the Trans Proteomic Pipeline v.4.6. Search parameters included a mass tolerance of 15 ppm for the parental ions and 0.1 Da for fragment ions, and allowed up to two missed cleavages. The following fixed peptide modifications were used: carbamidomethylation of cysteine residues (+57.02 Da) and dimethylation of lysine $\epsilon$-amines (+28.03 Da). N-terminal dimethylation of library carryover peptides (+28.03 Da), methionine oxidation (+15.99 Da), and thioacylation of protease-generated neo N-termini (+88.00 Da) were set as variable modifications. Peptide lists from both search engines were combined within the Trans Proteomic Pipeline for further analysis. A web-based bioinformatics tool [87] was used to reconstruct the non-prime side of each identified unique cleavage site [52, 81]. Obtained cleavage sites were aligned along the scissile peptide bond and visualized as heat maps in GnuPlot (www.gnuplot.info) and iceLogos [88].

### 3.2.4 Quenched fluorometric assays

Quenched fluorometric assays were performed by **Iain Wallace** (University of Waterloo) and **Dr.Ulrich Eckhard** (University of British Columbia).

Stock solutions (1.0 mM) of synthetic quenched fluorescence (QF) peptide substrates (GenScript Inc.) were dissolved in DMSO and working stocks (100 $\mu$M) were prepared

using the molar extinction coefficient of the conjugated quencher, (2,4)-dinitrophenyl, of 6.985 $cm^{-1}mM^{-1}$ at 400 nm. Purified recombinant FliA(H)-HVR region protease was incubated at a final concentration of 0.5 $\mu$M with 10 $\mu$M QF-peptide substrate in 100 $\mu$L of 50 mM Hepes, 150 mM NaCl, 10 mM $CaCl^2$, pH 7.5, in presence of HALT protease inhibitor cocktail (Life Technologies) plus/minus 20 mM EDTA at 37°C. The inhibitor cocktail was present to inhibit activity for contaminating proteases. Three different quenched fluorescent peptides (PLG - L, Mca-Pro-Leu-Gly-Leu-Dpa-Ala-Arg; ALG - L, Mca-Ala-Leu-Gly-Leu-Dpa-Ala-Arg; and PLG-V, Mca-Pro-Leu-Gly-Val-Dpa-Ala-Arg) were tested using a multi-wavelength fluorescence scanner (POLARstar OPTIMA, BMG Labtechnologies). The excitation and emission wavelengths were 320 and 405 nm, respectively. Fluorescence was measured in arbitrary units over 30 minutes at 45 seconds intervals. Values are means of triplicate measurements.

## 3.3 Results

### 3.3.1 Recombinant FliA(H)

As mentioned earlier in the chapter, the recombinant FliA(H) from *C. haemolyticum* was designed to only express the HVR containing the candidate metallopeptidase domain. Details regarding the construct are shown in figure 3.1. The recombinant protein was expressed and purified to high purity. This was confirmed by LC-MS/MS (Figure 3.2). It also confirmed the absence of contaminating proteases.

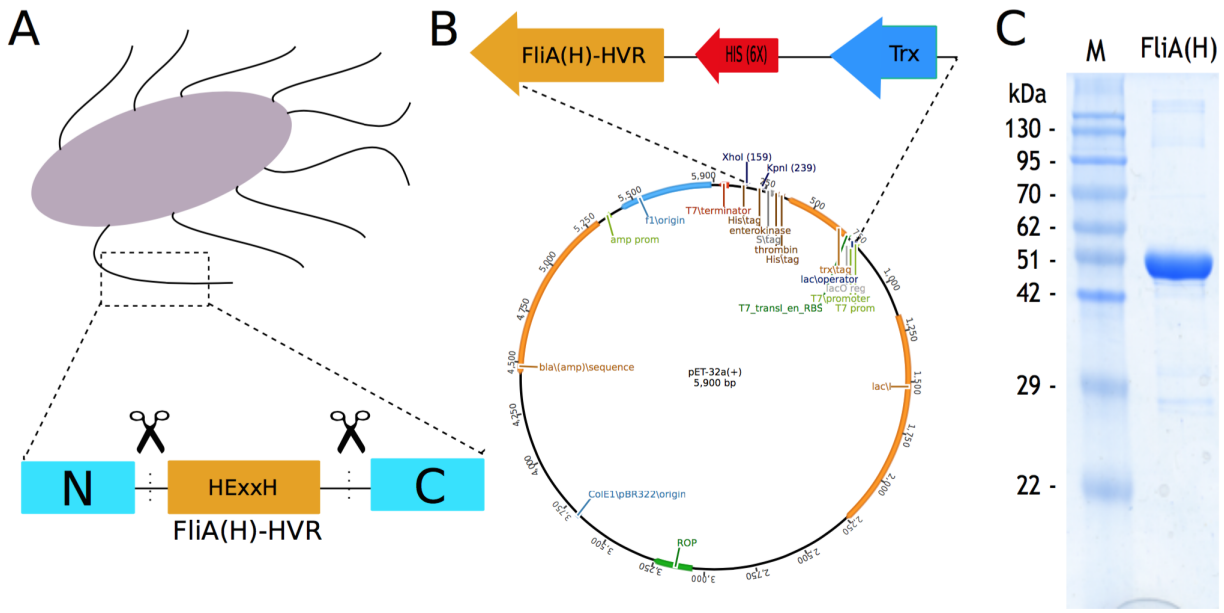


Figure 3.1: **Design and expression of Trx-FliA(H)-HVR**

**(A) The recombinant flagellin-associated protease was designed by trimming the N and C termini from FliA(H) from *Clostridium haemolyticum*. (B) The HVR was cloned into the expression vector pET32a(+) to allow for translational fusions with thioredoxin and His tags. (C) Expression of a soluble purified product was verified using SDS-PAGE.**

**a** **LC-MS/MS analysis of Trx::FliA(H)-HVR by trypsin**

92% sequence coverage

```
  1 MSDKIIHLTD DSFDTDVLKA DGAILVDFWA EWCGPCKMIA PILDEIADEY
 51 QGKLTVAKLN IDQNPGTAPK YGIRGIPTLL LFKNGEVAAT KVGALSKGQL
101 KEFLDANLAG SGSGHMHHHH HHSSGLVPRG SGMKETAAAK FERQHMDSPD
151 LGTDDDDKMK GLKTGWIEKS VENIKTAYGI EPTGANKLKV TISDDGAYGV
201 LASVTPKTGE FELHIDSSDF EKGDGESGNN IHGKLYDDRI IQHEMTHAVM
251 NDALGIDKMN DLHDKNKLWF IEGTAEAMAG ADERVKDIIG NDTQTGIDNT
301 KLSKLATRAD ALLNGVSWNS SDEDYAAGYL MVKYIASKGI DLKAVMKEIK
351 NTGASGLDNK IDLTNLKIDF KNNLENYIKD ISKVHLDWDD DEKDVGSILG
401 SDHGHGDIKA EDVVKGTTPE KEQPLDKFKI IWPDDNSDNT TGKIQLQVGA
451 NEGQSITILE
```

Blue: trx-domain; Green: hexahistidine-tag; Red: enterokinase cleavage site; Black: peptidase domain of C. haemlyticum FliA(H). **Peptides in bold were identified by MS/MS analysis at a false discover rate (FDR) of 1%.**

**b** **Identified proteins at a protein false discovery rate of 1%**

Proteins identified with >10 peptide spectrum matches are shown.

| UniProt | Gene | protein description | tot indep spectra | percent share of spectrum ids |
|---|---|---|---|---|
| Q8RR94 (mod) | Trx::FliA(H)-HVR | Trx::FliA(H)-HVR | **11630** | **95.1** |
| P17169 | GLMS_ECOLI | Glutamine-fructose-6-phosphate transaminase | **242** | **2.0** |
| P0A6Y8 | DNAK_ECOLI | Chaperone protein DnaK | **131** | **1.1** |
| P0A9K9 | SLYD_ECOLI | FKBP-type peptidyl-prolyl cis-trans isomerase | **105** | **0.9** |
| P0A6T5 | GCH1_ECOLI | GTP cyclohydrolase 1 | **32** | **0.3** |
| P77398 | ARNA_ECOLI | Bifunctional polymyxin resistance protein ArnA | **21** | **0.2** |
| P0ACJ8 | CRP_ECOLI | cAMP-activated global transcriptional regulator CRP | **22** | **0.2** |
| P0A7M2 | RL28_ECOLI | 50S ribosomal protein L28 | **17** | **0.1** |
| P76270 | MSRC_ECOLI | Free methionine-R-sulfoxide reductase | **12** | **0.1** |
| P64588 | YQJI_ECOLI | Transcriptional regulator YqjI | **11** | **0.1** |
| P0AFD1 | NUOE_ECOLI | NADH-quinone oxidoreductase subunit E | **12** | **0.1** |

Figure 3.2: **LC-MS/MS**

**(a) LC-MS/MS results for the recombinant FliA(H)-HVR protein, indicating 92% sequence coverage, and (b) illustrating together with our in-gel digests the high sample purity.**

### 3.3.2 PICS peptidase specificity assays

Protease activity of the gluzincin metalloprotease domain containing recombinant protein Trx-FliA(H)-HVR was confirmed using the PICS peptide cleavage assay. The recombinant protein cleaved 391 peptides in the trypsin-generated *Escherichia coli* K-12 library and 498 peptides in the GluC generated library. Two different proteome libraries were used to exclude bias. The peptides cleaved in both libraries were analyzed to assess the amino acid preference for prime and non-prime subsites flanking the hydrolyzed scissile peptide bond i.e., P1-P1'. The cleaved sequences were aligned at the scissile peptide bond and used to generate a logo (Icelogo) representing the amino acid abundance at each position. This data is also represented as heat maps (Figure 3.3). The recombinant protein was shown to be a protease with substrate specificity similar to that of mammalian MMPs i.e. a preference for a leucine residue in the P1' position and other hydrophobic amino acids in the P2' and P3' positions. Though this differs from the predicted homology with clostridial collagenases, it does agree with the structural model that does not contain the 'double-glycine' motif upstream of the HExxH motif. This motif is a key determinant of substrate specificity at the P1' position for clostridial collagenases [50, 48]. Furthermore, it is also in concert with deep pocket S1' subsite usually found in MMPs (Figure 3.5).

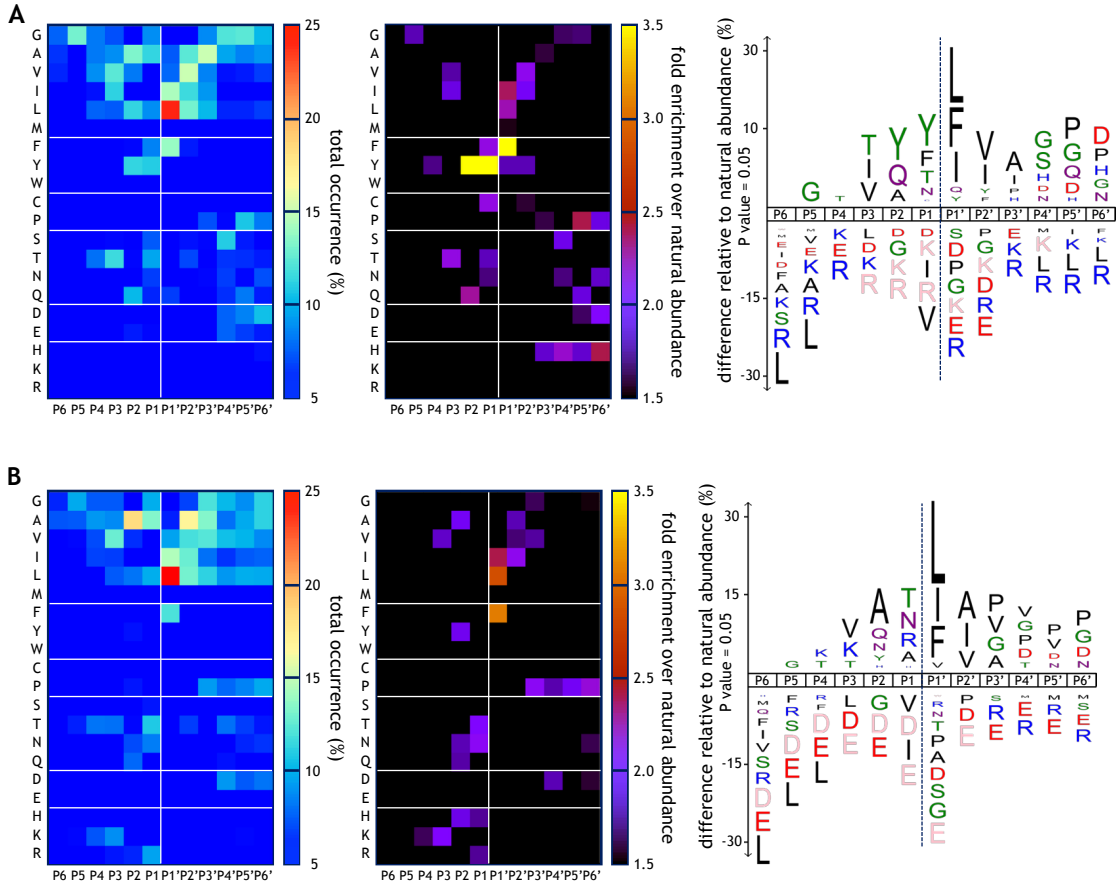


Figure 3.3: **PICS assay with recombinant Trx-FliA(H)-HVR.**

**(A) PICS assay carried out with a trypsin digested *Escherichia coli* K-12 proteome library. (B) PICS assay carried out with a GluC digested *Escherichia coli* K-12 proteome library. The analysis of cleaved biological peptides have been shown at prime and non-prime positions flanking the scissile peptide bond. The total occurrence and fold enrichment for each amino acid at each position have been represented as heat maps. The logo generated shows the preference for leucine at P1' position and other hydrophobic amino acids at P2' and P3' positions in each library.**

### 3.3.3 Quenched fluorometric peptide cleavage assays

The QF peptide cleavage assay was carried out to confirm substrate specificity. It was carried out with three different fluorophore conjugated peptides which are as follows: (i) ALG-L, to test dependency of Leucine in the P1' position as determined by the PICS assay; (ii) PLG-L, a classical MMP substrate and (iii) PLG-V. Fluorescence emission progress curves shown in Figure 3.4 shows that there was an increase in fluorescence with time for the two peptides that had Leucine in the P1' position (ALG-L, PLG-L). Increase in fluorescence means that the recombinant protease was able to cleave the two. When leucine was replaced with valine in the P1' position (PLG-V), there was no change in fluorescence thus showing that it was unable to cleave the peptide. Furthermore, peptide cleavage activity was also knocked out in the presence of EDTA when ALG-L was incubated with Trx-FliA(H)-HVR thus showing that the proteolytic activity is dependent on divalent cations. It is important to note that assays with inhibitors of all other protease classes did not have an effect on activity. Additionally, as shown in the structural model represented in Figure 1.5, Asp206 (subsitutes Glu430 in clostridial collagenase H) along with the backbone oxygens of Asn241, Leu244, and Asp247 residues, forms a calcium-binding site in close proximity to the active site.

To further strengthen this claim, QF assay was performed with purified mutant FliA(H) in which the catalytic residue Glu234 was changed to alanine. There was a loss of activity when mutant FliA(H) was incubated with the QF-peptide ALG-L. This confirms MMP-like activity from recombinant FliA(H).

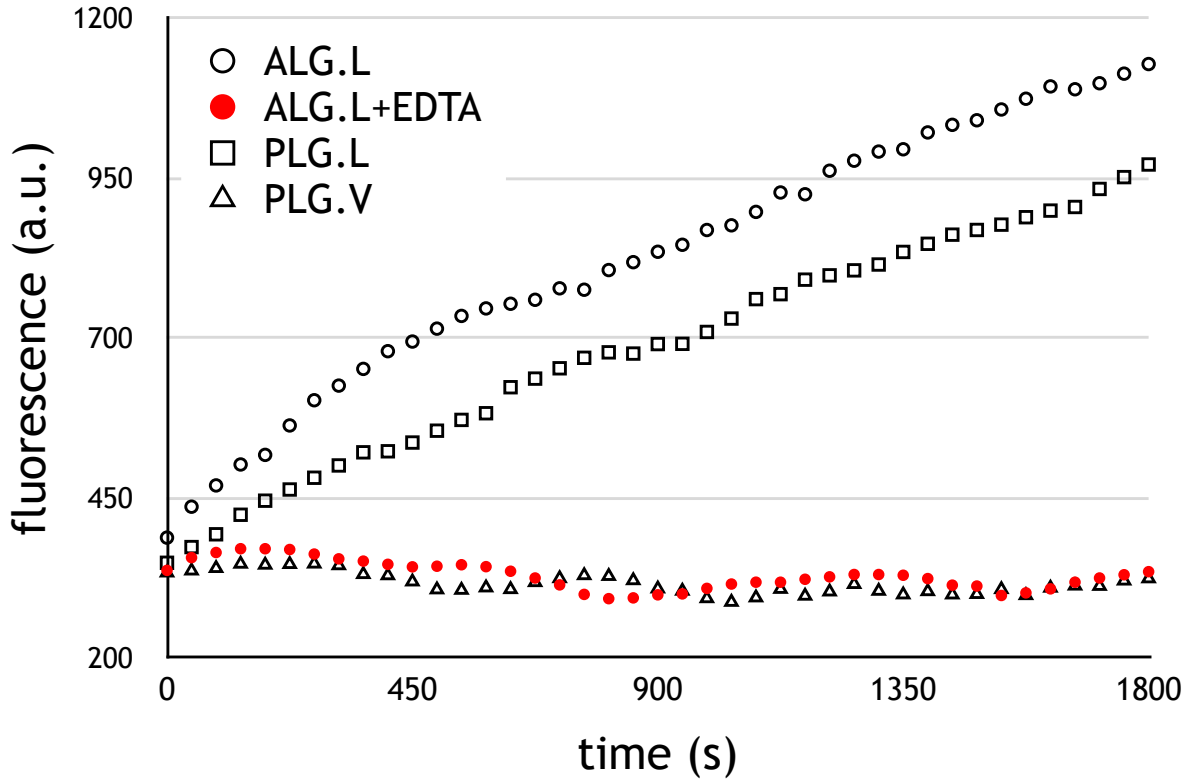


Figure 3.4: **Quenched Fluorometric peptide cleavage assay with Trx-FliA(H)-HVR.**

**QF assays were carried out with 3 different synthetic peptides to verify the results obtained via PICS. Increase in fluorescence over time is shown for peptides with leucine in the P1' position, however the fluorescence stays constant for the peptide with valine in the P1' position. Divalent cation dependent activity was shown when activity was abolished for ALG.L in the presence of EDTA.**

### 3.3.4 Peptide Docking Model

Performed by **Dr. Ulrich Eckhard**.

High resolution peptide docking (figure 3.5) was done using the homology model of the

flagellin-associated protease (as shown in Figure 1.5) and a nonapeptide which mimics the conserved specificity motif as predicted by PICS (AVTYY-LVIA). The peptide fits perfectly within the predicted active site. The model also highlights the deep hydrophobic S2 and S1' pockets in the active site.

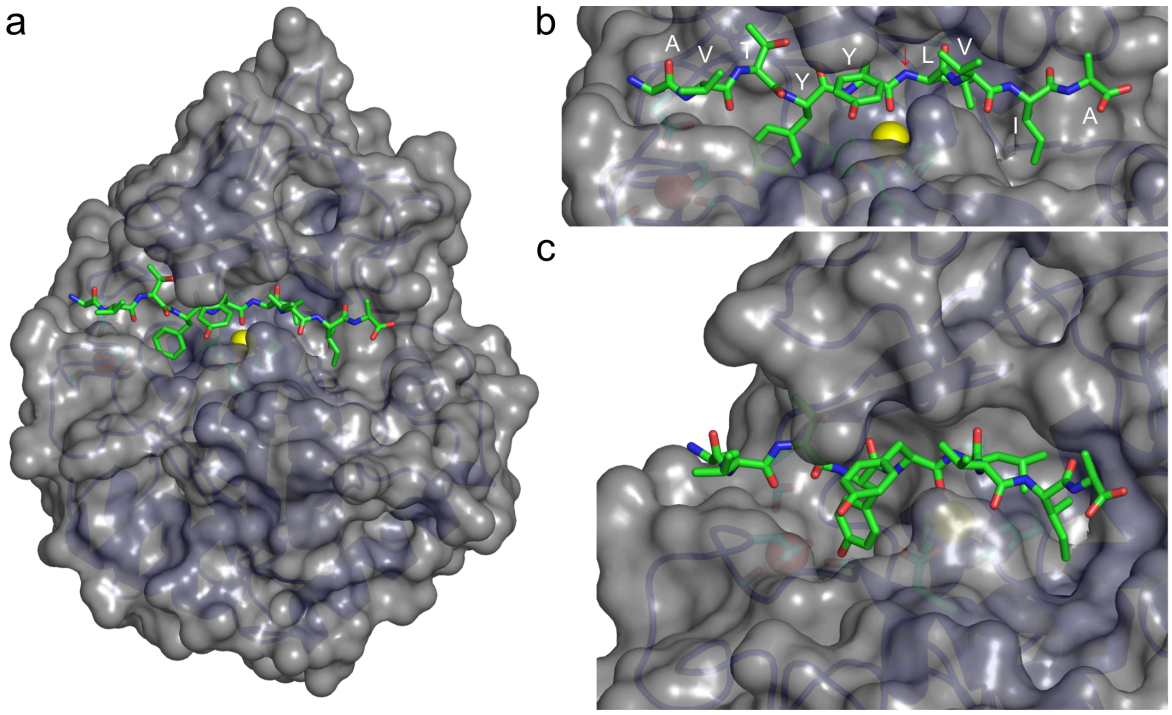


Figure 3.5: **Peptide docking model**

**Peptide docking model using the predicted structure of the protease domain of *Clostridium haemolyticum* FliA(H) and the (AVTYY-LVIA) nonapeptide based on the consensus specificity determined by PICS. (a) Whole view of FliA(H) protease domain-peptide docking model. (b) A zoomed view of the peptide-binding groove with peptide residues labeled in white and the cleavage site indicated by a red arrow. (c) A second zoomed view rotated 25 degrees counterclockwise revealing steric complementarity between the peptide and binding groove (Eckhard et al. 2017, in review)**

## 3.4 Discussion and future directions

The recombinant protein Trx-FliA(H)-HVR was shown to have protease activity by using PICS and quenched fluorometric assays. As outlined in section 3.2.3, the PICS assay was carried out with two *Escherichia coli* K-12 derived peptide libraries, digested with GluC and Trypsin respectively. The recombinant protein cleaved 391 peptides in the trypsin-generated library and 498 in the GluC derived library. The peptides cleaved in both libraries were analyzed to assess the amino acid preference on the prime and non-prime sides of the hydrolyzed scissile peptide bond. The cleaved sequences were used to generate a logo (Icelogo) representing the amino acid abundance at each position. The recombinant protein was shown to be a protease with substrate specificity more similar to that of mammalian MMPs (a preference for a leucine residue in the P1' position and other hydrophobic amino acids in the P2' and P3' positions) than clostridial collagenases. Though this differs from the predicted homology with clostridial collagenases, it does agree with the structural model that has a 'double-glycine' motif upstream of the HExxH motif. This motif is responsible for determining the substrate specificity at the P1' position [50, 48]. The substrate specificity was confirmed by quenched fluorometric assays. As outlined in 3.2.4, the assays were carried out with 3 synthetic peptides. Trx-FliA(H)-HVR with ALG-L and PLG-L showed an increase in fluorescence with time. This was further validated with the no change in fluorescence when PLG-V was used as well as ALG-L with the metal chelator EDTA. This shows that the presence of leucine in the P1' position is important for protease activity. Loss of activity in the presence of EDTA emphasizes the importance of divalent cations for its activity.

Biochemical characterization of the recombinant protein was the first step towards understanding the biological function of the protein. An important future direction currently underway in the laboratory of **Dr. Todd Holyoak** at the University of Waterloo, is obtaining the structure of FliA(H)-HVR protein via X-ray crystallography. Understanding the tertiary and quarternary configuration of this protein would give us more insight into the substrates it may prefer and therefore, elucidate its broader biological function. As outlined in 3.2.1, we have designed constructs for purification and crystallization of FliA(H)-HVR. These constructs will allow the recombinant protein to be expressed as a fusion with a cleavable SUMO tag. Having a SUMO tag assists high expression of recombinant protein by increasing protein solubility. A SUMO affinity tag is also advantageous because it evades the need for a linker peptide since it forms a covalent bond directly with the recombinant protein. The tag is also easily cleaved by SUMO protease which doesn't cleave the target protein [77]. Additional His.Tag®allows for its purification by Ni-NTA affinity chromatography. Additionally the catalytic motif

(HExxH) within the cloned HVR of FliA(H) has been modified by replacing the zinc-binding glutamic acid residue with alanine (E234A). This prevents the possibility of self-proteolysis while maintaining the structural integrity of the protein.

.

# Chapter 4

# Investigating protease activity *in vivo* from *Clostridium haemolyticum*

## 4.1 Hypotheses and Rationale

Following from Section 3.3, the recombinant FliA(H) protein was shown to have catalytic activity. Though the prediction based on structural modelling and sequence homology indicated that the recombinant Trx-FliA(H)-HVR protease is similar to clostridial collagenases, part of the gluzincin clan, experimental evidence based on PICS and QF assays showed that the substrate specificity is more similar to that of human MMPs. Given that this flagellin variant from an animal pathogen is catalytically active *in vitro* and exhibits MMP-like activity, there are two follow-up questions.

**I. Are proteolytic flagellins components of the polymerized flagella?**

To understand the role of this catalytic variant within the polymerized flagella, we aimed to determine its localization within flagellar filaments and the relative abundance. This was achieved by purifying the flagellar filament proteins by ultracentrifugation directly from *C. haemolyticum*. These proteins were then isolated using SDS-PAGE and sequenced using LC-MS to semi-quantitatively assess the relative abundance of FliA(H).

To visualize the presence of the proteolytic variant within the polymerized flagella, immunogold TEM was carried out with purified flagellar filaments from *C.haemolyticum*

Polyclonal primary antibodies against FliA(H)-HVR were produced in rabbit (Biobasic Inc.) and used to label the proteolytic flagellin. ECL-based western blotting was used for validation of the specificity of the primary antibody for both the recombinant FliA(H)-HVR as well as full-length proteolytic flagellin from isolated flagellar proteins. Anti-rabbit colloidal gold-conjugated secondary antibodies were then used to visualize these variants using TEM.

**II. Does the polymerized flagella exhibit predicted activity?**

To assess whether the polymerized flagella from *Clostridium haemolyticum* exhibits catalytic activity, PICS and quenched fluorometric assays with isolated flagellar filaments were carried out.

Addressing these two questions would confirm the existence of an active flagella-associated protease. Biochemical characterization with the assays mentioned above would highlight the substrate specificity of the protease and thus bring us closer to understanding the biological role of this protease in *Clostridium haemolyticum*.

## 4.2 Methodology

### 4.2.1 Isolation of flagellar proteins from *C. haemolyticum*

The flagellar filaments from *Clostridium haemolyticum* was isolated and purified by **Dr. John W. Austin** at Health Canada.

Cultures of *Clostridium haemolyticum* were grown overnight, harvested and re-suspended in 1/20th of it's original volume (0.1 M Tris-HCl pH 7.0). Flagella was removed from whole cells by mechanical shearing using a Teflon tissue homogenizer. The flagellar proteins were then separated from the homogenate by two low-speed centrifugations (5,000 g; 15 min) followed by a round of ultracentrifugation at 130,000 g for 1 hr (70 Ti rotor, Beckman Coulter, Mississauga, Ontario, Canada). The pellets were washed once in ultrapure water and followed by a final centrifugation at 130,000 g for 1 hr. The pellets were resuspended in 50% glycerol and stored at -20°C [89]. Purification was confirmed by SDS-PAGE and quantified using by BCA protein assay as outlined in section 2.5.2.

### 4.2.2 Peptide sequencing by LC-MS/MS

SDS-PAGE (2.5.1) using a 4 (stacking) -12% (separating) Bis-Tris Plus (Invitrogen$^{TM}$) was used to separate the purified flagellar proteins based on size. Protein bands seen at approximately 62 kDa and 46 kDa on the gel were excised. These were further processed by in-gel trypsin digestion and sequenced using mass spectrometry, LC-MS/MS (Advanced Analysis Centre, University of Guelph, Guelph, Ontario, Canada).

### 4.2.3 Immunogold Transmission Electron Microscopy

Gold-conjugated secondary antibodies were supplied by **Dr. Simon Chuong** at the University of Waterloo.

#### Preparation of antibodies and protein samples

The concentration of antibodies and isolated flagellar proteins used for immunogold TEM was optimized using enhanced chemiluminescent (ECL)-western blotting with purified recombinant Trx-FliA(H)-HVR protein, isolated flagellar proteins and primary anti-FliA(H) antibody produced in rabbit (Biobasic Inc.). The secondary antibody used was an HRP-conjugated anti-rabbit IgG produced in goat. The detection of protein occurs via a chemiluminescence reaction, in which horseradish peroxidase (HRP) catalyzes the oxidation of luminol (ECL substrate). This reaction causes the substrate to decay and as a result, emit light proportional to the amount of protein present.

#### Western Blot

Varying concentrations (0.5 ng, 1 ng, 10 ng, 100 ng, 250 ng) of purified recombinant protein were separated by SDS-PAGE using a 10/4% polyacrylamide gel. The gel was then transferred onto a 0.2 $\mu$m PVDF Blotting membrane (Amersham$^{TM}$Hybond$^{TM}$) using a Trans-Blot®SD Semi-Dry Electrophoretic Transfer cell (Bio-Rad). The transfer was run at 20V for 25 minutes. Equal loading of protein was determined by ponceau staining. The membrane was then blocked with 5% skim milk (Oxoid®) in 1 $\times$ TBS-T for 1 hour at room temperature. The membrane was incubated in primary Anti-FliA(H) antibody (Biobasic Inc.) overnight at 4°C with gentle shaking. The membrane was then washed three times, each 10 minutes with 2% skim milk in 1 X TBS-T. The membrane was then incubated with HRP-conjugated anti-rabbit IgG secondary antibody diluted 1:100,000 in 1 X TBS-T with 2% skim milk for 2 hours at room temperature with gentle shaking. The membrane was washed and then incubated with ECL for 5 minutes. Excess ECL was removed and the blot was analyzed using a gel-imager.

**Immunohistrochemistry and TEM**

50 ng of isolated flagella from *Clostridium haemolyticum*was plated on the formvar-coated side of a nickel grid (Ted Pella, inc. Lot# 221014; 200 mesh, Ni; Prod No. 01800N). The protein was allowed to dry for 15 minutes after which it was blotted dry using edges of a Whatman$^{TM}$filter paper. The grid was then incubated in 100 $\mu$L of blocking buffer for 2 hours before incubating it with 100 $\mu$L of anti-FliA(H) primary antibody diluted 1:100 from stock concentration of 2mg/ml in wash buffer. The primary antibody incubation was carried out overnight in a wet chamber at 4°C after which the grid was washed three times with wash buffer. For immunogold labelling, grids were incubated in anti-Rabbit IgG-Gold secondary antibody produced in goat, diluted 1:100 in wash buffer, for 1 hour. The grid was then washed 3 times, once with wash buffer, 1 X PBS (Fisher BioReagents$^{TM}$) and deionized water each.The grids were then negatively stained with 1% w/v ammonium molybdate. TEM was carried out using a Philips CM10 microscope running at 60 kV at the University of Waterloo.

### 4.2.4 PICS Assay with purified flagellar filaments

Please refer to section 3.2.3 for methodology.

### 4.2.5 Quenched fluorometric assays

Please refer to section 3.2.4 for methodology.

## 4.3 Results

### 4.3.1 LC-MS/MS

Flagellar proteins were isolated from *Clostridium haemolyticum*, separated by SDS-PAGE and analyzed by LC-MS/MS. As shown in figures 4.1 and 4.2, the proteolytic flagellin FliA(H), forms the second dominant component of the polymerized flagella. The proteolytic flagellin was identified with 79% coverage and contains the HExxH motif. The most dominant component was identified as a non-proteolytic "structural flagellin" (NCBI accession WP_039229459) with 82% sequence coverage.

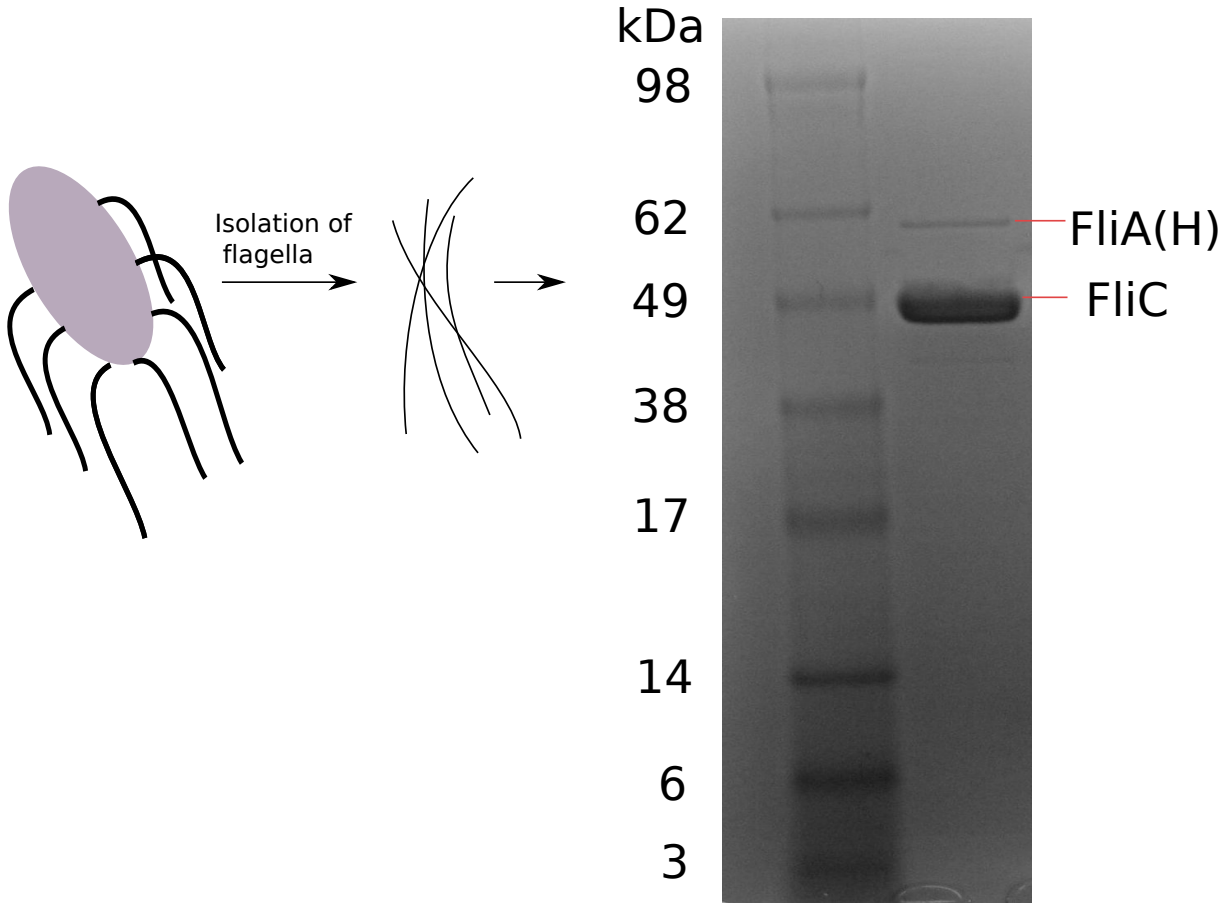


Figure 4.1: **Isolation of flagellar filaments from *Clostridium haemolyticum***

**Flagellar filaments were isolated from *Clostridium haemolyticum* and separated using SDS-PAGE. The protein bands seen on the gel were further analyzed using LC-MS/MS. The primary component was identified to be a non-proteolytic "structural" flagellin (NCBI accession WP_039229459) and the second most dominant component was identified to be FliA(H).**

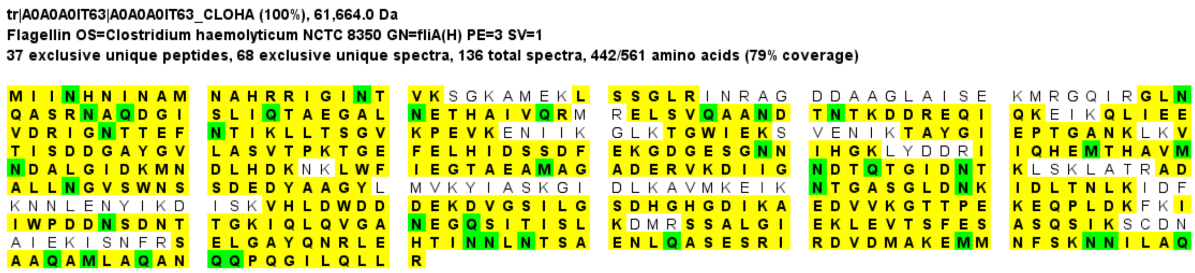


Figure 4.2: **Sequence coverage of the *Clostridium haemolyticum* FliA(H) obtained through mass spectrometry analysis**

### 4.3.2 Immunogold TEM

To investigate the localization of proteolytic flagellins on assembled flagellar filaments, immunogold TEM was performed on purified filaments. Polyclonal anti-FliA(H) primary antibodies were raised and purified against the HVR of FliA(H). The specificity of these antibodies was verified using ECL-based western blotting for both the recombinant Trx-FliA(H)-HVR protein and the purified flagellar filaments (Figure 4.3). Anti-FliA(H) primary antibodies along with gold-conjugated anti-rabbit antibodies were used to label the filaments. Immuno-labelled filaments were visualized using TEM imaging. This showed that the proteolytic flagellins localize uniformly along the length of both filament bundles and individual filaments (figure 4.4).

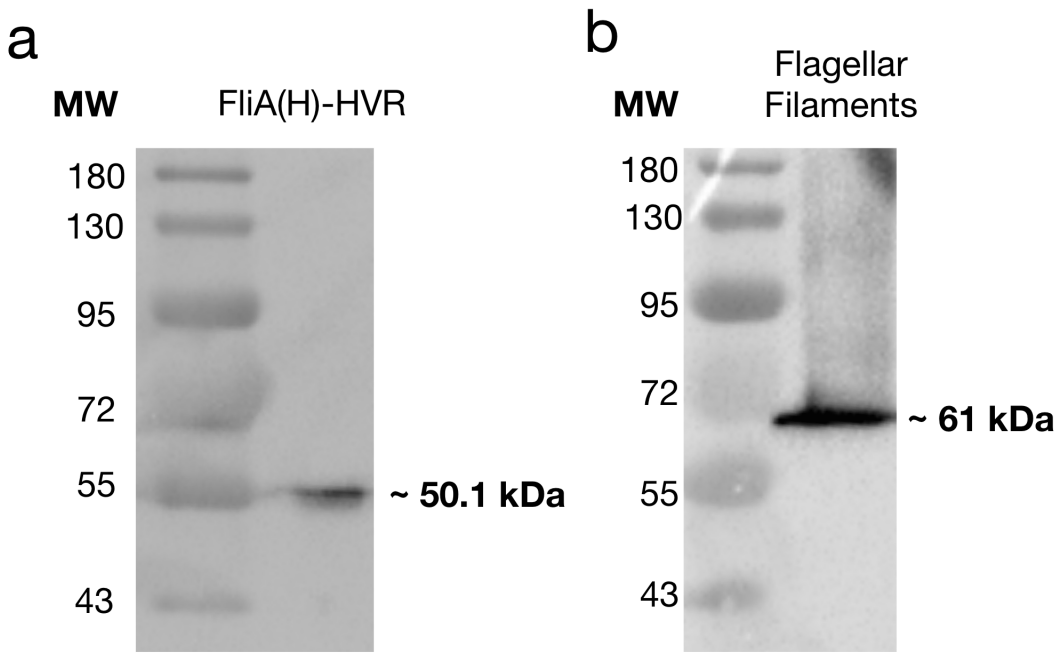


Figure 4.3: **ECL-based western blot for verification of anti-FliA(H) primary antibodies**

**Primary antibodies were raised against the HVR of FliA(H) in rabbit. (A) Detection of recombinant FliA(H)-HVR at 50.1 kDa is shown. (B) Detection of the flagellin variant at 61 kDa verifies the specificity of anti-FliA(H) primary antibodies.**

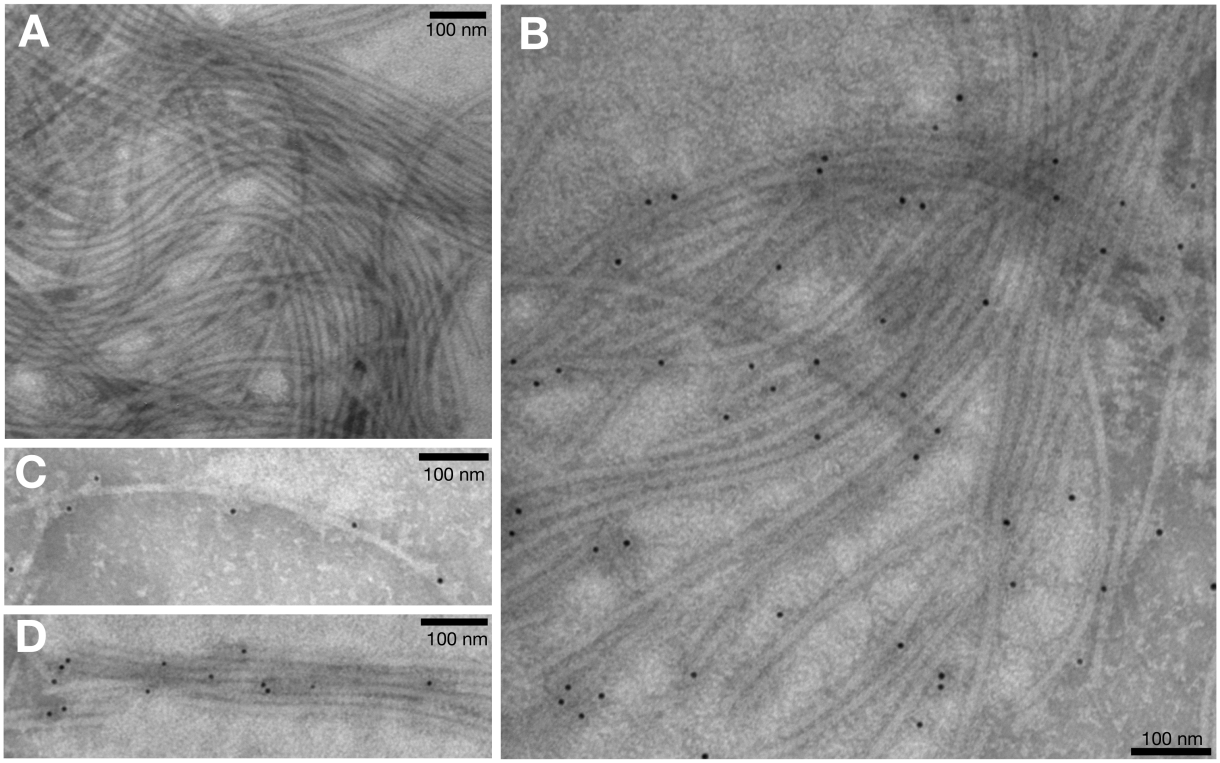


Figure 4.4: **Immunogold TEM with flagellar filaments**

**(A) TEM with isolated flagellar filaments from *Clostridium haemolyticum* (B)(C)(D) Transmission electron micrograph shows gold-labelled FliA(H) localizing along the flagellar filaments in *Clostridium haemolyticum*.**

### 4.3.3 PICS and QF peptide cleavage assays with purified flagellar filaments

To investigate whether native flagella from *Clostridium haemolyticum* exhibited proteolytic activity, PICS and QF peptide cleavage assays were carried out with purified flagellar filaments. PICS assay revealed that these filaments do exhibit peptidase activity as they cleaved 269 peptides in the trypsin digested *Escherichia coli* K-12 proteome library with specificity similar to that of recombinant FliA(H) (Figure 4.5). QF assays also re-

vealed that the filaments cleave the peptide ALG-L. As shown in Figure 4.6, the activity was abolished in all controls. Furthermore, when the filaments were ultracentrifuged, the supernatant only had residual activity (12.8%) thus confirming that the protease activity was associated with flagella as opposed to other proteases.

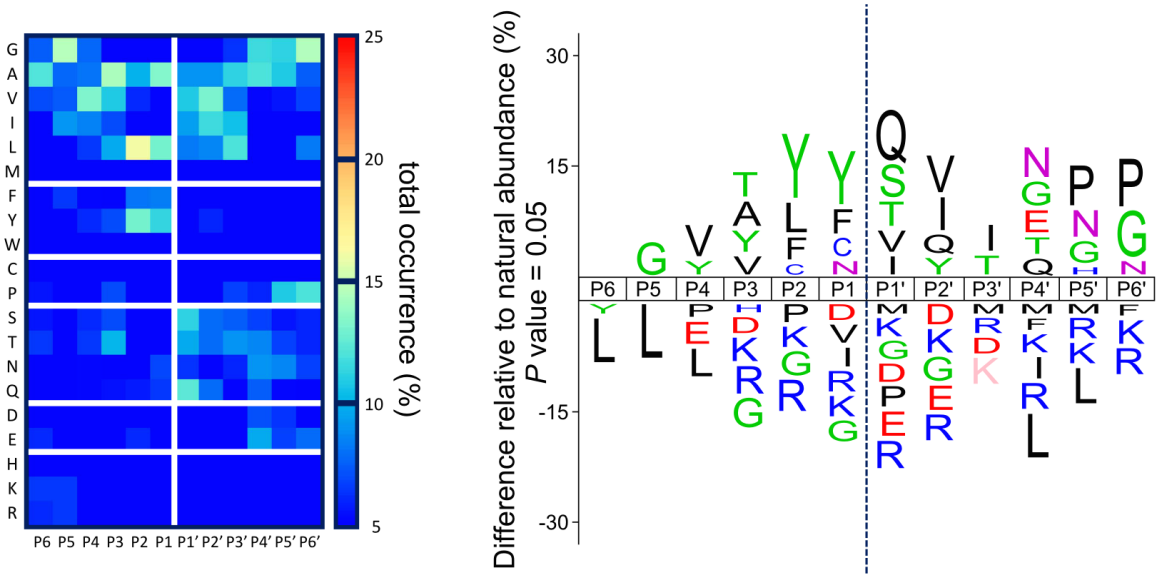


Figure 4.5: **PICS assay with isolated flagellar filaments**

**PICS assay was performed with isolated flagellar filaments from *Clostridium haemolyticum* to profile the protease *in vivo*. Cleaved peptides from the trypsin digested *Escherichia coli* K-12 proteome library were analyzed. The total occurrence for each amino acid at both prime and non prime positions flanking the scissile peptide bond are represented as heat maps. The Icelogo shows substrate specificity similar to that of MMPs i.e. preference for hydrophobic amino acids at the P1' position.**

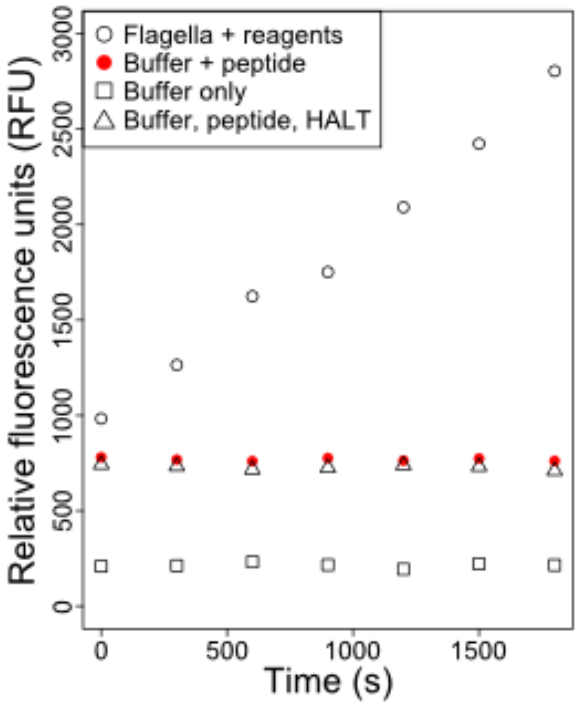


Figure 4.6: **QF assay with isolated flagellar filaments**

**The plot shows an increase in fluorescence over time with *Clostridium haemolyticum* flagella. The activity is abolished in the two negative controls in the absence of flagella. The activity is also abolished in the presence of Halt Protease Inhibitor cocktail.**

## 4.4 Discussion and future directions

Analysis of flagellar proteins from *Clostridium haemolyticum* shows that the catalytically active flagellin variant is the second-most dominant component of the polymerized flagella. Considering that *Clostridium haemolyticum* was grown in normal conditions i.e. not in the presence of the hypothesized substrate, this variant seems to be expressed constitutively. This is still speculative considering that further experimentation would be into the regulation of this protein is required.

Immunogold TEM with isolated flagellar proteins gave us additional evidence for the presence of the variant. The arrangement of 10 nm colloidal gold particles on the transmission electron micrograph indicate that the flagellin-associated protease localizes along the surface of the flagella. To obtain more precise evidence regarding how this protease is arranged within the polymerized flagella, immunogold TEM with fixed whole *Clostridium haemolyticum* cells would be ideal. This is one of the future directions this project will take. In summary, the both, peptide sequencing and immunogold TEM, give us preliminary evidence about the presence of FliA(H) in *Clostridium haemolyticum*.

Evidence for catalytic activity from polymerized flagella *in vivo* was obtained by repeating the PICS and quenched fluorometric assays with isolated flagellar proteins. The PICS assay was repeated with a peptide library derived from trypsin digested K-12 *Escherichia coli* proteome. The purified filaments cleaved 269 peptides with a specificity similar to that of the Trx-FliA(H)-HVR recombinant protein i.e. the protease profile was similar to that of zinc-dependent MMPs. This is because the profile showed a preference for hydrophobic amino acids at the P1' position. Quenched fluorometric assays with the synthetic peptide, ALG-L showed an increase in fluorescence over time whereas there was no apparent activity in the negative controls. An additional control was used to ensure that the activity was not from contaminating proteases but solely from the filaments. This was done by ultracentrifuging the sample containing the filaments and comparing the protease activity from the supernatant compared to the original. It was found that the supernatant resulted in protease activity 12.8% of the original thus confirming flagellar-protease specific activity as well as the absence of contaminating proteases.

Chapter 4 summarizes the evidence that the proteolytic flagellin FliA(H) forms a component of the native polymerized flagella from the animal pathogen *C.haemolyticum* as well as exhibits activity similar to that of the Trx-FliA(H)-HVR recombinant protein and therefore MMPs. This brings us closer to understanding the biological function of this acquired domain by flagellins of various bacterial species.

Two primary future directions underway for this project explore the questions asked in

4.1 more deeply. The first involves the analysis of immunogold labelled FliA(H) within the polymerized flagella of fixed *Clostridium haemolyticum* cells via TEM. The second involves exploration of loss-of-function phenotypes in *Clostridium haemolyticum*. ClosTron mutagenesis [90, 91] will be used to make *in vivo* mutants in which the glutamic acid residue within the catalytic motif HExxH (in the HVR of FliA(H)) will be replaced with alanine to result in loss of zinc-dependent MMP-like proteolytic activity. Repeating PICS and QF assays with this mutant will provide strong validation for our discovery.

# Chapter 5

# Additional insights into the biological roles of proteolytic flagellins

## 5.1 Summary

This thesis outlines the bioinformatics-driven discovery of the first naturally occurring proteolytic flagella which was experimentally validated using FliA(H) from the animal pathogen, *Clostridium haemolyticum*.

As outlined in chapter 3, the HVR from FliA(H), without the N and C termini was cloned into an expression vector to investigate whether the HExxH containing HVR produces a functional enzyme. This was done via the expression and purification of the recombinant Trx-FliA(H)-HVR protein. Subsequent proteolytic profiling of this putative flagellin protease with PICS revealed it cleaves peptides with a specificity similar to that of the gluzincin family of zinc-dependent MMPs. This was confirmed by QF assays with fluorophore-conjugated synthetic peptides. Peptides that had the hydrophobic amino acid, leucine, in the P1' position showed an increase in fluorescence with time. This activity was knocked out when leucine was replaced with valine. QF assays with this flagellin protease also demonstrated calcium dependent activity that was knocked out the the presence of the metal chelator EDTA. In summary, *in vitro* experiments paved way for exploring biological activity *in vivo*. X-ray crystallography of purified Trx-FliA(H)-HVR to obtain a 3D structure of this protein is an important future direction that is

currently underway. Studying the 3D protein structure relative to other protein structures in the database may give us more insight into the biochemistry and therefore the biological function of this flagellin-associated protease.

Chapter 4 outlines the exploration of the proteolytic flagellin *in vivo*. Flagellar proteins were isolated from *Clostridium haemolyticum* as outlined in chapter 4. Separation by SDS-PAGE and subsequent peptide sequencing of flagellar proteins revealed that the proteolytic flagellin FliA(H) formed the second most dominant component of the polymerized flagella. The structural flagellin FliC, on the other hand, was the most dominant component. As revealed by immunogold TEM, the proteolytic flagellin localizes along the filament of native polymerized flagella. Proteolytic profiling with PICS and QF assays provided evidence for MMP-like activity from purified native flagellar filaments. Further investigation into how this proteolytic flagellin is assembled within the polymerized flagella can be obtained via immunogold TEM of whole cells with fixed flagella. Furthermore, engineering mutant strains that have a modified HExxH motif within FliA(H) may allow exploration of loss of function phenotypes *in vivo*. These phenotypes can be investigated via peptidase profiling of isolated filaments from both wild-type and mutant strains by using the PICS assay. Further investigation of phenotypes associated with the candidate flagellin protease can be done via experimentation relative to substrates that mimic the natural biological ECM. This may allow for the generation of more streamlined hypotheses regarding the broader biological function of this organism's proteolytic flagella.

## 5.2 Insights into the molecular evolution of proteolytic flagellin HVR

Bioinformatic analysis reveals that amongst previously uncharacterized domains in the surface exposed flagellin HVRs, a metallopeptidase domain (M9 Peptidase) was found to be conserved in 74 phylogenetically diverse bacterial species (Tables 5.2). These species are mostly distributed amongst Proteobacteria and Firmicutes. The highly scattered phylogenomic distribution of the proteolytic domain within bacterial species (Figure 5.1) suggests a high degree of lateral gene transfer between species and phyla.

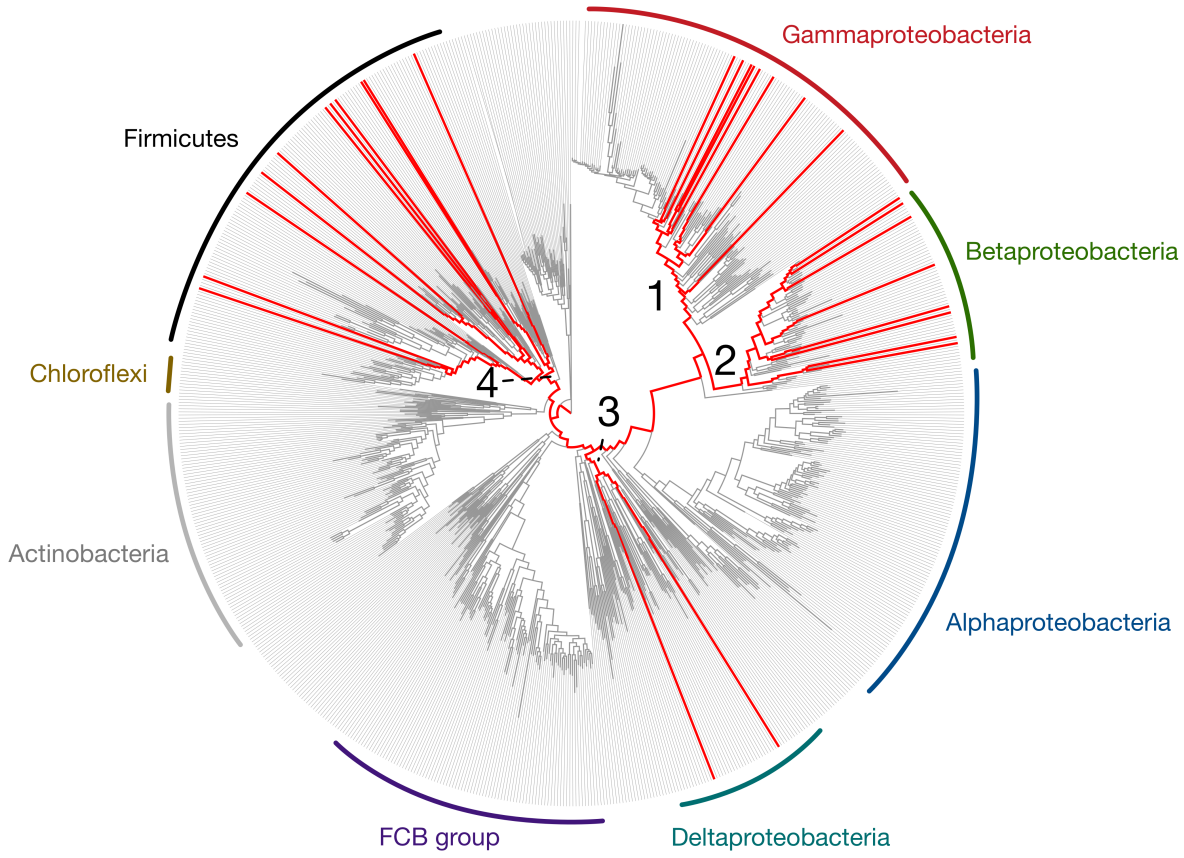


Figure 5.1: **Phylogenomic distribution of proteolytic flagellins on the bacterial tree of life**

**Genera containing proteolytic flagellin genes are indicated by red lines. The presence/absence pattern is highly scattered indicative of extensive lateral gene transfer. Proteolytic flagellins were identified in four major lineages (numbered 1-4). The phylogenetic tree was derived from Hug et al. 2017 (Eckhard et al. 2017, in review)**

**All flagellin domain architectures analyzed contain Flagellin_N and Flagellin_C domains**
**Note: Flagellin architectures with putative enzymatic domains in their hypervariable region are indicated in bold**

| Sequences | Taxonomy | Taxid | Accessions | Domain Architecture | Predicted Protease |
|---|---|---|---|---|---|
| 21667 | Bacteria | 2 | pfam00669~pfam00700 | Flagellin_N~Flagellin_C | |
| 3286 | cellular organisms | 131567 | pfam00669~pfam07196~pfam00700 | Flagellin_N~Flagellin_IN~Flagellin_C | |
| 715 | Proteobacteria | 1224 | pfam00669~pfam08884~pfam00700 | Flagellin_N~Flagellin_D3~Flagellin_C | |
| 417 | Proteobacteria | 1224 | pfam00669~pfam12445~pfam00700 | Flagellin_N~FliC~Flagellin_C | |
| 155 | Rhizobiales | 356 | pfam00669~pfam07482~pfam00700 | Flagellin_N~DUF1522~Flagellin_C | |
| 67 | Burkholderiaceae | 119060 | pfam00669~pfam12613~pfam00700 | Flagellin_N~FliC_SP~Flagellin_C | |
| **35** | **Bacteria** | **2** | **pfam00669~pfam01752~pfam00700** | **Flagellin_N~Peptidase_M9~Flagellin_C** | **Y** |
| 24 | Bacteria | 2 | pfam00669~pfam11863~pfam00700 | Flagellin_N~DUF3383~Flagellin_C | |
| 21 | Bacteria | 2 | pfam00669~cl00057~pfam00700 | Flagellin_N~vWFA~Flagellin_C | |
| **17** | **Acinetobacter sp. CAG:196** | **1262690** | **pfam00669~pfam05342~pfam00700** | **Flagellin_N~Peptidase_M26_N~Flagellin_C** | **Y** |
| **12** | **Bacteria** | **2** | **pfam00669~cl14813~pfam00700** | **Flagellin_N~GluZincin~Flagellin_C** | **Y** |
| **10** | **Desulfovibrionales** | **213115** | **pfam00669~cl21531~pfam00700** | **Flagellin_N~Sialidase~Flagellin_C** | |
| 10 | Pseudomonas | 286 | pfam00669~pfam04004~pfam00700 | Flagellin_N~Leo1~Flagellin_C | |
| 9 | Bacteria | 2 | pfam00669~pfam07196~cl00160~pfam00700 | Flagellin_N~Flagellin_IN~LbetaH~Flagellin_C | |
| 9 | Proteobacteria | 1224 | pfam00669~cl14106~pfam00700 | Flagellin_N~Rifin_STEVOR~Flagellin_C | |
| 7 | Bacteria | 2 | pfam00669~cl21456~pfam00700 | Flagellin_N~Periplasmic_Binding_Protein_Type_2~Flagellin_C | |
| 7 | delta/epsilon subdivisions | 68525 | pfam00669~cl00160~pfam00700 | Flagellin_N~LbetaH~Flagellin_C | |
| 5 | Bacteria | 2 | pfam00669~cl21454~pfam00700 | Flagellin_N~NADB_Rossmann~Flagellin_C | |
| 5 | Enterobacteriaceae | 543 | pfam00669~cl12013~pfam00700 | Flagellin_N~BAR~Flagellin_C | |
| 5 | Proteobacteria | 1224 | pfam00669~cl19731~pfam00700 | Flagellin_N~SipA~Flagellin_C | |
| **5** | **Proteobacteria** | **1224** | **pfam00669~cl16912~pfam00700** | **Flagellin_N~MDR~Flagellin_C** | |
| 4 | Proteus mirabilis | 584 | pfam00669~cl02760~pfam00700 | Flagellin_N~NEAT~Flagellin_C | |
| 4 | Bacteria | 2 | pfam00669~pfam12211~pfam00700 | Flagellin_N~LMWSLP_N~Flagellin_C | |
| 4 | Vibrio | 662 | pfam00669~pfam07196~pfam14925~pfam00700 | Flagellin_N~Flagellin_IN~HPHLAWLY~Flagellin_C | |
| **4** | **Methylibium** | **316612** | **pfam00669~cl00261~pfam00700** | **Flagellin_N~PLPDE_III~Flagellin_C** | |
| 4 | Butyrivibrio | 830 | pfam00669~pfam06156~pfam00700 | Flagellin_N~DUF972~Flagellin_C | |
| 4 | Bacteria | 2 | pfam00669~cl15685~pfam00700 | Flagellin_N~Wzt_C-like~Flagellin_C | |
| 3 | Bacteria | 2 | pfam00669~cl07060~pfam00700 | Flagellin_N~NPCBM~Flagellin_C | |
| 3 | Bacteria | 2 | pfam00669~cl21487~pfam00700 | Flagellin_N~OM_channels~Flagellin_C | |
| 3 | Bacteria | 2 | pfam00669~pfam13861~pfam00700 | Flagellin_N~FlgD_tudor~Flagellin_C | |
| 3 | Gammaproteobacteria | 1236 | pfam00669~TIGR02597~pfam00700 | Flagellin_N~TIGR02597~Flagellin_C | |
| **3** | **Roseburia inulinivorans** | **360807** | **pfam00669~cl00342~pfam00700** | **Flagellin_N~Trp-synth-beta_II~Flagellin_C** | |
| 3 | Vibrio | 662 | pfam00669~pfam07196~cl21456~pfam00700 | Flagellin_N~Flagellin_IN~Periplasmic_Binding_Protein_Type_2~Flagellin_C | |
| **3** | **Clostridium tyrobutyricum** | **1519** | **pfam00669~cl21455~pfam00700** | **Flagellin_N~P-loop_NTPase~Flagellin_C** | |
| 3 | Carnobacterium sp. 17-4 | 208596 | pfam00669~cl01110~pfam00700 | Flagellin_N~Sdh5~Flagellin_C | |
| 3 | Bacillales | 1385 | pfam00669~pfam04574~pfam00700 | Flagellin_N~DUF592~Flagellin_C | |
| 3 | Gluconobacter oxydans | 442 | pfam00669~cl11960~pfam00700 | Flagellin_N~Ig~Flagellin_C | |
| 2 | Bacteria | 2 | pfam00669~pfam07559~pfam00700 | Flagellin_N~FlaE~Flagellin_C | |
| **2** | **Bacteria** | **2** | **pfam00669~cl18945~pfam00700** | **Flagellin_N~AAT_I~Flagellin_C** | |
| 2 | Enterobacteriaceae | 543 | pfam00669~cl05878~pfam00700 | Flagellin_N~Trak~Flagellin_C | |
| **2** | **Arcobacter butzleri** | **28197** | **pfam00669~cl18951~pfam00700** | **Flagellin_N~Amidase~Flagellin_C** | |
| 2 | Chitinibacter | 230666 | pfam00669~cl09928~pfam00700 | Flagellin_N~Molybdopterin-Binding~Flagellin_C | |
| **2** | **Bacteria** | **2** | **pfam00669~cl00192~pfam00700** | **Flagellin_N~ribokinase_pfkB_like~Flagellin_C** | |
| **2** | **Azoarcus sp. CIB** | **198107** | **pfam00669~pfam09249~pfam00700** | **Flagellin_N~tRNA_NucTransf2~Flagellin_C** | |
| 2 | delta proteobacterium MLMS-1 | 262489 | pfam00669~pfam07196~cl00160~cl02073~pfam00700 | Flagellin_N~Flagellin_IN~LbetaH~DUF3422~Flagellin_C | |
| 2 | Bacillus alcalophilus | 1445 | pfam00669~pfam11589~pfam00700 | Flagellin_N~DUF3244~Flagellin_C | |
| 2 | Borreliella lusitaniae | 100177 | pfam00669~pfam14796~pfam00700 | Flagellin_N~AP3B1_C~Flagellin_C | |
| 2 | Nitrospira defluvii | 330214 | pfam00669~cl21542~pfam00700 | Flagellin_N~EthD~Flagellin_C | |
| 2 | Firmicutes | 1239 | pfam00669~cl23771~pfam00700 | Flagellin_N~Big_3_4~Flagellin_C | |
| **2** | **Rhodopirellula** | **265488** | **pfam00669~cl10557~pfam00700** | **Flagellin_N~Dak1~Flagellin_C** | |
| **2** | **Gammaproteobacteria** | **1236** | **pfam00669~cl11964~pfam00700** | **Flagellin_N~CYTH-like_Pase~Flagellin_C** | |
| **2** | **Bacillus sp. SG-1** | **161544** | **pfam00669~cl14647~pfam00700** | **Flagellin_N~GH43_62_32_68~Flagellin_C** | |
| **2** | **Bacteria** | **2** | **pfam00669~cl14876~pfam00700** | **Flagellin_N~Zinc_peptidase_like~Flagellin_C** | **Y** |
| 2 | Enterobacter cloacae | 550 | pfam00669~cl22495~pfam00700 | Flagellin_N~Gp23~Flagellin_C | |
| 2 | Cellvibrio sp. BR | 1134474 | pfam00669~cl01077~pfam00700 | Flagellin_N~SIMPL~Flagellin_C | |
| 2 | Idiomarina xiamenensis | 1207041 | pfam00669~cl00278~pfam00700 | Flagellin_N~CCC1_like~Flagellin_C | |
| 2 | Rhodospirillaceae | 41295 | pfam00669~pfam10983~pfam00700 | Flagellin_N~DUF2793~Flagellin_C | |
| 1 | Bdellovibrio bacteriovorus | 959 | pfam00669~cl01193~pfam00700 | Flagellin_N~DUF463~Flagellin_C | |
| 1 | Paenibacillus alginolyticus | 59839 | pfam00669~cl21457~pfam00700 | Flagellin_N~TIM_phosphate_binding~Flagellin_C | |
| 1 | Nematostella vectensis | 45351 | pfam00700~pfam00669 | Flagellin_C~Flagellin_N | |
| 1 | Lactobacillus capillatus | 480931 | pfam00669~cl01389~pfam00700 | Flagellin_N~Phage_sheath_1~Flagellin_C | |
| 1 | Escherichia fergusonii | 564 | pfam00669~cl00262~pfam00700 | Flagellin_N~TroA-like~Flagellin_C | |

Table 5.1: **Flagellin associated domain architectures and their abundance in the NCBI database as detected using the Conserved Domain Database.**

| GI | Ref | Name | Species |
|---|---|---|---|
| 738243862 | WP_036199097.1 | hypothetical protein | Lysinibacillus sinduriensis |
| 916691384 | WP_051298475.1 | hypothetical protein | Marinobacterium litorale |
| 518372292 | WP_019542499.1 | hypothetical protein | Selenomonas bovis |
| 916573195 | WP_051180286.1 | hypothetical protein | Selenomonas ruminantium |
| 504238503 | WP_014425605.1 | putative flagellin | Selenomonas ruminantium |
| 917028657 | WP_051635369.1 | hypothetical protein | Selenomonas sp. ND2010 |
| 916926872 | WP_051533584.1 | hypothetical protein | Anaerovibrio sp. RM50 |
| 917657302 | WP_052211964.1 | hypothetical protein | Anaerovibrio lipolyticus |
| 653148213 | WP_027397410.1 | hypothetical protein | Anaerovibrio lipolyticus |
| 507777340 | EOS35535.1 | hypothetical protein | C808_04738 Lachnospiraceae bacterium M18-1 |
| 551000896 | WP_022745605.1 | hypothetical protein | Dorea sp. 5-2 |
| 507824683 | EOS80891.1 | hypothetical protein | C817_01227 Dorea sp. 5-2 |
| 769144802 | WP_044921500.1 | hypothetical protein | Lachnospiraceae bacterium MA2020 |
| 910018461 | WP_049973154.1 | hypothetical protein | Lachnospiraceae bacterium MC2017 |
| 910018462 | WP_049973155.1 | hypothetical protein | Lachnospiraceae bacterium MC2017 |
| 651905107 | WP_026658295.1 | hypothetical protein | Butyrivibrio sp. AC2005 |
| 697062782 | WP_033152383.1 | hypothetical protein | Pseudobutyrivibrio ruminis |
| 769166148 | WP_044939094.1 | hypothetical protein | Pseudobutyrivibrio sp. LB2011 |
| 651903221 | WP_026657293.1 | hypothetical protein | Butyrivibrio sp. AC2005 |
| 551017073 | WP_022761310.1 | flagellin | Butyrivibrio sp. AD3002 |
| 651387362 | WP_026499388.1 | hypothetical protein | Butyrivibrio sp. WCD2001 |
| 651394250 | WP_026505968.1 | hypothetical protein | Butyrivibrio sp. NC3005 |
| 551028220 | WP_022772304.1 | flagellin | Butyrivibrio sp. AE2015 |
| 916590076 | WP_051197167.1 | hypothetical protein | Butyrivibrio sp. XBB1001 |
| 490739930 | WP_004602238.1 | flagellin/flagellar hook associated protein | Eubacterium cellulosolvens |
| 916920396 | WP_051527108.1 | hypothetical protein | Eubacterium cellulosolvens |
| 517856800 | WP_019027008.1 | hypothetical protein | Colwellia piezophila |
| 766745642 | WP_044831744.1 | hypothetical protein | Thalassomonas actiniarum |
| 766760745 | WP_044838529.1 | hypothetical protein | Thalassomonas viridans |
| 654652764 | WP_028113749.1 | hypothetical protein | Ferrimonas kyonanensis |
| 916608721 | WP_051215812.1 | hypothetical protein | Ferrimonas futtsuensis |
| 497523220 | WP_009837418.1 | flagellar protein | Pseudoalteromonas tunicata |
| 654852782 | WP_028305148.1 | hypothetical protein | Oceanospirillum maris |
| 654847818 | WP_028300311.1 | hypothetical protein | Oceanospirillum beijerinckii |
| 518442882 | WP_019613089.1 | hypothetical protein | Psychromonas ossibalaenae |
| 655481215 | WP_028863050.1 | hypothetical protein | Psychromonas aquimarina |
| 917474576 | WP_052080993.1 | hypothetical protein | Pseudomonas sp. ML96 |
| 917193838 | WP_051800550.1 | hypothetical protein | Pseudomonas oleovorans |
| 1057533868 | WP_068829844.1 | hypothetical protein | Pseudomonas sp. BMS12 |
| 759519778 | WP_043240123.1 | hypothetical protein | Pseudomonas alcaligenes |
| 544803667 | WP_021220822.1 | hypothetical protein | Pseudomonas alcaligenes |
| 919350262 | WP_052807587.1 | hypothetical protein | Pseudomonas sp. FeS53a |
| 758832821 | KIV71406.1 | Flagellin protein FlaB | Pseudomonas sp. FeS53a |
| 495134114 | WP_007858923.1 | flagellin | Acidovorax sp. CF316 |
| 950225299 | WP_057269442.1 | flagellin | Acidovorax sp. Root219 |
| 948088328 | WP_056747048.1 | flagellin | Acidovorax sp. Root568 |
| 947398799 | WP_056064769.1 | flagellin | Acidovorax sp. Root402 |
| 1023918667 | WP_063462804.1 | flagellin | Acidovorax sp. GW101-3H11 |
| 498146157 | WP_010460313.1 | flagellin | Acidovorax radicis |
| 950228284 | WP_057272374.1 | flagellin | Acidovorax sp. Root267 |
| 950175389 | WP_057223076.1 | flagellin | Acidovorax sp. Root275 |
| 493342448 | WP_006299301.1 | flagellin | Hylemonella gracilis |
| 737635795 | WP_035605832.1 | flagellin | Hylemonella gracilis |
| 1054768843 | WP_066532860.1 | flagellin | Comamonas terrigena |
| 759662642 | WP_043380247.1 | flagellin | Comamonas aquatica |
| 835203907 | WP_047395168.1 | hypothetical protein | Chitinibacter sp. ZOR0017 |
| 654999239 | WP_028448441.1 | hypothetical protein | Chitinibacter tainanensis |
| 517221226 | WP_018410044.1 | hypothetical protein | Methyloversatilis thermotolerans |
| 1059956041 | WP_069037803.1 | hypothetical protein | Methyloversatilis sp. RAC08 |
| 1059488308 | AOF81425.1 | hypothetical protein BSY238_565 | Methyloversatilis sp. RAC08 |
| 522143982 | WP_020655191.1 | hypothetical protein | Massilia niastensis |
| 916699479 | WP_051306570.1 | hypothetical protein | Massilia alkalitolerans |
| 947778910 | WP_056440654.1 | hypothetical protein | Massilia sp. Root335 |
| 738284338 | WP_036238512.1 | hypothetical protein | Massilia sp. JS1662 |
| 947464650 | WP_056129008.1 | MULTISPECIES: hypothetical protein | Massilia |
| 500090247 | WP_011766260.1 | flagellin | Azoarcus sp. BH72 |
| 1043078958 | WP_065340783.1 | hypothetical protein | Azoarcus olearius |
| 847184073 | WP_047967967.1 | hypothetical protein | Vogesella sp. EB |
| 938316768 | WP_054620994.1 | hypothetical protein | Rhodocyclaceae bacterium Paddy-1 |
| 737307010 | WP_035289894.1 | hypothetical protein | Clostridium sp. KNHs214 |
| 746179135 | WP_039238750.1 | hypothetical protein | Clostridium novyi |
| 746201658 | WP_039258875.1 | hypothetical protein | Clostridium botulinum |
| 746169334 | WP_039229452.1 | hypothetical protein | Clostridium haemolyticum |
| **19910973** | **BAB87738.1** | **flagellin protein FliA(H)** | **Clostridium haemolyticum** |
| 746189539 | WP_039247767.1 | hypothetical protein | Clostridium novyi |
| 737261571 | WP_035245515.1 | hypothetical protein | Desulfobulbus mediterraneus |
| 492846401 | WP_006000355.1 | flagellin-like | Desulfuromonas acetoxidans |
| 550893434 | WP_022654199.1 | flagellin | Aquaspirillum serpens |
| 1062631149 | WP_069313537.1 | hypothetical protein | Piscirickettsia sp. Y2 |
| 737346898 | WP_035329181.1 | hypothetical protein | Bacillus firmus |
| 766666229 | WP_044826104.1 | hypothetical protein | Clostridium aceticum |
| 738759210 | WP_036652630.1 | hypothetical protein | Paenibacillus wynnii |
| 504121263 | WP_014355249.1 | flagellin | Acetobacterium woodii |
| 503951982 | WP_014185976.1 | flagellin/flagellar hook associated protein | Desulfosporosinus orientis |
| 505207458 | WP_015394560.1 | flagellin | Clostridium saccharoperbutylacetonicum |
| 1011282853 | WP_062200331.1 | hypothetical protein | Bacillaceae bacterium mt8 |

Table 5.2: **Predicted "proteolytic flagellins" identified in the NCBI database.**

**The figure shows the 86 predicted proteolytic flagellins idenified in the NCBI database. All possess a putative gluzincin-family metallopeptidase domain in their hypervariable region.**

Considering that this domain was predicted to have collagenase-like activity, the origin of this domain may have been a *Clostridia* species that natively use collagenases for nutrient acquisition. A collagenase-like domain may have been acquiried by the flagellin HVR via intragenic recombination and then undergone a shift in substrate specificity. As suggested earlier in the thesis, the double-glycine motif is absent from the HVR-associated protease, which is a key determinant of substrate specificity of *Clostridial* collagenases. *Clostridia* species as a source of the original gluzincin proteolytic flagellin gene is further supported by the fact that the proteolytic flagellin family clusters separately from all other gluzincin proteases. After initial acquisition, this domain may have spread to other species by lateral gene transfer (Figure 5.2). Examples of proteolytic flagellin gene insertions are directly evident by comparative analysis of gene clusters containing the proteolytic flagellin gene. For instance, in Comamonas aquatica DA1877, a single proteolytic flagellin gene has been inserted within a conserved gene cluster present in other Comamonadaceae (Figure 5.2). To gain insights into the broader functions of the proteolytic flagellin family, the functions of neighbouring genes were examined. Interestingly, most gene clusters containing a proteolytic flagellin gene also include other flagellar genes, especially within *Clostridium* and several specific lineages (Figure 5.2). These proteolytic flagellins are therefore likely to be co-regulated with, and components of, the flagellar machinery. In other genomes, genes flanking the proteolytic flagellin include other peptidases, membrane proteins, ABC transporters, and in a few cases transposase or transposon-related sequences which may reflect a mobile mechanism of transfer (Figure 5.2).

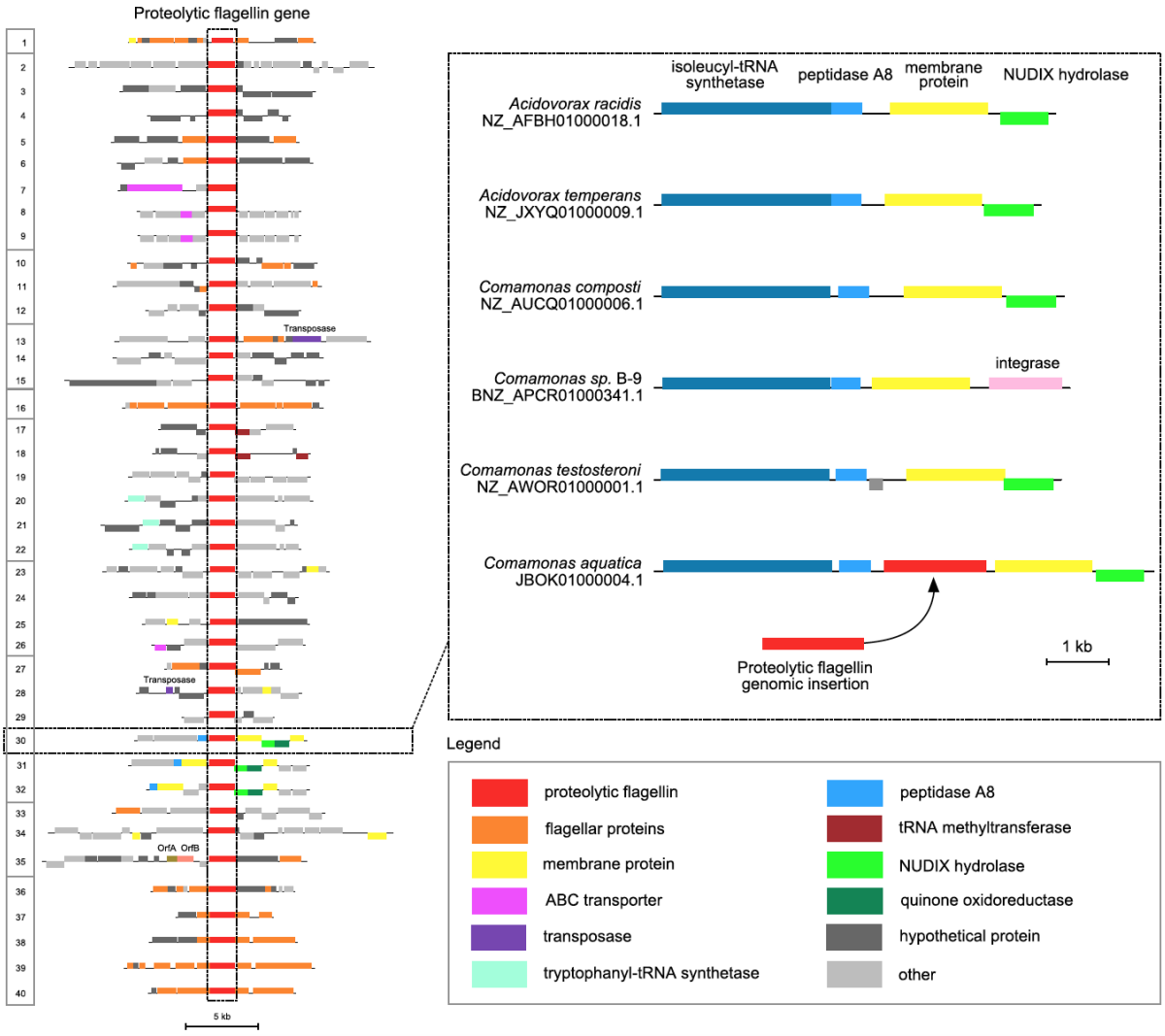


Figure 5.2: **Genomic context of proteolytic flagellin gene**

**Genomic context surrounding the proteolytic flagellin gene in 40 genomes (left) and evidence for a proteolytic flagellin gene insertion into conserved Comamonadaceae gene cluster (right). The 40 gene clusters are ordered into ten clusters based on a phylogeny of the flagellin protease domain (Eckhard et al. 2017, in review)**

## 5.3 Insights into the biological function of proteolytic flagellins

The presence of proteolytic flagella in the animal pathogen, *Clostridium haemolyticum*, has several biological implications. Protease activity from a flagella of an organism burrowing through the ECM may have a much wider range than a comparable secreted protease. Additionally, considering that flagellated bacteria existed before their eukaryotic hosts, it suggests that flagella didn't exist as virulent organelles, but instead evolved to play pivotal roles in various facets of pathogenesis [22]. These roles have been outlined in 1.2.2.To speculate further, the expression and regulation of this protease could be coupled to one of the many signals utilized by bacteria e.g. chemical gradients for chemotaxis that further assists roles such as host colonization. Other examples include, alteration of flagellins for host evasion. The numerous possible roles of this flagellin HVR-associated protease has been highlighted in Figure 5.3. Another important consideration is that these putative proteases are encoded within approximately 5-10% of surface exposed HVR of flagellin monomers and approximately 20,000 of them assemble to form a flagellar filament. This means that each flagellar filament of *Clostridium haemolyticum* may have about one to two thousand active sites. This may be advantageous for an animal pathogen such as this one, because it may be capable of penetrating deeper tissue structures. The modularity of bacterial flagellin has made it an ideal candidate for engineering *in vitro* and has been described in 1.3. Considering that this has become available to synthetic biologists, it seems viable for evolution to favour selection of protease domains within the HVR of bacterial flagellins.

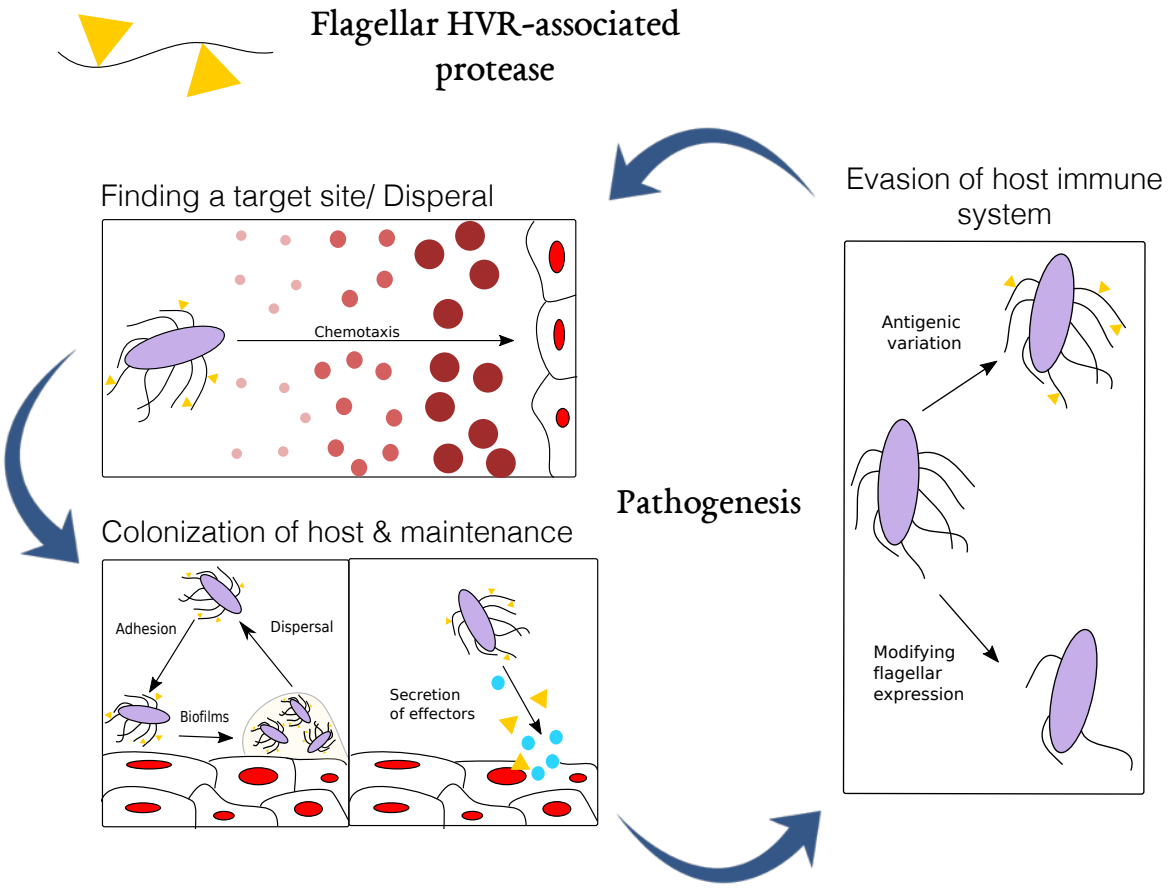


Figure 5.3: **Hypothesized function of the flagellin-HVR associated protease**

**Proteolytic flagellins may play a role in various facets of pathogenesis. It may assist in finding favourable target sites and/or dispersal, biofilm formation and/or host evasion. This figure is modified from [22].**

## 5.4 Conclusion

As mentioned above, the protease domain was found to be conserved in various species amongst which is the animal pathogen, *Clostridium botulinum* C/D str. DC5 as well as

avid biofilm former, *Pseudoaltermonas tunicata*. The phylogenetic diversity is indicative of lateral gene transfer to be the mechanism that has allowed the protease domain to spread. To discover the biological function of this HVR-associated proteolytic domain may require studying its role in other species as well. Experimental characterization of the flagellin variant in *Pseudoalteromonas tunicata* is currently underway in Dr. Andrew Doxey's laboratory. The domain may confer broad-spectrum protease activity which may assist different species for different functions. While it may be helping *Clostridium haemolyticum* to burrow through the ECM, it may be helping *Pseudoalteromonas tunicata* in being efficient in the formation of multi-species biofilms.

To conclude, this thesis summarizes evidence for the flagellin FliA(H) from *Clostridium haemolyticum* to be catalytically active. The MMP-like activity has been shown both *in vitro* and *in vivo*. This discovery is the first instance of a naturally-occurring HVR associated protease and may have evolved to play vital roles in various flagellated bacteria.

# References


[1] H. C. Ramos, M. Rumbo, and J. Sirard. Bacterial flagellins: mediators of pathogenicity and host immune responses in mucosa. *Trends in Microbiology*, 12(11):509–517, November 2004.

[2] S. A. Beatson, T. Minamino, and M. J. Pallen. Variation in bacterial flagellins: from sequence to structure. *Trends in Microbiology*, 14(4):151–155, April 2006.

[3] D. R. Thomas, N. R. Francis, C. Xu, and D. J. DeRosier. The three-dimensional structure of the flagellar rotor from a clockwise-locked mutant of *Salmonella enterica serovar Typhimurium*. *Journal of Bacteriology*, 188(20):7039–7048, October 2006.

[4] F. V. Chevance and K. T. Hughes. Coordinating assembly of a bacterial macromolecular machine. *Nature Reviews Microbiology*, 6(6):455–465, June 2008.

[5] C. J. Jones, M. Homma, and R. M. Macnab. L-, P-, and M-ring proteins of the flagellar basal body of *Salmonella typhimurium*: gene sequences and deduced protein sequences. *Journal of Bacteriology*, 171(7):3890–3900, July 1989.

[6] Howard C. Berg. The rotary motor of bacterial flagella. *Annual Review of Biochemistry*, 72:19–54, 2003.

[7] H. Matsunami, C. S. Barker, Y. Yoon, M. Wolf, and F. A. Samatey. Complete structure of the bacterial flagellar hook reveals extensive set of stabilizing interactions. *Nature Communications*, 7:13425, November 2016.

[8] Y. Mimori-Kiyosue, I. Yamashita, Y. Fujiyoshi, S. Yamaguchi, and K. Namba. Role of the outermost subdomain of *Salmonella* flagellin in the filament structure revealed by electron cryomicroscopy. *Journal of Molecular Biology*, 284(2):521–530, November 1998.

[9] D. Apel and M. G. Surette. Bringing order to a complex molecular machine: the assembly of the bacterial flagella. *Biochimica Et Biophysica Acta*, 1778(9):1851–1858, September 2008.

[10] Q. Duan, M. Zhou, L. Zhu, and G. Zhu. Flagella and bacterial pathogenicity. *Journal of Basic Microbiology*, 53(1):1–8, January 2013.

[11] J. Haiko and B. Westerlund-Wikstrm. The Role of the Bacterial Flagellum in Adhesion and Virulence. *Biology*, 2(4):1242–1267, October 2013.

[12] F. A. Samatey, K. Imada, S. Nagashima, F. Vonderviszt, T. Kumasaka, M. Yamamoto, and K. Namba. Structure of the bacterial flagellar protofilament and implications for a switch for supercoiling. *Nature*, 410(6826):331–337, March 2001.

[13] S. Merino and J. M. Toms. Gram-Negative Flagella Glycosylation. *International Journal of Molecular Sciences*, 15(2):2840–2857, February 2014.

[14] S. D. Reid, R. K. Selander, and T. S. Whittam. Sequence diversity of flagellin (fliC) alleles in pathogenic *Escherichia coli*. *Journal of Bacteriology*, 181(1):153–160, January 1999.

[15] K. Yonekura, S. Maki-Yonekura, and K. Namba. Growth mechanism of the bacterial flagellar filament. *Research in Microbiology*, 153(4):191–197, May 2002.

[16] S. Kojima, A. Shinohara, H. Terashima, T. Yakushi, M. Sakuma, M. Homma, K. Namba, and K. Imada. Insights into the stator assembly of the *Vibrio* flagellar motor from the crystal structure of MotY. *Proceedings of the National Academy of Sciences of the United States of America*, 105(22):7696–7701, June 2008.

[17] L. L. McCarter. Regulation of flagella. *Current Opinion in Microbiology*, 9(2):180–186, April 2006.

[18] T. G. Smith and T. R. Hoover. Deciphering bacterial flagellar gene regulatory networks in the genomic era. *Advances in Applied Microbiology*, 67:257–295, 2009.

[19] B. R. Phillips, J. A. Quinn, and H. Goldfine. Random motility of swimming bacteria: Single cells compared to cell populations. *AIChE Journal*, 40(2):334–348, February 1994.

[20] J. P. Armitage. Classic Spotlight: Seeing Is Believing. Imaging the Active Bacterial Flagellar Filaments. *Journal of Bacteriology*, 198(18):2391–2392, September 2016.

[21] D. B. Kearns. A field guide to bacterial swarming motility. *Nature reviews. Microbiology*, 8(9):634–644, September 2010.

[22] B. Chaban, H. V. Hughes, and M. Beeby. The flagellum in bacterial pathogens: For motility and a whole lot more. *Seminars in Cell & Developmental Biology*, 46:91–103, October 2015.

[23] B. Misselwitz, N. Barrett, S. Kreibich, P. Vonaesch, D. Andritschke, S. Rout, K. Weidner, M. Sormaz, P. Songhet, P. Horvath, M. Chabria, V. Vogel, D. M. Spori, P. Jenny, and W. Hardt. Near surface swimming of *Salmonella Typhimurium* explains target-site selection and cooperative invasion. *PLoS pathogens*, 8(7):e1002810, 2012.

[24] L. A. De Weger, C. I. van der Vlugt, A. H. Wijfjes, P. A. Bakker, B. Schippers, and B. Lugtenberg. Flagella of a plant-growth-stimulating *Pseudomonas fluorescens* strain are required for colonization of potato roots. *Journal of Bacteriology*, 169(6):2769–2773, June 1987.

[25] T. Morooka, A. Umeda, and K. Amako. Motility as an intestinal colonization factor for *Campylobacter jejuni*. *Journal of General Microbiology*, 131(8):1973–1980, August 1985.

[26] A. L. Erdem, F. Avelino, J. Xicohtencatl-Cortes, and J. A. Girn. Host protein binding and adhesive properties of H6 and H7 flagella of attaching and effacing *Escherichia coli*. *Journal of Bacteriology*, 189(20):7426–7435, October 2007.

[27] A. Tasteyre, M. Barc, A. Collignon, H. Boureau, and T. Karjalainen. Role of FliC and FliD Flagellar Proteins of *Clostridium difficile* in Adherence and Gut Colonization. *Infection and Immunity*, 69(12):7937–7940, December 2001.

[28] D. Rao, J. S. Webb, and S. Kjelleberg. Competitive Interactions in Mixed-Species Biofilms Containing the Marine Bacterium *Pseudoalteromonas tunicata*. *Applied and Environmental Microbiology*, 71(4):1729–1736, April 2005.

[29] A. Mai-Prochnow, F. Evans, D. Dalisay-Saludes, S. Stelzer, S. Egan, S. James, J. S. Webb, and S. Kjelleberg. Biofilm Development and Cell Death in the Marine Bacterium *Pseudoalteromonas tunicata*. *Applied and Environmental Microbiology*, 70(6):3232–3238, June 2004.

[30] S. Egan, S. James, C. Holmstrm, and S. Kjelleberg. Inhibition of algal spore germination by the marine bacterium *Pseudoalteromonas tunicata*. *FEMS Microbiology Ecology*, 35(1):67–73, March 2001.

[31] A. Blocker, K. Komoriya, and S. Aizawa. Type III secretion systems and bacterial flagella: Insights into their function from structural similarities. *Proceedings of the National Academy of Sciences*, 100(6):3027–3030, March 2003.

[32] B. Coburn, I. Sekirov, and B. B. Finlay. Type III Secretion Systems and Disease. *Clinical Microbiology Reviews*, 20(4):535–549, October 2007.

[33] G. Kuwajima. Flagellin domain that affects H antigenicity of *Escherichia coli* K-12. *Journal of Bacteriology*, 170(1):485–488, January 1988.

[34] F. Hayashi, K. D. Smith, A. Ozinsky, T. R. Hawn, E. C. Yi, D. R. Goodlett, J. K. Eng, S. Akira, D. M. Underhill, and A. Aderem. The innate immune response to bacterial flagellin is mediated by Toll-like receptor 5. *Nature*, 410(6832):1099–1103, April 2001.

[35] A. Klein, B. Tth, H. Jankovics, A. Muskotl, and F. Vonderviszt. A polymerizable GFP variant. *Protein engineering, design & selection: PEDS*, 25(4):153–157, April 2012.

[36] J. Tanskanen, T. K. Korhonen, and B. Westerlund-Wikstrm. Construction of a Multihybrid Display System: Flagellar Filaments Carrying Two Foreign Adhesive Peptides. *Applied and Environmental Microbiology*, 66(9):4152–4156, September 2000.

[37] P. S. Daugherty. Protein engineering with bacterial display. *Current Opinion in Structural Biology*, 17(4):474–480, August 2007.

[38] N. Cerd-Costa and F. X. Gomis-Rth. Architecture and function of metallopeptidase catalytic domains. *Protein Science : A Publication of the Protein Society*, 23(2):123–144, February 2014.

[39] S. S. Apte and W. C. Parks. Metalloproteinases: A parade of functions in matrix biology and an outlook for the future. *Matrix Biology*, 4446:1–6, May 2015.

[40] N. D. Rawlings, M. Waller, A. J. Barrett, and A. Bateman. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Research*, 42(D1):D503–D509, January 2014.

[41] C. M. Seibert and F. M. Raushel. Structural and Catalytic Diversity within the Amidohydrolase Superfamily. *Biochemistry*, 44(17):6383–6391, May 2005.

[42] S. Das, M. Mandal, T. Chakraborti, A. Mandal, and S. Chakraborti. Structure and evolutionary aspects of matrix metalloproteinases: a brief overview. *Molecular and Cellular Biochemistry*, 253(1-2):31–40, November 2003.

[43] L. Polgr. Basic Kinetic Mechanisms of Proteolytic Enzymes. In Erwin E. Sterchi and Prof Dr Walter Stcker, editors, *Proteolytic Enzymes*, Springer Lab Manual, pages 148–166. Springer Berlin Heidelberg, 1999. DOI: 10.1007/978-3-642-59816-6_10.

[44] B. L. Vallee and H. Neurath. CARBOXYPEPTIDASE, A ZINC METALLOPROTEIN. *Journal of the American Chemical Society*, 76(19):5006–5007, October 1954.

[45] B. L. Vallee and D. S. Auld. Zinc coordination, function, and structure of zinc enzymes and other proteins. *Biochemistry*, 29(24):5647–5659, June 1990.

[46] B. W. Matthews. Structural basis of the action of thermolysin and related zinc peptidases. *Accounts of Chemical Research*, 21(9):333–340, September 1988.

[47] F. X. Gomis-Rth, T. O. Botelho, and W. Bode. A standard orientation for metallopeptidases. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1824(1):157–163, January 2012.

[48] U. Eckhard, E. Schnauer, P. Ducka, P. Briza, D. Nss, and H. Brandstetter. Biochemical characterization of the catalytic domains of three different *Clostridial* collagenases. *Biological Chemistry*, 390(1):11–18, 2008.

[49] A. S. Duarte, A. Correia, and A. C. Esteves. Bacterial collagenases  A review. *Critical Reviews in Microbiology*, 42(1):106–126, January 2016.

[50] U. Eckhard, E. Schnauer, D. Nss, and H. Brandstetter. Structure of collagenase G reveals a chew-and-digest mechanism of bacterial collagenolysis. *Nature Structural & Molecular Biology*, 18(10):1109–1114, September 2011.

[51] U. Eckhard, E. Schnauer, and H. Brandstetter. Structural Basis for Activity Regulation and Substrate Preference of *Clostridial* Collagenases G, H, and T. *Journal of Biological Chemistry*, 288(28):20184–20194, July 2013.

[52] U. Eckhard, P. F. Huesgen, H. Brandstetter, and C. M. Overall. Proteomic protease specificity profiling of clostridial collagenases reveals their intrinsic nature as dedicated degraders of collagen. *Journal of Proteomics*, 100:102–114, April 2014.

[53] F. X. Gomis-Rth. Catalytic Domain Architecture of Metzincin Metalloproteases. *The Journal of Biological Chemistry*, 284(23):15353–15357, June 2009.

[54] P. E. Van den Steen, B. Dubois, I. Nelissen, P. M. Rudd, R. A. Dwek, and G. Opdenakker. Biochemistry and molecular biology of gelatinase B or matrix metalloproteinase-9 (MMP-9). *Critical Reviews in Biochemistry and Molecular Biology*, 37(6):375–536, December 2002.

[55] P. T. G. Elkington, C. M. O'Kane, and J. S. Friedland. The paradox of matrix metalloproteinases in infectious disease. *Clinical and Experimental Immunology*, 142(1):12–20, October 2005.

[56] J. Vandooren, P. E. V. Steen, and G. Opdenakker. Biochemistry and molecular biology of gelatinase B or matrix metalloproteinase-9 (MMP-9): The next decade. *Critical Reviews in Biochemistry and Molecular Biology*, 48(3):222–272, May 2013.

[57] C. M. Overall. Molecular determinants of metalloproteinase substrate specificity. *Molecular Biotechnology*, 22(1):51–86, September 2002.

[58] E. D. Janzen, J. P. Orr, and A. D. Osborne. Bacillary Hemoglobinuria Associated with Hepatic Necrobacillosis in a Yearling Feedlot Heifer. *MyScienceWork*, December 1981.

[59] P. J. Hauer, T. J. Yeary, and R. F. Rosenbusch. Cloning and molecular characterization of the beta toxin (phospholipase C) gene of *Clostridium haemolyticum*. *Anaerobe*, 10(4):243–254, August 2004.

[60] Y. Sasaki, N. Takikawa, A. Kojima, M. Norimatsu, S. Suzuki, and Y. Tamura. Phylogenetic positions of *Clostridium novyi* and *Clostridium haemolyticum* based on 16s rDNA sequences. *International Journal of Systematic and Evolutionary Microbiology*, 51(3):901–904, 2001.

[61] S. Nakamura, I. Kimura, K. Yamakawa, and S. Nishida. Taxonomic Relationships among Clostridium novyi Types A and B, *Clostridium haemolyticum* and *Clostridium botulinum* Type C. *Microbiology*, 129(5):1473–1479, 1983.

[62] B. Lobb and A. C. Doxey. Novel function discovery through sequence and structural data mining. *Current Opinion in Structural Biology*, 38:53–61, June 2016.

[63] A. C. Doxey, M. D. J. Lynch, K. M. Mller, E. M. Meiering, and B. J. McConkey. Insights into the evolutionary origins of clostridial neurotoxins from analysis of the *Clostridium botulinum* strain A neurotoxin gene cluster. *BMC Evolutionary Biology*, 8:316, 2008.

[64] L. Y. Geer, M. Domrachev, D. J. Lipman, and S. H. Bryant. CDART: protein homology by domain architecture. *Genome Research*, 12(10):1619–1623, October 2002.

[65] R. C. Edgar. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32(5):1792–1797, 2004.

[66] A. M. Waterhouse, J. B. Procter, D. M. A. Martin, M. Clamp, and G. J. Barton. Jalview Version 2–a multiple sequence alignment editor and analysis workbench. *Bioinformatics (Oxford, England)*, 25(9):1189–1191, May 2009.

[67] L. Bordoli and T. Schwede. Automated protein structure modeling with SWISS-MODEL Workspace and the Protein Model Portal. *Methods in Molecular Biology (Clifton, N.J.)*, 857:107–136, 2012.

[68] J. Yang, R. Yan, A. Roy, D. Xu, J. Poisson, and Y. Zhang. The I-TASSER Suite: protein structure and function prediction. *Nature Methods*, 12(1):7–8, January 2015.

[69] L. A. Kelley and M. J. E. Sternberg. Protein structure prediction on the Web: a case study using the Phyre server. *Nature Protocols*, 4(3):363–371, 2009.

[70] Z. Yang, K. Lasker, D. Schneidman-Duhovny, B. Webb, C. C. Huang, E. F. Pettersen, T. D. Goddard, E. C. Meng, A. Sali, and T. E. Ferrin. UCSF Chimera, MODELLER, and IMP: an Integrated Modeling System. *Journal of structural biology*, 179(3):269–278, September 2012.

[71] E. R. LaVallie, E. A. DiBlasio, S. Kovacic, K. L. Grant, P. F. Schendel, and J. M. McCoy. A thioredoxin gene fusion expression system that circumvents inclusion body formation in the E. coli cytoplasm. *Bio/Technology (Nature Publishing Company)*, 11(2):187–193, February 1993.

[72] R. G. Taylor, D. C. Walker, and R. R. McInnes. *E. coli* host strains significantly affect the quality of small scale plasmid DNA preparations used for sequencing. *Nucleic Acids Research*, 21(7):1677–1678, April 1993.

[73] F. W. Studier and B. A. Moffatt. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology*, 189(1):113–130, May 1986.

[74] J. Sambrook and D. W. Russell. SDS-Polyacrylamide Gel Electrophoresis of Proteins. *Cold Spring Harbor Protocols*, 2006(4):pdb.prot4540, September 2006.

[75] S. Costa, A. Almeida, A. Castro, and L. Domingues. Fusion tags for protein solubility, purification and immunogenicity in *Escherichia* coli: the novel Fh8 system. *Frontiers in Microbiology*, 5, February 2014.

[76] L. Corsini, M. Hothorn, K. Scheffzek, M. Sattler, and G. Stier. Thioredoxin as a fusion tag for carrier-driven crystallization. *Protein Science : A Publication of the Protein Society*, 17(12):2070–2079, December 2008.

[77] M. E. Kimple, A. L. Brill, and R. L. Pasker. Overview of Affinity Tags for Protein Purification. *Current protocols in protein science*, 73:Unit–9.9, September 2013.

[78] O. Schilling, P. F. Huesgen, O. Barr, U. auf dem Keller, and C. M. Overall. Characterization of the prime and non-prime active site specificities of proteases by proteome-derived peptide libraries and tandem mass spectrometry. *Nature Protocols*, 6(1):111–120, January 2011.

[79] U. Eckhard, P. F. Huesgen, O. Schilling, C. L. Bellac, G. S. Butler, J. H. Cox, A. Dufour, V. Goebeler, R. Kappelhoff, Ulrich A. D. Keller, T. Klein, P. F. Lange, G. Marino, C. J. Morrison, A. Prudova, D. Rodriguez, A. E. Starr, Y. Wang, and C. M. Overall. Active site specificity profiling of the matrix metalloproteinase family: Proteomic identification of 4300 cleavage sites by nine MMPs explored with structural and synthetic peptide cleavage analyses. *Matrix Biology*, 49:37–60, January 2016.

[80] O. Barr, A. Dufour, U. Eckhard, R. Kappelhoff, F. Bliveau, R. Leduc, and C. M. Overall. Cleavage Specificity Analysis of Six Type II Transmembrane Serine Proteases (TTSPs) Using PICS with Proteome-Derived Peptide Libraries. *PLOS ONE*, 9(9):e105984, September 2014.

[81] G. Marino, P. F. Huesgen, U. Eckhard, C. M. Overall, W. P. Schrder, and C. Funk. Family-wide characterization of matrix metalloproteinases from *Arabidopsis thaliana* reveals their distinct proteolytic activity and cleavage site specificity. *The Biochemical Journal*, 457(2):335–346, January 2014.

[82] Rebekah L. Gundry, Melanie Y. White, Christopher I. Murray, Lesley A. Kane, Qin Fu, Brian A. Stanley, and Jennifer E. Van Eyk. Preparation of Proteins and Peptides for Mass Spectrometry Analysis in a Bottom-Up Proteomics Workflow. *Current protocols in molecular biology / edited by Frederick M. Ausubel ... [et al.]*, CHAPTER:Unit10.25, October 2009.

[83] J. Rappsilber, Y. Ishihama, and M. Mann. Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics. *Analytical Chemistry*, 75(3):663–670, February 2003.

[84] R. Craig and R. C. Beavis. TANDEM: matching proteins with tandem mass spectra. *Bioinformatics (Oxford, England)*, 20(9):1466–1467, June 2004.

[85] A. Keller, A. I. Nesvizhskii, E. Kolker, and R. Aebersold. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Analytical Chemistry*, 74(20):5383–5392, October 2002.

[86] Pitter F. Huesgen, Philipp F. Lange, Lindsay D. Rogers, Nestor Solis, Ulrich Eckhard, Oded Kleifeld, Theodoros Goulas, F. Xavier Gomis-Rth, and Christopher M. Overall. LysargiNase mirrors trypsin for protein C-terminal and methylation-site identification. *Nature Methods*, 12(1):55–58, January 2015.

[87] O. Schilling, U. auf dem Keller, and C. M. Overall. Factor Xa subsite mapping by proteome-derived peptide libraries improved using WebPICS, a resource for proteomic identification of cleavage sites. *Biological Chemistry*, 392(11):1031–1037, November 2011.

[88] N. Colaert, K. Helsens, L. Martens, J. Vandekerckhove, and K. Gevaert. Improved visualization of protein consensus sequences by iceLogo. *Nature Methods*, 6(11):786–787, November 2009.

[89] C. J. Paul, S. M. Twine, K. J. Tam, J. A. Mullen, J. F. Kelly, J. W. Austin, and S. M. Logan. Flagellin Diversity in *Clostridium botulinum* Groups I and II: a New Strategy for Strain Identification. *Applied and Environmental Microbiology*, 73(9):2963–2975, May 2007.

[90] S. A. Kuehne and N. P. Minton. ClosTron-mediated engineering of *Clostridium*. *Bioengineered*, 3(4):247–254, July 2012.

[91] J. T. Heap, S. A. Kuehne, M. Ehsaan, S. T. Cartman, C. M. Cooksley, J. C. Scott, and N. P. Minton. The ClosTron: Mutagenesis in *Clostridium* refined and streamlined. *Journal of Microbiological Methods*, 80(1):49–55, January 2010.

# APPENDICES

# Appendix A

## A.1 Molecular Biology

1. Luria-Bertani (LB) Medium, For 1 Litre:
    - 10g Bacto-tryptone
    - 2g Yeast Extract
    - 10g NaCl
2. LB agar
    - LB with 15g agar

All solutions were autoclaved before use

## A.2 Protein purification

1. Lysis Buffer, For 1 Litre:
    - 25 mM HEPES (5.96g)
    - 300 mM NaCl (17.6g)
    - 10 mM Imidazole (0.68g)
    - 10% Glycerol (100g) pH adjusted to 7.5
2. Elution buffer, for 100 mL:

- 25 mM HEPES (0.596g)
- 300 mM Imidazole (2.04g) pH adjusted to 7.5

3. 25 mM HEPES buffer, for 1 Litre:
    - 5.96g HEPES pH adjusted to 7.5
4. 750 mM imidazole buffer, for 1 Litre:
    - 51.06g Imidazole
5. 1M IPTG, for 10 mL:
    - 2.383g IPTG filter sterilized and store at -20°C

## A.3 Western Blotting

1. 10X Tris buffered Saline stock solution, for 1 Litre:
    - 24g Tris Base
    - 88g NaCl pH adjusted to 7.6
2. TBS-T Solution, for 1 Litre:
    - 1X TBS solution
    - 0.1% Tween 20
3. Blocking buffer
    - 5% skim milk in 1X TBS-T
4. Wash buffer
    - 2% skim milk in 1XTBS-T

## A.4 Immunogold TEM

1. 10X Phosphate Buffered Saline stock solution (Fisher Bioreagents$^{TM}$)
2. Blocking buffer
    - 1X PBS
    - 0.1% Triton X
    - 1% BSA
3. Wash buffer
    - 1X PBS
    - 0.8% BSA