Characterization of Novel Protease Flagellinolysin in *Pseudoalteromonas tunicata*

by

Benjamin George Llewelyn Jenkins

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Biology

Waterloo, Ontario, Canada, 2023

© Benjamin George Llewelyn Jenkins, 2023

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Dr. Nikhil Thomas Professor, Department of Microbiology & Immunology, Division of Infectious Diseases Dalhousie University
Supervisor	Dr. Andrew Doxey Associate Professor, Department of Biology University of Waterloo
Internal Member	Dr. Trevor Charles Professor, Department of Biology University of Waterloo
Internal Member	Dr. Brendan McConkey Associate Professor, Department of Biology University of Waterloo
Internal-external Member	Dr. Brian Ingalls Professor, Department of Applied Mathematics University of Waterloo

Author’s Declaration

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

Flagellinolysins are a family of novel enzymes that act both as structural flagellins and proteases. Although previously identified and shown to be both structurally integrated into filaments and proteolytically active, flagellinolysin's biological function has remained unknown. The size of the bacterial flagellum and its prevalence in the natural world mean an enzymatic flagellin could have a significant effect in shaping the microscopic and macroscopic environments in ways that have as of yet gone undiscovered. This thesis represents a multipronged approach to characterize flagellinolysin in the marine bacterium *Pseudoalteromonas tunicata*, elaborate upon its context, and attempt to uncover both its role and possible substrates.

The first avenue of exploration relies upon building an understanding of flagellinolysin's context in the natural world. This includes a phylogenetic survey of flagellinolysin's presence across the bacterial tree of life, the identification of a likely regulatory method, and its presence within a flagellar gene cluster. Flagellinolysin is also identified as being expressed through both qrtPCR and proteomics, and shown to be integrated into mature filaments by antibody-based techniques. Flagellinolysin's nature in *P. tunicata* as a metalloprotease is also affirmed here by demonstrating abolition of activity through mutation.

With this information in hand, two mutant strains were generated: a catalytically inert point mutant (E238A) and a whole gene knockout mutant (ΔFlaMP). These mutants were shown to be capable of motility and to produce full-length flagella capable of generating movement, but were found to also produce defective biofilms that produce hyperaggregated architecture, indicative of difficulties in spreading across surfaces.

Defective biofilms led to an investigation to try and identify a potential substrate and determine a mechanism for the biofilm defect phenotype. Multiple proteomics-based methods were deployed, eventually pointing to a putative surface adhesin (VCBS) as a promising candidate. This has culminated in the development of a possible model of flagellinolysin function similar to the LapG/A

system and opened the possibility of an entirely new chapter in flagellar adhesion, flagellar sheaths, and biofilm regulation.

Acknowledgements

Science is a collaborative discipline; alike to a grand edifice built from innumerable hands, each contributing a few fine grains of sand. If this thesis represents a humble collection of specks added to that collection, it has only been possible from many hands carrying them. Here, I would like to acknowledge a few of them.

I wish to extend my deepest and most heartfelt gratitude to my supervisor, Dr. Andrew Doxey, in whose hands I have been clay. The fingerprints of his enthusiasm and boundless curiosity will mark me long after the fires of this trial have cooled. I would also like to extend my thanks to my committee members, Dr. Brendan McConkey and Dr. Trevor Charles, for their guidance and support over these long years.

I wish to thank the University of Waterloo, its staff, and administrators for the opportunities and funding it has provided, along with the governments of Canada and Ontario for the grants and scholarships they have likewise provided. I can only hope to repay such generosity with time.

I wish to express thanks to the many collaborators I have worked with over the years, with particular thanks to Dr. Ulrich Eckhard for his assistance with proteomics, and Sura Ali, whose familiarity with *P. tunicata* rivals my own. I also wish to acknowledge Dr. Suhelen Egan, who provided us with the initial strains, and Dr. Jiujun Cheng, whose skill with mutagenics rises to wizardry. To this list of names, I also add unnumbered undergraduate volunteers and 499 students, and consultations with my fellow grad students.

Dedication

If I have the gift of prophecy and can fathom all mysteries and knowledge, and if I have strength that can move mountains, but have not love, I am nothing. All that I am and all that I have done would never have been possible without the help of those who love me, whether with a kind word or material support.

I dedicate this work to my mother, Gwen, for whose kindness and support words fail me, and to my father, Iwan, without whom attending university would've been a pipe dream.

Table of Contents

Examining Committee Membership ii

Author's Declaration iii

Abstract iv

Acknowledgements vi

Dedication vii

List of Figures xiii

List of Tables xv

Chapter 1 Introduction 1

1.1 The history of the flagellum 1

1.2 Bacterial flagella 2

1.2.1 Structure of the bacterial flagellum 2

1.2.2 Evolution of the flagellum 8

1.2.3 Variation within the flagellar structure 10

1.2.4 Flagellar sheaths 13

1.3 Role of flagella in the environment 15

1.3.1 Flagellar motility 15

1.3.2 Flagella and surface attachment, colonization 16

1.4 Bacterial biofilms 18

1.4.1 Structure of a biofilm and stages of biofilm development 20

1.4.2 Biofilm emergent properties 21

1.4.3 Flagella within biofilms 22

1.5 Flagellinolysin 23

1.5.1 Metalloproteases 23

1.5.2 Discovery and characterization of flagellin variants containing protease domains 24
1.6 *Pseudoalteromonas tunicata* 29
1.7 Research objectives 30
Chapter 2 General Protocols & Methods 31
2.1 Strains and plasmids 31
2.2 Construction of mutant strains 33
2.3 Culturing media 35
2.4 SDS-PAGE 37
2.5 Constructs 40
2.6 Western blot and antibody construction 42
2.6 Flagellar preparation 44
Chapter 3 Identification, expression, and localization of flagellinolysin in *Pseudoalteromonas tunicata* 47
3.1 Introduction 47
3.2 Methods 49
3.2.1 Bioinformatic identification and analysis of flagellinolysin in *P. tunicata* 49
3.2.2 Experimental detection of flagellinolysin expression 51
3.2.3 Recombinant expression, purification, and enzymatic characterization of flagellinolysin 54
3.2.4 Immunolabeled transmission electron microscopy (ImmunoTEM) 55
3.3 Results 56
3.3.1 Identification of flagellinolysin genes across the bacterial tree of life 56
3.3.2 *P. tunicata* flagellinolysin is encoded within a flagellar gene cluster 60
3.3.3 *P. tunicata* flagellinolysin is likely under the control of FliA 61

3.3.4 *Predicted structure of P. tunicata* flagellinolysin is similar to *C. haemolyticum* flagellinolysin 63
3.4 Experimental confirmation of flagellinolysin expression in *P. tunicata* 66
3.4.1 Detection of flagellinolysin expression by qrtPCR 66
3.4.2 Mass-spectrometry based detection of flagellinolysin in extracellular samples 67
3.4.3 Flagellinolysin is detected in *P. tunicata* flagellar samples 69
3.4.4 Flagellinolysin is incorporated into *P. tunicata* flagellar filaments 70
3.4.5 *P. tunicata* flagellinolysin has protease activity 72
3.5 Discussion 74
Chapter 4 Investigating flagellinolysin function in *P. tunicata* by characterization of mutant phenotypes 77
4.1 Introduction 77
4.2 Methods 81
4.3 Examination of flagellinolysin mutant phenotypes in *P. tunicata* 81
4.3.1 Gelatinase assays 81
4.3.2 Soft agar motility assay 81
4.3.3 Confocal light microscopy 82
4.3.4 Liquid cultures 82
4.3.5 Culturing of pellicle biofilms 82
4.3.6 Culturing of surface-adherent biofilms 83
4.3.7 Laser scanning confocal microscopy of biofilms in static conditions 84
4.3.8 Transmission electron microscopy 84
4.3.9 Flow cell growth assays and confocal imaging 85
4.3.10 Growth curves 85

4.3.11 Competition assays 86
4.3.12 Scanning electron microscopy 88
4.3.13 Crystal violet assay 88
4.4 Results 90
4.4.1 Flagellinolysin mutant strains exhibit normal swimming motility and produce flagella with no visible defects 90
4.4.2 Identification of biofilm-related differences in *P. tunicata* flagellinolysin mutants 92
4.4.3 Focused investigation of flagellinolysin E238A mutant biofilms 95
4.5 Discussion 99
Chapter 5 Identification of flagellinolysin proteolysis targets via proteomics 104
5.1 Background and rationale 104
5.1.1 PICS 105
5.1.2 TAILS 107
5.1.3 Differential abundance from comparative shotgun proteomics 108
5.2 Methods 109
5.2.1 PICS 109
5.2.2 Shotgun mass spectrometry 110
5.2.3 TAILS and differential abundance analysis 110
5.3 Results 111
5.3.1 PICS provides insights into flagellinolysin substrate specificity 111
5.4 TAILS reveals differences in proteolysis between WT and E238A biofilms 113
5.5 Differential protein abundance reveals a plausible protein target of flagellinolysin 118
5.6 Analysis of the *P. tunicata* secretome reveals additional support for VCBS as a flagellinolysin target 122

5.7 Discussion 125
Chapter 6 Discussion 131
6.1 Brief summary 131
6.2 Flagellinolysin's context within the bacterial world 132
6.3 Defective flagellinolysin produces defective biofilms 134
6.4 New model for flagellinolysin in biofilm development 135
6.5 Conclusion 139
References 141
Appendix A Additional Figures 155

List of Figures

Figure 1.1. Diagram of a “standard” bacterial flagellum. 3
Figure 1.2. A diagram outlining the sequential assembly of flagellinolysin 4
Figure 1.3. 3D diagrams of a flagellar filament 6
Figure 1.4. A network diagram showing the relationships of the core set of flagellar proteins 9
Figure 1.5. Illustrations of different orientations and structures of the bacterial flagella 11
Figure 1.6. 3D models of the S. typhimurium flagellin and Bermanella marisrubri Red65 flagellin 12
Figure 1.7. An electron micrograph of a *P. tunicata* flagellar filament inside of a sheath. 13
Figure 1.8. A pair of images of immuno-labeled flagellar sheaths. 14
Figure 1.9. A cartoon illustration of an idealized attachment and escape cycle 17
Figure 1.10. An illustration outlining the emergent properties of biofilms. 19
Figure 1.11. Five stages of biofilm development 21
Figure 1.12. Summary of bioinformatic analysis of flagellinolysins 25
Figure 1.13. Immunoelectron microscopy showing the localization of *C. haemolyticum* flagellinolysin in assembled flagellar filaments 27
Figure 1.14. IceLogo visualizations of peptide substrate preferences for the *C. haemolyticum* and *H. gracilis* flagellinolysin. 28
Figure 2.1. A plasmid map for pK19mobsacB 33
Figure 2.2. Cartoon showing the process of isolating flagellar filaments. 45
Figure 3.1. Phylogenetic distribution of flagellinolysin-containing species across the bacterial tree of life. 59
Figure 3.2. Genomic context of flagellinolysin in *Pseudoalteromonas tunicata* and other flagellinolysin-containing species within related species (*Alteromonadales*) 61
Figure 3.3. Identification of putative FliA binding sites in non-coding promoter regions upstream of flagellinolysin. 63
Figure 3.4. Four structural models representing various flagellinolysins. 65
Figure 3.5. qrt-PCR analysis of mRNA transcript levels for 16S rRNA, FlaMP, and flagellin transcripts. 67
Figure 3.6. Detected abundance of flagellinolysin and other proteins in *P. tunicata* cellular and extracellular fractions by shotgun mass-spectrometry. 68
Figure 3.7. Western blot detection of flagellinolysin from purified *P. tunicata* flagellar filaments. 70

Figure 3.8. A transmission electron micrograph of concentrated *P. tunicata* flagella after immunolabeling with gold-conjugated anti-flagellinolysin antibody. 71
Figure 3.9. SDS-PAGE and casein zymography analysis of recombinant flagellinolysin from *P. tunicata*. 73
Figure 4.1. Cartoon illustrating the methods for collecting biofilm. 83
Figure 4.2. A series of figures showing the rate and extent of swimming motility in P. tunicata wild type and flagellinolysin mutants. 91
Figure 4.3. A series of images showing that all three strains are capable of producing intact flagella. 92
Figure 4.4. Confocal microscope images showing premature aggregation in *P. tunicata* flagellinolysin mutants. 93
Figure 4.5. Images of *P. tunicata* pellicle biofilms taken over a 1-3 day period. 95
Figure 4.6. Confocal imaging of surface-attached and pellicle biofilms in static conditions. 97
Figure 4.7. CLSM images showing stained cells in a flow cell. 99
Figure 4.8. A comparison of *P. tunicata* wild type and flagellinolysin mutant phenotypes against a figure reproduced with permission from Klausen et al., 2003. 101
Figure 4.9. Differences in biofilm architecture between WT and flagellinolysin mutant strains of *P. tunicata* suggest a general function of flagellinolysin. 102
Figure 5.1. A collection of consensus sequence logos showing the target preferences for flagellinolysin using PICS. 112
Figure 5.2. Summary of TAILS results 114
Figure 5.3. Icelogos based on high confidence substrates from TAILS terminomics. Extracellular motifs are markedly different between wildtype and E238A mutant. 116
Fig 5.4. Differential protein abundance identifies VCBS as a top flagellinolysin target protein. 119
Figure 5.5. Analysis of the WT versus E238A secretome by mass-spectrometry reveals flagellinolysin-mediated release of VCBS into the supernatant. 124
Figure 5.6. Hypothetical model illustrating flagellinolysin's method of action for processing sheath-associated proteins. 128
Figure 7.1. Flagellinolysin "stick and release" model for aiding in surface-related spread. 137

List of Tables

Table 2.1. A list of all strains used in this research. 31
Table 2.2: List of oligonucleotide primers used in generating the E238A mutant. 34
Table 2.3: LB & LMB broth recipe 36
Table 2.4: VNSS broth recipe[78] 36
Table 2.5: 10% separating gel 37
Table 2.6: 18% stacking gel 38
Table 2.7: Recipe for 10 mL of 4X SDS-PAGE sample loading buffer. 38
Table 2.8: Fixing solution 39
Table 2.9: Coomassie Blue staining solution 39
Table 2.10: Renaturation buffer composition 41
Table 2.11: Developing buffer composition 41
Table 2.12: Towbin buffer recipe 43
Table 2.13: Tris-buffered saline (TBS) recipe 43
Table 2.14: Ponceau stain recipe 44
Table 3.1: A list of the primers developed for use in qrtPCR. 52
Table 4.1: A list of strains used as a part of competition 87
Table 5.2: List of membrane signal-bearing proteins with high fold-change in the wild type with high confidence from Z-scores. 117
Table 5.1. Protein differences in *P. tunicata* biofilms. VCBS is bolded as the top increased protein in E238A biofilms. 120
Table 5.2. A list of peptides identified in the excised high molecular weight band identified in Figure 5.5c. 123

Chapter 1
Introduction

1.1 The history of the flagellum

The bacterial flagellum is one of the most iconic features of the natural world; biology textbook covers and classroom posters across the planet include the familiar cartoon diagram of a rod-shaped bacterial cell followed by the whirling, twirling shape of the flagellum. Its existence was inferred at the birth of microscopy by Antony van Leeuwenhoek in 1676, when he said that he was certain his little animalcules had some manner of "paw" to facilitate their movement, though he could not see it[1]. It would be 200 years before that dream would be made real by Robert Koch, who published the first flagellar stain protocol in 1877[2]. It was 1958 when the modern silver stain method for flagellar staining was be developed, and 1977 - three centuries after Leeuwenhoek inferred the existence of flagella – when this method would be simplified[3,4].

For the majority of human knowledge of its existence, the flagellum has been the subject of much speculation but few answers. Until 1973, the predominant assumptions around flagella assumed a waveform (i.e., side to side) rather than rotational propagation of movement, when a series of papers conclusively demonstrated the latter as the better model[5]. While different overall flagellar structures were observed and their differences in motility confirmed, it wasn't until 1983 that the inherency of the corkscrew structure of the flagellum was demonstrated to be a product of the protein structure and not induced by motion[6]. While the basal motor of the flagellum was imaged for the first time in 1967 using electron microscopy[7], it wasn't until the late '80s and early '90s that detailed studies of the proteins revealed the Mot complexes that act as stators in the motor[8–11].

All of these names and dates are offered to give insight into how recent and shallow our collective knowledge of flagella may truly be. Since these findings, new discoveries have continued to expand our understanding of the role of the bacterial flagellum; implicating it in processes like surface sensing, biofilms, and virulence[12–14]. In 2017, the discovery by Eckhard and colleagues of a proteolytically active structural flagellin in the filaments of *Clostridium haemolyticum* added the first line to another new chapter in the study of flagella. This thesis is an attempt to fill in the rest of the page; characterising the function of this novel protein and finding its place in the tree of life.

1.2 Bacterial flagella

1.2.1 Structure of the bacterial flagellum

The bacterial flagellum is one of the largest and most complicated organelles in the natural world. Comprised of upwards of sixty different proteins, some with copy numbers in the tens of thousands, it is often longer than the organism which produces it [15,16]. Like the bacteria which express them, flagella also display a great deal of diversity in structure and organization. Involved in diverse processes, from motility to chemotaxis to host colonization, flagella are frequently multi-functional and contain adaptations suited to the environment within which the organism lives [13,17,18].

In 2007, an analysis by Liu and Ochman identified a set of 24 genes that comprise a core flagellar set[16]. These core flagella genes form the basis of six necessary components for the construction and/or operation of the flagellar filament: The basal body, a motor and stator, a switching system, an export apparatus, a hook, and a filament.

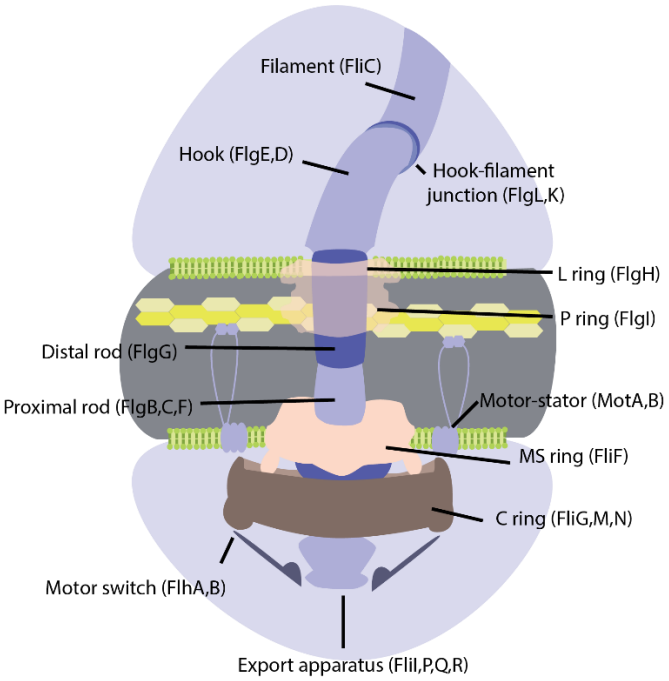


Figure 1.1. Diagram of a "standard" bacterial flagellum. The major components have been labeled, along with their respective protein components. The L ring and P ring have been included but are rendered semi-transparent for ease of viewing. The L ring and P ring are also not considered a part of the standard complement, due to their absence in the *Firmicutes* and other Gram-positives. Figure based on Liu & Ochman, 2007, and Echazarreta & Klose, 2019.

Basal body

Simultaneously acting as mount, export apparatus, and engine, the basal body contains the majority of protein diversity within the flagellum. It consists of all structures in the cytoplasm and extending outwards to the hook. Its innermost portions are built upon the MS ring – a circular structure typically consisting of 34 copies of the protein FliF, embedded in the cytoplasmic membrane – that acts as a scaffold on which the subsequent structure is built[19]. The export apparatus is a complex of ATP-using proteins that aid in the passage of other proteins through the MS ring. It forms on the MS ring's interior surface and provides the motive machinery for assembling the proximal and distal rod, and for the subsequent assembly of the hook and filament. The motor complex is formed by the motor-stator proteins MotA and MotB, while the rotor is formed by the C ring, which mounts onto the MS ring. The proximal and distal rods extend through the peptidoglycan layer and outer membrane (where cells possess such), linking the two and conveying rotary force to the hook and filament.

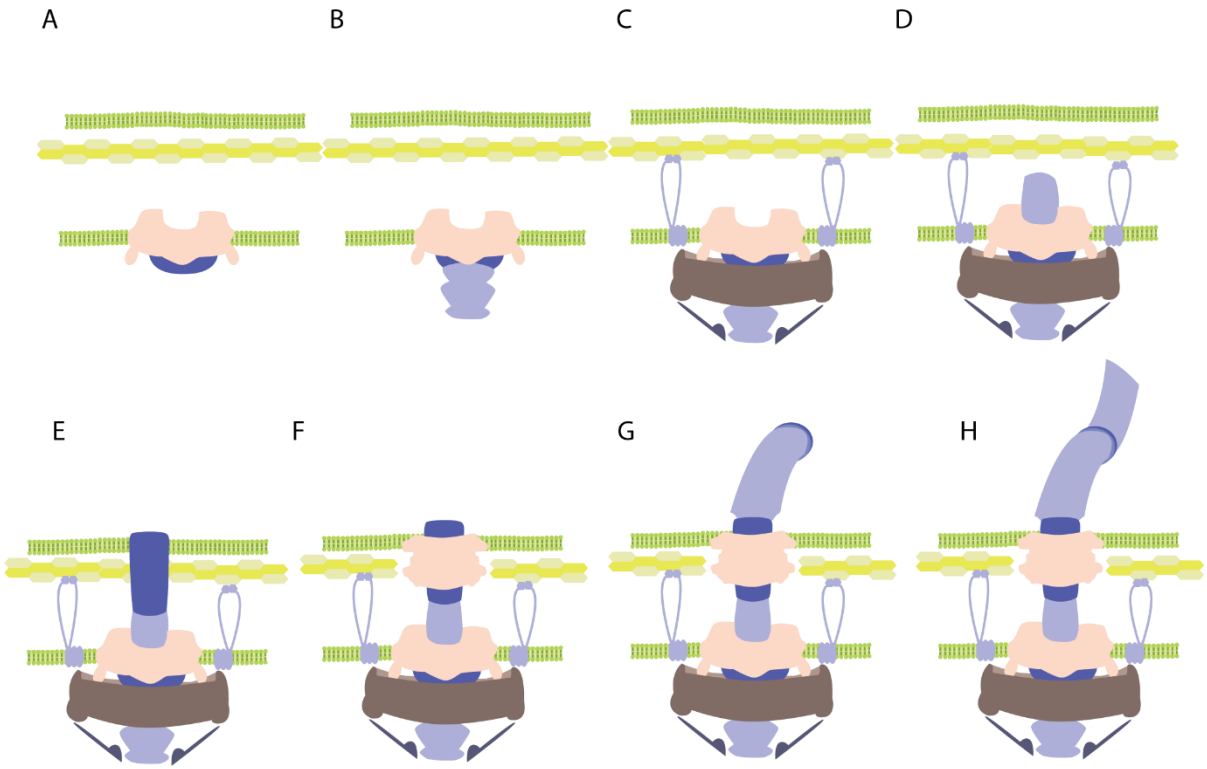


Figure 1.2. A diagram outlining the sequential assembly of flagellinolysin, adapted from Macnab, 2004. (A) The MS ring assembles on the inner membrane. (B) The export apparatus assembles on the cytoplasmic face of the MS ring. (C) The motor switch assembles around the MS ring. (D,E) Type III export begins with the assembly of the proximal and then distal rods. (F) Sec-mediated export sees the assembly of the P and L rings around the distal rod. (G) The hook forms atop the distal rod, capped by the filament junction. (H) The flagellar cap (not pictured) assembles atop the hook-filament junction and extends to form the filament.

The flagellar motor is driven by gradients across the inner membrane: The transfer of H^+ or $Na^=$ charge potential across the membrane through linked channel proteins (usually the MotAB complex) generates work, causing the flagellum to rotate. H^+-driven motors have been detected at rotation speeds of approximately 300 Hz, whereas Na^+-driven motors can operate at over five times those speeds[20]. The C ring, composed of FliGMN, is able to control the direction of flagellar rotation[21]. This integrates the flagellum to the cell's chemotactic sensors: the sensor kinase CheA phosphorylates the regulatory protein CheY, which interacts with the C ring to adjust flagellar motion from counterclockwise to clockwise. This causes tumbling in cells with peritrichous flagella, flicking in

unipolar cells, and flexing in *Spirochetes*[22–24]. In all cases, the result is a random change in direction, allowing the organism to reorientate in its pursuit or flight.

In most Gram-negative bacteria, the flagellar basal body also includes two or more additional ring structures. The most common of these (found in almost all Gram negatives) are the L and P rings[16]. Composed of FlgI and FlgH respectively, these proteins encircle and stabilize the flagellar rod; analogous to bushings in an electric motor [21].

Hook

The flagellar hook is made of approximately 120 copies of FlgE in a helical endecameric pattern. The tubular structure balances flexibility and rigidity, due to the stresses applied during flagellar rotation[21]. The D0 domain of FlgE forms the central channel through which further proteins are exported[21].

The assembly of the hook is regulated by FlhB and FliK[19]. FliK plays no structural role in the flagellum but is crucial to the switch between hook assembly and filament assembly. *ΔfliK* mutants produce a polyhook phenotype; producing extended FlgE polymers that are non-motile[19]. Upon reaching a predetermined length (around 55 nm in most cases), FliK is exported through the hook and the system switches from hook assembly to filament assembly. This begins with FlgK and FlgL, which form the hook-filament junction, and is followed by the export of the filament proteins. Although the regulatory roles of FlhB and FliK are well-attested from experimental evidence, the signal for hook completion and the initiation of the transition remains unknown.

Filament

The flagellar filament is the largest portion of the flagellum. Usually, it is composed of eleven protofilaments of the protein flagellin (FliC), with each protofilament containing thousands of copies. The structure of these proteins causes the filament to adopt a supercoiled structure – left-handed during CCW (forward) motion, right-handed during CW (tumbling) motion[21].

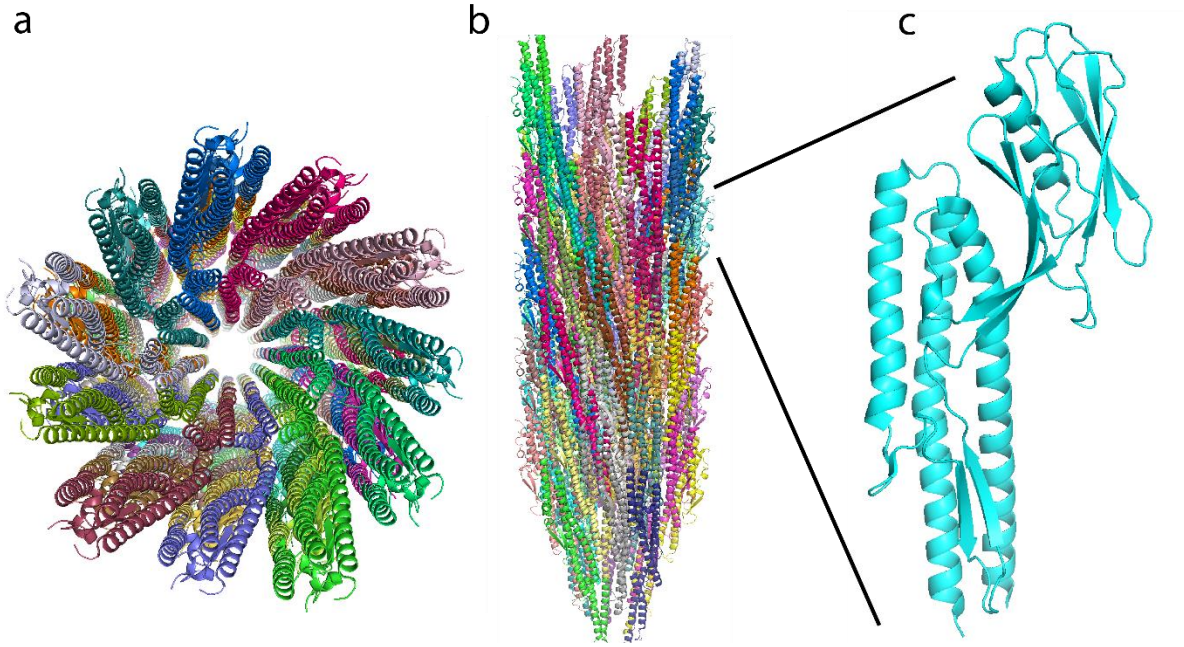


Figure 1.3. 3D diagrams of a flagellar filament (a,b; PDB ID: 8CXM[25]) and individual flagellin (c; PDB ID: 4NX9[26]). (a) A cross-section of a portion of a supercoiled flagellar filament produced by *Escherichia coli* K12. The filament uses only ten protofilaments, instead of eleven. The D0/D1 terminal domains cluster in the centre of the filament, while the D3 hypervariable domain faces to the outside. (b) The same filament viewed vertically. (c) A single flagellin monomer. The D3 hypervariable region is oriented towards the top right.

FliC contains four domains: The highly conserved D0/D1 domains and the much more variable D2/D3. The D0 and D1 domains are found at the N- and C-terminal regions of the flagellin protein and form the interior core of the filament. These domains are vital to maintaining the overall structure of the filament, and so contain a high degree of conservation of structure across the bacterial tree of life. These domains are also highly immunogenic, activating TLR5 in humans and stimulating the primary immune response[15,27]. By contrast, the D2 linker domain and D3 exterior-facing domain show more variability, with the D3 also referred to as the hypervariable region or hypervariable domain. This hypervariability shields the interior-facing D0/D1 from detection, reducing the ability of macrophages or possible predators from using the flagellum to detect the cell[28–30].

Many cells maintain multiple copies of FliC, which may serve a multitude of different purposes. Since the supercoiling of flagella is an innate product of the FliC protein structure, different variants of FliC may be expressed to produce different angles or tightness of coiling, which may adapt the filament to provide greater thrust in more or less viscous environments. Some cells express multiple variants simultaneously at time-varied degrees, producing composite flagella. *Salmonella typhimurium* undergoes phase variation, with individual cells undergoing periodic transition between expressing different types of FliC, creating multiple distinct sub-populations that are mutually unintelligible to the immune responses of host organisms[12]. The ability to therefore regulate not just the amount of flagellin expressed but also the type and even relative admixture of flagellins allows for a greater range of variability – and thus adaptability – for the flagellar filament.

Flagellar cap

The flagellar cap is a structure found in almost all flagellated bacteria, save for a small handful of *Alphaproteobacteria* that likely represent secondary losses[16]. As its name implies, the flagellar cap is a structure constructed at the end of filament, preventing the loss of exported flagellins and aiding the integration of said flagellins into the filament structure[31]. The cap itself is formed from a complex consisting of a single protein – FliD – whose closest homologs are the structural flagellin FliC and the hook-filament junction protein FlgK[16]. Although initially identified as a pentameric structure in *Pseudomonas*, both the size of FliD proteins and the resulting stoichiometry of the cap are variable, and may even vary within the same organism with environmental condition shifts[31].

The flagellar cap begins as an assemblage atop the hook-filament junction and aids in the assembly of the growing filament, though, at present, the exact mechanism by which it helps flagellins integrate into the growing filament and properly fold is not well-explained[31]. However, some explanatory models at present suggest the flagellar cap's component proteins undergo conformational change to admit FliC molecules and guide their integration through interactions through their terminal domains[31].

The flagellar cap is also implicated in processes of adhesion. *Pseudomonas* FliD, for example, is shown to be directly capable of binding mucin[32]. However, as mutants deficient in FliD are incapable of forming flagellar filaments, it is difficult to parse out how much of a contribution FliD makes to adhesion independent of FliC, which has been shown to have its own contributing role to adhesion[15].

1.2.2 Evolution of the flagellum

The evolution of the flagellum has been a subject of considerable controversy[33]. One of the most complex structures of the microbial world, it has been the subject of court cases debating the validity of evolutionary theory itself (e.g., the Kitzmiller v. Dover School District trial in 2005). Despite this, it is clear from numerous studies that the flagellum – like all structures in all living creatures – has evolved from earlier evolutionary antecedents[16,33]. For example, the flagellar export machinery is homologous to that found in the type III secretion system (T3SS) of the injectisome, not only at a general structural level but at the level of individual protein components[34]. These homologies indicate that the flagellar and injectisome T3SS descended from a common ancestor and have since diverged into separate roles[34–36].

The distribution of 24 common flagellar genes across the bacterial branch of the tree of life is indicative of an ancient origin to the bacterial flagellum, predating the extant divisions of the broad phyla. An analysis by Liu and Ochman comparing 25 universally-distributed genes against the distribution of the 24 common flagellar core genes showed large congruence between the two trees, with only a few notable exceptions[16], likely a result of later lateral distribution of whole gene complexes. The 24 core flagellar proteins in *E. coli*, when run through BLAST against the *E. coli* proteome, also show their strongest homologies to other core flagellar proteins. These patterns are consistent with a paralogous pattern of development: Ancestral flagellar genes dividing from a core set of one or more genes that diverged in function over time[16].

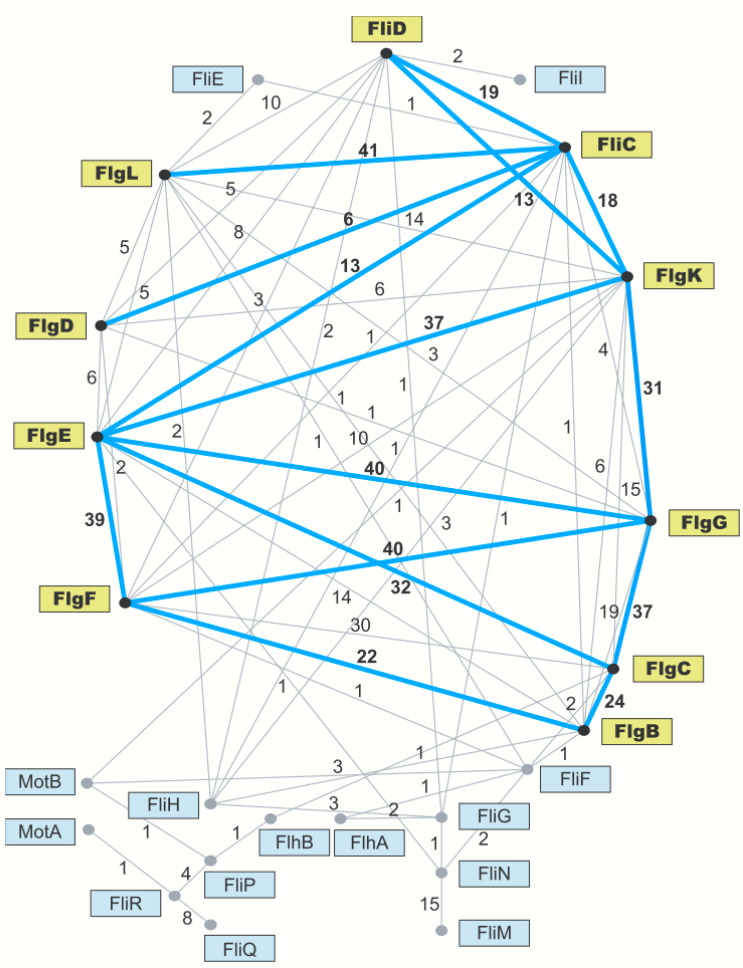


Figure 1.4. A network diagram showing the relationships of the core set of flagellar proteins. The number above each line represents the number of instances in which the two proteins were identified to possess homology in pairwise comparisons between the 40 species under interrogation, with an *E*-value cutoff of 10^{-4}. Figure reproduced with permission from Liu & Ochman, 2007.

The subsequent development of other flagellar features likely followed similar patterns of duplication and divergence, though the exact pathway for each is a subject of debate. The genus *Vibrio* possesses three additional ring structures – the H, T, and O rings – composed of at least three different proteins, but their origins are presently unknown. The protein FlgT is known to be a part of the H ring, and FlgT knockout mutants have difficulty penetrating the outer membrane[37]. *ΔflgT* mutants form periplasmic flagella, which are incapable of motion, or even have the filaments pierce the outer

membrane, damaging or potentially killing the cell. Organisms with the core complement of proteins do not produce periplasmic flagella, despite the lack of this protein (spirochetes excepted, having specially evolved to utilize them). This raises the question as to whether flagella evolved as a component of the outer membrane and then moved inwards with the development of the inner membrane, with other organisms later losing unnecessary components, or whether the flagellum evolved as an inner membrane organelle from the start[37]. The role of these proteins will be discussed in more detail in 1.2.4, in relation to discussing flagellar sheaths. For now, although it seems more plausible that the *Vibrio* ring structures are later additions, the dearth of evidence on their origins is another indicator of how young the study of the bacterial flagellum is.

1.2.3 Variation within the flagellar structure

Despite incomplete knowledge regarding the evolutionary history of the bacterial flagellum, the range of structural and functional diversity known to exist within the bacterial flagellum is at times breathtaking. Amongst the earliest discoveries were that flagella varied in number and location. Bacteria are capable of possessing one flagellum, none, or dozens, and the same organism may possess different numbers in different circumstances[15]. Flagella might be arranged at the poles of the cell or scattered across its surface or even embedded within the periplasm[34]. Their filaments may be constructed of one type of flagellin or several[38]. They may use sodium or proton gradients to drive their rotation[20]. They may be sheathed or unsheathed[39,40]. They can show considerable variation in how their clutch (the portion of the motor that transmits power) is structured[41].

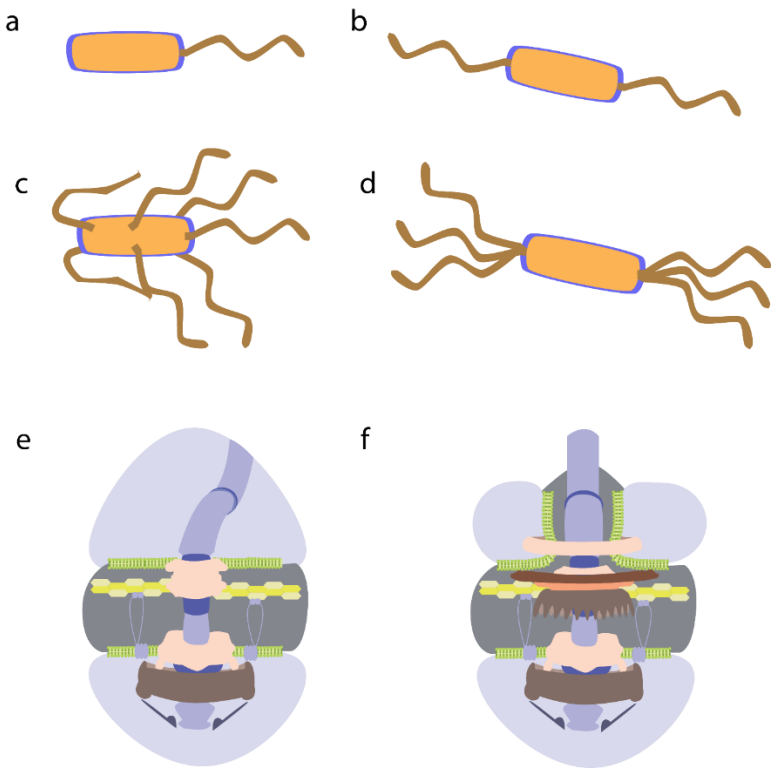


Figure 1.5. Illustrations of different orientations (a-d) and structures (e,f) of the bacterial flagella. The orientations shown here are not exhaustive, and include (a) monotrichous, (b) amphitrichous, (c) peritrichous, and (d) cephalotrichous. The flagellar structures are (e) the standard flagellin structure shown in an earlier figure and (f) a model of the *Vibrio* flagellar motor. This includes the three additional ring structures and shows the formation of the flagellar sheath around the filament.

By the 1990s, phase switching in organisms like *S. typhimurium* was an understood phenomenon, but organisms maintaining multiple flagellins for role switching or even for integrating into the same flagellum was thought to be a comparative rarity. The widening world of genetics studies and advances in sequencing saw that understanding begin to change[42]. More recently, new studies have added to this spectrum of flagellar diversity by revealing entirely new classes of flagellins. These include the recently discovered (at time of writing) "giant" flagellins[18] and flagellinolysins[43,44]. The latter will be discussed in greater detail in 1.5.

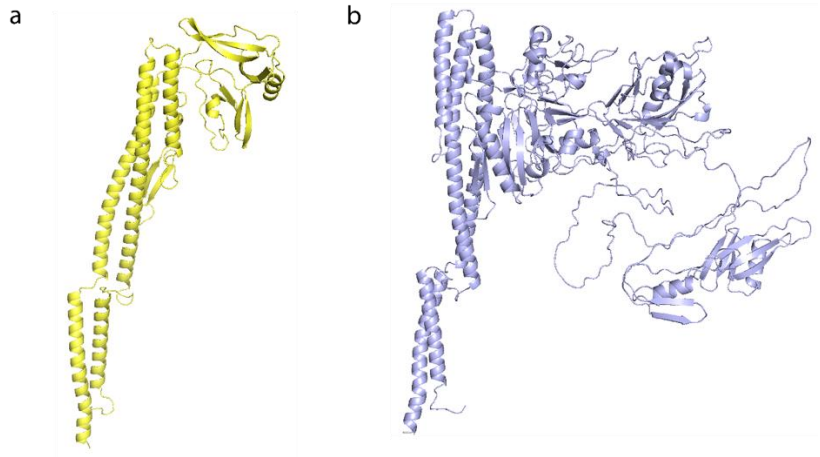


Figure 1.6. 3D models of the (a) *S. typhimurium* flagellin (UniProt accession id: Q9R4Q7) and (b) *Bermanella marisrubri* Red65 flagellin (UniProt accession id: Q1N2Y3) examined in Thomson et al., 2018[18]. The two 3D models were retrieved from the AlphaFold database. The D0/1 domains in the interior are of approximately similar size but the exterior-facing region of the *B. marisrubi* flagellin is more than twice the size of its counterpart.

The giant flagellins were identified by Thomson and colleagues in 2018 by searching the UniProtKB database for sequences that matched PFAM domain pf00669, which maps to the N-terminal helical region. The results returned 4,678 proteins longer than the *S. typhimurium* flagellin, and 92 that were longer than 1,000 amino acids; more than twice the size of a typical "standard" flagellin[18]. Subsequent bioinformatic analysis revealed that the giant flagellins are widely but intermittently spread across the bacterial tree of life, suggesting that giant flagellins are a product of either considerable lateral spread or multiple instances of convergent evolution.

The example of the giant flagellins illustrates the as-yet-undiscovered diversity that remains to be identified within flagellin protein family. Although large-scale sequencing and publicly available archives are more accessible than ever, the number of bioinformaticians sifting through the bulk of this data to find patterns remains comparatively small. Much likely remains to be discovered.

1.2.4 Flagellar sheaths

Flagellar sheaths are an extension of the outer membrane encompassing the filament of a flagellum. Although first described in 1950, struggles in producing mutants that can generate unsheathed and intact flagella has made studying the functions of sheaths difficult[39]. A study by Zhu and colleagues in 2017 found that *Vibrio alginolyticus* spontaneously produced occasional unsheathed flagella and compared the two structures in a flagellum-overproducing mutant[45]. By comparing the two under cryo-electron microscopy (cryo-EM), they were able to identify the O ring of *V. alginolyticus* as the only missing structure, implying it may play an important role in sheath formation. However, the scope of their study did not include identifying the protein composition of the structure, nor is the structure known to exist in all sheathed flagella.

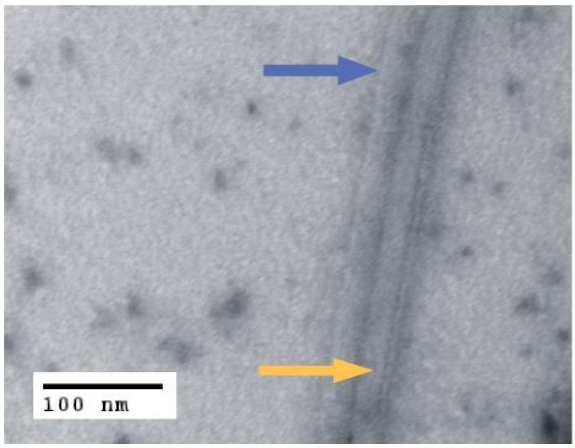


Figure 1.7. An electron micrograph of a *P. tunicata* flagellar filament (yellow arrow) inside of a sheath (blue arrow). The sheath is considerably larger than the filament it contains, approx. 53 nm across vs 21 nm. Samples were fixed in 2.5% glutaraldehyde and stained with 0.1% ammonium molybdate. Image is 64,000x magnification. Image taken by B. Jenkins (2020).

The phylogenetic distribution of flagellar sheaths is complicated by the absence of a core protein or set of proteins that are necessary for the formation of sheaths[39]. As of 2020, the best understanding of their distribution derives from lists of species whose production of sheaths have been confirmed by direct observation. Thus far, sheaths have been confined to being identified in five of the sub-classes of the Proteobacteria: Alpha-, Beta-, Gamma-, Delta-, and Epsilonbacteria[39]. The known distribution

is sporadic and further complicated by variation even within the same organism: the genus *Vibrio*, which has perhaps the best-characterized sheaths, produces sheathed polar flagella but unsheathed lateral flagella[39]. This implies the possibility that well-studied organisms thought to only possess unsheathed flagella may possess sheaths at certain times and not at others, and such features may simply have been missed.

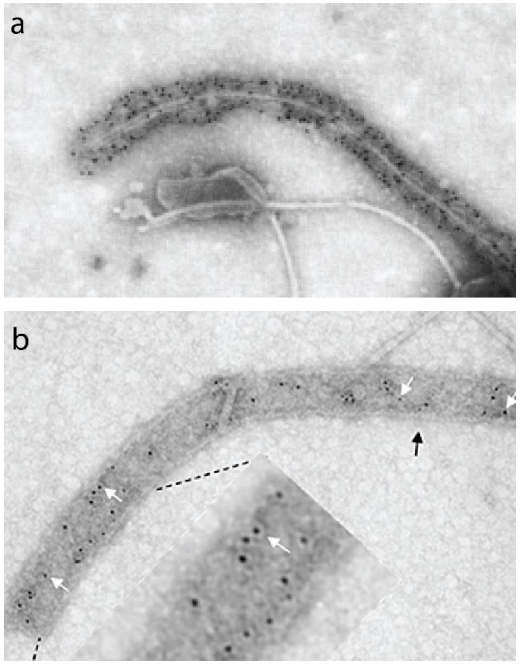


Figure 1.8. A pair of images of immuno-labeled flagellar sheaths. (a) An *H. pylori* sheathed flagellar filament next to one where the sheath has been stripped. The filament is labeled with immunogold antibody raised against HpaA (*H. pylori* adhesin A), showing strong affinity to the sheath[39,46,47]. Image from Luke & Penn, 1995. (b) A flagellar sheath from *Magnetococcus marinus*, showing immuno-labeling to a protein dubbed M1, which shares homology with known cadherin proteins[48]. Image from Lefèvre et al., 2010.

The understanding of the structure of flagellar sheaths has also changed with time. While the sheath does represent an extension of the outer membrane, its composition is not identical. The sheath is often enriched or depleted of fatty acids – *Bdellovibrio bacteriovorus* is noted as having an approximately 17-fold reduction in β-hydroxymyristic acid[39], a membrane component used as an industrial marker for endotoxin presence. Sheaths are also reported to be lower in protein than the rest

of the outer membrane, and appear to possess a different protein complement than the rest of the cell[39]. The flagellar sheath has also been identified as possessing several unique proteins – in particular, several porins and sheath-specific adhesins[39,48]. The roles of these proteins have not yet been elucidated in great detail but point to sheath-specific roles in adhesion and other processes.

1.3 Role of flagella in the environment

1.3.1 Flagellar motility

The primary utility of bacterial flagella is to facilitate locomotion. Potential energy stored in ion gradients (H^+ or Na^+) is transformed into work by the motor-stator complex crossing the membrane[20]. Changes in the conformation of these proteins at their point of interaction with the MS-C ring causes the rings to turn, with the direction of the turn determined by the phosphorylation of associated CheY. Under normal circumstances, flagella rotate counter-clockwise, with clockwise motion triggering tumbling or flicking, allowing the cell to reorient itself[20]. This rotation is quite rapid – anywhere from 15-300 Hz in *E. coli*[49]. This variability is owed both to the cell and the environment: The amount of energy needed to operate a flagellum at higher rotation speeds is not linear, and bacteria show the ability to regulate the number of stators in response to changes in environment or available energy[20].

The speed generated by flagella also varies considerably, with the overall structure, cell shape and size, media viscosity, energy source, and motor regulation all contributing[12]. Perhaps unintuitively, the number of filaments does not seem to necessarily correlate to overall speed: polyflagellated *E. coli*, for example, has an average speed of 25 μm/s, while the unipolar bacterial predator *B. bacteriovorus* is capable of reaching speeds of up to 160 μm/s[12].

While motion within media is the best-understood mode of flagellar motion, flagella also play a role in surface-driven motility via swarming. Unlike swimming motility, swarming motility is not assumed to be well-understood. The patchy distribution of swarming in the tree of life remains a mystery, for example, where swarming is found in the *Alphaproteobacteria, Gammaproteobacteria*, and *Firmicutes*, but not in the intermediate *Delta-, Epsilon*, or *Betaproteobacteria*[23]. Bacteria

cultivated in laboratories often also swiftly lose swarming capability, either because it is largely unnecessary or because microbiologists simply prefer neat little microcolonies and unwittingly select for swarming-deficient mutants[23]. Many bacteria capable of swarming will not do so until or unless certain environmental conditions are met[23].

Despite these issues, certain prerequisites are known as requirements for swarming motility: A small amount of surface moisture, enhanced flagellation, cell-cell interactions, and a surfactant[23]. The degree of enhanced flagellation varies considerably, with organisms like *P. aeruginosa* expressing only a single additional flagellum and even that only in half of the cells undergoing swarming[50]. While the speed of swarming does correlate with the degree of flagellation, the reason for its requirement is as yet unknown[23]. These uncertainties invite the consideration of the possibility of uniflagellate swarming, though at time of writing it has not been detected.

1.3.2 Flagella and surface attachment, colonization

In addition to its role in motility, flagella also play a key role in surface attachment and sensing of the cell's environment[12–14]. This may come as something of a surprise: The flagellar *flhDC* master regulatory system is often regulated in an opposite fashion to the expression of biofilm genes[51]. The late-stage flagellar sigma factor FliA has been shown to downregulate the expression of adhesins in some strains, including pathogenic *E. coli*[15]. However, the same strain with the *fliC* gene knocked out was less adhesive than the wild-type.

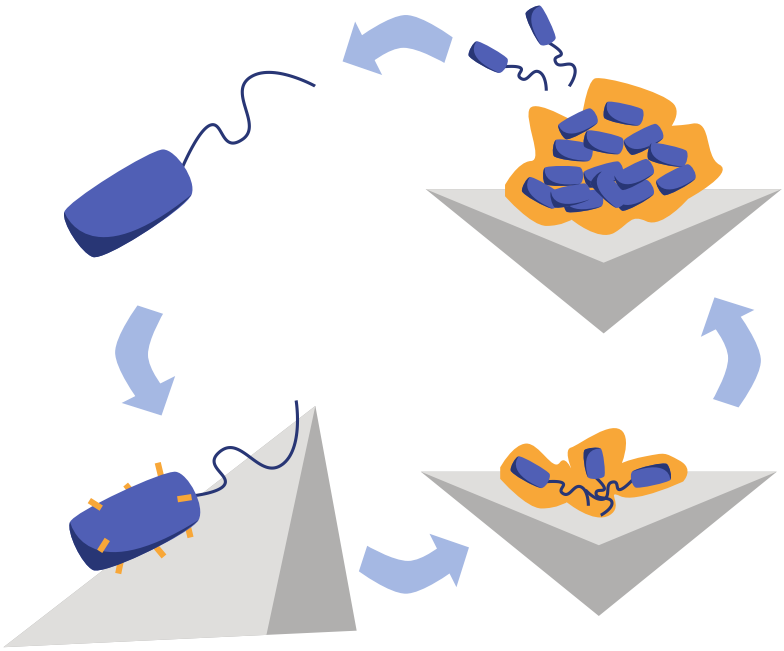


Figure 1.9. A cartoon illustration of an idealized attachment and escape cycle. Free-swimming bacteria are drawn towards surfaces. On receiving a signal of approaching or making contact with a surface, the cell produces adhesion proteins. Once transiently attached, the cells produce protein, ECM material, and attachment pili and become permanently attached. As the biofilm grows and develops, some cells redevelop flagella and escape to restart the cycle.

The traditional "stick or swim" model[13,52] of flagellar and adhesin regulation is complicated by this. In fact, the flagellum plays a part in transitioning between the two modes of existence and beginning biofilm formation. Deficiency in flagellar function is often an excellent predictor of deficiency in biofilm functionality[13].

Flagella and biofilms are mutually but oppositely regulated by cyclic-di-GMP (c-di-GMP)[51]. In *P. aeruginosa*, an upregulation of c-di-GMP triggers the expression of biofilm genes. FleQ, a flagellar regulatory protein, inhibits the expression of *pel* genes that are necessary for exopolysaccharide (EPS) synthesis[13,51]. C-di-GMP binds and inhibits FleQ, relieving inhibition of *pel* expression. The regulation of c-di-GMP is integrated into the flagellar system itself via the diguanylate cyclase SadC (surface attachment deficient), which moderates c-di-GMP in response to an unknown but viscosity-

dependent environmental signal. This also controls the rate of flagellar reversals, with mutants showing heightened reversal in high viscosity solutions; a prelude to biofilm formation.

In *Vibrio*, flagellar rotation is necessary for biofilm formation. It is thought that the halting of flagellar rotation may cause a rise in charge in the intermembrane space, resulting in a hyperpolarized membrane[13]. This difference in charge acts as a signal for attachment, which triggers the shift from transient to permanent attachment. This model is supported by experiments that dissipate the membrane potential, preventing the cells from transitioning from transient to permanent attachment[13,53].

1.4 Bacterial biofilms

Bacterial biofilms are best defined as an aggregate of bacterial cells, embedded in a self-produced matrix of extracellular polymeric substances (EPS) that allows cells to adhere to each other and/or a surface. Biofilms can exist either anchored to a surface or free-floating as floc, though the number of forms that either mode can take can be diverse. They are ubiquitous throughout the natural world and responsible for a large portion of biogeochemical cycling in water, soil, sediment, and subsurface environments[54].

Biofilms are everywhere, including on and within multicellular organisms. They are responsible for the purification (and fouling) of drinking water, the breakdown of waste products (including solid and liquid waste), and are utilized in a wide variety of industrial processes for the manufacture of chemical and agricultural goods[54]. Of particular note, biofilms are a constant issue in dealing with chronic infections and in managing the risk of medical implants, as both short- and long-term medical implants may become colonized by bacterial biofilms[54,55]. The presence of biofilms in medical circumstances is of great concern because of the resilience of biofilms to eradication by the usual methods of the immune system and medical profession, including antibiotics, oxidizing molecules, and dessication[54,55].

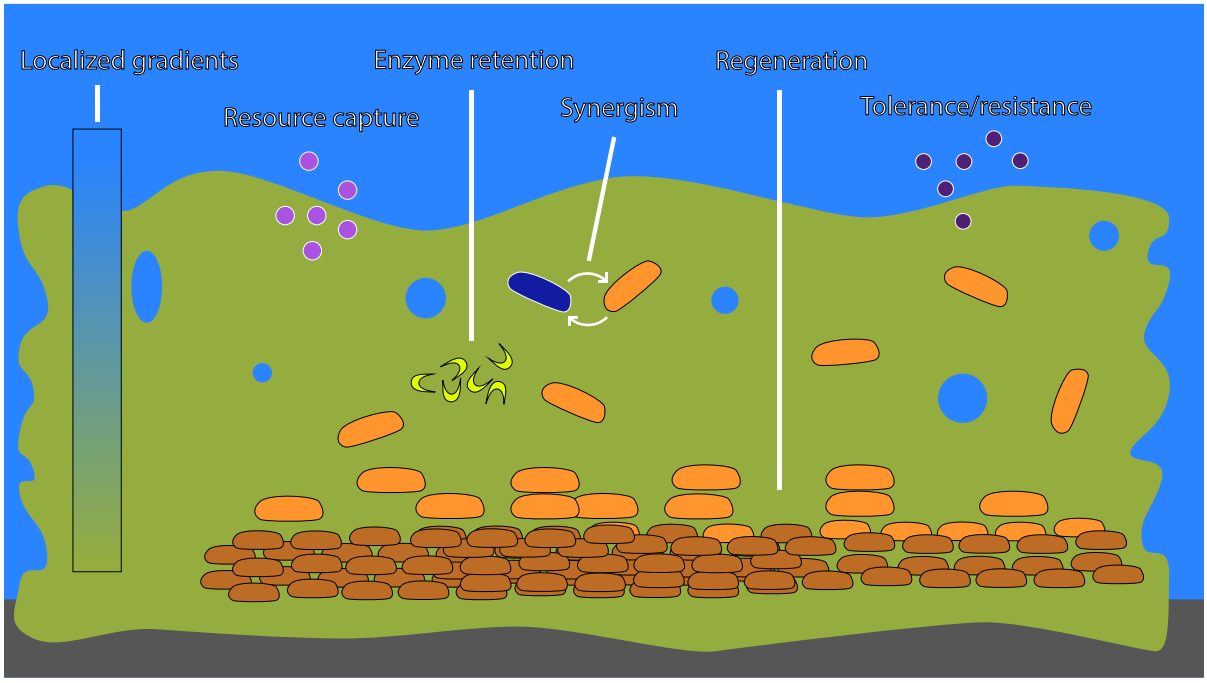


Figure 1.10. An illustration outlining the emergent properties of biofilms. Biofilm matrix allows for the control of flow of material along localized gradients and its retention, concentrating trace elements and maximizing their use. Enzymes are retained in the matrix to assist in processes like digestion. Different organisms may specialize in processing the metabolic by-products of other organisms, producing synergistic or mutualistic effects. The deep vault of a biofilm also acts as a reserve of cells with slowed metabolisms that are highly resistant to attack, and which can constantly replenish the film. Figure adapted from Flemming et al., 2016.

Biofilm-associated bacteria cannot simply be thought of as identical to a collection of free-swimming cells. The structure of a biofilms and the conditions within them produce emergent properties not found in the free-swimming lifestyle of the organisms that inhabit it[54,55]. These properties are the product of the spatial organization of the biofilm and its resultant effects of restricting material flow, enabling cell-cell interactions, and moderating the metabolisms of bound organisms[54,55]. Some organism-specific effects are also noteworthy.

1.4.1 Structure of a biofilm and stages of biofilm development

The formation of a biofilm begins with the attachment of a bacterium to a surface (Figure 1.11). The ability of bacteria to colonize a surface is determined by a number of factors, primarily related to attractive/repulsive forces associated with the surface[55,56]. Many bacteria are attracted to preconditioned surfaces – surfaces already pre-coated in a layer of organic film – such as tooth enamel coated in "acquired pellicle", the mixture of proteinaceous material found in the oral cavity[56]. Flagellated bacteria possess an advantage in being able to actively swim towards surfaces, to overcome repulsive/hydrodynamic forces that might prevent surface adhesion[55].

Once attached to a surface, cells entering a biofilm state begin to secrete a layer of EPS that cements them to the surface (Figure 1.11). The material composition of this layer varies between species, and can include protein, polysaccharides, lipopolysaccharides, and external DNA[54]. The resulting structure consists of matrix-bound cells in microcolonies divided by microchannels that allow for the exchange of media[56]. There is no common singular architecture to all biofilms, varying across space, time, and species boundaries, with mixed-species biofilms adding even greater complexity to the interactions[56].

Once established, cell division and coordinated cell death contribute to the structure of the biofilm[57,58]. These allow for the creation of more complicated structures, for the removal of weak or injured cells, and the freeing up of resources for their neighbours. Periodic rearrangement of the biofilm may also be achieved by reactivation of flagellar function, leading to a tunneling behaviour that opens channels and allows the flow of fresh media into a biofilm and allows the outflow of metabolic wastes[54]. The biofilm may also be disrupted by environmental sheer forces, grazing, etc., or by the action of cells escaping the film in response to an environmental signal[59]. The production of enzymes (e.g., dispersins[60–62]) and similar agents allow for cells to break down their own matrix and facilitates their release into the environment as free-swimming cells.

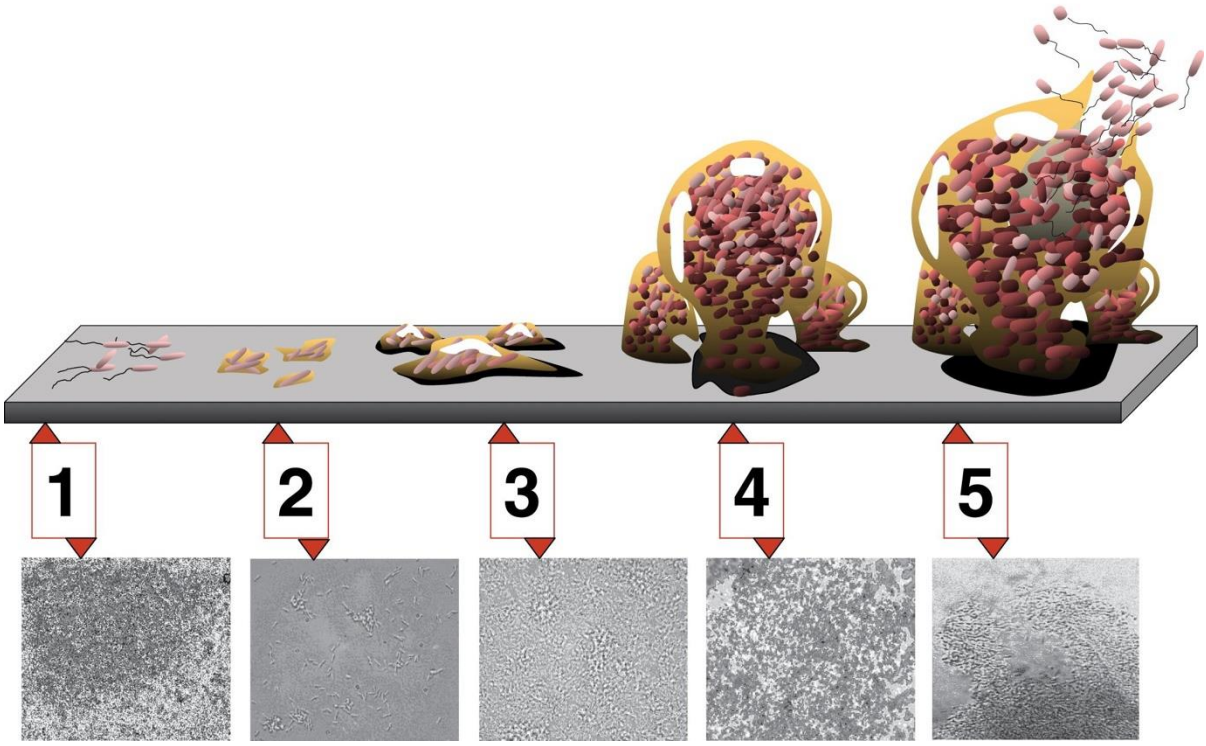


Figure 1.11. Five stages of biofilm development. (1) initial attachment; (2) irreversible attachment; (3 and 4) maturation; (5) dispersal. Photographs below show corresponding micrographs of *P. aeruginosa* biofilms at each stages. Figure reproduced with permission from Monroe (2007)[63].

1.4.2 Biofilm emergent properties

The effects of biofilms give rise to a number of properties that are not found in single cells. The biofilm matrix, by slowing the diffusion of materials into and out of the space surrounding the cells, triggers shifts in the proteomes and metabolisms of the organisms within the matrix[54]. The ability of the matrix to capture proteins reduces the need for the organism to maintain a constant internal suite of digestive enzymes, which can instead be exported to the matrix itself[54]. The matrix's alternating hydrophobic and hydrophilic nature can act as a moisture barrier, protecting the organisms against desiccation[54,55]. Most prominently in medical settings, the matrix's effect of slowing metabolism and intake of outside toxins, and the reserve of cells within its structure, allows for a "sacrificial layer" of surface cells to absorb the brunt of attempts to eradicate the film[55,64].

Emergent properties can also arise from specific combinations of cells. Cooperation between different mutualist species allows biofilms to efficiently maximize processes like nitrification, with nitrite oxidizers and ammonia oxidizers utilizing each other's waste products as an energy source[54]. Heightened rates of DNA exchange within the matrix can also enhance speed of adaptation to new threats and increase the spread of mutualistic phage throughout a film[54]. This makes biofilms a natural point for the spread of beneficial genes.

1.4.3 Flagella within biofilms

As detailed above, flagella are typically counter-regulated with biofilms. As also mentioned above, however, the idea of these systems as binary is inaccurate. Even after attachment, flagella play important roles in regulating biofilm structure and development, both via aiding its spread (see Section 1.3.1.) and by regulating its structure. Alterations to flagellar structure and function directly impact biofilm development: *flgE* deletion mutants in *P. aeruginosa* show increased biofilm growth under flow conditions than wild-type cells, reduced expression of genes for producing biofilm matrix proteins, and increased antibiotic resistance under flow cell conditions[65]. These contrasted with microtiter growth conditions, where the mutants showed patchy wall adherence compared to the wild-type's denser pellicle biofilm, consistent with a lack of motility. Wild-type cells also formed thin, evenly distributed biofilms in flow cell conditions, whereas the mutants formed "mushroom" structures: dense microcolonies with minimal distribution across the surface. These findings all indicate how much flagellar function can regulate the overall structure of a biofilm.

In 2012, Houry and colleagues reported the discovery of motile bacteria within *Bacillus thuringiensis* biofilms[17]. This team of researchers used confocal laser microscopy to track the periodic reactivation of cells within biofilms. This revealed that, at any given time, approximately 0.1-1.0% of the total population of cells within a biofilm reactivated their flagellum and swam through the biofilm matrix. This swimming action generated short-lived tunnels as the cell bored through the matrix, each lasting around 2-5 seconds, after which the tunnels collapsed. These transient pores are expected to allow the periodic refreshing of the films; exporting wastes and importing fresh media, and triggering local dispersal. This motion was deemed to be flagella-driven, as flagella-deficient or motion-incapable mutants did not display the behaviour.

In polyflagellated species, the flagellar filament can also form a major portion of the biofilm matrix. While flagellar filaments regularly play an important role in initial adhesion, in *E. coli*, flagellar filaments (alongside curli fibres) make up the major proteinaceous component of the biofilm matrix[66]. This maintenance of flagella as a matrix component also means that cells are capable of reactivating flagellar filaments swiftly on escape, rather than needing to build the entire flagellar apparatus from the beginning.

1.5 Flagellinolysin

1.5.1 Metalloproteases

Metalloproteases and metallopeptidases are a broad category of catalytic enzymes, capable of degrading protein through the use of a metallic catalyst. Metalloproteases are found across the tree of life and are not unique to any one branch. Despite very different sequences and a lack of homology, disparate and independently evolved metalloproteases do often contain structural similarity when degrading common targets or using a given residue as a ligand for a metallic cofactor[67]. There are at the time of writing over a hundred metalloprotease families (denoted with M and a number, such as M1, M2, etc.) in the MEROPS database, including a number of subfamilies and other overarching categorizations. This includes division into broad clans. Within the metallopeptidases, clan MA is the largest, consisting of 82 of the 106 listed metallopeptidase families in the MEROPS database[68]. The unifying feature of this clan is the presence of the HEXXH motif.

The His-Glu-Xaa-Xaa-His (HEXXH) motif is common among metalloproteases. A simple pentapeptide motif, it is capable of evolving independently multiple times without need for reference to a common ancestor[67]. X-ray crystallography has confirmed that the motif helps bind a metallic cofactor (usually zinc) in the catalytic pocket of the protein[67]. In the MA clan, the histidine residues act as zinc ligands and the glutamic acid aids in catalysis. In studies of thermolysins, the glutamic acid residue promotes nucleophilic attack from a water molecule against the carbonyl carbon atom of the substrate peptide, aided by an additional glutamic acid residue 20 residues away in sequence[67].

In bacteria, MA clan proteases play diverse roles. Many are secreted but others remain attached to cell surfaces and modify cell-adjacent structures, such as host tissues[67]. The M9 family of metalloproteases is of particular interest to medicine, as it contains bacterial collagenases. These secreted proteins degrade host tissue matrices, allowing for deeper infiltration and tissue colonization by bacterial pathogens[69].

1.5.2 Discovery and characterization of flagellin variants containing protease domains

As described earlier, flagellin hypervariable regions form surface-exposed domains found on the exterior of flagellar filaments and therefore define the surface properties of extracellular flagella (Figure 1.12a). In 2008, a study by Doxey and colleagues.[70] reported the identification of several unusual flagellins found in *C. haemolyticum* (NCBI accession ID: BAB87738), *Azoarcus olearius* (YP_934037, now WP_011766260), *Desulfuromonas acetoxidans* (then ZP_01312630, now WP_006000355), and *Pseudoalteromonas tunicata* (then ZP_01132756, now EAR29544). These flagellins were unique by possessing a MA family metalloprotease domain predicted within the flagellin hypervariable region, with a conserved HEXXH catalytic motif, which showed homology to the clostridial collagenases. At the time, it was suggested that these "collagenase-like" flagellins could be a means of altering host surfaces in flagellar proximity, as native collagenases were already utilized.

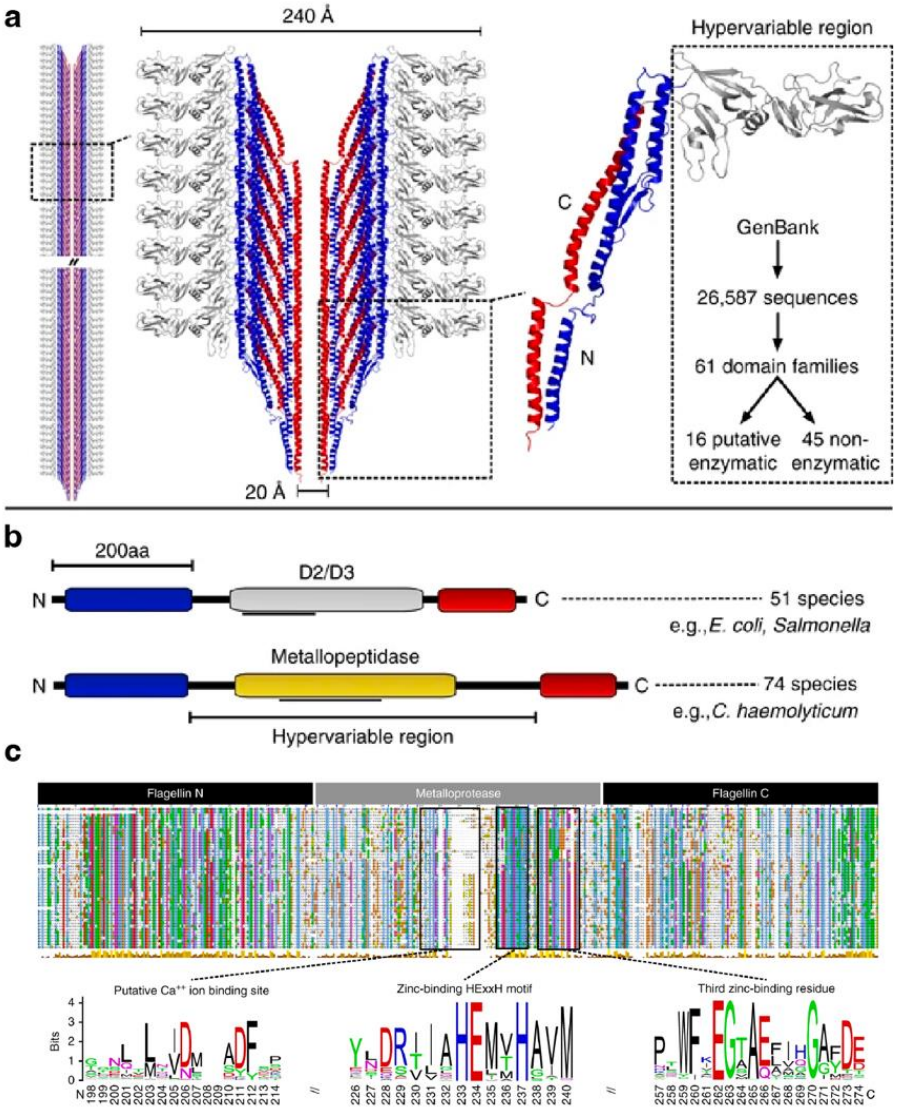


Figure 1.12. Summary of bioinformatic analysis of flagellinolysins. Reproduced with permission from Eckhard et al. (2017). (a) a schematic showing a flagellar filament composed of flagellin molecules. Example flagellins with D2/D3 domains (gray) in the hypervariable region are shown, and these form surface-exposed domains on the outermost layer of flagellar filaments. (b) Protein domain architecture of flagellins containing D2/D3 domains and flagellins containing predicted metallopeptidase domains in the hypervariable region. (c) Sequence alignment of identified flagellins containing metallopeptidase hypervariable regions, highlighting the conserved HEXXH zinc-binding site and Ca-binding site predicted to be involved in catalytic activity.

A more recent 2017 investigation by Eckhard and colleagues. sought to characterize these new collagenase-like flagellins in greater detail[43]. This study used a bioinformatic survey across existing genomes, and identified 86 flagellin sequences across 74 species and 35 genera containing a metalloprotease domain within the HVR (Figure 1.12b). All of the identified flagellin homologs contained the HExxH motif indicating conservation of protease function in these flagellin variants (Figure 1.1c). Putative zinc-binding and Ca-binding sites could be identified based on homology with other zinc metallopeptidases (Figure 1.12c). It is also worth noting that protease domains were identified in an even greater number of species than the well-characterized D2/D3 domain found in flagellins from *E. coli* and *Salmonella*. No protease-domain containing flagellin had ever been studied experimentally before.

Eckhard and colleagues then selected *Clostridium haemolyticum* as a model to study these newly identified protease domain containing flagellins, owing to its importance as an animal pathogen and its known ability to degrade host tissues. Based on immunogold electron microscopy (immunoEM), SDS-PAGE, and proteomic analysis of intact flagella, the protease domain containing flagellin from *C. haemolyticum* was found to be incorporated into flagellar filaments (Figure 1.12b). SDS PAGE and immunoEM revealed that the protease domain containing flagellin was not the only component of filament, and that it was second in abundance after the major structural flagellin of *C. haemolyticum*. Interestingly, all species that contain homologs of protease domain containing flagellins, also contain other flagellins, which is consistent with the idea that it may function as one of the flagellin components of filaments, rather than representing the sole component or even the major component of flagella in some species.

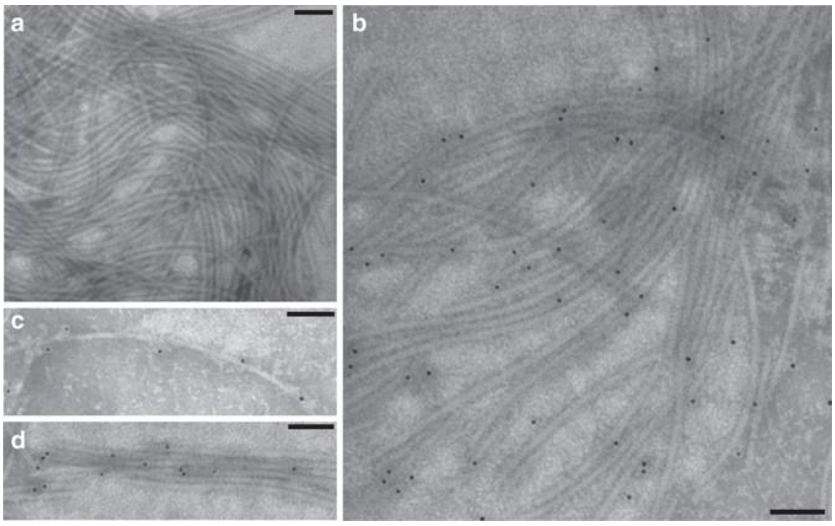


Figure 1.13. Immunoelectron microscopy showing the localization of *C. haemolyticum* flagellinolysin in assembled flagellar filaments. (a) a control sample of isolated filaments; (b-d) immunoEM micrographs of gold-labeled flagella from *C. haemolyticum* with anti-flagellinolysin antibodies. Image is reproduced from Eckhard et al., 2017.

The protease domain containing flagellin from *C. haemolyticum* was then confirmed to be proteolytically active through quenched fluorescence assays using peptides as well as through high-throughput proteomic identification of protease cleavage sites (PICS) assays[43], demonstrating degradation of peptides derived from trypsin- and GluC-generated libraries (Figure 1.14a). The PICS method works through the labeling of cleavage sites, allowing for the differentiation of cleavage sites derived from a protease of interest from those produced by trypsin or GluC, and is discussed in greater detail in Chapter 5. The PICS-derived consensus sequence was then used to produce tagged peptides for use in a quenched fluorescence assay, confirming a leucine preference in the binding site[43]. Protease activity was also abolished in the presence of EDTA, informing its identification as a metallopeptidase[43]. Its function as a protease confirmed, the new group of metallopeptidase flagellins was given the name “flagellinolysins” and the symbol “FlaMP”. It was entered as family M101 in the MEROPS database[68].

Isolated flagella were also confirmed to be active via quenched fluorescence assays, but characterization of proteolytic activity against whole proteins was limited to observations of autolytic

activity and cleavage of bovine fibronectin[43]. Ultimately, the native targets of *C. haemolyticum* flagellinolysin and its biological function still remain unclear, and no studies have yet been done using knockout mutants or other assays to investigate this question further. Eckhard et al. (2017) concluded with reasonable speculation that flagellinolysin may take part in the degradation of host tissues, aiding further infiltration of the host.

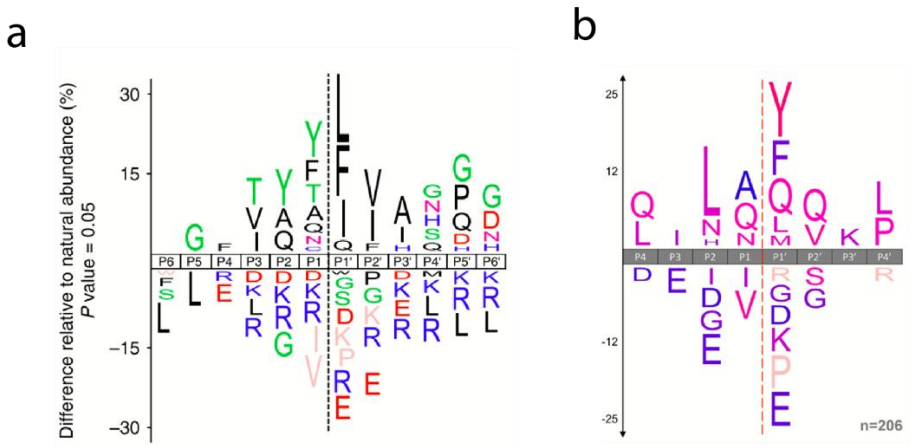


Figure 1.14. IceLogo visualizations of peptide substrate preferences for the *C. haemolyticum* and *H. gracilis* flagellinolysin (HVR construct). (a) The consensus sequence for target specificity for *C. haemolyticum* HVR construct against a trypsin-derived peptide library. (b) The combined consensus sequence for the *H. gracilis* HVR construct against multiple substrates. Figures are from Eckhard et al., 2017, and Eckhard et al., 2020, respectively, and are reproduced with permission.

A second paper in 2020, again by Ekchard and colleagues, characterized a second flagellinolysin in *Hylemonella gracilis*[44]. *H. gracilis* is a freshwater bacterium, noteworthy for its ability to bypass water filters and suppression of pathogenic bacteria[44]. Expressed as a full-length protein, *H. gracilis* flagellinolysin was tested against bovine casein and flagellin from another species of bacterium but produced only minor degradation of the bovine casein[44]. HVR-only constructs fully degraded both substrates but also showed minor autolytic activity, removing a short segment of amino acids from the C-terminus. This allowed for substrate specificity profiling via PICS in the same manner as in *C. haemolyticum.*

A substrate preference for large, hydrophobic residues in the P1' position is present in both flagellinolysins from *C. haemolyticum* and *H. gracilis* (Figure 1.14). A preference for leucine is also exhibited, though it is in the P1' position in *C. haemolyticum* and the P2 position in *H. gracilis.* However, both represent data from constructs, which means they may not match biological conditions. Efforts to isolate full-length flagellinolysin also failed, owing to the protein forming inclusion bodies. The *H. gracilis* construct also engaged in autolysis, cleaving a 1-2 kDa C-terminal segment of itself but leaving the overall HVR region intact. This is intriguing, as *H. gracilis* also degraded the terminal domains of flagellinolysin from *P. tunicata* but left the HVR region untouched[44]. Despite these findings, however, the biological function of flagellinolysin remained unelucidated.

The unique domain architecture of flagellinolysin raises questions regarding the mechanism of its origin. Bioinformatic analysis suggests that flagellinolysins have likely arose as a result of protein domain recombination, followed by significant lateral transfer[43]. Flagellins are frequent sites of intragenic recombination and lateral transfer, as a mechanism for maintaining exterior domain diversity and reducing the ability of the filament to be targeted by antibodies, predators, and so forth[43,71–73]. This also makes plausible the possibility that flagellinolysin has arisen independently in different species in different environments and spread laterally therein. If so, flagellinolysins may serve a variety of niche-specific purposes, rather than being organism-specific.

1.6 Pseudoalteromonas tunicata

Pseudoalteromonas tunicata is a benign marine bacterium, first formally characterized in 1998 by Holmström and colleagues[74]. It is facultatively anaerobic, possesses a single sheathed polar flagellum, and undergoes organized cell death in a biofilm state to organize its biofilms[58,74]. Originally isolated from the interior surface of the tunicate *Ciona intestinalis*, it is ubiquitous in the marine environment, having been detected in samples taken from mussels, pufferfish, sponges, and a range of marine plants[74]. It is also of import to humanity, due to its production of anti-fouling compounds and its ability to control phytotoxic algal blooms[74,75].

In studying a novel biological system, it is always advantageous to begin with the simplest example of it available, preferably in an organism that is otherwise well-studied. C. haemolyticum, where flagellinolysin was initially characterized, is part of a well-studied genus but requires strict conditions for its growth, is polyflagellated, and is a biohazard. This made it undesirable as a test bed for future research.

P. tunicata, by contrast, is also well-characterized but produces only a single unipolar flagellum, is amenable to growth in laboratory conditions, and does not represent a hazard to humans. Having been discovered in the original *Clostridium* neurotoxin study[70], its flagellinolysin also has a reasonable degree of homology (33% across its HVR region) to the clostridial flagellinolysin. This makes it likely that it is an active protease and that the findings of the original study may be generalized to *P. tunicata*.

In addition, the original paper hypothesized that flagellinolysin may interact with hosts or alter its environment. *P. tunicata* is known to colonize many eukaryotic host organisms (though is generally non-pathogenic) and is an excellent biofilm former and invader. These qualities make it an excellent model organism for testing the host interaction and environmental interaction hypotheses.

1.7 Research objectives

The detection of the first flagellinolysins was an important discovery in flagellar biology as it represented the first identification of enzymatic activity present within flagellar filaments[43]. Furthermore, even a rare inclusion of flagellinolysin into flagellar filaments would potentially make flagella among the largest enzymatic complexes known in nature. However, many questions remain unanswered beyond the initial study relating to the biological function of flagellinolysin, including which process(es) it is implicated in. This thesis represents a comprehensive investigation of the function of flagellinolysin in a single organism, beginning with bioinformatics, then biochemical analysis, phenotypic characterization using mutant strains, and finally comparative proteomics to identify possible protease targets.

Chapter 2
General Protocols & Methods

The following chapter describes general methods that are used repeatedly throughout the thesis and which form the basis for later chapters. These included methods for experimental construction of mutant strains, primer sequences, lists of reagents, gel-based assays, etc.

2.1 Strains and plasmids

All strains are maintained in the Doxey Lab's cryogenic freezer at -80ºC, in 15-25% glycerol.

Table 2.1. A list of all strains used in this research.

Strain	Description	Reference
Pseudoalteromonas tunicata D2	Wild-type parent	Holmstrom et al., 1998[74]
P. tunicata E238A	Point mutant; inactivated protease	-
P. tunicata ΔFlaMP	Whole gene deletion mutant	-
E. coli βDH10B	DAP-deficient strain for triparental mating	Rowe-Magnus, 2009[76]
E. coli W3110	Strain used for competition assay	Bachmann, 1972[77]
Ochrobactrum anthropi ATCC 49188	Strain used for competition assay	Chain et al., 2011[78]
Agrobacterium tumefaciens C58	Strain used for competition assay	Wood et al., 2001[79]
Vibrio cholerae (Bang-2)	Strain used for competition assay	Pearson et al., 1993[80]
Vibrio harveyi	Strain used for competition assay	Austin & Zhang, 2006[81]
Salmonella typhimurium ATCC 1331	Strain used for competition assay	McNee, 1917[82]
Salmonella pullorum	Strain used for competition assay	Rettger, 1901[82]

Pseudomonas putida KT2440	Strain used for competition assay	Bagdasarian et al., 1981[83]
Pseudomonas aeruginosa ATCC 10145	Strain used for competition assay	Schroeter, 1872[84]
Staphylococcus aureus Newman	Strain used for competition assay	Newman, 1952[85]
Staphylococcus epidermidis ATCC 12228	Strain used for competition assay	Rosenbach, 1884[86]
Bacillus subtilis	Strain used for competition assay	Ehrenberg, 1835[87]
Streptomyces coelicolor	Strain used for competition assay	Muller, 1908[88]

P. tunicata D2 was provided by the Egan Lab at the University of New South Wales. It expresses the flagellinolysin protein EAR29544.1.

Mutant strains were produced by Dr. Jiujun Cheng as described below. All mutants were generated via homologous recombination using pK19mobsacB[89].

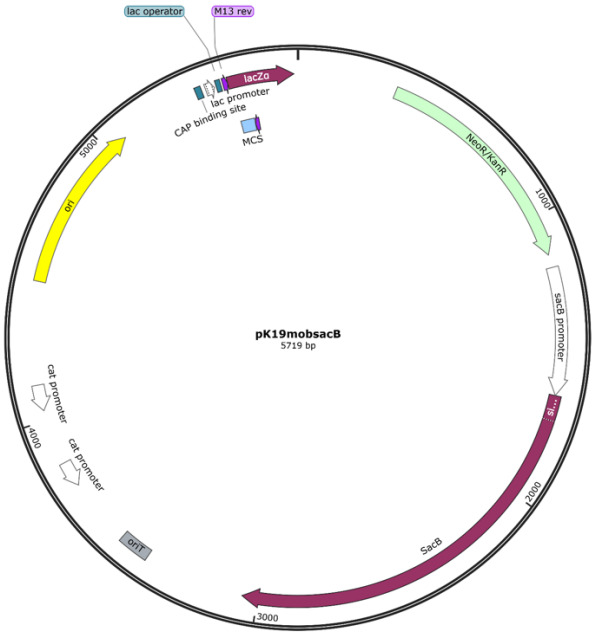


Figure 2.1. A plasmid map for pK19mobsacB (NCBI accession ID: LC257601). The plasmid possesses a neomycin/kanamycin resistance gene (NeoR/KanR) that lacks a promoter. This gene will not express itself until the plasmid is integrated into the host genome by a single crossover event. It also possesses an incP origin of transfer (oriT) to allow it to be transferred with the aid of a helper plasmid. It possesses a sacB gene which produces the enzyme levansucrase, which converts sucrose to levans (high weight fructose polymers). This is lethal in Gram-negative bacteria in the presence of sucrose.

2.2 Construction of mutant strains

P. tunicata mutants were cultured at 30º C for 2 days with shaking at 180 rpm. Cells were collected by centrifugation and lysed with SDS and proteinase K. Phenol-chloroform extraction was performed to separate aqueous DNA from cell debris. RNA was removed with RNase A and the resulting mixture was quantified using a NanoDrop 2000 spectrophotometer (ThermoFisher Scientific). Digested segments were run on 0.8% TAE agarose gel to confirm identity. A number of primers (Table 2.2) were generated to aid in mutagenesis.

Table 2.2: List of oligonucleotide primers used in generating the E238A mutant. The site of the mutation codon is in bold and underlined text.

Oligo ID	Sequence
JC368	CGGCAAGCTTTTATTAATGCCTAAGGCTGTGT
JC369	ATTAGAGGCCATGGTGGCATGCACCAT**AGC**ATGTAATAAGGTTTCATTTAACGCCAA
JC370	TTGGCGTTAAATGAAACCTTATTACAT**GCT**ATGGTGCATGCCACCATGGCCTCTAAT
JC371	GACCGAATTCTGCTGAGGCGCTATTAAGGTTTGT
JC374	GGTTTAATGGGTCGAGGCAC
JC375	GGCATCACTCCTTGAAGATC

To construct a suitable point mutation, PCR was done using primer pairs JC368 and JC369, and JC370 and JC371. The former represents the region of the flagellinolysin gene upstream from the GAA codon representing the glutamic acid residue in the catalytic HEXXH motif. The latter, the downstream region. JC369 and JC370 contain a mutation, replacing the GAA codon with GCT, transforming the glutamic acid into an alanine. These fragments were then purified and pooled in equal quantity to act as PCR template using JC 368 and JC371 as primers. The resulting PCR product was purified and digested with HindIII and EcoRI and inserted in the same sites in pK19mobsacB. The resulting plasmid was dubbed pJC192. The same process was used to generate deletion mutants, using oligonucleotide primers targeting regions upstream and downstream of the flagellinolysin gene.

pJC192 was transformed into *E. coli* βDH10B. This strain has had its *dapA* gene disrupted, requiring exogenous diaminopimelic acid (DAP) to be provided in its growth medium[76]. This allows for absolute counter-selection by culturing the potential transformants on media without DAP. An additional strain of βDH10B carrying helper plasmid pRK600 (Cm^R) was also produced to allow for triparental mating.

P. tunicata was cultured with the transformant and helper strains overnight at 30°C, in LB broth with DAP. The cells were then collected by centrifugation, resuspended in sterile saline, and spotted on

marine agar plates containing 50 µg/mL kanamycin. As only organisms with the plasmid integrated into their genome could produce kanamycin resistance, growth was limited to cells with single crossovers.

After incubation, a single colony was collected and incubated in marine broth without kanamycin. The culture was then serially diluted and streaked on marine agar containing 5% sucrose. This selected against any organisms still hosting the plasmid. The resulting colonies were then tested for kanamycin sensitivity to confirm loss of plasmid.

To confirm successful mutation, E238A and ΔFlaMP mutants had their target regions amplified by PCR and fragment length checked on a gel. The bands were then excised and sent for Sanger sequencing to confirm the mutation.

2.3 Culturing media

Difco marine broth & agar 2216

Unless specified otherwise, all culturing was done in Difco marine broth 2216 or on marine agar 2216. This commercially available media provides optimal growth for marine bacteria by imitating the marine environment's isotonicity and contains trace elements necessary for growth by some organisms. Its composition is available online via Difco and its distributors.

LB & LMB

Lysogeny broth (LB) is a standard medium used for the culturing of *E. coli* and other enteric bacteria. Variations on the recipe abound, including LMB – marine LB broth – to aid in the cultivation of other strains. The variants used here are found in Table 2.3.

Table 2.3: LB & LMB broth recipe

Ingredients	Quantity per L
Peptone	10 g
Yeast extract	5 g
NaCl	5 g (20 g in LMB)

Väätänen Nine Sterile Salts solution (VNSS)

VNSS (Table 2.4) is a common medium used for the culturing of marine bacteria, developed by Väätänen as an alternative to ZoBell's medium. It serves as a useful culturing medium for many organisms but is not as complete as 2216 marine broth.

Table 2.4: VNSS broth recipe[90]

Ingredients	Quantity per L
Peptone	1.0
Yeast extract	0.5
D-glucose	0.5
Soluble starch	0.5
$FeSO_4 . 7H_2O$	0.01
Na_2HPO_4	0.01

2.4 SDS-PAGE

Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) is a standard and common protein protocol throughout the biological sciences. Operating off of the principle of using SDS to correlate protein size with charge, the method allows for the differentiation of proteins by size.

Gel assembly

Gels were assembled using the ingredients listed in Tables 2.5 & 2.6. Separating gel was loaded first and covered with 70% ethanol and allowed to set for an hour. The ethanol was then poured off and replaced with the stacking gel, with the comb inserted and allowed to set again for an hour. Gels were then either used immediately or wrapped in wet paper towel and plastic wrap and stored at 4°C for use the next day.

Table 2.5: 10% separating gel

Ingredient	Amount per gel
Acrylamide (40% v/v)	1.5 mL
1M Tris-HCl (pH 8.8)	2.25 mL
Deionized water	2.16 mL
10% SDS	60 uL
10% APS	30 uL
TEMED	3 uL

Table 2.6: 18% stacking gel

Ingredient	Amount per gel
Acrylamide (40% v/v)	0.3 mL
1M Tris-HCl (pH 6.8)	0.3125 mL
Deionized water	1.95 mL
10% SDS	25 uL
10% APS	12.5 uL
TEMED	2.5 uL

Sample preparation and gel electrophoresis

Protein samples are usually prepared first by lysing cells, which may be done by sonication, glass bead shearing, French press, ethanol, or heat. Different methods were each tested with *P. tunicata* and lysis by heat (10 minutes at 95°C in denaturing buffer) in a single step was found effective for SDS-PAGE. Once the protein is extracted, it is quantified via NanoDrop and up to 20 µg is mixed in a 3:1 ratio with 4X loading buffer, the recipe for which is below.

Table 2.7: Recipe for 10 mL of 4X SDS-PAGE sample loading buffer.

Ingredient	Volume/quantity
1M Tris HCl, pH 6.8	2.5 mL
ddH_2O	0.7 mL
SDS	1.0 g
0.1% Bromophenol Blue	0.8 mL
100% glycerol	4 mL
14.3M β-mercaptoethanol	2 mL

The sample is then heat-denatured for ten minutes by heating in a heat block at 95°C. After that time, the samples are transferred into the wells of the SDS-PAGE gel.

Once loaded onto the gel, the samples are run for 20 minutes with voltage held constant at 90V, to allow the samples to stack. The samples are then run with voltage held constant at 140V for 60 minutes or until the dye front reached the end of the gel, whichever came first.

Coomassie Blue staining

After running, the gel was prepared either by fast prep or slow prep. Fast prep gels were produced more quickly but had higher background and less distinct bands, making them incapable of detecting any but the thickest and most evident bands. Slow prep requires additional steps but produces clearer, cleaner bands. All stains were done using Coomassie Blue. Recipes for this protocol are as follows:

Table 2.8: Fixing solution

Ingredients	Volume
Ethanol	500 mL
Acetic acid	100 mL
Deionized water	400 mL

Table 2.9: Coomassie Blue staining solution

Ingredients	Volume/powder weight
Methanol	500 mL
Acetic acid	100 mL
Deionized water	400 mL
Coomassie Blue powder	1 g

Destaining solution is identical to staining solution in composition, except it lacks the Coomassie Blue dye.

Fast preparation gels

Fast prep gels were briefly rinsed with deionized water and then submerged in staining solution. The submerged gel was microwaved for 45 seconds on high and then shaken gently for 15 minutes before being transferred to destain solution with shaking overnight. Absorbent tissue wadding was added to absorb released stain.

Slow preparation gels

Slow prep gels were washed in fixing solution for an hour with gentle shaking, followed by washing in fresh solution overnight. The next day, the gel are immersed in staining solution for 3-4 hours, followed by at least three rounds of destaining for 30 minutes each. Once background is minimized and bands are visible, the gel is submerged in 5% acetic acid solution for at least an hour to equilibrate and regain its original size.

2.5 Constructs

Gene and plasmid constructs were generated in collaboration with Dr. Ulrich Eckhard and through services provided by Biobasic Inc.

Full-length flagellinolysin (PtuFlaMP-FL; M1-Q605), metallopeptidase domain alone (PtuFlaMP-HVR; S142-G497), and full length minus the terminal flagellin domains (PtuFlaMP-FL2; D42-R563) construct sequences were developed and codon-optimized to eliminate common restriction sites, followed by appending of His6-SUMO tag sequences. These genes were inserted into pET21a and transformed into *E. coli* BL21 DE3 cells for expression as tagged proteins. Cells were then lysed and the proteins were refolded on nickel affinity columns using decreasing concentrations of urea. The flow-through was treated with SUMO protease. The recombinant proteins were then isolated via

rechromatography and size exclusion chromatography. Catalytically inert variants were also generated for all three protein products. These inert variants had the glutamic acid residue in the HEXXH zinc metalloprotease motif replaced with an alanine.

Zymography gel assays

Zymography gel assays are variants of SDS-PAGE. Proteins were run on a 12% SDS-PAGE gel containing 1.0 mg/mL bovine casein. Zymogram gels were then washed with renaturation buffer (Table 2.10), followed by incubation in developing buffer (Table 2.11) for two hours at room temperature. The zymograms were then imaged and observed for degradation of the casein, denoted by zones of clearance in the gel. These zones of clearance are taken as positive signals of protease activity.

Table 2.10: Renaturation buffer composition

Ingredients	Concentration
Triton X-100	2.5% (v/v)
Tris (pH 8.0)	50 mM
NaCl	200 mM
EDTA	5 mM

Table 2.11: Developing buffer composition

Ingredients	Concentration
Brij-35	0.02% (v/v)
Tris (pH 8.0)	50 mM
NaCl	200 mM
$CaCl_2$	20 mM
$ZnCl_2$	20 mM

2.6 Western blot and antibody construction

Antibody generation

Antibodies were generated through BioBasic Inc. To generate antibodies, PtunFlaMP-HVR (D42-R563) construct was used to immunize two New Zealand rabbits in four 500 µg/immunization doses over the course of 7 weeks. Sera were tested after the third immunization to detect successful antibody generation via ELISA, with the rabbits exsanguinated one week (week 8) after the final dose. Final sera titers were tested again by ELISA and purified antibody was collected via a chromatography column of HVR bound to CNBR-activated Sepharose. Antibody was then eluted by elution buffer, followed by rechromatography and final titer was measured by ELISA.

Secondary antibody was goat anti-rabbit IgG (H+L) HRP (horseradish peroxidase) conjugate, provided commercially by Bio-Rad. The peroxidase interacts with the luminol in the Clarity Western ECL developer fluid to produce chemluminescence.

Western blotting

Western blots were used to detect the presence of a single protein within a sample. The sample of interest is run on an SDS-PAGE gel as described in section 2.4. Rather than proceeding to staining, however, the protein contents are transferred onto a blotting membrane and mixed with the primary antibody and a secondary detecting antibody. The secondary antibody provides a measurable signal to indicate if a target is present.

Once a sample of interest was run in a gel, the gel was briefly rinsed with deionized water and transferred to a solution of Towbin buffer (Table 2.12) and allowed to equilibrate for 15 minutes. Thick filter paper pads and nitrocellulose membrane were also equilibrated in buffer for at least five minutes. The “transfer sandwich” was assembled in a Trans-Blot SD transfer system (Bio-Rad) and current was run with constant voltage at 15V for 30 minutes.

After transfer, the membrane is washed in TBS (Table 2.13) for ten minutes and blocked in blocking buffer for an hour. Excess blocking buffer is then removed with two washes in TBS-T (TBS + 0.5% Tween 20). Primary antibody was applied at a concentration of 1.25 µg/mL and the membrane was incubated for either an hour or overnight in the primary antibody. The membrane was then washed six times for ten minutes each in TBS-T to remove any unbound antibody, followed by incubation in secondary antibody at a concentration of 1:3,000 in blocking buffer. Six washes follow in TBS-T, followed by a final wash in TBS. The membrane is then incubated in Clarity Western ECL substrate (BioRad) for five minutes before imaging.

Buffer recipes

Table 2.12: Towbin buffer recipe

Ingredient	Volume/quantity per 1 L
Tris base	3.03 g
Glycine	14.4 g
Deionized water	800 mL
Methanol	200 mL

Table 2.13: Tris-buffered saline (TBS) recipe

Ingredient	Volume/quantity
Tris base	4.84 g
NaCl	58.48 g
Deionized water	1.5 L

TBS requires its pH to be adjusted to 7.5, using 4M HCl. The final volume can then be adjusted to 2 L with deionized water. TBS-T is made by adding 0.5 mL of Tween 20 per 1 L of TBS. Blocking/antibody buffers are produced by adding 5 g of non-fat powdered milk to 100 mL of TBS.

All primary antibody quantities were kept constant – and repeated freeze-thaw cycles avoided - via the production of one-time use antibody tubes. Each tube contained 125 µg of antibody in 250 uL of dilution buffer.

Ponceau staining

Post-transfer, nitrocellulose membranes were placed in Ponceau stain (Table 2.14) with gentle shaking for five minutes. The membrane was then gently rinsed with distilled water to remove the excess stain. The membrane was then imaged and destained. Destaining was done by washing five times with TBS-T for 10 minutes each. After this, the membrane was blotted as described in the previous section.

Table 2.14: Ponceau stain recipe

Ingredient	Quantity per 1 L
Deionized water	950 mL
Glacial acetic acid	50 mL
Ponceau Red dye powder	1 g

2.6 Flagellar preparation

Flagellar preparation involves the shearing and concentration of bacterial flagellar filaments for more detailed analysis (Figure 2.2). Owing to their existing on the exterior of the cell and their considerable size, this separation and concentration can be achieved by using shear forces to separate the flagella from the cells, followed by sequential centrifugation.

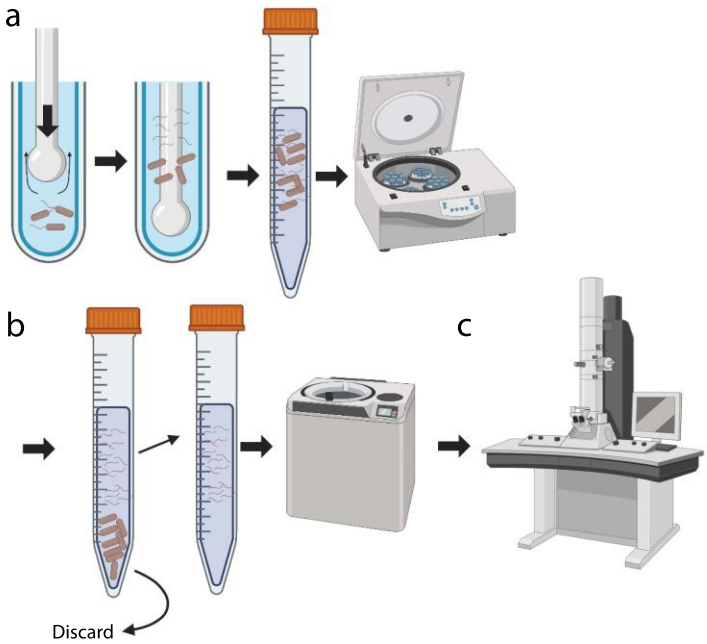


Figure 2.2. Cartoon showing the process of isolating flagellar filaments. (a) Flagella are sheared from cells using a Dounce homogenizer the mixture of cells and flagella are centrifuged to remove the cells. (b) Cell pellets are discarded and the supernatant containing the flagella is collected and ultracentrifuged to force flagellar sedimentation. (c) The concentrated flagella are applied to grids and stained (not shown) and viewed under an electron microscope.

100 uL of overnight cultures were subcultured into 600 mL of fresh marine broth and cultured with shaking for eight hours to produce a log-phase culture. These cultures were then centrifuged in an Eppendorf 5910 R refrigerated centrifuge at 4°C for 18 minutes at 4,500x *g*, using an S-4x Universal rotor adaptor. The cell-free supernatant was discarded and the cell pellet was resuspended in 50 mL of TBS with G-Biosciences ProteaseArrest protease inhibitor and EDTA. These were stored on ice.

25 mL aliquots of cells were added to an ethanol-sterilized 30 mL Dounce tissue homogenizer. The raising and lowering of the homogenizer generates shear forces along the edge of the Teflon homogenizer, removing filaments from cells. Each aliquot was pumped ten times total with the homogenizer.

The tissue-separated aliquots were centrifuged at 5,000x *g* at 4°C for 15 minutes to pellet cells and cell debris. The supernatant was collected and recentrifuged for the same time in the same conditions to remove any remaining cells. The supernatants were then ultracentrifuged at 110,000x *g* for one hour, concentrating the flagella into a pellet. These pellets were resuspended in 100 uL of buffer with inhibitor and vigorously pipetted and hand-warmed to encourage dissolution.

Chapter 3

Identification, expression, and localization of flagellinolysin in *Pseudoalteromonas tunicata*

This chapter covers the bioinformatic identification and structural modeling of flagellinolysin in *Pseudoalteromonas tunicata*, the investigation of its genetic context and promoter regulatory motifs, the experimental confirmation of its mRNA expression, and the use of immunoelectron microscopy to determine its localization.

3.1 Introduction

The bacterial flagellum is an organelle of remarkable complexity that is responsible for motility and chemotaxis. In flagellated bacteria, a membrane-spanning molecular motor formed by ~20 different proteins drives rotation of an extracellular filament to provide swimming or swarming-type motility. Flagella also mediate cell attachment to external substrates and host cells[13,15], facilitate mechanosensing, and enable complex tunnelling-like behaviours to explore and disperse from surfaces[17]. Flagella are thus not only critical to the locomotory lifestyle of planktonic cells, but play multifaceted roles in biofilm formation and development.

The largest component of any flagellum is the filament, which can be considerably longer than the organism itself. The production of flagella is thus a metabolically expensive and tightly regulated process, with as much as 2% of the cell's entire metabolic output devoted to flagellar assembly[51]. Flagellar filaments are composed of tens of thousands of copies of the structural protein, flagellin[12], which in turn is composed of three protein domains: The highly conserved helical N- (Pfam ID: PF00669) and C-terminal (PF00700) domains and a central, hypervariable region (HVR). Once polymerized, the N and C-terminal domains of different flagellins interact with each other to form the interior core of the filament, while the HVR faces outward and is exposed on the filament surface. HVR domains are - as the name implies - extremely variable in sequence, and mostly

uncharacterized, but critical to the properties of filaments and how they interact with the external environment and host organisms[91,92].

In 2017, Eckhard and colleagues discovered a highly unusual variant of bacterial flagellin called flagellinolysin, which is the only flagellin identified to date with enzymatic function[43]. Flagellinolysins are unique from typical flagellins by possessing a conserved protease domain within the HVR region that resembles M9 zinc metallopeptidases, sharing the catalytic HExxH zinc-binding motif. Flagellinolysins have since been classified as their own family of proteases, forming M101 metalloproteases in the MEROPS classification, and demonstrated to possess proteolytic activity[43,44]. Recombinantly produced flagellinolysins have been studied in both gram-positive bacteria (the animal pathogen, *Clostridium haemolyticum*) and gram-negative bacteria (the freshwater organism, *Hylemonella gracilis*). Additional homologs of flagellinolysin are present in additional genomes of species across the bacterial tree, suggesting that flagellinolysins may be a widespread component of flagella in many bacterial lineages.

Despite the broad occurrence of flagellinolysin across bacteria, the function of flagellinolysins remain elusive. It has been speculated that flagellinolysins may facilitate extracellular proteolysis, function in inter-species interactions, virulence, and other activities[43]. Protease assays to study flagellinolysin substrate specificity have been performed to uncover potential protein targets that may hint at their biological role[43,44]. However, these studies have revealed a relatively broad target specificity with a general preference for hydrophobic residues, making it difficult to narrow down their true physiological targets. In *H. gracilis*, the recombinant flagellinolysin protease domain has even been shown to be capable of cleaving the flagellins of other species and will undergo self-cleavage of its N- and C-terminal domains in the absence of any other substrate. The physiological protein targets of flagellinolysins, and their biological role, therefore, is currently unknown.

Although flagellinolysin from *C. haemolyticum* had been investigated experimentally in Eckhard et al. (2017), the pathogenic nature of this organism and its anaerobic requirements limits the use of *C. haemolyticum* as a model organism for further investigations of flagellinolysin function. Thus, a new

flagellinolysin-containing species is needed for biological studies. *Pseudoalteromonas tunicata* was one of the four species in which flagellinolysin was initially discovered[70], and represents a promising candidate as a model organism for examining flagellinolysin function. As described in Chapter 1, *P. tunicata* has been thoroughly investigated for its role in biofilm development and host tissue colonization. It is conceivable that the unique abilities of *P. tunicata* in biofilm invasion and interspecies competition could be consistent with a role for flagellinolysin given the critical involvement of flagella in biofilm development. Furthermore, P. tunicata is a non-pathogenic organism, can be cultured aerobically, and has been relatively well studied with its genome/proteome and major microbiological characteristics defined [57,58,74,93–96].

In this chapter, I establish the marine organism *Pseudoalteromonas tunicata* as a new model species for examining flagellinolysin function. First, I identify flagellinolysin within the *P. tunicata* genome, predict its structure, and examine its genomic context and regulatory motifs. Then, I develop a recombinant construct of the HVR (protease domain) from *P. tunicata* flagellinolysin to perform biochemical assays on the recombinant protein. Then, in cultures of P. tunicata I examine the mRNA and protein expression of flagellinolysin and detect it as an actively expressed gene and protein. Finally, using an antibody raised against the flagellinolysin HVR, I perform immunoelectron microscopy to characterize its localization.

3.2 Methods

3.2.1 Bioinformatic identification and analysis of flagellinolysin in *P. tunicata*

Detection and distribution of flagellinolysins across the bacterial tree of life

To gather a list of candidate flagellinolysins, three rounds of PSI-BLAST were performed against the NCBI's nr (non-redundant) database, using the following HVR region of *P. tunicata* flagellinolysin as the query. Note: the proteolytic HEXXH motif is shown in bold/underline.

> EAR29544.1 (residues 110 to 534)

INEEFIELKREINRISDATSFGSKKLFQKASGSLVDTVERDMVKALQRTWLSESEDIIQDQLGLMGRGTL
KIDLEYVDGDGGTAAYVKPAYSGGQAVQMVMVIDKGDFATSADIHDDSSLALNETLL**HEMVH**ATM
ASNFQLTNLPIWFLEGTAEAIRGADSRVAQDSAGLGLKNIIDDGAAGGNNSLKDLQGVMPFSPGTAAE
VAAVYSGGYLAMRYMYQEIGQNGLQSLMSELTTGQSFDASLNTASGAIWSNEADFFTDMMTTGEDP
LDATKTRLQIFVENKMDLTNLDNGALGGKDASDGEVRENTMQGTGSGRLDGSKGFQEFLVLDDNDN
SRSDFDTGATYDKPGGGQIRLQDYETQVHGAGGEYINLQVGARSQQVIGFTLGSFSTENLGLEYTELV
DQPQFAIFAIDDALTIVDKQRA

The search was performed with an E-value threshold of 0.05 and a limit of 1,500 sequences in the first round, 5,000 in the second, and 10,000 in the third. Results were collected at each round, then consolidated and filtered for redundancies.

The consolidated sequence list was scanned using HMMER's hmmscan PFAM-alignment tool against the PFAM database of protein families. The results were then filtered to only those which contained at least one hit for a flagellar PFAM alignment and the HExxH proteolytic motif. This produced a final total of 732 proteins. This final list was used in conjunction with AnnoTree to assemble a phylogenetic tree, providing a broad overview of the distribution of flagellinolysin in the bacterial branch of the tree of life.

Gene neighbourhood analysis

The above analysis was repeated to identify flagellinolysin homologs in the Genome Taxonomy Database (GTDB) release 95 using Annoview v1.0 (annoview.uwaterloo.ca). The identified flagellinolysin-containing genomes were then further filtered to organisms within the *Aeromonadales* order, containing the species of interest *Pseudoalteromonas tunicata*. Gene neighborhoods surrounding putative flagellinolysin genes (10 kb upstream and downstream) were then visualized using AnnoView and arranged vertically based on increasing phylogenetic divergence from *P. tunicata*. The KEGG, TIGRFAM, and PFAM annotations were then used to assign gene functions and names.

Upstream promoter analysis

Upstream sequences of 100 nt (nucleotides) were collected from 262 identified flagellinolysin genes from sequenced genomes available in the NCBI and run through MEME (Multiple Em for Motif Elicitation) with zoops (zero or one instance per sequence) distribution. The results were run through Tomtom against the CollecTF database and examined for any hits with significance.

Structure prediction by homology modeling

The structure of *P. tunicata* flagellinolysin was predicted using two independent methods, ColabFold[97] and Phyre2[98].

ColabFold is an AlphaFold 2-derived tool that uses sequence data aligned to a database of MSAs (multiple sequence alignments) to provide a rapid guide to structural prediction for AlphaFold 2. The process is described in greater detail in a paper by Mirdita and colleagues, describing the tool itself[97]. The *P. tunicata* flagellinolysin sequence was used with ColabFold to generate a predictive model.

Phyre2 modeling was done in collaboration with Dr. Ulrich Eckhard. A structural model of the PtuFlaMP hypervariable region was generated using the intensive modeling procedure of Phyre2[98,99] with the peptidase domains of clostridial collagenases H (PDB entry 4ar1), T (PDB entry 4ar8), and G (2y50) as templates[100,101]. Energy minimization and secondary structure assignment were performed within the UCSF Chimera package[102], and molecular graphic figures were prepared using the molecular visualization system PyMOL (www.pymol.org)[103].

3.2.2 Experimental detection of flagellinolysin expression

qrtPCR measurement of flagellinoysin mRNA expression

Primers were generated for five target genes (Table 3.1) in *P. tunicata* D2: Flagellinolysin, two structural flagellins, and two constitutive genes – 16s rRNA and erythrose-4-phosphate dehydrogenase. Primers were generated using Primer3, with product size ranges limited to 50-200

nucleotides and primer size limited to 18-23 nucleotides with an optimum target of 20. Melting temperature (Tm) was set for a minimum of 54ºC and a maximum of 63ºC and an optimum of 60ºC. GC% minimum was set to 20%, maximum to 80%. Estimated annealing oligo concentration was set to 50 nM. Maximum self-complementarity was left at default. Each search produced five candidates, with the ideal chosen based on melting temperature.

Table 3.1: A list of the primers developed for use in qrtPCR.

Oligo ID	Target mRNA (accession ID)	Sequence	Melt. temp
PtD2-16s-L	16s rRNA (Z25522.1)	TGATAAACCGGAGGAAGGTG	57ºC
PtD2-16s-R	-	CGTAAGGGCCATGATGACTT	-
PtD2-Flysin-L	Flagellinolysin (EAR2544.1)	ACAGGGCAAAGTTTTGATGC	56ºC
PtD2-Flysin-R	-	GCTTCGTTGCTCCAAATAGC	-
PtD2-Flag1-L	Flagellin (EAR29563.1)	CGGTGCAAAACCGTTTATCT	55ºC
PtD2-Flag1-R	-	TTCTTGATGCTGACGCATTC	-
PtD2-Flag2-L	Flagellin (EAR29565.1)	GTACAGCACTAGGCGCACAA	59ºC
PtD2-Flag2-R	-	CCCATATTGGCTCGTTGACT	-
PtD2-E4D-L	Erythrose-4-phosphate dehydrogenase (AXT31331.1)	CTTGATTCACCTGCCGATTT	55ºC
PtD2-E4D-R	-	GCATGCTCACGAGAATGAAA	-

Samples were collected from overnight cultures (marine broth with shaking at 24°C) and lysed in RLT buffer using an Omni TH rotor-stator homogenizer (Omni International). RNA was collected using an RNeasy kit (Qiagen) and reverse-transcribed using a QuantiNova reverse-transcription kit (Qiagen). The resulting cDNA was quantified and 100 ng of each sample was loaded in triplicate on a PCR plate in SYBR master mix (Qiagen) and measured using CFX Maestro (BioRad). Where kits were used, their specifications were followed. Their manuals are all available online from Qiagen.

Shotgun mass spectrometry

In-depth mass spectrometry (mass spec) services were provided by SickKids Hospital in Toronto and by Bioinformatics Solutions in Waterloo, while initial characterization was provided by the University of Guelph's mass spectrometry facility. Liquid culture samples were cultured overnight, while biofilm samples were cultured for 24, 48, or 72 hours. All liquid cultures were sent for high sensitivity mass spec. Flagellar samples were collected as described in Section 2.1.6 and SDS-PAGE gel bands in the 60-65 kDa range (the range for *P. tunicata* flagellins) were cut out and sent to the University of Guelph for rapid identification.

Western blotting

Cells were cultured in marine broth with shaking overnight and, with subculturing, to 12, 24, and 48 hour time points before total protein was harvested. When this did not yield detection, purified flagellar samples were used instead, cultured and harvested as outlined in Section 2.1.6. All samples were prepared and run in SDS-PAGE gels as outlined in Section 2.1.4 and blots were run as outlined in 2.1.5.

3.2.3 Recombinant expression, purification, and enzymatic characterization of flagellinolysin

Note: The following work was done in collaboration with Dr. Ulrich Eckhard and using services provided by Biobasic Inc.

Variants of the flagellinolysin from *Pseudoalteromonas tunicata* (PtuFlaMP; UniProt ID A4C6E0) were recombinantly expressed as His6-SUMO tagged proteins in BL21 DE3 cells as inclusion bodies, refolded on metal affinity columns using decreasing concentrations of urea, treated with home-made SUMO protease for tag-removal, followed by re-chromatography and size exclusion chromatography, as described for the homologous protein from *Hylemonella gracilis*[44] (HgrFlaMP). Both active (HE^{238}xxH) and mutationally inactivated variants were produced, including full-length flagellinolysin (PtuFlaMP-FL; M1-Q605), the central metallopeptidase domain of the hypervariable region alone (PtuFlaMP-HVR; S142-G497), and a full-length variant lacking the terminal D_0-domains (PtuFlaMP-FL2; D42-R563). The enzymatically "dead" mutants were produced as HA^{238}xxH variants, where the general base glutamate of the catalytic HExxH motif is exchanged to an alanine residue, rendering the enzyme inactive.

SDS PAGE and casein zymography

Proteins were separated using 12% SDS-PAGE in the presence of β-mercaptoethanol and the PageRuler Prestained Protein Ladder (Thermo Scientific) as a molecular weight reference, and stained using Coomassie Brilliant Blue G250 (BioRad). For casein zymography, flagellinolysin samples were separated by SDS-PAGE containing 1.0 mg/ml bovine casein (Sigma Aldrich). Zymograms were then washed with renaturation buffer (2.5% Triton X-100 in 50 mM Tris, 200 mM NaCl, 5 mM EDTA, pH 8.0) and incubated in developing buffer (0.02% Brij-35 in 50 mM Tris, 200 mM NaCl, 20 mM $CaCl_2$, 20 µM $ZnCl_2$, pH 8.0) for approximately 2 hours at room temperature[44]. Of note, similar to HgrFlaMP, PtuFlaMP does not contain any cysteine residues. Thus, the protein behaves identical on reducing and non-reducing SDS-PAGE.

3.2.4 Immunolabeled transmission electron microscopy (ImmunoTEM)

Transmission electron microscopy relies on the deflection of electrons fired from a tungsten filament by dense, electron-rich atoms bound to a formvar mesh atop a nickel grid. The transmitted electrons are received by sensors and used to generate an image by representing the areas of greater deflection as darker and greater transmission as lighter, not dissimilarly to a light microscope. The key difference is that an electron has a much smaller wavelength than visible light, allowing far greater magnification.

ImmunoEM uses primary-secondary antibody labeling of samples, similar in principle to Western blots. Here, the secondary antibody is labeled with a sphere of colloidal gold. The electron-dense gold deflects the electron beam, producing a visible dark spot on the detector image. This allows for direct detection of the protein or other substance of interest in the sample.

To prepare samples for immunoEM, 10 uL of sample was applied to a formvar-coated nickel grid and allowed to attach for 2 minutes before having excess drained by tapping the edge of the grid gently to filter paper. The samples were then applied face-down to 20 uL droplets of blocking buffer (see 2.1.6 for more details on buffer recipes) for 30 minutes. This was followed by applying the grids to droplets of primary antibody in antibody buffer for 30 minutes (for cellular samples) or overnight (for flagellar filament samples). The grids were then washed five times in TBS-T with 5% milk protein by applying each grid to a new droplet for one minute each. Secondary immunolabeling was performed by applying samples to droplets with secondary antibody for 30 minutes at a 1:300 concentration. The samples were then washed three times in TBS-T with milk protein and three times in sterilized deionized water.

The grids were fixed and background stained by dipping each for 30 seconds into a 0.1% ammonium molybdate solution. Any remaining molybdenum stain was removed by touching to filter paper. To maximize drying prior to use, the grids were then placed in a sample box and covered with KimWipe paper towel, and left in a fume hood to finish drying overnight.

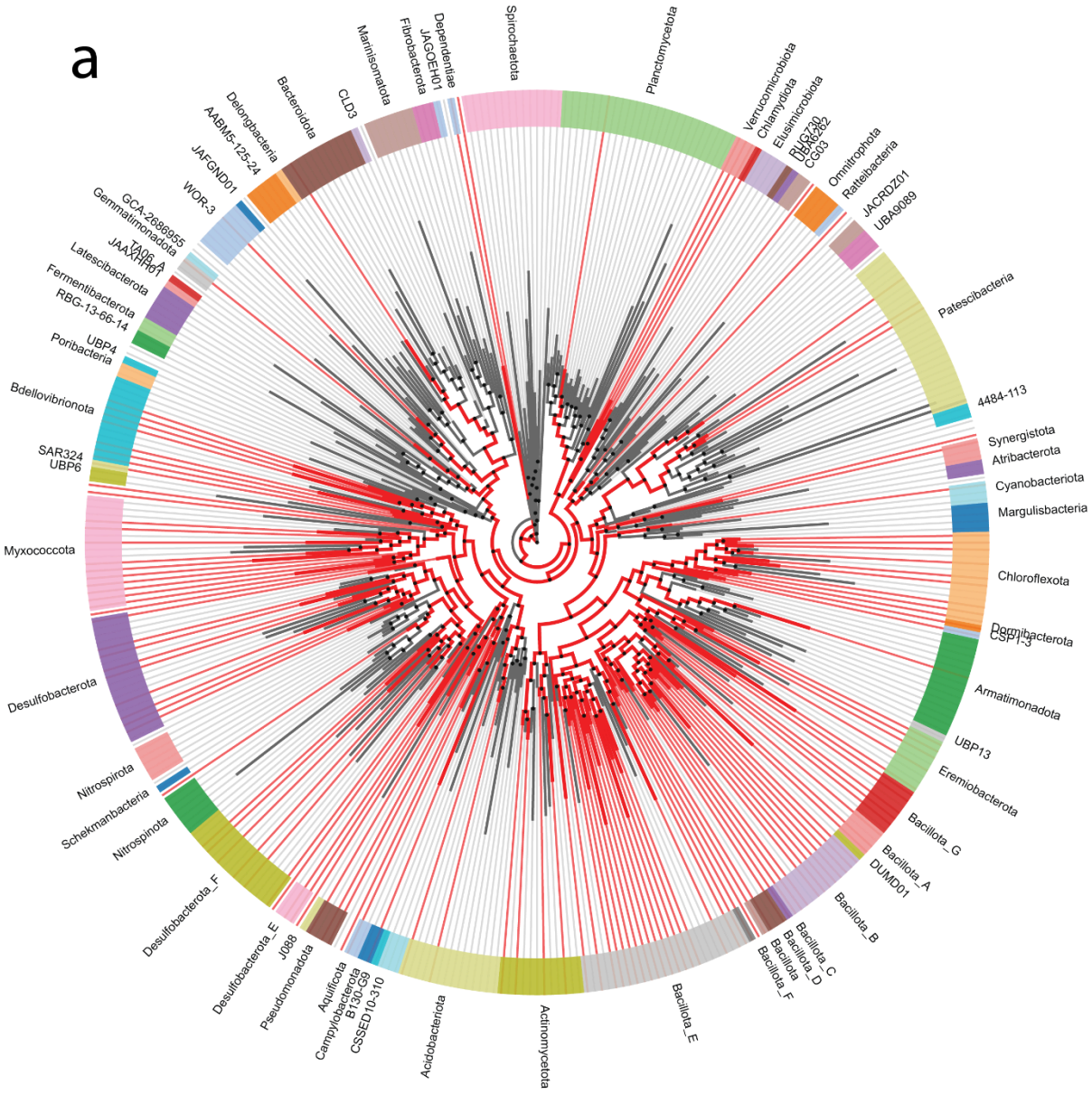
3.3 Results

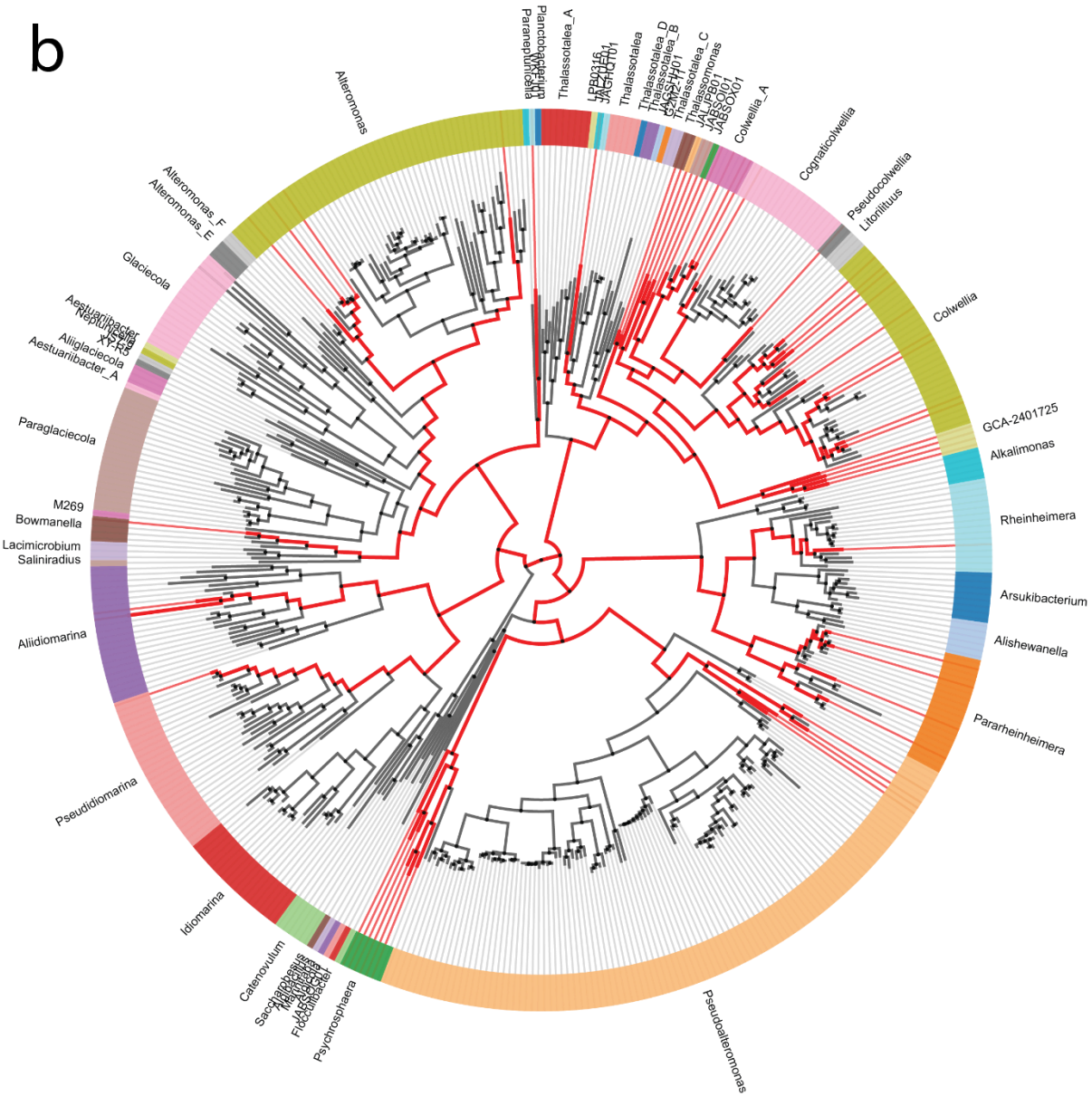
3.3.1 Identification of flagellinolysin genes across the bacterial tree of life

To produce a list of candidate flagellinolysin-containing species, three rounds of PSI-BLAST were performed against the NCBI's nr (non-redundant) database, using the HVR region of flagellinolysin as a query. The identified species were then mapped onto AnnoTree for visualization (Figure 3.1a).

The 732 identified flagellinolysins have a widespread taxonomic distribution, including in groups not previously noted as possessing flagellinolysins (Figure 3.1). The largest number of flagellinolysins of any given order are found in the *Pseudomonadota* (synonym *Proteobacteria*), followed by the *Firmicutes*, which accords with the findings from Eckhard et al. (2017). The *Gammproteobacteria* constitute the majority of identified flagellinolysins within the *Pseudomonadota*. New to this tree are the finding of a large presence in the *Myxococcota*, as well as sporadic distribution into other taxa. This also includes the first detections of flagellinolysins in the FCB group, identified within a collection of aquatic and sub-crustal aquatic members of the *Flavobacteriales*, found as far afield as Lake Baikal (an inland lake near central Mongolia) and the North Pond (subsurface sediments of the deepest portions of the Atlantic ocean).

While most clades have at least one instance of an identified flagellinolysin, the abundance within a given clade varies considerably, with some having only one or two genome hits amongst hundreds or thousands of database genomes and others showing hits in >60% of database genomes (e.g., the Bacillota G group). Even where flagellinolysin is common, however, its distribution is typically patchy, with an organism's closest relative frequentl missing the gene but a more distant one possessing it.


a
Spirochaetota
Planctomycetota
Verrucomicrobiota
Chlamydiota
Elusimicrobiota
RIF-GZ20
UBA6262
CG03
Omnitrophota
Rattelbacteria
JACRDZ01
UBA9089
Patescibacteria
4484-113
Synergistota
Atribacterota
Cyanobacteriota
Margulisbacteria
Chloroflexota
Dormibacterota
CSP1-3
Armatimonadota
UBP13
Eremiobacterota
Bacillota_G
Bacillota_A
DUMD01
Bacillota_B
Bacillota_C
Bacillota_D
Bacillota_F
Bacillota_E
Actinomycetota
Acidobacteriota
CSSED10-310
B130-G9
Campylobacterota
Aquificota
Pseudomonadota
J088
Desulfobacterota_E
Desulfobacterota_F
Nitrospinota
Schekmanbacteria
Nitrospirota
Desulfobacterota
Myxococcota
SAR324
UBP6
Bdellovibrionota
Poribacteria
UBP4
RBG-13-66-14
Fermentibacterota
Latescibacterota
JAAXTH01
TA06_A
Gemmatimonadota
GCA-2696955
WOR-3
JAFGND01
AABM5-125-24
Delongbacteria
Bacteroidota
CLD3
Marinisomatota
Fibrobacterota
JAGOEH01
Dependentiae


b
Alteromonas
Alteromonas_F
Alteromonas_E
Glaciecola
Aestuaribacter
Aestuaribacter_A
Aliiglaciecola
Paraglaciecola
M269
Bowmanella
Lacimicrobium
Saliniradius
Aliidiomarina
Pseudidiomarina
Idiomarina
Catenovulum
Psychrosphaera
Pseudoalteromonas
Pararheinheimera
Alishewanella
Arsukibacterium
Rheinheimera
Alkalimonas
GCA-2401725
Colwellia
Litorilituus
Pseudocolwellia
Cognaticolwellia
Colwellia_A
Thalassotalea_C
Thalassotalea_B
Thalassotalea_D
Thalassotalea
Thalassotalea_A
Planctobacterium
Paraneptunicella

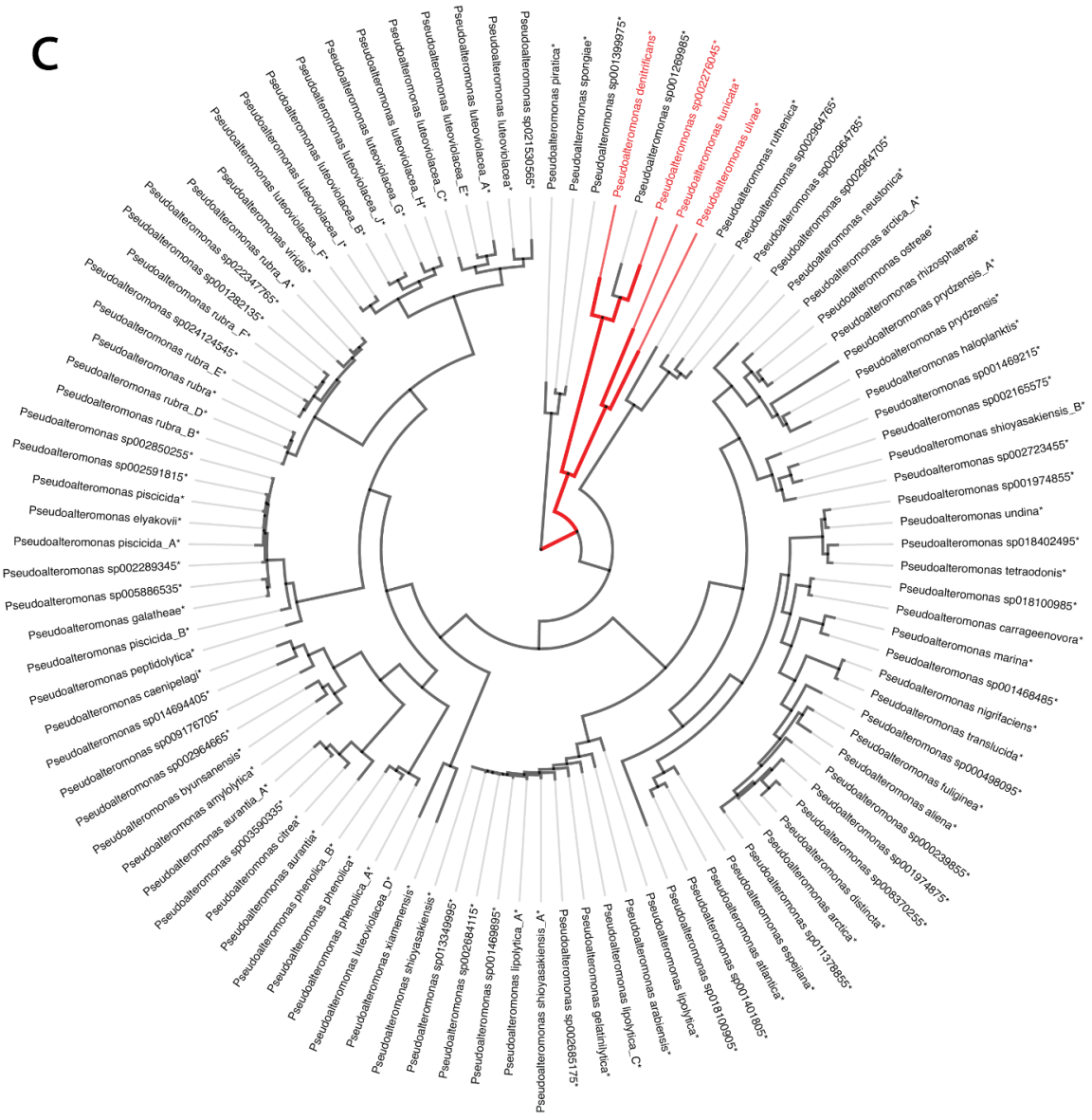


Figure 3.1. Phylogenetic distribution of flagellinolysin-containing species across the bacterial tree of life. (a) A phylum-level tree derived from the AnnoTree[104] database with flagellinolysin-containing lineages marked in red. Flagellinolysin is found across a wide range of phyla. It is most abundant in the *Pseudomonadota* but more highly represented in others (e.g., 67.5% of genomes in the *Bacillota* E group). (b) A genus-level tree, showing genera within the *Alteromonadales*. Flagellinolysin is most prevalent in *Colwellia*, though a few taxa have higher representation (e.g. Psychrosphaera, 85.7% of genomes hit). (c) A species tree within the genus *Pseudoalteromonas*. Only four species were identified within the clade as possessing flagellinolysin, all closely-related.

Literature analysis of the identified flagellinolysin-containing species was unable to reveal a common environment or niche. Flagellinolysin-possessing bacteria are found in soil, freshwater environments, marine water, sub-surface sediments, and on and in a variety of host organisms. Some are pathogenic, many are benign. Its presence in *Myxococcota* – not normally known to be capable of producing flagella – may be indicative of pseudogenization in some circumstances or even adaptation to other T3SS-related purposes.

Pseudoalteromonas tunicata is a fairly typical example within this distribution pattern. Although found within the *Gammaproteobacteria*, the distribution within the family *Alteromonas* is patchy, with no characterized genus carrying flagellinolysin in all known instances. This suggests *P. tunicata* represents an example of flagellinolysin being laterally transferred at some point before it diverged from its relative, *P. ulvae*.

3.3.2 *P. tunicata* flagellinolysin is encoded within a flagellar gene cluster

Analysis of the flagellinolysin gene cluster within *P. tunicata* reveals that flagellinolysin is surrounded by other flagellar genes (Figure 3.2), including the flagellar chaperone *fliS* encoded immediately upstream, another flagellin gene immediately downstream, and the biofilm/flagellar regulator *fleQ*. In *Pseudomonas*, *fleQ* has been demonstrated to control the transition from free-swimming flagellated states to the biofim state by suppressing or activating transcription of biofilm-associated genes[60,105]. There are also a number of genes for hook-associated proteins and, further downstream, a cluster of basal body-associated genes.

Interestingly, upstream of flagellinolysin there are also genes that have been implicated in flagellar filament glycosylation. The leftmost gene upstream of flagellinolysin (*neuB*, Figure 3.2) encodes an N-acetylneuraminate synthase family protein. This protein family includes synthases of pseudaminic acid and legionaminic acid, which are known to be incorporated into flagellins of some bacterial species via O-linkage[106–109]. The presence of these genes in the genome suggests that *P. tunicata* flagellar filaments contain glycosylated flagellins.

Similar flagellinolysin-containing gene clusters can also be found in the genomes of related species including other *Pseudoalteromonas* species and members of the *Alteromonadales*. For instance, the two upstream genes (which includes *fliS*) and the downstream 8 genes shown in Figure 3.2, are also conserved in *P. ulvae*, and these regions are also highly conserved in the genomes of *Colwelia* species. The genomic context becomes more divergent in *Thalassomonas* and *Psychrosphaera* (Figure 3.2). In general, however, other flagellar genes are the most common types of genes found nearby flagellinolysin genes not only within the *Alteromonadales* but across the bacterial tree of life (data not shown).

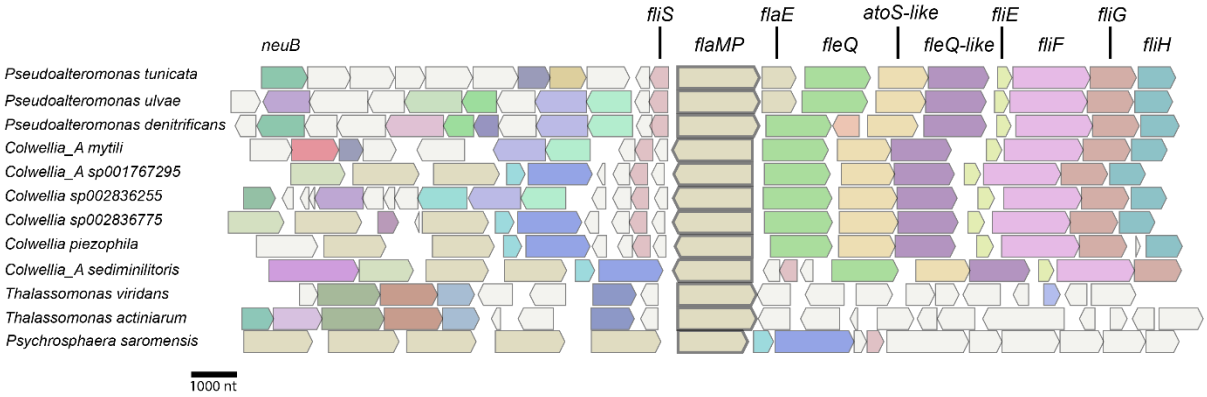


Figure 3.2. Genomic context of flagellinolysin in *Pseudoalteromonas tunicata* and other flagellinolysin-containing species within related species (*Alteromonadales*). Gene neighborhood visualizations and functional annotations are derived from AnnoView version 1.0 (annoview.uwaterloo.ca). Genes in the flagellar cluster are labeled. The gene labeled *atoS*-like is a gene of unknown function but its homology to *atoS* (a sensor kinase involved in fatty acid metabolism) suggests it likely represents a part of a two-component sensor kinase system. It bears no significant similarity to *P. tunicata*'s *cheY* gene, the typical flagellar-related mediator (see Section 1.2.1). The *fleQ*-like gene bears 39% identity to the adjacent *fleQ* gene.

3.3.3 *P. tunicata* flagellinolysin is likely under the control of FliA

To identify potential regulatory motifs within the flagellinolysin promoter, we performed a bioinformatic approach similar to "phylogenetic footprinting". Upstream promoter regions of flagellinolysin genes were extracted and used as input for the MEME motif discovery algorithm (see Methods). MEME identified a single motif that was highly prevalent, found in 102 of the 262

sequences (Figure 3.3), and highly significant with an E-value of 1.9e-157. The predicted motif was compared to known motifs from Tomtom's combined prokaryotes motif database, which identified a strong match to the known transcription factor binding motif of the FliA sigma factor (p = 1e-03). As described in Chapter 1, FliA is the flagellar sigma factor (σ28), which is required for initiation of transcription of late stage flagellar genes including *fliC* (flagellin)[40].

Three matches to this motif within the upstream region of the *P. tunicata* flagellinolysin gene (EAR29544) are shown in Figure 3.3b, including one that is immediately upstream of the ATG start codon. These results suggest that *P. tunicata* flagellinolysin may be under transcriptional control by FliA (which is found in the *P. tunicata* genome under accession ID # EAR29521).

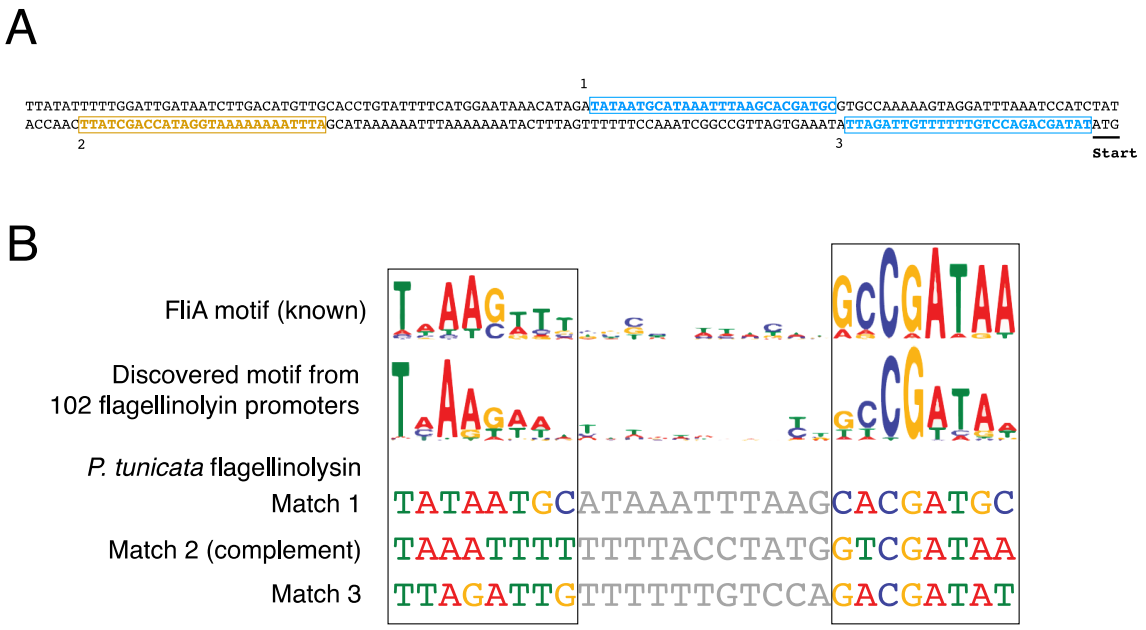


Figure 3.3. Identification of putative FliA binding sites in non-coding promoter regions upstream of flagellinolysin. (A) Three matches to FliA binding sites in the *P. tunicate* flagellinoysin upstream promoter region (209 contiguous base pairs). (B) Several motifs are shown. The upper motif is a position weight matrix representing the experimentally determined binding preferences of FliA from the MEME database (Combined Prokaryotes DB). The second motif shown is the top motif identified by the motif discovery algorithm MEME when applied to the promoter regions of all flagellinolysin homologs. This motif was identified in 102 flagellinolysin promoters. The three aligned matches to these motifs in the *P. tunicata* flagellinolysin promoter are shown below. Match 2 presents the strongest match (p = 1.95e-07).

3.3.4 *Predicted structure of P. tunicata* flagellinolysin is similar to *C. haemolyticum* flagellinolysin

Sructure prediction methods were used to generate 3D models of *P. tunicata* flagellinolysin. Alphafold (CoLab) was used to produce a general model of the full-length flagellinolysin protein, whereas Phyre2 homology-modeling was used to perform more detailed modeling of the hypervariable region (protease domain) that is of most interest given its proteolytic role. The predicted Alphafold structure shows the flagellin N and C coiled coil domains as seen in other flagellins, but the flagellin contains a unique exterior-facing domain that represents the predicted

metalloprotease, with a visible binding pocket. The protease domain structure (Figure 3.4b) is highly similar to that predicted for the *C. haemolyticum* flagellinolysin, despite only sharing 33% identity (Figure 3.4c). Areas of lower confidence were detected in the flexible disordered portions of the linker region. Disorder is expected in linker regions, to allow them to bend under the stresses of rotation. Areas identified as binding pockets show high confidence in all models, with different variations on the full flagellinolysin model showing the HVR orientated in different directions from equally valid orientations of the linker.

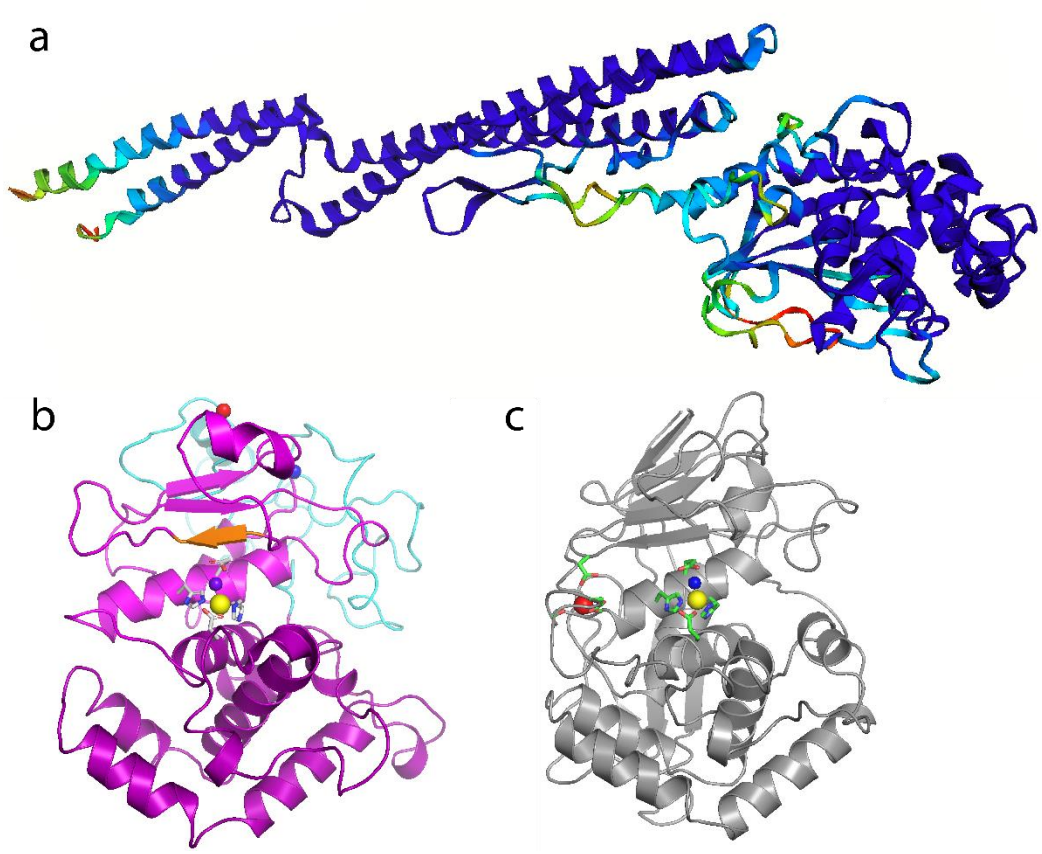


Figure 3.4. Four structural models representing various flagellinolysins. (a) A model of *P. tunicata* flagellinolysin, generated by ColabFold. Model is coloured by lDDT (Local Distance Difference Test), indicating the plausibility of the configurations. Dark blue is indicative of high confidence, shifting towards red for lowest confidence. (b) The Phyre2-derived model of the flagellinolysin HVR, including zinc atom at the centre of the active site. The upper and lower sub-domain are shown in light and deep magenta, respectively, while the putative catalytic-helper subdomain is depicted in cyan. The substrate guiding edge strand is highlighted in orange, and protein substrates bind in antiparallel fashion to the edge strand and along the central HE^{238}xxH-harboring active site helix. The side chains of the active site residues are shown in stick mode, and the active site zinc and water molecule as yellow and blue sphere, respectively. Importantly, the N- and C-terminus of the structural model co-locate at the rear back, and are indicated in blue (Leu155) and red (Val484), respectively. (c) The model of the *C. haemolyticum* flagellinolysin HVR from Eckhard et al (2017)[43].

The homology model of the flagellinolysin peptidase domain (Figure 3.4b,c) is shown in standard orientation of metallopeptidases[110], with a horizontally-aligned active-site cleft and substrate binding from left (non-primed side) to right (primed side).

3.4 Experimental confirmation of flagellinolysin expression in *P. tunicata*

Based on bioinformatic analyses, it appears that flagellinolysin is encoded within a flagellar gene cluster in *P. tunicata* and may be co-regulated with other flagellar genes. Next, I aimed to test this experimentally by culturing *P. tunicata* cells and measuring flagellinolysin transcript expression using qrt-PCR assays.

3.4.1 Detection of flagellinolysin expression by qrtPCR

PCR primers were designed flanking a portion of the HVR region of the *P. tunicata* flagellinolysin gene, corresponding to residues 345 to 365, and were validated by gel electrophoresis and the use of a deletion mutant (see Chapter 4 and Methods for more information). Primers were also generated corresponding to two structural flagellins (EAR29563 and EAR29565) for comparing relative abundance, and to *P. tunicata*'s 16S rRNA as an abundant housekeeping gene. A primer was developed to target another housekeeping gene, E4D, but this was later found to be prone to off-target binding.

P. tunicata cells were cultured in marine broth overnight, and to 24, 48, and 72 hour time points with shaking before harvesting. Cells were lysed using a rotor-stator homogenizer and total RNA extracted (see Methods). Attempts were also made to harvest mRNA from biofilm samples but the biofilm matrix material interfered with available extraction methods.

qPCR analysis reveals flagellinolysin is expressed as mRNA in *P. tunicata* (Figure 3.5). The range of Cq values across samples overlapped and the Cq means were within 0.5 of one another, indicating very similar levels of expression (i.e. <2-fold difference in levels of expression). ΔFlaMP mutants showed amplification of structural flagellar primers but not of the flagellinolysin primers, indicating

that they are specific to flagellinolysin. PCR products produced only a single band on a gel and were of an expected size.

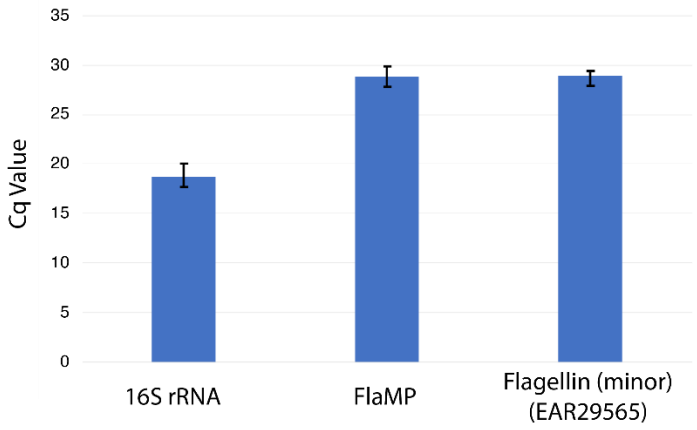


Figure 3.5. qrt-PCR analysis of mRNA transcript levels for 16S rRNA, FlaMP, and flagellin transcripts. Error bars show the standard deviation. Cq values are logarithmic, meaning a difference in Cq value of +1 means a halving of relative quantity of transcript.

16S expression was considerably higher than flagellinolysin and flagellin expression, as expected. Relative quantity (ΔCq) was also taken as a measure within samples and compared across them. Flagellinolysin most closely matched that of EAR29565, the less abundant of the two structural flagellins (see section 3.4.2 below), dubbed the minor flagellin. Flagellinolysin was expressed at a level approximating 72% of the minor flagellin, with a standard deviation of 12%. Differences in expression that fall below a 2-fold difference are usually considered similar.

3.4.2 Mass-spectrometry based detection of flagellinolysin in extracellular samples

In concert with qrtPCR, samples of *P. tunicata* grown to the same time points were harvested and sent for shotgun mass spectrometry to look for flagellinolysin's relative abundance. In addition, overnight samples were also centrifuged for collection, creating a pellet and supernatant sample set.

In all samples, EAR29563, dubbed the major flagellin, was the most abundant of the filament proteins, followed by EAR29565, the minor flagellin (Figure 3.6). In all samples, these two flagellins existed in approximately a 2:1 ratio. Spectral hits do not measure objective abundance but the ratio of spectral hits between proteins is an indicator of their relative abundance within a sample and they can be normalized using a constitutive protein.

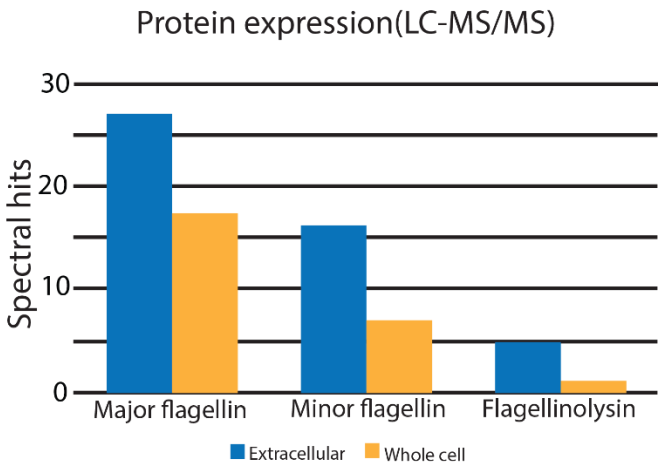


Figure 3.6. Detected abundance of flagellinolysin and other proteins in *P. tunicata* cellular and extracellular fractions by shotgun mass-spectrometry. The abundance is shown as # spectral hits, and does not necessarily represent the true proportional abundance in the sample. Flagellinolysin was detected but at a very weak level in cellular and extracellular samples, indicating a potentially low relative abundance in the proteome.

Flagellinolysin's measured abundance is much lower than either of the two structural flagellins, and it was frequently not detected at all. *P. tunicata* also produces a number of extremely high abundance proteins, such as Slr4, its S-layer protein, which frequently accounts for >10% of total spectral hits in all samples by itself[111]. These make detection of lower abundance proteins difficult and necessitate the use of other techniques to parse out the actual relative abundance of *P. tunicata* in protein samples.

3.4.3 Flagellinolysin is detected in *P. tunicata* flagellar samples

As the attempts to detect flagellinolysin using mass-spectrometry yielded only a weak detection (Figure 3.6), further attempts were made to detect flagellinolysin using immunoassays. An antibody to the flagellinolysin-HVR was generated (see Methods), and used for immunoblot assays with the same time points and conditions used above (overnight, extracellular, 24, 48, 72 hours).

Western blot attempts with cellular fractions consistently returned no detection of flagellinolysin. This is not surprising, as Western blots often require considerable fine tuning, but greatly enhancing concentration caused a similar enhancement in background staining. It was hypothesized that flagellinolysin may be weakly abundant and localized within flagellar filaments, and so purification of filaments might be required for more sensitive detection of flagellinolysin. Flagellar filaments were therefore purified by ultracentrifugation (see Methods) and verified by TEM, as shown in the electron micrographs below (Figure 3.7a).

In these concentrated filaments, flagellinolysin was positively detected via Western blot using primary antibody concentrations of 1.25 µg/mL and secondary antibody concentrations of 1:3,000 (Figure 3.7b,c). Bands were cut from the band region from an identically loaded gel run at the same time and sent for mass spectrometry to confirm the presence/absence of flagellinolysin, confirming the results. Flagella from the ΔFlaMP mutant were used as a negative control.

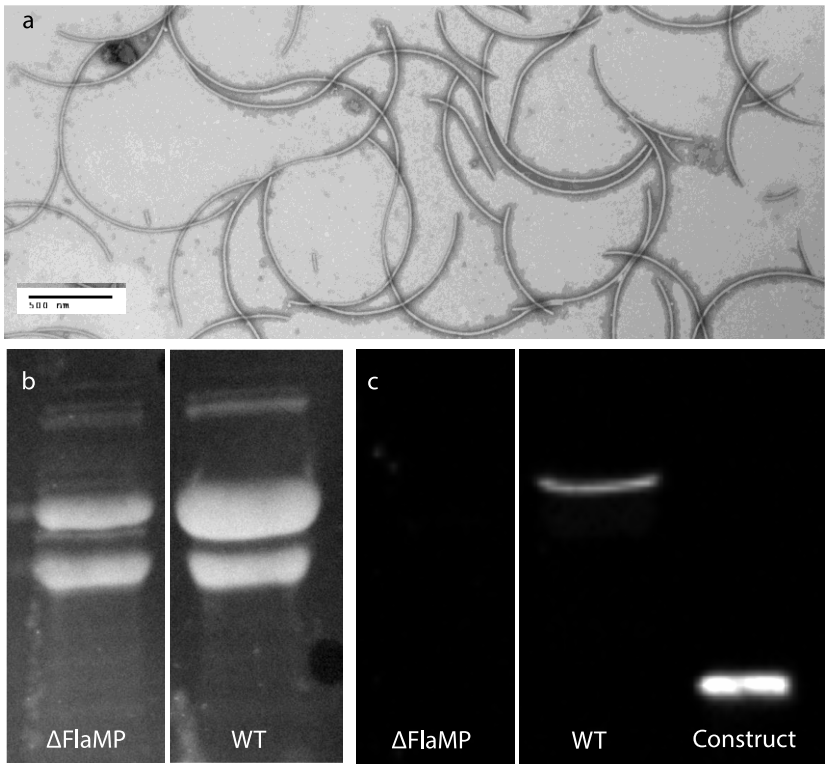


Figure 3.7. Western blot detection of flagellinolysin from purified *P. tunicata* flagellar filaments. (a) TEM image of purified flagellar filaments from *P. tunicata*. (b) Ponceau stain of gel transfer onto nitrocellulose membrane. The large central bands are the range for flagellins. The lower bands are S-layer protein, while the uppermost bands are likely flagellar sheath proteins. (c) Western blot of the same membrane. A sample of the HVR construct used to generate the antibody was included as a positive control. <200 ng of construct was used to run the gel, so it is not visible under the Ponceau stain.

3.4.4 Flagellinolysin is incorporated into *P. tunicata* flagellar filaments

Next, in order to examine the specific localization of flagellinolysin within assembled flagellar filaments, immunoelectron microscopy was performed. Samples of concentrated flagella (collected by ultracentrifugation, see Methods) were applied to a nickel formvar grid suitable for immunostaining. These samples were then blocked with solution containing milk protein and incubated overnight with milk protein and primary antibody, to maximize the time for the antibody to find its target. Samples were then washed, incubated with secondary antibody, and washed again

before being fixed with ammonium molybdate and allowed to dry overnight. Deletion mutant filaments were used as a negative control.

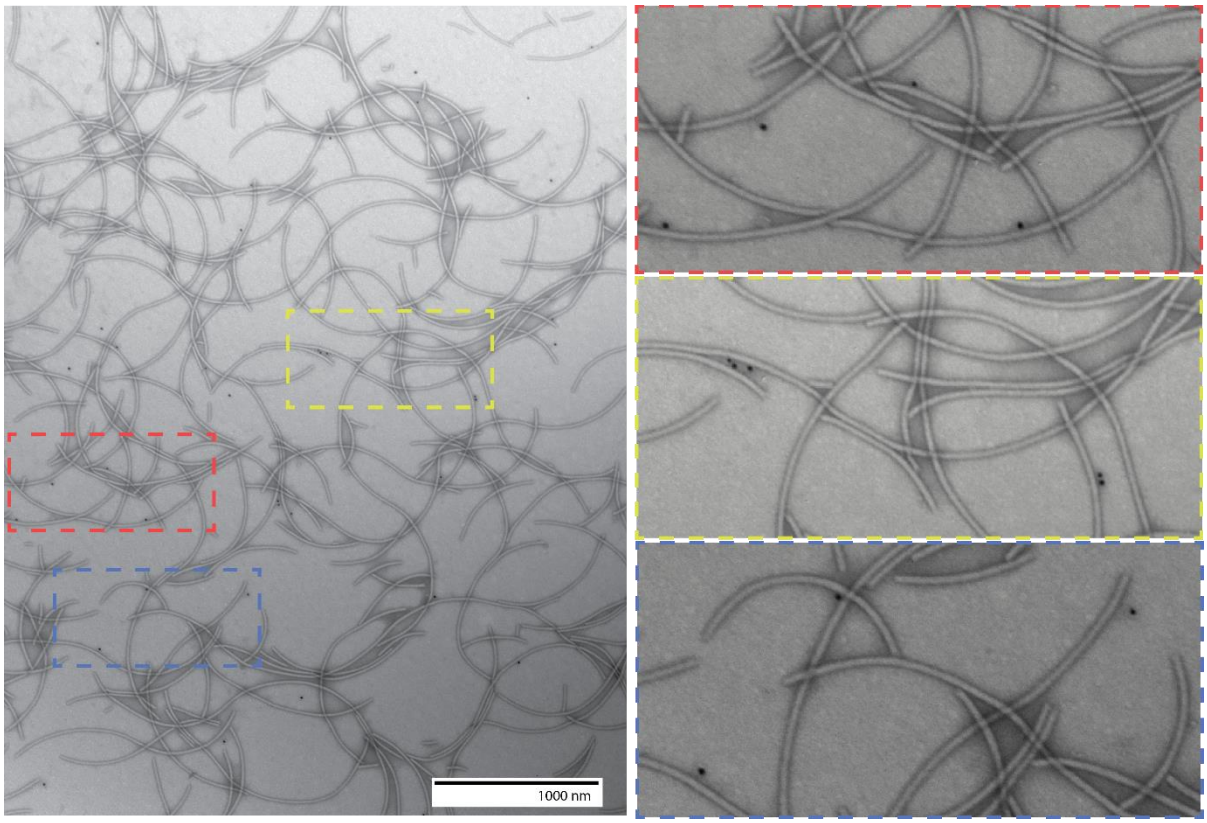


Figure 3.8. A transmission electron micrograph of concentrated *P. tunicata* flagella after immunolabeling with gold-conjugated anti-flagellinolysin antibody. Image on the left shows a wide field at 25,000x magnification. Scale bar is 1000 nm. Each dotted box is shown magnified on the right with colour-coding. Each black dot is a colloidal gold-bound secondary antibody.

Thirty fields of view of wild type and deletion mutant flagellar filaments were examined for instances of colloidal gold dots in association with filaments. In general, gold particles were observed alongside the edges of flagellar filaments within the WT samples. This pattern was almost non-existent in the deletion mutant sample as expected. Representative TEM images showing the immunolocalization of

flagellinolysin antibodies in assembled filaments are shown in Figure 3.8. Antibodies were seemingly randomly scattered along filaments and did not show a specific localization within a particular structure or region of flagella. Some gold particles were clustered together, whereas most occurred singly.

3.4.5 *P. tunicata* flagellinolysin has protease activity

In collaboration with Dr. Ulrich Eckhard and Biobasic Inc., a full-length construct (M1-Q605), HVR (protease domain) and linker (D42-R563), and the HVR region only (S142-G497) of flagellinolysin (PtuFlaMP) was recombinantly expressed in *E. coli* and purified first by affinity column chromatography and then by size exclusion chromatography. Several mutated versions of PtuFlaMP were also generated, including an E238A variant of each construct. These mutants have the glutamic acid of their HEXXH catalytic motif replaced with an alanine, leaving the structure intact but incapable of binding zinc ions and thus catalytically inert (see Figure 3.9a).

Although a more detailed characterization of protease activity and substrate preferences is provided in Chapter 5 using proteomics-based approaches, here an initial investigation of PtuFlaMP's predicted protease activity was performed using a zymography assay (Figure 3.9b).

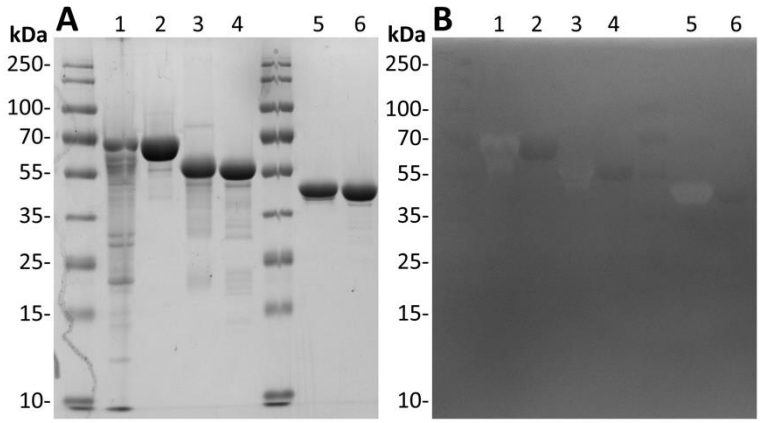


Figure 3.9. (A) SDS-PAGE and (B) casein zymography analysis of recombinant flagellinolysin from *P. tunicata*. Protein preparations of PtuFlaMP-FL (lanes 1 and 2, full length), PtuFlaMP-FL2 (lanes 3 and 4. HVR and linker), and PtuFlaMP-HVR (lanes 5 and 6, HVR only) are shown as both active (uneven numbers) and mutationally inactivated variants. Of note, proteolytic activity in the casein zymogram was only observed for the active $HE^{238}xxH$ proteins.

Zymograms of expressed constructs of flagellinolysin, including full-length, partial length, and HVR-only all showed successful degradation of casein within the gel (Figure 3.9b). Mutated variants (e.g,. E238A) each failed to degrade casein, indicating that they are proteolytically inactivated. This is consistent with the expectation that replacing the charged sidechain with an uncharged one will prevent the binding of zinc ions and halt catalysis. Thus, zymography assays confirm PtuFlaMP as a protease and provide biochemical support for several predicted active site residues required for its metalloprotease activity.

3.5 Discussion

In this chapter, I bioinformatically characterized a novel flagellinolysin protein from *Pseudoalteromonas tunicata*, and performed a preliminary experimental biochemical investigation of its expression, function, and localization.

Numerous pieces of evidence support the idea that *P. tunicata* flagellinolysin functions as component of the flagellar filament. First, flagellinolysin is found within a flagellar gene cluster and contains an FliA motif directly upstream in its promoter. This motif is highly conserved across other flagellinolysins as well, implying a conserved mode of regulation. The qrt-PCR results also accord with the predictive data regarding promoter sequences, showing similar levels of expression for both flagellins and flagellinolysins. Finally, immunoTEM demonstrates its integration into mature flagellar filaments.

Interestingly, *P. tunicata*'s flagellinolysin gene is also found nearby several genes encoding FleQ-like proteins. FleQ represents a master regulator of the transition from the flagellated state to biofilm[13,51]. It is therefore possible that the gene locus containing flagellinolysin and other flagellins in *P. tunicata* may be under additional type of regulatory control during the biofilm transition or at specific stages of biofilm development.

Protease assays confirm the predicted function of *P. tunicata* flagellinolysin's HVR as a metalloprotease. An E238A mutation within the HEXXH catalytic site abolished protease activity in a zymography assay, as did two other mutations within the predicted Ca-binding site based on similar work done in flagellinolysin from *C. haemolyticum*[43]. The predicted structure of flagellinolysin using Alphafold also supports its molecular function as a proteolytic enzyme, as it closely resembles the flagellinolysin protease from *C. haemolyticum* despite sharing only 30.12% identity across the protease domain.

Some caution about these conclusions must be found in that the structures produced are only predicted, not experimentally verified. Ongoing work by collaborators are attempting to crystallize

flagellinolysin from *P. tunicata* and other species, and this will be necessary to confirm the structural modeling done here. Interestingly, several attempts made by collaborators to recombinantly express flagellinolysin in *E. coli* for crystallization and structural determination have revealed the tendency of flagellinolysin to form inclusion bodies (data not shown). Flagellin proteins often require assistance from FliS chaperones to maintain an unfolded structure in the interior of the cell and only fold into their final shape when integrated into the tip of the filament. The location of the flagellinolysin gene immediately downstream of FliS is therefore intriguing as it suggests it may require co-regulation with a chaperone. Flagellinolysin expression may also require flagellar cap proteins to assist in this process. Future work may include fine-tuning the expression of flagellinolysin in controlled conditions to ensure proper folding or the use of cryo-EM to assess the structure directly as a part of whole filaments.

In the foundational paper by Eckhard and colleagues, it was speculated that flagellinolysin's spread is likely a product of simultaneous lateral transfer and convergent evolution. The results shown here also support this model for development. The updated bioinformatic survey of flagellinolysin presented here (which includes many more genomes than what was available in 2017) revealed a widespread but also patchy distribution. This distribution – combined with the identification of common promoter elements (FliA motif), occurrence in flagellar gene clusters, and common domain architecture – points to a protein which has experienced extensive lateral transfer and/or emerged independently multiple times. Since a true independent emergence is a rare event, lateral transfer is the most likely explanation for its diverse spread. The patchy distribution and repeated emergence of flagellinolysin in different lineages across the tree of life also suggests a potentially simple mode of gene regulation. Complex regulatory programs would pose greater taxonomic limitations on the ability of a gene to spread, as it would require more genes and regulatory regions to be laterally transferred between organisms for it to spread successfully and may require clustering of regulatory genes with flagellinolysin to be successful[112,113]. Co-occurrence analysis performed using the PhyloCorrelate tool[114] (data not shown) revealed no strong correlation(s) other than known flagellar genes, and the promoter analysis adds to the idea that the protein does not require any additional support systems to perform its function beyond what it is already common to flagellated cells. These features are all suggestive of a protein that is useful to a process common to a wide variety of organisms across many environments.

One factor that complicates flagellinolysin's potential function in *P. tunicata* is the presence of its sheathed flagellum. The flagellar sheath would conceivably inhibit flagellinolysin's access to the extracellular environment if it is integrated into the flagellar filament. One possibility is that *P. tunicata* produces unsheathed flagella in certain situations (which has not been documented previously to my knowledge), which might allow flagellinolysin to act on external substrates, or that flagellinolysin becomes accessible when the sheath is partially shed so as to expose the filament. A second possibility is that flagellinolysin proteins may be released from flagellar filaments (e.g., through autoproteolysis) and able to act on extracellular substrates. A third possibility is that the flagellar sheath is porous and permits substrate access. A fourth hypothesis is that the target(s) of *P. tunicata*'s flagellinolysin may be a proteinaceous component of the flagellar sheath itself. There is precedent for this idea since many bacterial proteases are membrane proteases that are capable of cleaving targets that are attached to or embedded within membranes[115,116]. If flagellinolysin's target(s) is/are sheath associated, this would have broad implications since many flagellinolysin-containing species do not have sheathed flagella. Thus, rather than serving a singular purpose across many species, flagellinolysin may be capable of serving multiple purposes and cleaving different substrates.

Further work using flagellinolysin-deficient strains to identify mutant phenotypes (Chapter 4) as well as comparative proteomics (Chapter 5) to identify potential protease targets will be critical to answer some of these questions and provide additional insights into the biological role of flagellinolysin in *P. tunicata* and related species.

Chapter 4
Investigating flagellinolysin function in *P. tunicata* by characterization of mutant phenotypes

The previous Chapter identified the flagellinolysin protein within *Pseudoalteromonas tunicata,* confirmed its predicted biochemical function as a metalloprotease, and determined its localization within assembled flagellar filaments. Despite this, major questions still remain regarding the biological role of flagellinolysin and the fitness advantage that it provides to the many bacterial species that produce it. In this chapter, I focus on a series of experimental investigations of flagellinolysin mutant strains, with the goal of identifying phenotypes related to the inactivation of flagellinolysin to develop insight into its biological function.

4.1 Introduction

Reverse genetics is a well-established technique in the biological sciences. As opposed to forward genetics, where random mutations are induced to produce a phenotype of interest which is then related to the affected genes, reverse genetics mutates or knocks out a gene of interest and examines the resulting phenotype to determine the gene's function[117]. In bacteria, this method relies on techniques pioneered by Blomfield and colleagues, using selectable and counter-selectable plasmids[117,118]. By inserting a plasmid bearing a mutated version of the gene of interest into a recombinase-positive bacterium, the mutant gene is able to replace the functional copy. This can also be done using an insert consisting solely of the flanking regions of the site of the gene, producing a deletion mutation.

Once a mutant is established, characterization of phenotype is the next step in elucidating biological function. This is not a simple or straight-forward process: Since the process or role in which the protein of interest takes part is unknown, the potential number of phenotypes it can take part in is theoretically infinite. Practically, however, homology can guide the researcher to infer likely phenotypes.

In constructing a mutant strain of flagellinolysin to investigate mutant phenotypes, we first need to consider the structure and biology of flagellinolysin in the context of a flagellum. Flagellinolysin is both a structural protein and an enzyme, integrated into the flagellar filament (see Figure 3.8 and Eckhard et al., 2017). This presents an added problem for investigation: as structural flagellins contribute to the structure, integrity, and curvature of the flagellar filament, removing flagellins can affect the motility of the organism in its natural medium[119,120] and thus have potentially broad impacts on phenotype. Therefore, while a knockout of a full-length flagellinolysin gene might have numerous phenotypic impacts, it could complicate the goal of isolating the biological role of the proteolytic activity of flagellinolysin, which is limited to the protease domain only. For this purpose, a key substitution that preserves structure but inactivates proteolytic function may be more useful. For instance, the zinc-binding HEXXH catalytic motif is necessary for flagellinolysin function. By replacing the glutamic acid residue in flagellinolysin with an alanine, the protease should be inactivated while the overall structure of the protein should remain intact. For this reason, replacement of H**E**XXH to H**A**XXH has been used in previous studies – or even evolved naturally – to inactivate zinc metalloproteases while preserving overall 3D structure[44,121,122]. Of course, it is also possible that such mutations may lead to other unintended functional consequences. For example, an inactivated protease may cause binding of a substrate without cleavage[44,123,124], which could result in increased substrate affinity, and lead to potential impacts on phenotype. As such, it is necessary to produce both knockout and inactivation mutants.

Possible biological roles of flagellinolysin and phenotypes for exploration

In reference to flagellinolysin's features as both a flagellin and a protease, a number of potential functions have been proposed (see Chapter 1 for a more thorough discussion). These include digesting extracellular proteins for nutritional purposes, degrading host tissues to aid in infiltration, rearranging biofilm matrix, or even aiding in competition with other bacteria by degrading proteinaceous parts of the cell wall[43,44]. In addition, flagellum-related defects often result in detectable changes in the structure and development of biofilms[13]. Disruption in any of these processes should produce distinct phenotypes due to growth, biofilm, or competition-related functions.

Growth models

The most straight-forward model, and the easiest to detect, is that flagellinolysin degrades extracellular protein to provide a direct nutritional benefit to the host organism. In this model, flagellinolysin degrades extracellular proteins and large peptides into smaller chains that are able to be absorbed by the host organism. Bacteria secreting proteases to the extracellular environment is not a new concept[125], making it a feasible mechanism for flagellinolysin to act, with the added benefit of the flagellar filament ensuring that the proteases are retained in proximity of the organism. In this way, the filament might play a role not dissimilar to the biofilm matrix in the retention of enzymes[54], reducing the metabolic cost of producing them.

To test for a growth deficiency, growth must be measured. The standard measure for this is the growth curve: Cultivating cells and monitoring them, either for optical density in liquid culture or for colony growth on static plates. Differences in rates or maximum capacity for growth while being grown on or in a protein-rich medium can indicate that the organism is deficient in the ability to break down and absorb peptides.

Motility models

Bacterial motility (discussed in greater detail in Chapter 1) is the primary purpose of the bacterial flagellum. By rotating the flagellar filament, the organism is able to propel itself through a liquid medium at high speeds, moving towards new food sources and away from dangers. Flagella also play a role in surface-associated motility, such as swarming and near-surface swimming. Bacteria are also capable of other forms of surface-associated motility, such as twitching (using Type IV pili to pull themselves along a surface) and gliding (a mode of movement that remains mysterious).

Bacterial swimming motility can be confirmed visually by light microscopy. Live free-swimming cells can be placed under a microscope and observed directly to detect the presence or absence of swimming motility. Once confirmed, swimming motility can be measured quantitatively using soft agar assays. By inserting cells into low percentage agar (below 0.4%), the resulting spread can be measured to assess the speed with which the bacteria spread. Swarming motility can be tested by

applying bacteria to the surface of moist semi-solid agar (>0.4% and <1.0% agar) and observing the spread of the cells across the surface. Swarming motility often produces characteristic colony phenotypes, such as radiating zones of consolidation.

Twitching motility can be tested on solid agar (>=1.5% agar), producing a distinct, hazy “halo” effect around the central colony. This halo is a consequence of an extremely thin layer of bacteria coating the surface. Halos in bacteria are not always a consequence of twitching cells, however, and may instead be a result of secretion of various pigments, antibiotics, or other substances. To confirm twitching directly, cells must be applied to a very thin layer of nutrient coating a glass slide, allowing for direct visualization of cell motility in the absence of moisture or surfactant. Notably, twitching motility does not typically cause the central colony to expand, merely the halo. Swimming or swarming, meanwhile, cause the whole colony to spread and grow, making them visually distinct.

Biofilm models

Biofilm phenotypes are theoretically amongst the easiest phenotypes to detect in bacteria, because they are often some of the most visible. Changes in signaling, the timing of events, alterations to microstructure, and so forth within a biofilm can all result in phenotypes such as reduced colony size, change in (or absence of) pigmentation, ability to spread across (or adhere to) a surface, and so on. Each of these should produce phenotypes visible to the naked eye. In practice, identifying such a phenotype still requires subjecting the organism to a wide range of conditions, since the biofilm-related process the novel protein of interest may be involved in might only have significant impacts at certain times or in certain environments. For example, a protein moderating heat adaptation within a biofilm is only going to produce an effect under high temperature.

This means that it is necessary to culture the bacterium in a wide variety of conditions, including adjusting temperature, shaking speed (or lack thereof), surface material, and so on. Depending on the organism, plate assays, static growth assays, and crystal violet assays can all be appropriate tools to examine colony development, pellicle formation, and total biofilm cell count respectively.

In this chapter, I conducted a comprehensive phenotypic investigation of wild type, flagellinolysin knockout and point mutant (E238A) strains of *P. tunicata*, in order to gain insights into flagellinolysin biological function. I first investigated the strains for motility and growth-related phenotypic differences since motility-related phenotypes in particular are expected for flagellar mutants. I then examined other phenotypes such as gelatin hydrolysis, interspecies competition, and surface adhesion. Finally, I explored biofilm-related phenotypes, and observed marked differences in the spatial organization and architecture of flagellinolysin mutant biofilms. These biofilm-related differences provide new insights into the function of flagellinolysin.

4.2 Methods

4.3 Examination of flagellinolysin mutant phenotypes in *P. tunicata*

4.3.1 Gelatinase assays

To test for reductions in the ability to degrade large exterior proteins, droplets of *P. tunicata* overnight culture were applied to nutrient plates containing 15 g/L of agar and 8 g/L of gelatin. Cultures were allowed to grow overnight at room temperature. The following day, they were coated with saturated ammonium sulfate to enhance the visibility of zones of clearance.

4.3.2 Soft agar motility assay

Soft agar motility assays were used as a quantifiable method for determining the rates of swimming motility. Overnight cultures of *P. tunicata* wild type and mutant strains were subcultured into 4 mL of fresh media and incubated with shaking at 28^{o}C until they reached an OD at 600 nm of between 0.4 and 0.6. Cultures were then standardized by dilution to an OD600nm of 0.4 in fresh broth and 1 uL from each culture was inserted into 0.35% marine broth agar (Difco marine broth 2216 with 0.35% agar). Measurements of the spherical zone of growth were taken daily.

4.3.3 Confocal light microscopy

Observations of whole cultures with the naked eye or of fixed cells under the electron microscope provide insights into aggregate behaviour and cell structure but are not equivalent to direct observations of live cells. Confocal microscopy uses the removal of out-of-focus light to enhance resolution. Using it in combination with differential interference contrast (DIC) produces contrast in images that allows for the observation of unstained, otherwise nearly transparent objects. This allows for live visualization of unstained, untreated bacterial cells in liquid media. Here, these methods were used to image *P. tunicata* and mutant strains under a variety of conditions.

Confocal microscopy was performed using a Zeiss LSM 700 confocal microscope, in the University of Waterloo Microscopy Center, hosted in the Spafford Lab.

4.3.4 Liquid cultures

Examination of free-swimming cells in liquid culture allows for the observation of swimming motility. Liquid cultures were collected from either overnight or from subcultures after 2 and 4 hours of growth. Samples were applied to standard glass microscope slides for immediate viewing.

4.3.5 Culturing of pellicle biofilms

Overnight cultures of *P. tunicata* were subcultured at a 1:1,000 ratio into 100 mL of fresh broth in Erlenmeyer flasks. These flasks were incubated for eight hours with shaking and then allowed to stand at room temperature for 24, 48, or 72 hours. At those time points, flasks were gently tilted, separating biofilm from its anchor on the walls of the flask and a P1000 pipette tip was gently slid underneath the biofilm and raised gently out of the solution, allowing a layer of film to wrap around the pipette tip. The pipette tip was then rolled slowly along the surface of a microscope slide, unfurling the adhered biofilm along its surface (see Figure 4.1 for a diagram).

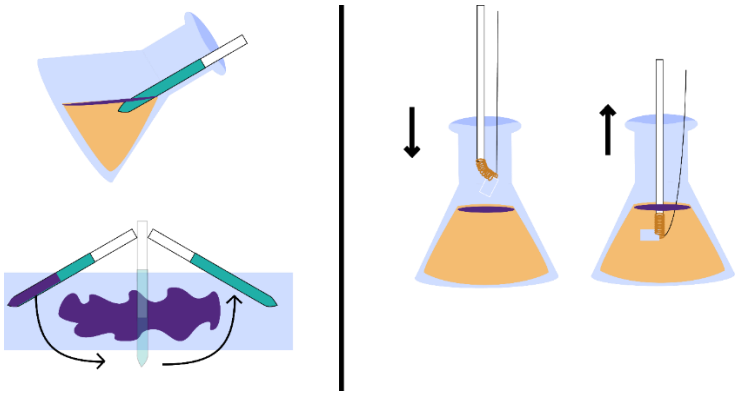


Figure 4.1. Cartoon illustrating the methods for collecting biofilm. Pipette collection method (left) and cover slip collection method (right).

Due to the fragility of E238A mutant biofilms, this method rarely produced intact biofilms for the E238A mutant. An alternative method was developed by Sura Ali with the assistance of University of Waterloo technical staff for use with atomic force microscopy slides, which I utilized here as well. For this method, a metal dowel is affixed with a tightly coiled spring. A glass cover slip is placed between the coils of the spring, the end of which is attached to a piece of fishing line. This allows the slide to be gently inserted under the film and then raised, bringing a portion of biofilm with it. This method suffers from the larger surface area generating more upward force but, when samples are successfully collected, does not require their unrolling.

4.3.6 Culturing of surface-adherent biofilms

To observe surface-adherent biofilms, I developed a method using slide staining vessels. Since these vessels are made of glass that can be autoclaved, untreated glass microscope slides can be placed inside the vessel cradle and the vessel sealed. The vessel can then be autoclaved to render it sterile. Approximately 30 ml of marine broth is then poured into the sterilized vessel until the glass slides are half-submerged. Overnight culture is subcultured into this vessel with brief shaking before being allowed to settle. At 24 hour, 48 hour, and 72 hour time points, slides were removed and rinsed with PBS to remove non-adherent cells and the slide covered with a cover slip. Cells outside the cover slip were then removed with an alcohol wipe to prevent any contamination of microscope surfaces. After

testing the effectiveness of the method myself, undergrad honours thesis students Mercedes Di Bernardo and Sadie Graves performed it at my direction.

4.3.7 Laser scanning confocal microscopy of biofilms in static conditions

Laser scanning confocal microscopy uses excitement of fluorophores by lasers to produce 3D maps of structures under the microscope. It can elucidate the structure of biofilms in a 3D manner, as well as indicate whether cells present are live or dead. Confocal laser scanning was done in collaboration with an undergraduate student, Krista Herbacko.

Overnight cultures were subcultured 1:1,000 into fresh media in CultureWell chambered coverslips (Grace BioLabs). These wells were cultured at room temperature for 24, 48, or 72 hours, at which point the culture was gently aspirated off and replaced with formalin. After fixing, the films were stained with SYTO9, a fluorescent dye that stains DNA. The LSM 700's scanning laser was then used to scan cross-sections of fields of visibility, measuring maximum, minimum, and average depth of cell coverage.

4.3.8 Transmission electron microscopy

Samples for TEM were prepared by applying a 10 uL droplet to a formvar-coated nickel grid. After two minutes, excess sample was removed by gently touching the edge of the grid to absorbent filter paper. The grid was then immersed in 0.1% ammonium molybdate for 30 seconds, followed by draining excess again with filter paper and storage in the fume hood overnight to maximize drying time before viewing.

Sample imaging was done using a Phillips CM10 100 kV transmission electron microscope, with imaging done using an AMT HR 11-megapixel side-mounted camera.

4.3.9 Flow cell growth assays and confocal imaging

As flow cell biofilm systems were not available locally, we collaborated with the lab of Dr. Zhenyu Cheng Lab at Dalhousie University to assist with these experiments. Bacterial cultures were grown overnight in VNSS broth overnight at room temperature and then pelleted by centrifugation at 10,000 rpm for ten minutes at 4^{o}C. Cells were then washed three times in FSMW (filter-sterilized marine water) before being resuspended in fresh VNSS and standardized to an OD600 of 0.5. 250 uL of cell suspension was then inoculated into the ethanol-sterilized injection port in the flow cell via a 1 mL syringe with a 27-gauge needle. The cell was then run at 0.2 mm/s linear flow at room temperature, with sampling every ten minutes. Cell dispersal was measured in CFU/mL from dispersed samples.

After two hours, the flow cell samples were stained by injection of SYTOX Orange and SYTO13 fluorescent dyes. 250 uL of one of the dyes was injected as above, followed by inversion of the cell (to protect the dye from sunlight) for 15-30 minutes. After the dye was deemed to have been sufficiently infused, the second dye was injected. Samples were then scanned using a Zeiss LSM 510, with an oil immersion 63x objective, an argon laser (488 nm) and a helium laser (543 nm). The 488-nm Ar laser and a 500- to 640-nm band pass emission filter were used to excite and detect the stained cells. Two color images (n=10) were scanned sequentially at 488 nm and 543 nm for biofilms after treatments and processed using Zen 2009 image software). Bacterial clusters were identified as any bacterial assemblages more than 10 μm thickness. Therefore, biofilm thickness was detectable as the distance between the base of the biofilm cell cluster at the substratum and the apex of the biofilm cell cluster at the bulk-liquid interface farthest.

4.3.10 Growth curves

Growth curves were generated using a Bioscreen C automated microbiology growth curve analysis system (Thermo Labsystems), hereafter referred to as Bioscreen C. Cell culturing conditions were varied, including the condition and quantity of starting cell culture, temperature of culturing, and whether cells were cultured with or without shaking. Optical density was measured using both OD600 and wideband settings, with the former tracking transmission of light at 600 nm frequency and the latter across the entire visual spectrum.

4.3.11 Competition assays

Competition assays were performed by Dr. Jiujin Cheng and myself. Strains for the competition assay were provided from the Charles Lab's collection.

Lawn competition assays are an established method for demonstrating competitive advantage between bacteria[126,127]. By streaking out a lawn of culture on a plate and adding droplets of the bacterium of interest to it, zones of clearance may be observed that indicate a competitive dominance of one organism over the other. This method is frequently used as a test for antimicrobial genes, as they represent a primary method of clearing competition. Here, we test a possibility that flagellinolysin either acts directly as some kind of toxin or indirectly by degrading food particles or otherwise providing an advantage that makes organisms carrying it more capable of competing than their competitors. These assays were performed by Dr. Cheng.

As *P. tunicata* does not grow readily on LB agar, tests were performed to identify an optimal mix of LB and marine agar to allow co-cultivation of diverse bacteria. A 1:2 LB:Marine ratio was decided upon owing to it being the only ratio in which *P. tunicata*'s production of tambjamines and violacein was observed. All other strains grew readily on all tested media, including the 1:2 mix.

Test strains were cultured overnight at 30^{o}C. The next day, 100 uL of the overnight culture was spread on an individual plate and allowed to form a lawn. *P. tunicata* cultures were grown in marine broth at 20^{o}C for two days, with 3 uL of this culture applied as a droplet to a plate. These culture were then incubated at 20^{o}C for 4 days, after which they were observed for the presence of a zone of inhibition.

Table 4.1: A list of strains used as a part of competition

Strains	Notable traits
Escherichia coli W3110	-
Ochrobactrum anthropi ATCC 49188	Opportunistic human pathogen, commonly associated with hospital-acquired infections[128].
Agrobacterium tumefaciens C58	Plant pathogen, also used as a tool to deliver DNA to transform plants[129].
Vibrio cholerae (Bang-2)	Non-toxigenic derivative of *V. cholerae*[130].
Vibrio harveyi	Marine bacterium, pathogen of fish[131].
Salmonella typhimurium ATCC 1331	Human pathogen, major cause of food poisoning[132].
Salmonella pullorum	Poultry pathogen, also a cause of food poisoning[133].
Pseudomonas putida KT2440	-
Pseudomonas aeruginosa ATCC 10145	Opportunistic human pathogen[134].
Staphylococcus aureus Newman	Opportunistic human pathogen[135].
Staphylococcus epidermidis ATCC 12228	-
Bacillus subtilis	-
Streptomyces coelicolor	-

Due to concerns about *P. tunicata*'s temperature and salinity preferences, I also attempted a competition assay. In this assay, 20 uL droplets of *P. tunicata* cultures were applied to spots on plates across from equivalent droplets of *E. coli* cultures. These droplets were then given a "path" to meeting by gentle application of a moistened, sterile wooden stick, creating a groove in the gel for liquid to travel along without causing the droplets to merge. Cultures were then observed to look for signs of one culture outcompeting the other.

4.3.12 Scanning electron microscopy

Scanning electron microscopy was performed collaboratively by myself and Mercedes Di Bernardo at the Molecular and Cellular Imaging Facility.

Scanning electron microscopy (SEM) is a variant of electron microscopy. Whereas TEM transmits a beam of electrons through a sample and assembles an image by level of transmission, SEM reads deflections of electrons off of a metal-coated surface. These deflections are reassembled to form an image, allowing for high resolution imaging of surfaces, though without detail as to what is going on underneath. This method was chosen to provide a closer look at the surface of *P. tunicata* agar colony biofilms, which are too dark to be visualized using conventional light microscopy and too dense to be observed with TEM.

Sterile filter disks were placed on the surface of marine agar and droplets of overnight culture were gently applied. These disks were then allowed to grow overnight at room temperature to form colonies. The filter disks were then gently lifted to remove the colonies from the plate and transfer them to staining dishes where they were fixed in osmium tetroxide fixing solution. After 60 minutes, the colonies were gently floated off of the filter paper using cacodylate buffer and onto the adhesive pad. This pad was attached to the metal insertion button. The sample was then dehydrated and inserted into a Desk V TSC (Denton) for sputter coating with gold. The gold-coated sample was then imaged using a Quanta FEG 250 (FEI).

4.3.13 Crystal violet assay

Crystal violet assays were performed by undergraduate student Alexander Stavropoulos under my supervision.

Crystal violet is a triarylmethane dye that is vividly purple in colour and produces a peak light absorbance at the 550 nm wavelength[136]. The dye is soluble in water and will bind DNA and protein, leading to its use as a total stain by Hans Gram as a primary dye in the Gram stain method. This

binding is reversible in absence of a mordant by lowering the pH of the solution, allowing the dye to be retrieved again from the sample.

Crystal violet microtiter assays allow for the quantification of biofilm development in bacteria. Relying on the ability of biofilm-associated cells to adhere to surfaces, the method uses the uptake and release of stain as an indirect measure for quantifying the amounts of biofilm present. By removing free-swimming cells through washing, only attached cells remain in the sample. Staining the attached cells with crystal violet and then reversing the binding through washing with acid allows the total absorbed stain can then be quantified via spectrophotometry[136]. The method for it used here was drawn from O'Toole, 2011, but was modified for use with *P. tunicata* by Sura Ali of the Doxey Lab.

Overnight cultures were grown in marine broth and subcultured at 1:1,000 into fresh broth. These cultures were then grown to log phase (OD600 0.3-0.4) and standardized by dilution. These cultures were then diluted 1:100 in 1 mL of broth in microtubes in batches of three replicates per time point. Collection points were set at 24, 48, and 72 hours.

After incubation, tubes were overturned to pour out their contents and watched twice by gentle immersion in deionized water, followed by pouring out as before. Following this, the tubes were filled with 1 mL of 0.1% crystal violet solution and incubated for 15 minutes at room temperature. Following incubation, unbound crystal violet was washed out by immersion as above. The tubes were then inverted and blotted and allowed to dry. The stained tubes were photographed at this point to visualize the effect.

After drying, 1 mL of 30% acetic acid was added to the tubes to solubilize the crystal violet and incubated for 15 minutes. The mobilized stain was then transferred to cuvettes and measured by spectrophotometry at 550 nm.

4.4 Results

4.4.1 Flagellinolysin mutant strains exhibit normal swimming motility and produce flagella with no visible defects

First, confocal light microscopy was used to visually examine the motility behaviour of mutant and WT strains in liquid media. Free-swimming *P. tunicata* wild type and mutants both showed motility in liquid after 4 hours of growth. No differences in speed were noted between the strains, though the E238A mutant did show the presence of microcolonies by 4 hours, something not observed in the wild type (explored in later section). These results are inconsistent with defects in swimming motility.

As both a qualitative and quantitative assay for swimming motility, a soft agar assay was then performed by stabbing cultures below the surface of low percentage agar (0.35%) plates cultured at 28^{o}C and examining the diameter of motility-related growth through the agar over time. Cells cultured in soft agar revealed no observable difference in measured diameter over time (Figure 4.2a). No significant differences were detected by 72hrs ($p>0.05$) based on four true replicates.

These data suggest that the flagellinolysin mutants (E238A and ΔFlaMP) are capable of swimming motility.

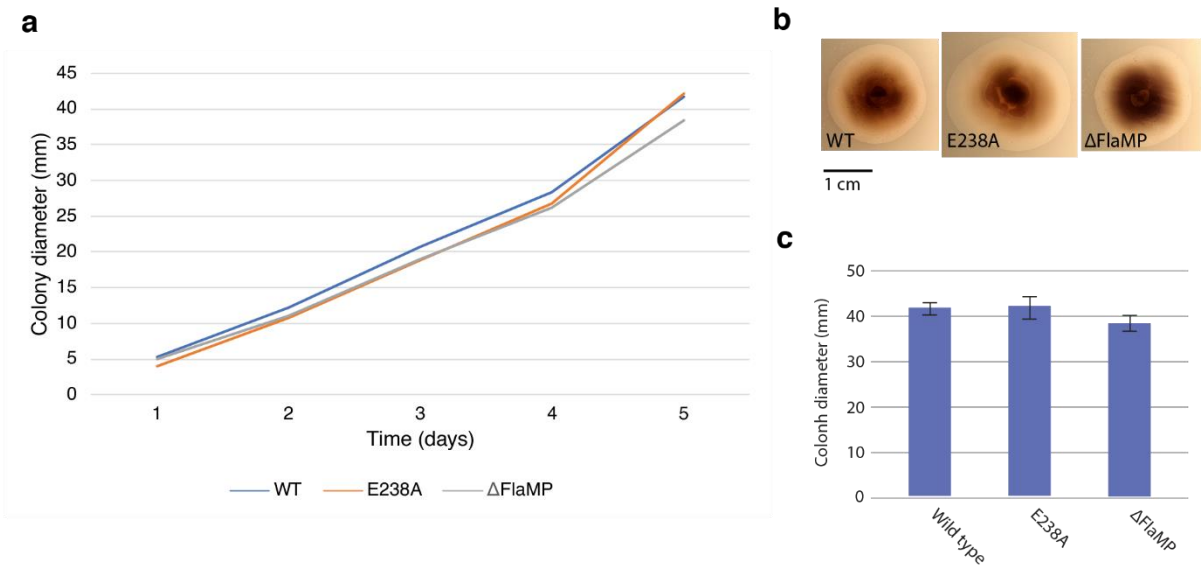


Figure 4.2. A series of figures showing the rate and extent of swimming motility in *P. tunicata* wild type and flagellinolysin mutants. (a) A diameter-time chart, showing the averaged diameter of swimming distance measured daily for five days. (b) Representative images of wild type (left), E238A (middle), and ΔFlaMP strains colonies grown in soft agar. (c) Box plots showing final growth diameters after five days. Bars represent standard deviation.

Observations of cells under the TEM similarly revealed no defects in the presence, width, length, or shape of flagellar filaments. Wild type and both mutant strains of *P. tunicata* successfully produced full-length flagellar filaments, without visible defects (Figure 4.3). Flagellar sheaths were also visible in all strains. This is consistent with observations of swimming motility in all strains.

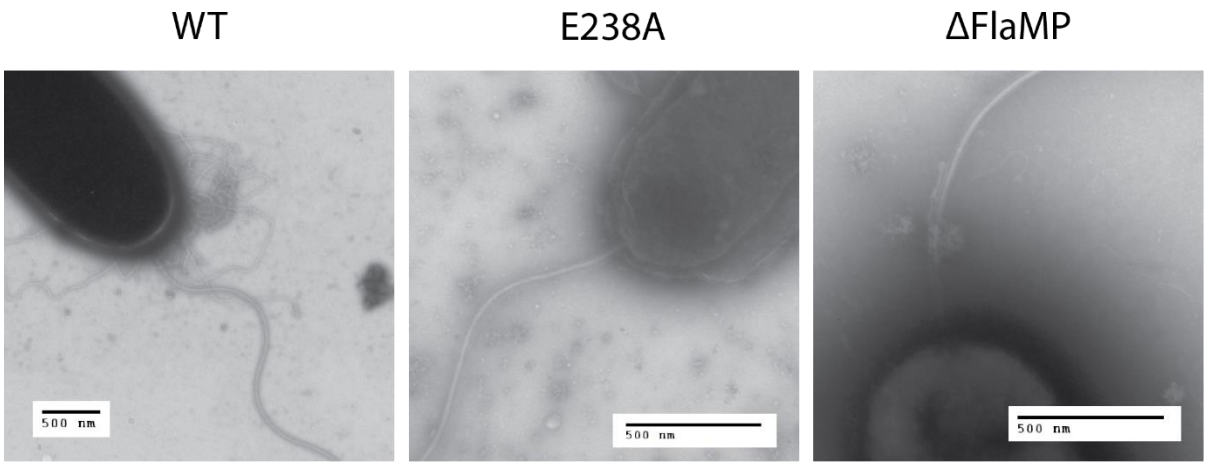


Figure 4.3. A series of images showing that all three strains are capable of producing intact flagella. Wild type and E238A images were taken by myself (2021, 2020). The deletion mutant image was collected by Denise Le (2022), under my supervision.

Additional experiments were also performed that provided negative results. WT and mutants were both gelatinase positive (data not shown) and no strains were capable of interbacterial killing with a subset of target strains (see Appendix, Figure A.1).

4.4.2 Identification of biofilm-related differences in P. tunicata flagellinolysin mutants

Given the critical roles of both flagella and proteases in biofilm development, it was hypothesized that flagellinolysin might play a role in this process. I therefore implemented strategies for cultivating *P. tunicata* biofilms on surfaces (glass slides) as well as air-liquid interfaces (pellicle biofilms).

Surface biofilms on glass slides

First, *P. tunicata* biofilms were cultured on glass slides by incubating glass microscope slides half-submerged in media containing subcultured *P. tunicata*. These slides were removed at the described time points and a cover slip applied. The cover slip was then sealed with nitrocellulose and its unsealed portions cleaned with ethanol wipes. These captured biofilms were imaged by confocal microscopy. Marked differences in the spatial architecture of cells could be observed in both mutant

strains compared to the WT as early as 24 hrs (Figure 4.4). These differences persisted and became more exaggerated over time (Figure 4.4). Whereas the WT biofilms form smooth sheets of roughly even depth across glass surfaces, the mutant biofilms are spatially heterogeneous with visible cell aggregates that fail to spread laterally and evenly across the surface. This pattern was reproducible in numerous independent confocal imaging experiments (N = 3).

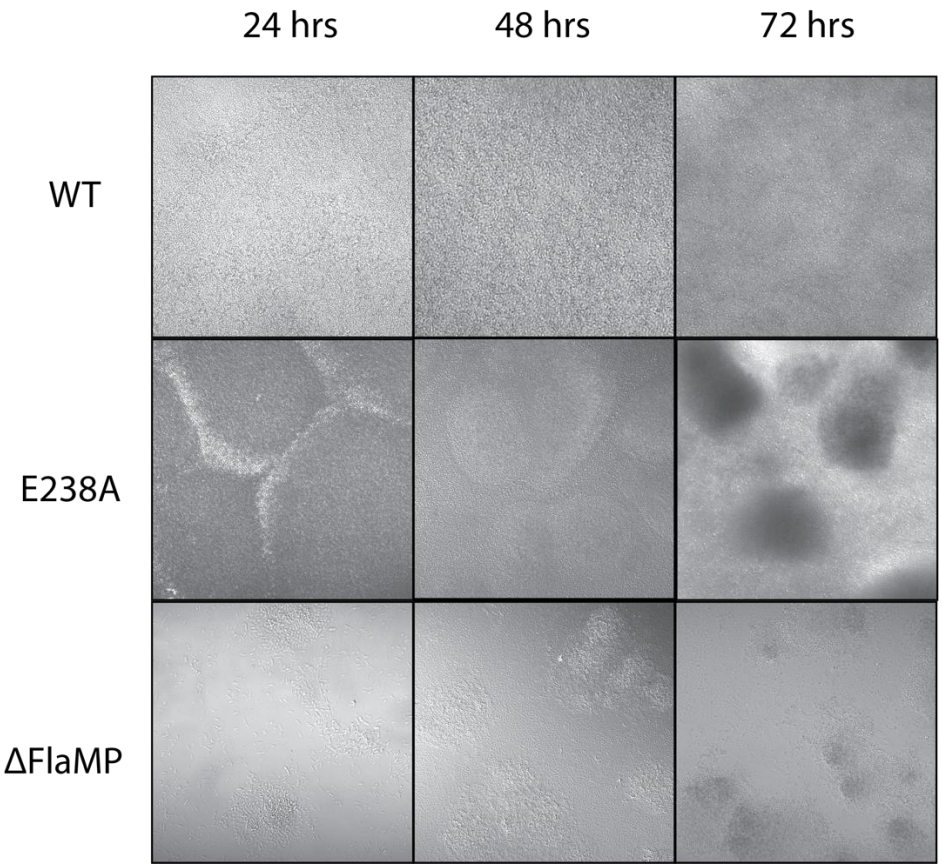


Figure 4.4. Confocal microscope images showing premature aggregation in *P. tunicata* flagellinolysin mutants. Biofilms were imaged at 24, 48, and 72 hours of growth. Images at 24 hours were taken at 63x magnification, due to low biofilm presence, whereas samples at 48 and 72 hours are at 40x magnification. Wild type cultures are smoothly and evenly distributed, while E238A mutants form dense clumps and ΔFlaMP mutants show a lack of spread across the surface and minimal biofilm formation.

Pellicle biofilms at air-liquid interfaces

Wild type *P. tunicata* was found to produce pellicle biofilms across the surface of liquid media in static (non-shaking) conditions. At 24 hrs, biofilms formed smooth distributions across the surface, which are typically uniform enough that they are difficult to see except as a faint discolouration (Figure 4.5). By 72 hours, WT biofilms were thickened and pigmented sufficiently to be a visible dark sheet across the surface of the medium. It was found that gentle agitation could detach these sheets of biofilm from where they are anchored against the container, but the sheet maintained cohesion.

Both flagellinolysin mutant (E238A and ΔFlaMP pellicle biofilms), however, were distinct from the WT (Figure 4.5), and these differences became more apparent over time. Mutant pellicles were more spatially heterogeneous and less organized, with visible clumps of biofilm material that were unevenly distributed. The E238A mutant showed the most dramatic difference in biofilm structure, and produced the largest aggregates, which were visible as early as 48hr (Figure 4.5).

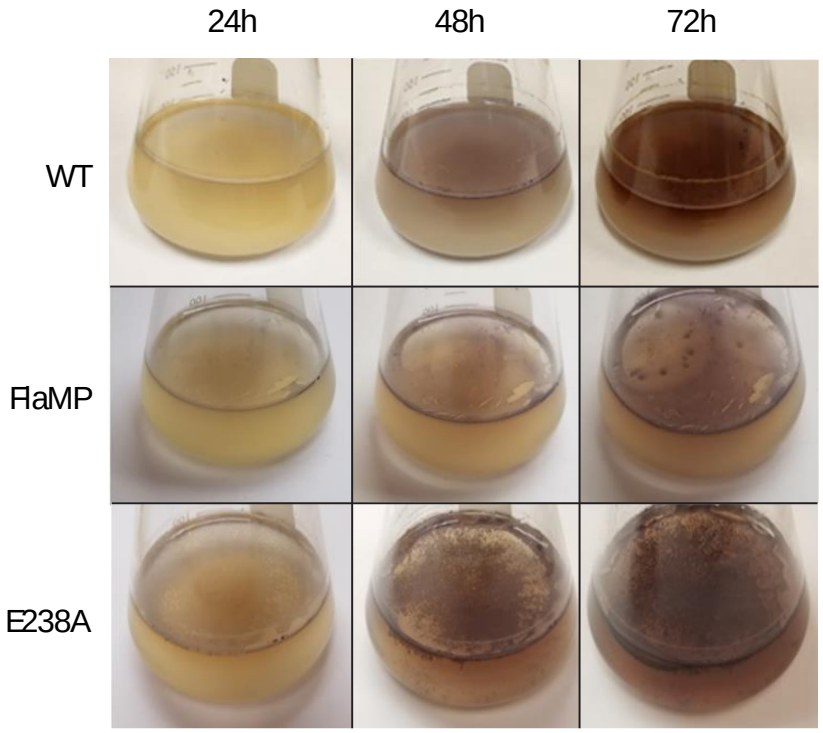


Figure 4.5. Images of *P. tunicata* pellicle biofilms taken over a 1-3 day period. Wild-type biofilms form smooth and even consistency throughout and maintain cohesion to themselves and adhesion to the glass of the flask. Gentle agitation does not disrupt the wild type films. Both mutant strains show disruption in the pellicle, with zones of greater and lower density and reduced integrity. They also show greater susceptibility to further disruption through agitation.

4.4.3 Focused investigation of flagellinolysin E238A mutant biofilms

Given the substantial phenotypic differences observed in E238A biofilms, we performed a deeper investigation of E238A biofilms using a variety of techniques including further quantitative confocal imagining in flow and static (non-flow) conditions. Given our primary goal of isolating the biological role of proteolytic activity in flagellins, we wanted to focus on a minimal mutation that removed enzymatic activity while preserving flagellar structure.

Comparison of WT vs E238A glass-attached and pellicle biofilms using confocal techniques

Confocal imaging was performed of glass surface attached biofilms, which replicates the results presented earlier in Figure 4.4, showing hyperaggregation in the E238A mutant (Figure 4.6a below). Similar to glass-attached surface biofilms, E238A pellicle biofilms are also spatially heterogenous and show cell aggregation, whereas the wild type pellicle biofilms are smooth and evenly distributed (Figure 4.6b).

Next, we used z-stack confocal imaging to examine the biomass and depth profile of WT vs E238A biofilms, and used glass-attached biofilms for this purpose (Figure 4.6c). Again the WT is evenly distributed, and has higher cellular biomass, whereas the E238A mutant is uneven with the presence of vertical mounds of cells. This is shown quantitatively in Figure 4.6d, which reveals a significantly higher biomass of the WT despite a larger maximum thickness for the E238A mutant.

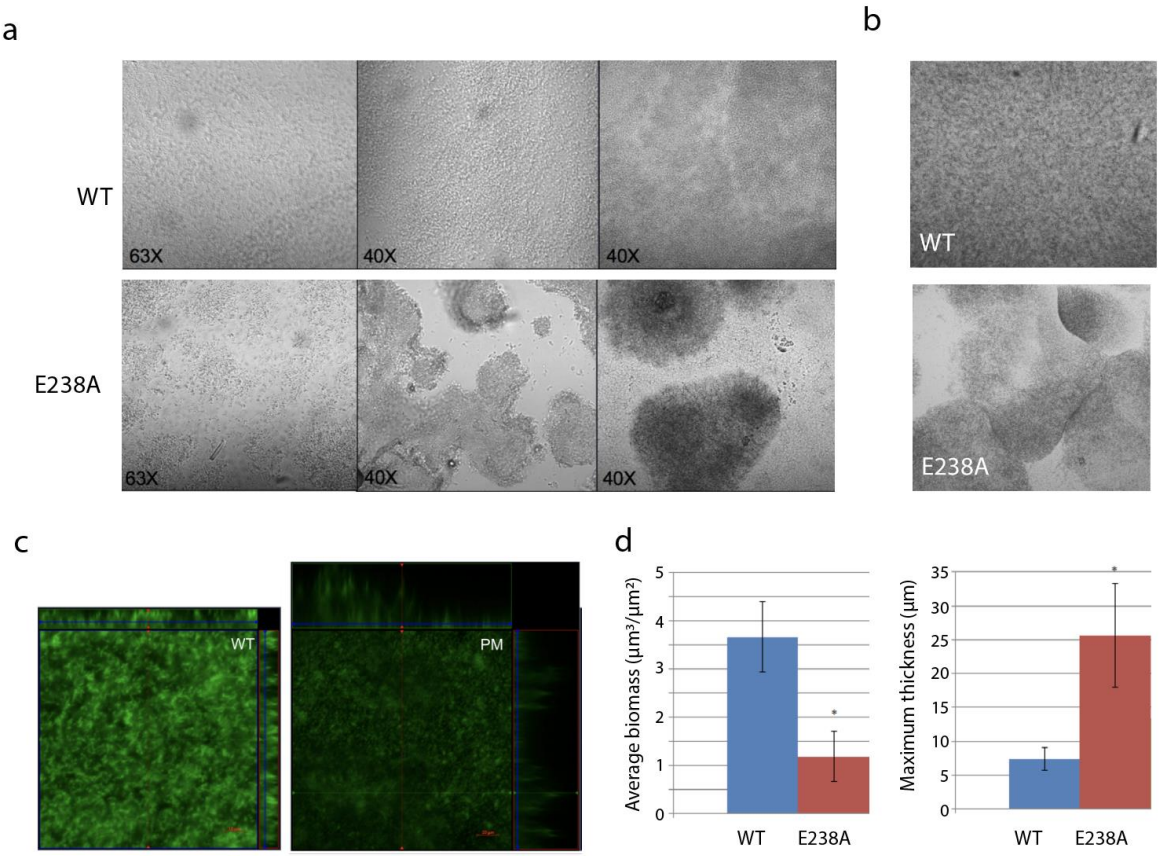


Figure 4.6. Confocal imaging of surface-attached and pellicle biofilms in static conditions. (a) Confocal micrographs of wild type and E238A *P. tunicata* biofilms cultured on glass slides at 24 (left), 48 (middle), and 72 (right) hour time points. (b) Pellicle biofilm after 72 hours from the wild type (top) and E238A mutant (bottm). (c) Fluorescent stains of wild type (left) and E238A mutant (right) biofilms. Areas outside the blue square show a height map of a cross-section of the image. (d) A set of box plots showing the average biomass (left) and maximum thickness (right) of the biofilms in 4.6c. Both measures are significant ($p < 0.05$).

Flow cell biofilms

Although the glass surface-attached biofilms provide a snapshot of *P. tunicata* biofilm development, they lack certain conditions common to natural environments, such as liquid flow that applies sheer forces and removes metabolic wastes/adds fresh media. This lack means that liquid biofilms can develop in ways that are not necessarily natural or die off prematurely, due to build-up of cells or suffocation under the burden of metabolic wastes. Flow cell systems provide a solution to this

problem, by allowing a continuous, gradual addition and removal of media and cells, letting the culture grow as it would in moving fluid.

Therefore, WT and E238A biofilms were analyzed in a flow cell system by collaborators (Zhenyu Cheng Lab) at Dalhousie University. Confocal images of *P. tunicata* wild type biofilms cultured in flow cells are shown in Figure 4.7, and maximum and average thickness measurements (taken from 21 slices in the WT and 18 in E238A) are summarized in Table 4.2. The WT resembles those previously observed by Mai-Prochnow and colleagues[58]. The cells are evenly distributed across the surface in a consistent layer, with the exception of voids that occur sporadically throughout the biofilm. The E238A mutant shows a deficiency in growth in the flow cell system.

Confocal laser scanning microscopy (CLSM) allowed for the quantification of these observations, scanning sections to produce measurements of average and maximum thickness, to compare populations of living and dead cells, and to further highlight the structure of these biofilms. Applying that technique to flow cell cultures showed the wild type produces greater maximum and average thickness under flow cell conditions. It also shows voids within the biofilm containing dead cells. This is normal in P. tunicata, which undergoes regulated cell death as a part of biofilm development, and suggests that the wild type biofilms are developing normally in this environment[58].

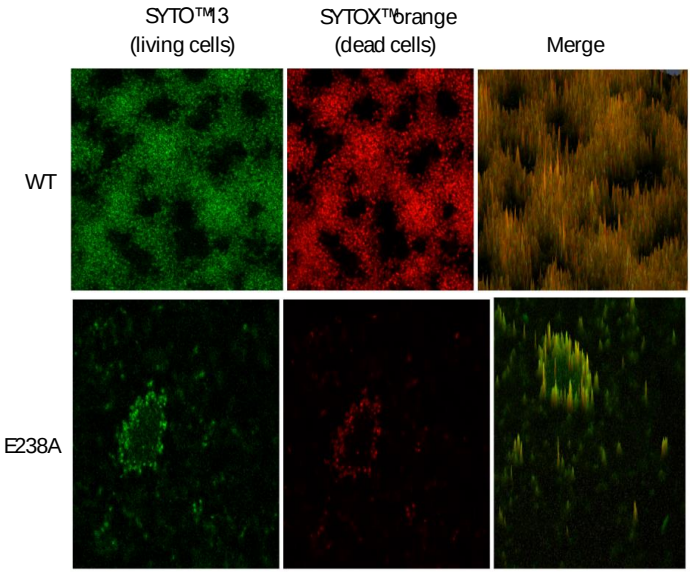


Figure 4.7. CLSM images showing stained cells in a flow cell. (top-bottom) The WT (top) shows a typical distribution, with cells distributed across the surface and the presence of voids and cell death. The E238A mutant (right) has fewer attached cells, with the ones in evidence concentrated in a dense ring of cells that may be a cell cluster. (left-middle-right) All samples were stained with two STYO stains for live (left) and dead (middle) cells. These colour contrasts were then merged (right) and projected in pseudo-3D. Sub-clusters of dead cells are concentrated within the core of larger clusters, as seen in Mai-Prochnow et al., 2004.

Table 4.2. Biofilm thickness measurements for *P. tunicata* wild-type and E238A in flow cells.

	Maximum thickness (µm)	Average thickness (µm)
Wild type	83.85	32.0±17.2
E238A mutant	38.00	20±7.1

4.5 Discussion

In this Chapter, I investigated two flagellinolysin mutants (ΔFlaMP and E238A) in order to uncover phenotypic differences that provide insights into flagellinolysin's biological function. Inactivation of

flagellinolysin resulted in no detectable differences in swimming motility, ability to produce flagella, or growth behaviour in marine media. However, flagellinolysin mutants were found to form biofilms with unique differences from the WT including an uneven cell distribution with aggregated microcolonies. A focused analysis of WT and E238A biofilms was then performed as E238A biofilms had dramatic phenotypic differences despite arising from only a single point mutation in flagellinolysin. Biofilm defects were detected in numerous experiments and conditions including glass-slide attached biofilms, pellicle biofilms, and flow-cell biofilms. In all cases, a functional flagellinoysin gene appears to be essential for proper biofilm formation and viability.

Defects in biofilm development are commonly reported for flagellar mutants[13]. The systems that control the activation and rate of reversal for flagella are usually connected to the expression of biofilm genes, typically mediated by a common regulatory molecule like c-di-GMP. Thus, it not too surprising that a defect in flagellinolysin was found to produce biofilm-related phenotypes. However, most flagellar biofilm-related defects are a result of immobility[13]; a defect preventing flagellar rotation causing the organism to have difficulty colonizing surfaces. The results presented here demonstrate that flagellinolysin is not directly related to swimming motility: Disabling or deleting the protein did not impede cell locomotion significantly based on swimming motility assays. TEM and live microscopic imaging of the mutant strains also revealed an intact flagellar structure strains capable of propulsion.

Interestingly, the mutant strains formed dense microcolonies that often grow above the plane of the field on which they are grown but show reduced ability to spread once established. Similar phenotypes have been reported previously and associated with inability to spread across surfaces once attached[137,138] or as a response to conditions that inhibit near-surface motility[139]. Measurements of the maximum heights of these microcolonies show variation between conditions of static biofilm and flow cell growth, which may be due to differences in material or media flow. Scans of stained colonies in flow cells show a greater maximum density for the WT. Scans of glass well slides show the opposite, with the maximum thickness greater in the mutants, though with a far lower average thickness due to the concentration within microcolonies. The greater propensity of E238A pellicles to

break apart with even mild agitation may offer a clue to this, suggesting that even the gentle flow of liquid in the flow cell is sufficient to disrupt and degrade mutant biofilms.

The inability of flagellinolysin mutants to distribute themselves across surfaces and the formation of dense cell clumps resembles phenotypes seen in in *P. aeruginosa* mutants with *ΔfliM* and *ΔpilA* mutations[137] (e.g., see Figure 4.8 below for a comparison). These mutations inactivate the ability of the bacterium's ability to engage in flagellar switching and to produce type IV pili, respectively. The resulting biofilms were thicker but also less evenly distributed than the wild type. This effect was attributed to the inability of the cells to move across the surface they were adhered to, either by pili contraction or by flagellar flicking.

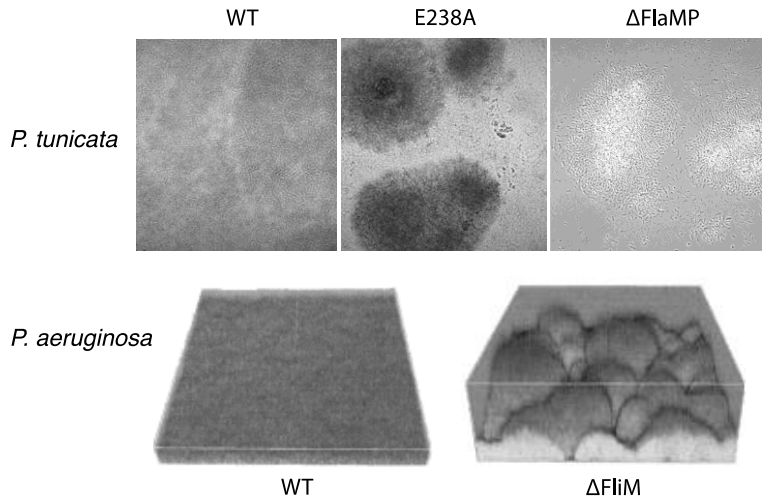


Figure 4.8. A comparison of *P. tunicata* wild type and flagellinolysin mutant phenotypes against a figure reproduced with permission from Klausen et al., 2003. All mutant strains produce an uneven, "lumpy" biofilm.

Disrupting flagellinolysin functionality in *P. tunicata* does produce a similar phenotype but the mechanism is unlikely to be similar. Flagellinolysin mutants are still capable of swimming locomotion, including tumbling. A different mechanism must necessarily be at work to produce a phenotype that disrupts surface spreading of attached cells.

One possible mechanism for flagellinoysin that is consistent with its widespread taxonomic distribution is the idea that it may proteolytically release flagella that are bound to cells or matrix material, as depicted below in Figure 4.9. Within a biofilm, flagellar filaments are known to aid in surface attachment and can become interlinked with other flagella, organic substrates, or biofilm proteins, creating a network that tethers cells together[15,31,140,141]. It can be hypothesized that flagellinolysin is involved in proteolysis of unknown target proteins that release flagella from a bound state during biofilm development, freeing cells to disperse and distribute evenly (Figure 4.9). According to this model, inactivation of flagellinolysin could result in the phenotypes observed in the E238A mutant and also ΔFlaMP.

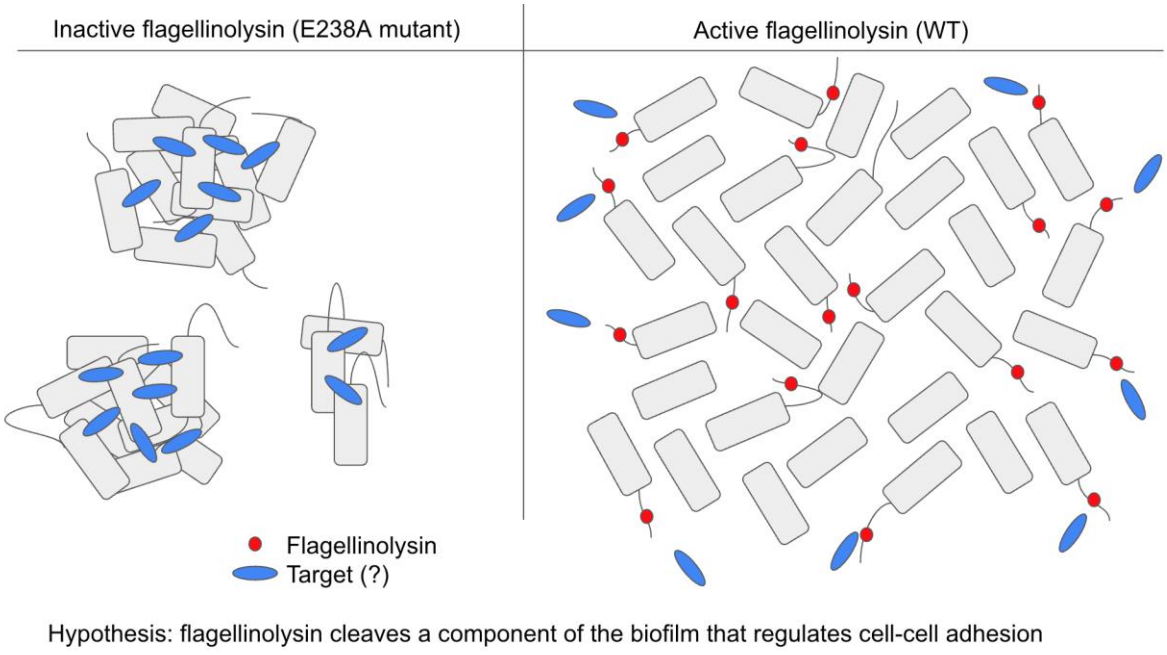


Figure 4.9. Differences in biofilm architecture between WT and flagellinolysin mutant strains of *P. tunicata* suggest a general function of flagellinolysin. The inability of mutant biofilms to spread suggests that flagellinolysin may proteolytically degrade target proteins within the biofilm matrix that maintain biofilm adhesion.

There remain a number of unanswered questions within this model, however. For example, if flagellinolysin is necessary for biofilm spread, why do organisms without flagellinolysin (e.g., *P. aeruginosa* as shown above) still manage to produce evenly distributed biofilms? If it is so crucial to this process, why is it so patchily distributed across the bacterial tree of life as shown in Chapter 3?

These questions are not without potential answers. It may be that organisms without flagellinolysin depend on different proteases to achieve the same end. Since flagellinolysin is likely a product of gene fusion or recombination, there may be a wide number of as-of-yet-undiscovered flagellinolysin-like proteases that possess different domain architectures. Alternatively, these organisms might rely on different forms of surface-associated motion and thus do not need to rely on flagella being able to free themselves. Any or all of these answers is plausible but all would require significant further experimentation.

To gain insights into these questions and provide additional support for the model of flagellinolysin-mediated detachment during biofilm development, further work is needed to identify biological substrates of the flagellinolysin protease. In Chapter 5, I attempt to investigate these questions by identifying candidate substrates in part using high-throughput proteomic approaches.

Chapter 5
Identification of flagellinolysin proteolysis targets via proteomics

In previous Chapters, I identified and characterized flagellinolysin from *P. tunicata* as an active metalloprotease localized to flagellar filaments. I then identified differences in the architecture of developing biofilms of flagellinolysin mutants, suggesting a role for flagellinolysin in aiding biofilm surface spreading via proteolytic cleavage of unknown biofilm targets. In this chapter, to identify potential targets of flagellinolysin, I use shotgun proteomics and other methods to compare the proteomes of WT and flagellinolysin mutant strains.

5.1 Background and rationale

A protease is commonly defined as "an enzyme that cleaves proteins by the catalysis of peptide-bond hydrolysis"[142]. Within this broad category, proteases are subdivided into five subcategories - aspartic, cysteine, metallo-, serine, or threonine – depending on the mechanism by which catalysis is achieved[142,143]. Of these, metalloproteinases represents the largest class, with 159 members at time of writing[142]. Metalloproteases – proteases dependent on metal ions to perform catalysis – are found across the tree of life, in all kingdoms, and may be found inside, outside, or attached to cells[144].

The revolution brought about by information technology has transformed the understanding of proteases over the past 50 years. Although proteases were first known and identified as far back as 1825, with the discovery of pepsin, proteases were still widely misunderstood as largely being secreted and responsible purely for the total degradation of proteins as late as the 1970s[143,145,146]. Since that time, advances in mass spectrometry, genome sequencing, and homology testing have revealed thousands of proteases which appear in a great variety of sizes and shapes, with functions that include regulating the lifespan, localization, and activity of other proteins[142,145]. In particular, the discovery of protease specificity and that proteases do not simply degrade proteins entirely but instead may engage in proteolytic processing both opened new doors and demanded new methods of investigation[142,145].

With the appearance of the discipline of proteomics in the 2000s came also the development of the field of degradomics: A discipline and methodology centred on determining the functions of proteases and other enzymes[142]. With the expansion of knowledge of proteases also came an awareness of the potential obscurity of their functions: Protease function may be regulated spatially, environmentally, through the use of activators or inhibitors, or through post-translational modification, making analysis of biological role difficult[142,145]. Degradomics focuses on the identification of substrates as a way to gain insight into the function of an enzyme. By combining activity profiling of specific enzymes with analyses of biases in cleavage of substrate libraries, researchers can identify potential targets within the organism's (or the host's, if a pathogen) proteome.

Flagellinolysin is a zinc-dependent metalloprotease embedded in flagellar filaments. As such, its most likely method of providing its function is to degrade either a host or environmental target protein found in the intermembrane or extracellular space. Finding flagellinolysin's target means providing a direct physical mechanism for its function, allowing a mechanistic explanation for why flagellinolysin mutants produce the phenotypes that they do. An ideal scenario for identifying a target is simply adding the protease to a sample of total protein and looking for a degraded target, but this scenario suffers from limitations. A candidate protein might not be available under all circumstances. It may be membrane-bound and be inaccessible to the protease due to vesicle formation, or misfold when the membrane is removed with detergent and render the cut site inaccessible. It may be so small a quantity that it is undetectable by conventional detection. All of these possibilities necessitate expanding the methods used to investigate potential targets.

5.1.1 PICS

The functionality of a protease is dependent on the recognition of peptide sequences. The active site on the protease is shaped to recognize specific sequences of amino acids based on features like size, charge, and hydrophobicity. This recognition can vary in its degree of perfection, with some enzymes degrading only a specific site on a specific protein and others actively degrading any protein they encounter[142]. Peptide library-based methods were developed to bypass this difficulty: Testing a protease against a library of peptides to look for those it cleaves[147]. A variety of library-based

methods exist, including phage display and fluorophore release, but all have advantages and drawbacks that can render them cumbersome and time-consuming.

Proteomic identification of protease cleavage sites (PICS) is a library-based method of identifying protease target site specificity, developed by Schilling and Overall in 2008 in humans and expanded to bacteria in 2011[147,148].

For bacteria, the library is usually generated from *E. coli*. Total protein is isolated from *E. coli* cultures and peptide libraries are generated using GluC, trypsin, and/or another endopeptidase with a known cleavage specificities. These proteases are then inactivated, disulfide bonds are reduced, and free thiol groups protected. Primary amines are blocked by demethylation. The resulting library contains the whole array of amino acids found within the organism it was generated from, representing a natural set of potential target peptides. This library can then be mixed with the protease of interest to allow it to cleave peptides. After this, the protease is inactivated and newly exposed primary amines may then be biotinylated. The biotinylated peptides can then be collected on a column and sequenced using LC-MS/MS.

Proteases are not necessarily exact in their substrate specificity, and are often capable of cleaving broad protein or peptide targets[142]. The large peptide library generated from total protein provides PICS with its strength in this regard: generating a consensus sequence from peptides cleaved avoids potential biases from using more limited sets, while still guiding target identification. Its use of natural peptides also ensures that cysteine is present (although it and lysine are blocked), something that most peptide libraries exclude[148].

Despite these advantages, PICS – and peptide library methods in general – do have some limitations and caveats. The method only works with endopeptidases, rather than exopeptidases, and aminopeptidases and carboxypeptidases are incompatible with elements of the method (the blocked N-termini and need for freed N-termini for biotinylation respectively)[148]. Most importantly, though, PICS cannot be reliably used to directly infer a target. Many proteases do not actually target optimal

cleavage sites and peptide libraries often contain peptides that would otherwise be shielded within the core of a protein, introducing potential false positives into the sample[148]. However, the consensus sequence can still guide selection of candidates identified by other methods, such as TAILS (see below)[148]. There are also extant PICS profiles for two other flagellinolysins, which may help guide the development of a function if there is a common profile[43,44].

5.1.2 TAILS

Identifying a protease's biological target is fraught with difficulty, as detailed above. While reverse genetics can indicate a process a protease is involved in, proteases have been demonstrated to regulate processes by innumerable means, including activation, inactivation, localization, and so forth[142]. Identifying a specific target can be a time-consuming process. Forward genetics looking for proteins involved in the same pathway relies on screening large numbers of mutants before even performing any protein-related work. This led to the development of methods investigating N-termini generated by proteases of interest[149]. Most of these methods rely on purifying peptides generated exclusively after exposure to the protease, excluding peptides that are not generated by interaction with the protease of interest. This has the issue of potentially producing false positives when done using a protein of broad specificity[149].

Terminal amine isotopic labeling of substrates (TAILS) is a method first reported in 2010 by Kleifeld and colleagues[149]. The method possesses the advantage of allowing analysis of the entire degradome, detecting N-terminal peptides both before and after exposure to the protease of interest.

TAILS utilizes the ability of mass spectrometry to detect slight variances in the mass of isotopes to differentiate between pre- and post-exposure N-terminal peptides[149]. Total protein isolated from the organism of interest is prepared in two batches: A control and experimental sample. The protease of interest is incubated with the experimental sample and then the protease is inactivated and both samples denatured and reduced. Each is then labeled with heavy (C^{13}) or light (C^{12}) formaldehyde; the formaldehyde reductively demethylates the N-terminal amines in the presence of sodium cyanoborohydride, transferring the carbon atom to the sample. The labeled samples are then merged

and trypsinized. Since the reductive demethylation also blocks lysine groups, this blocks some trypsin cleavage and produces a size difference between trypsin and non-trypsin peptides.

The resultant peptides are run through a polymer matrix composed of hyperbranched polyglycerol aldehydes (HPG-ALDs). This polymer matrix reacts with unblocked internal and C-terminal tryptic peptides but not with the labeled samples. Dimethylated lysines and isotopically labeled N-terminal peptides are unreactive and pass through the matrix where they may be collected by ultrafiltration. This means that all recovered peptides are those whose N-terminal cleavage sites were generated by proteases within the host organism, not trypsin. The resulting peptides can then be analyzed with mass spectrometry to generate a list of peptides and these mapped to known or predicted genome sequences within the target organism.

The end product of this method is a list of candidate proteins. Since native proteases can act within the proteome and protease target specificity can vary, it is not always possible to find an example of proteins degraded in the experimental sample but not in the control. As such, candidate lists must be subjected to hierarchical substrate winnowing[149]. This process involves identifying the candidates with the highest presence/absence ratios from either sample and examining them for signs of being potential targets, including features like colocalization, homology to targets of similar proteins, participation in a process the protease of interest is known to be associated with, consensus sequences from PICS, and so on.

5.1.3 Differential abundance from comparative shotgun proteomics

Simpler than either of the above, another effective method of identifying potential protease substrates is the use of shotgun proteomics to compare differential abundance of a protein across sets of samples. In theory, the inactivation of a protease may cause an accumulation of a target protein within the host organism. If this accumulation produces enough of a difference of abundance, the target may then be detectable by quantitative shotgun proteomics by looking for increases in normalized abundance or other metrics such as fold-change ratios.

The benefits of this method are its simplicity and directness, but these are also its drawbacks. While it is certainly possible that a protease acts solely to degrade a target protein and thus the protease's absence will lead to an overabundance, many proteases cleave targets to process them into an active form or to release them from the cell, or the target protein might simply be so low abundance that it does not appreciably accumulate. All of these considerations require varying the sampling based on known characteristics of the protease, such as its location within the cell, which proteins it shares homology with, etc., to help narrow search conditions to likely candidates.

This chapter details attempts to identify candidate proteins that are proteolytic targets of flagellinolysin by a series of proteomic methods, including PICS, TAILS, comparative proteomics, and SDS-PAGE gel based approaches.

5.2 Methods

5.2.1 PICS

PICS was performed in collaboration with Dr. Ulrich Eckhard, using the method described in Schilling et al., 2011.

In brief, an *E. coli* K12 peptide library was generated by proteolytically degrading total protein isolated from sonicated K12 samples with either trypsin or Glu-C. These degraded peptides were then N-terminally labeled and the proteases inactivated. This labeling makes the peptides distinct in mass spectrometry from unlabeled samples. The resulting peptide library was then incubated with flagellinolysin HVR construct and shotgun sequenced, building a consensus sequence of cleavage sites produced by the flagellinolysin construct.

PICS was also performed on bovine casein, inactivated flagellinolysin, and whole flagellar filaments. In these experiments, the samples were exposed to the flagellinolysin construct first and labeled, followed by subsequent trypsinization. In these cases, the labeled peptides were looked for, not unlabeled, but the principle remains the same.

5.2.2 Shotgun mass spectrometry

Samples for mass spectrometry were prepared under a variety of conditions to attempt to identify potential proteomic differences between the wild type and E238A mutant. This includes varying over time and growth condition (i.e. biofilm or free-swimming). In addition, samples of the extracellular medium were also taken and sent for mass spectrometry, to look for any potential differences in the strains' secretomes.

Shotgun mass spectrometry was performed as a service by Sick Kids spectrometry facility in Toronto and by Bioinformatics Solutions in Waterloo.

5.2.3 TAILS and differential abundance analysis

Samples of *P. tunicata* wild type and E238A mutant biofilm were prepared and harvested and sent on dry ice to the Overall Lab at the University of British Columbia, who runs a mass-spectrometry facility including TAILS analysis. Only the 48-hour samples passed quality control and were used for TAILS. Wild type and mutant samples were compared directly due to difficulties with sustaining protease (see below).

A detailed protocol for TAILS can be found in Klefield et al., 2010. Typically, TAILS is done using a protease of interest but the flagellinolysin HVR construct was rapidly degraded on addition to the sample, likely due to native proteases. Here, the protocol was modified due to the difficulty of sustaining the protease. In brief, sample replicates of *P. tunicata* wild type and E238A mutant were first analyzed by a generalized shotgun proteomics and then differentially labeled with light and heavy formaldehyde. The samples were then digested with trypsin and merged for shotgun sequencing. The resulting spectra were compared to look for unique splice sites for each sample and for differential abundance of peptides between samples. Fold-change between overall protein abundance was compared using the pre-TAILS samples.

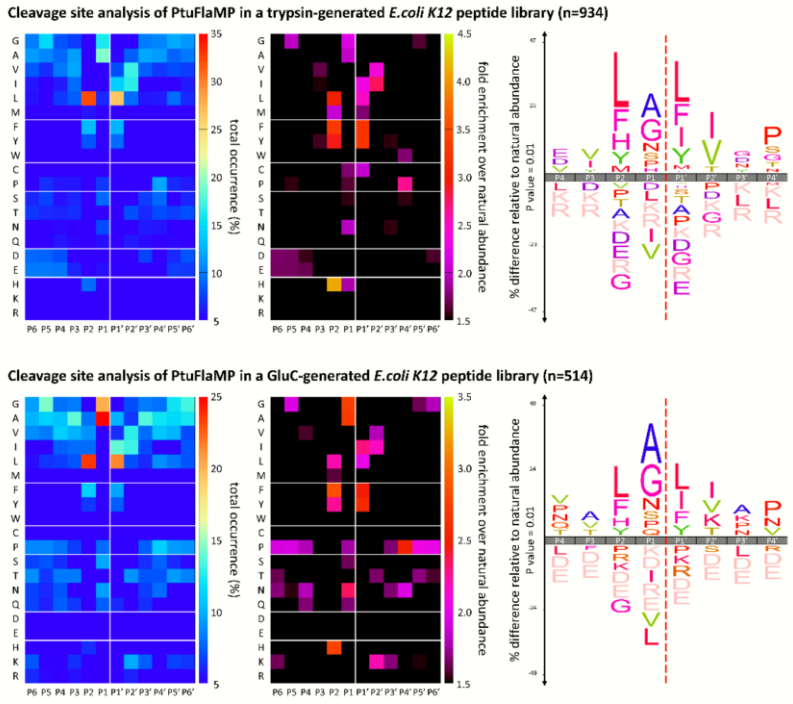
5.3 Results

5.3.1 PICS provides insights into flagellinolysin substrate specificity

PICS analysis was performed on *P. tunicata* flagellinolysin-HVR, which was recombinantly expressed using *E. coli* and purified. PICS analysis performed with *E. coli* peptide libraries produced a sequence logo showing a preference for a hydrophobic residues in the P2 and P1' positions, with a small, flexible amino acid in the P1 position (Figure 5.1a). This is similar to what was identified by Eckhard and colleagues in *H. gracilis*, which also showed a hydrophobic-flexible-hydrophobic preference in the P2-P1-P1' slots, but divergent from the *C. haemolyticum* flagellinolysin.

PICS analysis with whole proteins (Figure 5.1b) revealed a pattern with some similarities but also some differences to the PICS profile in Figure 5.1a. The hydrophobic preference in P2 and P1' are maintained but a preference for glutamine appears in P4, P1, P2', and P3'. A similar appearance of glutamine preference in total protein libraries appears in *H. gracilis,* though the positioning is different. This may be an artifact, owing to the fact that the whole proteins tested by PICS analysis (e.g., bovine casein) are rich in glutamine residues.

a



b

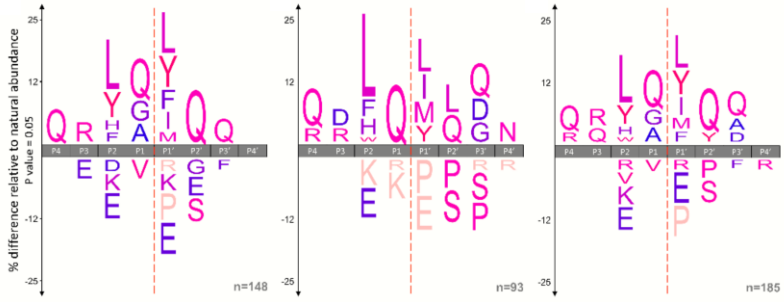


Figure 5.1. A collection of consensus sequence logos showing the target preferences for flagellinolysin using PICS. (a) Heat maps and sequence logos for trypsin-generated (top) and GluC-generated (bottom) *E. coli* K12 peptide libraries. (b) Sequence logos for the combined casein, inactive flagellinolysin, and flagellin libraries, showing pre-trypsin (left), post-trypsin (middle), and combined favoured cut sites.

5.4 TAILS reveals differences in proteolysis between WT and E238A biofilms

48-hour wild type and E238A pellicle biofilm samples were compared using the TAILS assay at the Overall Lab (University of British Columbia). High agreement and correlation between the two biological replicates (A and B) was observed in addition to high enrichment efficiency using TAILS as seen in Figure 5.2. A total of 1,326 unique peptides were identified in preTAILS across both experiments and 1207 unique peptides in TAILS. The enrichment of N-terminal peptides was 98% in TAILS while preTAILS revealed that only 12% of peptides were N-terminally blocked, consistent with prior TAILS reports. Plotting of $\log_2$ ratios of H/L between replicates A and B revealed that there is high concordance between both samples ($R \sim 0.75$ for preTAILS and 0.83 for TAILS) in addition to peptide ratios being normally distributed. Only the most confident TAILS identifications were further considered (the intersection of 545 quantified N-terminally dimethylated replicate pairs out of 553 total identified replicate pairs from both experiments), whereas for preTAILS, two or more peptides were required for protein quantification.

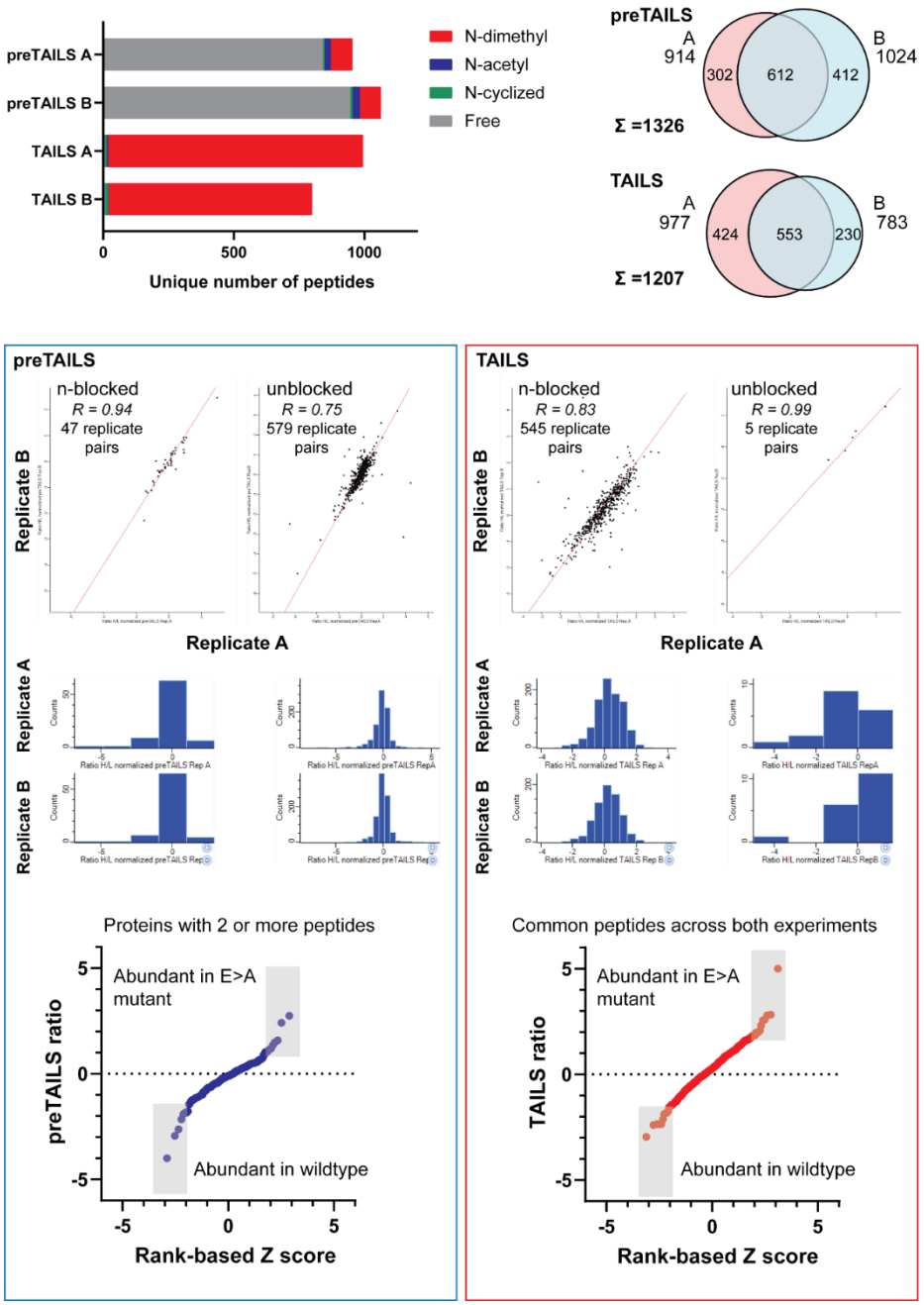


Figure 5.2. Summary of TAILS results. Enrichment by HPG-ALD polymer proved effective with high reproducibility between biological replicates.

Terminome analysis of wildtype and mutant strains of *P. tunicata* identified 112 N-terminal peptides more abundant in the mutant and 36 N-terminal peptides more abundant in the wildtype. These peptide identifications were further refined based on whether the peptide was the natural N terminus and by subcellular localization. Sequences were manually aligned to the *P. tunicata* proteome to check if it was the natural N terminus following signal peptide removal, and the 5 preceding amino acids recorded for all identifications. This yielded 44 extracellular substrates and 49 intracellular substrates highly abundant in the E->A mutant (Appendix Table 1) whereas there were 10 extracellular substrates and 7 intracellular substrates highly abundant in the wildtype (Appendix Table 2). Extracellular peptides more abundant in the wildtype represent possible cleavage substrates for flagella, and these were focused on in more detail.

Motif analysis was performed on wildtype and mutant TAILS substrates and categorized by subcellular localization. The motifs between wildtype and mutant are markedly different (Figure 5.3 below), with the peptides high in wildtype resembling the flagella cleavage motif previously determined by PICS (Figure 5.1). Moreover, the same glutamine preference is visible in the wild type samples, especially with proteins that have localization sequences that would see them incorporated into the outer membrane. The same QxQQ pattern seen in the whole protein PICS with the construct is repeated here, though the positioning has shifted to P2-x-P1-P1'.

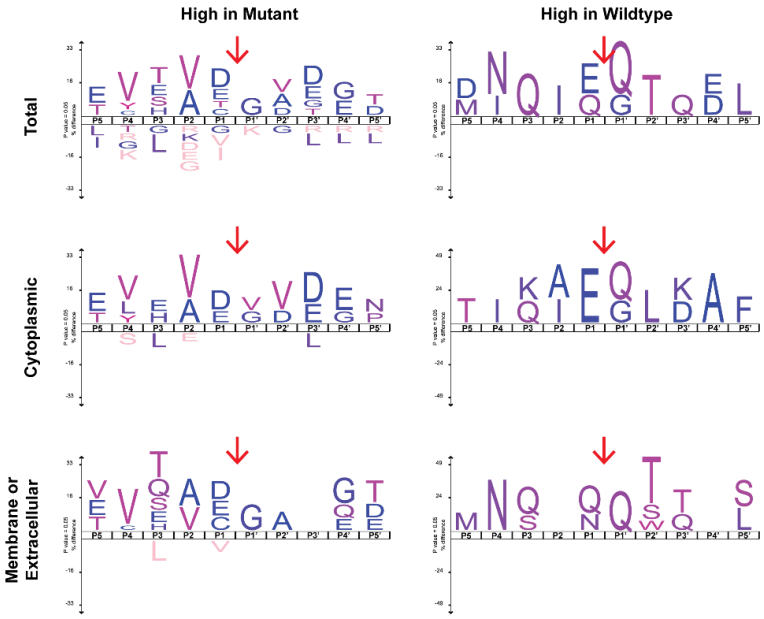


Figure 5.3. Icelogos based on high confidence substrates from TAILS terminomics. Extracellular motifs are markedly different between wildtype and E238A mutant.

TAILS peptides appearing abundant in the E->A mutant are those more prevalent with the loss of flagella proteolytic activity. As such, these represent substrates of proteases more active when flagellar proteolytic activity is lost. A possible explanation is that flagellinolysin might digest certain proteases, which result in their substrates being intact in wildtype conditions. These were not focused on furthered, but include TonB receptors or the TolQ proton channel.

TAILS peptides appearing abundant in the wildtype are those that are released in the presence of active flagella and may be primary or secondary substrates (see Table 5.1 below). Peptides here are not the annotated protein N termini (signal peptide removed, or initiator methionine removed, etc.) and are cleaved to generate fragments. Eight substrates are from proteins with signal peptides, while two (TonB2 and TonB) are membrane associated proteins.

Table 5.1: List of membrane signal-bearing proteins with high fold-change in the wild type with high confidence from Z-scores.

Protein name	Accession ID	# of unique preTAILS peptides	FC over mutant
DUF3450-containing protein	EAR29318.1	8	1.85
Pal (peptidoglycan-associated protein)	EAR26516.1	6	2.47
TonB2	EAR29314.1	3	2.45
MlaC-like transporter	EAR26697.1	2	2.19
Hypothetical protein PTD2_18595	EAR27859.1	-	3.58
DUF3103-containing protein	EAR28802.1	-	2.90
Hypothetical protein PTD2_02046	EAR30312.1	-	2.62
TonB	EAR30271.1	-	2.58
Zinc metalloprotease	EAR27159.1	-	2.32
YbgF	EAR26515.1	-	2.04

Of the top ten candidates, four are of unknown function. A fifth – EAR27159.1 – contains a domain of unknown function, despite sharing homology with zinc metalloproteases. Interestingly, within this list are several proteins associated with regulation of outer membrane stability. TonB-like proteins crucial for membrane integrity. Peptidoglycan-associated protein (Pal) belongs to the Tol-Pal system which has several roles including the stabilization of the outer membrane and underlying cell wall[150]. The protein YbgF is also included, and is encoded immediately next to Pal in the genome. This protein also is involved in periplasmic regulation of peptidoglycan crosslinking. These identifications suggest that flagellinolysin might be capable of cleaving outer membrane and potentially even periplasmic proteins. One possibility is that these proteins become accessible by flagellinolysin within the flagellar sheath, which is an extension of the outer membrane.

5.5 Differential protein abundance reveals a plausible protein target of flagellinolysin

Potential target proteins of a protease may also be identified based on significant differences in abundance between WT and mutant samples.

To investigate this, differential abundance analysis was performed for all proteins identified in WT versus E238A biofilm samples. Protein $\log_2$ ratios were calculated by averaging peptide $\log_2$ ratios across both replicates and then ordered by ascending value. A rank-based Z score was calculated for the list of proteins (263 proteins total) and used to construct a Q-Q plot against the $\log_2$ ratios to determine how well the data approach a normal distribution. This allowed the establishment of significance thresholds. Proteins with a high Heavy to Light log ratio (highly positive values) are those more abundant in the E238A mutant and conversely, proteins with a low Heavy to Light log ratio (highly negative values) are more abundant in the wildtype.

Figure 5.4 below ranks all detected proteins based on their increased abundance in E238A mutant biofilms. Proteins on the left of the plot represent antitargets, while proteins on the right of the plot represent potential candidates as these have accumulated most in abundance when flagellinolysin is inactive.

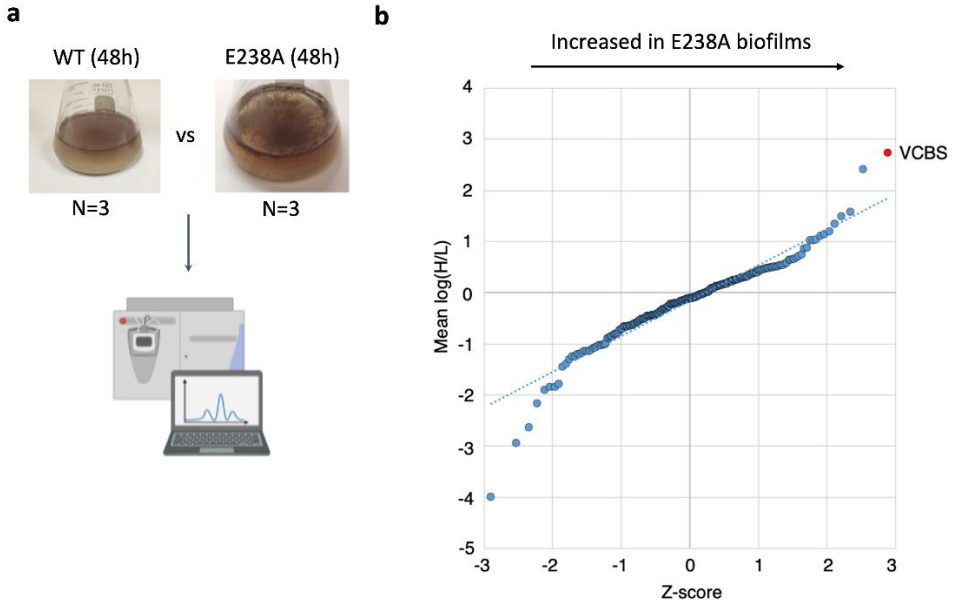


Fig 5.4. Differential protein abundance identifies VCBS as a top flagellinolysin target protein. (a) schematic of analysis workflow. WT versus E238A 48-hr biofilms were compared by shotgun proteomics (LC-MS/MS) to identify differentially abundant proteins. (b) Statistical analysis of differentially abundant proteins between WT and E238A biofilms. VCBS is highlighted as the top increased protein in E238A biofilms relative to WT samples, indicating that active flagellinolysin is associated with decreased VCBS levels in biofilms.

Antitargets

Proteins in the "More abundant in wildtype" category indicate that these are proteins whose abundance are high in regular conditions with an active flagella system. These can be interpreted as “anti-targets”. In other words, once flagella are proteolytically inactive, these proteins either get degraded or have their expression inhibited. Of note are the carbon storage regulator - a transcription factor that appears to mediate multiple phenotypes involving cell growth and carbon utilization. As this protein is not extracellular, it could be due to the presence of lysed cells in biofilms. Interestingly, lpxA (A4C5Z1) is the first step in lipid A biosynthesis - a component of LPS, and a critical molecule for membrane integrity in Gram negative organisms. This would suggest that once flagellinolysin is proteolytically inactive, *P. tunicata* may be defective in membrane integrity resulting in aberrant morphology or fluidity.

Potential flagellinolysin targets

Proteins in the "more abundant in E238A mutant" category indicate that these are proteins are in high abundance in conditions where flagellinolysin is proteolytically inactive. In other words, proteins the WT has lost due to flagellinolysin activity. As such, it can be interpreted as proteins that accumulate due to lack of degradation by flagellinolysin directly, or by secondary proteases activated by flagellinolysin. There are far fewer proteins in this category and these are strong candidate targets of flagellinolysin. These candidates are shown earlier in Figure 5.4 and also the top ranked candidates are listed below in Table 5.2

Table 5.2. Protein differences in *P. tunicata* biofilms. VCBS is bolded as the top increased protein in E238A biofilms.

FC over WT	Mean log H/L	Unique Pep. Count	Acc.	Name	Gene	Length
6.7	**2.74**	**6**	**EAR30327.1**	**VCBS**	**PTD2_02121**	**1600**
5.4	2.42	5	EAR29276.1	carbamoyl-phosphate synthase large subunit	carB PTD2_10689	1072
3	1.58	2	EAR29297.1	putative response regulator	PTD2_10794	334
2.8	1.5	2	EAR28947.1	HflC; HflKC is a membrane-associated complex	PTD2_07884	292
2.6	1.36	4	EAR26366.1	50S ribosomal subunit protein L11	rplK PTD2_04901	142
2.3	1.2	2	EAR29282.1	Hsp 24 DnaK nucleotide exchange factor; probable member of the DnaK/DnaJ/GrpE foldase complex	grpE PTD2_10719	194
2.2	1.11	6	EAR26991.1	50S ribosomal subunit protein L3	rplC PTD2_10433	211
2.2	1.15	3	EAR26987.1	30S ribosomal protein S19	rpsS PTD2_10413	92
2.1	1.04	5	EAR26984.1	50S ribosomal protein L16	rplP PTD2_10398	137
2.1	1.04	3	EAR28946.1	HflK complex with HflC	PTD2_07879	396
2	1.03	2	EAR28112.1	ribonucleotide-diphosphate reductase alpha subunit	PTD2_19892	760

The top differentially abundant protein in the E238A mutant is also the most promising candidate, and this is VCBS (EAR30327) detected at 6.7 fold increased abundance in the E238A mutant compared to the WT (Fig 5.4, Table 5.2). VCBS is a 1600 amino acid protein that is annotated in the NCBI as an "autotransporter adhesin" (https://www.ncbi.nlm.nih.gov/protein/EAR30327.1), consistent with the hypothesis put forth in the previous chapter.

Based on its predicted domain composition, VCBS contains predicted cadherin-like repeat domains. Interestingly, similar cell surface-associated biofilm adhesins have been identified in other species, which contain similar repetitive domain architectures, such as the VCBS/Cadherin-like repeat domain containing adhesin LapV in *Vibrio fischeri*[151]. In fact, a BLAST search of LapV against the *P. tunicata* proteome detected VCBS as the top match with an *E*-value of 8×10^{-9} and a percentage identity of ~28%. This strongly suggests that *P. tunicata* VCBS may be a surface adhesin that is related to large biofilm adhesins from other species.

It is also highly intriguing to note that, in *Vibrio fischeri,* the calcium-dependent cysteine protease, LapG, has been shown to proteolytically release the VCBS-containing adhesin, LapV, from the cell surface, as a mechanism for controlling biofilm formation and dispersal[151]. Similar mechanisms have been identified in other species; e.g., in *Pseudomonas fluorescens*, biofilm formation is promoted by the surface adhesin, LapA, and the activity of the LapG protease controls whether LapA is retained or released from cells, thereby acting as a biofilm switch[152].

Other candidate proteins that scored less highly than VCBS include HflKC is a membrane-bound proteolytic complex that negatively regulates FtsH (aka HflB, a ClpP-related protease) - also a metalloprotease. HflKC degrades FtsH. FtsH degrades sigma32 (rpoH, a transcriptional regulator that mediates multiple phenotypes including motility, tumbling switching, etc.), secY (important for folding of outer membrane proteins) and lpxC (involved in biosynthesis of LPS). FtsH tightly regulates the levels of those to maintain correct LPS amounts and folding of outer membrane proteins. Overexpression of HlfKC reduces levels of FtsH, and loss of HflKC in increased activity of FtsH.

5.6 Analysis of the *P. tunicata* secretome reveals additional support for VCBS as a flagellinolysin target

Given the extreme accumulation of VCBS in biofilms when flagellinolysin is proteolytically inactive, combined with previous reports of other protease-adhesin mechanisms for biofilm regulation, it was hypothesized that flagellinolysin may act similar to LapG and function in the release of VCBS during a non-biofilm (planktonic) state.

Wild type and E238A *P. tunicata* cultures were grown overnight with agitation to prevent pellicle biofilm formation, followed by subculturing for a further eight hours to ensure all cells were in a free-swimming planktonic state. Cultures were then centrifuged at low speed (<10,000x *g*) to pellet out any planktonic cells. Supernatant was collected, forming the supernatant sample, and cell pellet was resuspended in fresh solution, forming the cellular sample. Both were frozen and stored on dry ice to halt protease activity and cell growth. Samples were sent to SickKids mass spectrometry for analysis. Total protein from these fractions was then analyzed by LC-MS/MS and the abundance of all detected proteins was compared.

Consistent with the hypothesis that flagellinolysin controls the proteolytic release of VCBS from cells, the top protein with increased abundance in the WT extracellular (secretome) fraction was VCBS (Figure 5.5a). This can be seen more clearly by comparing the number of spectral hits for VCBS across all four samples (Figure 5.5b): VCBS is detected at relatively low levels in all samples except the supernatant of the WT sample.

Finally, as an additional independent experiment, we cultured WT, E238A, and the flagellinolysin deletion mutant strain (hypothesizing that is should also be deficient in release of VCBS into the supernatant), and analyzed the extracellular protein fraction by SDS-PAGE with silver stain (Figure 5.5c). A ΔVCBS mutant (constructed by Dr. Jiujun Cheng, Charles Lab) was included as a negative control. The gel reveals similar protein compositions across all strains, save for a high molecular weight band that is clearly visible and unique to the wild type extracellular sample only. The band

was excised and analyzed by mass spectrometry by the MS facility at the University of Guelph, and confirmed as VCBS. This is also confirmed by the lack of the band in the VCBS deletion mutant.

Table 5.3. A list of peptides identified in the excised high molecular weight band identified in Figure 5.5c.

Peptide sequence	UniProt identifier – Protein name
K.ATTVIDSVLNVK.F	A4C445 – VCBS
R.GASGSFDVPLGK.N	-
K.LFAVDLAAGTVR.E	-
K.NITYVLAR.D	-
K.PSLLSSINYR.I	-
K.VTLTGFATDADK.L	-
MKKIQK.A	A4C6J3 - glucose-1-phosphate uridylyltransferase
H.RRFAQWGSKLDIAPATLNSR.F	A4CBG9 – Pyrimidine 5'-nucleotidase
L.KITPNNIEKIR.V	A4C8I0 – Uncharacterized tetratricopeptide repeat protein

Ultimately, this result demonstrates that flagellinolysin activity is required for release (shedding) of the VCBS adhesin from cells. The inability to proteolytically shed VCBS may lead to excess accumulation of VCBS on cells leading to the mutant phenotypes observed in Chapter 4 (hyperaggregation and clumping of cells, premature biofilm formation, and defects in spatial architecture).

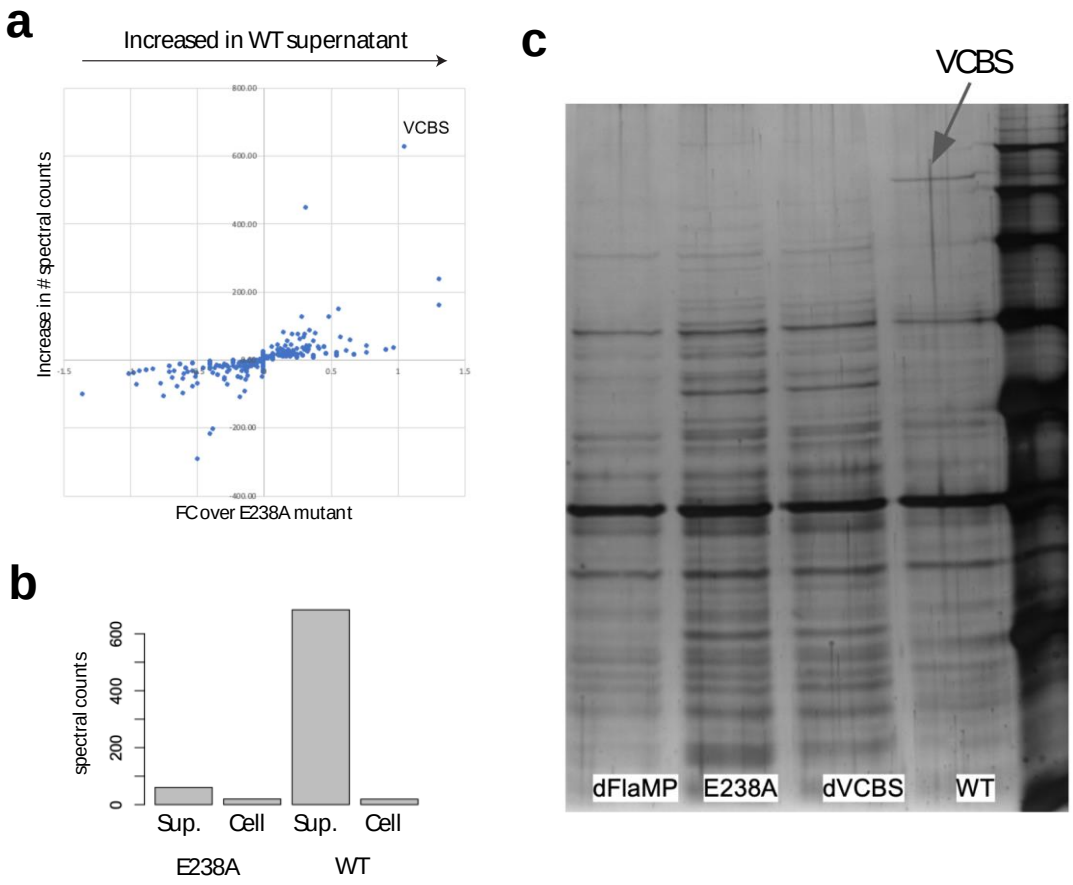


Figure 5.5. Analysis of the WT versus E238A secretome by mass-spectrometry reveals flagellinolysin-mediated release of VCBS into the supernatant. (a) Shotgun proteomic comparison of protein abundance in WT versus E238A supernatants. VCBS is the top protein of increased abundance in WT supernatants compared to E238A. (b) Abundance (spectral counts) of VCBS in four samples: WT and E238A cellular and extracellular (secretome) samples. (c) Silver-stained SDS gel of supernatant from deletion mutant (dFlaMP), E238A, VCBS deletion (dVCBS), and WT cells. The WT strain uniquely releases VCBS, which is visible as a unique high MW band. This band was confirmed as VCBS by mass spectrometry.

5.7 Discussion

In this chapter, a series of methods utilizing mass-spectrometry-based proteomics approaches were used to study the protease substrate specificity of flagellinolysin and identify potential biological substrates. Not only did these approaches reveal a peptidase specificity profile of flagellinolysin, but they also led to the discovery of a promising candidate target (the adhesin protein, VCBS), which is hypothesized to be proteolytically released by flagellinolysin to control biofilm development. This mechanism is reminiscent of other protease-adhesin systems in other species, albeit uniquely localized to the flagellum[151].

The PICS method was used to explore the substrate preferences of peptides capable of being cleaved by the flagellinolysin protease. Several PICS specificity profiles were generated: two based on trypsin and GluC-generated peptides from the *E. coli* peptide library, and three from peptide libraries based on whole proteins (flagellin, flagellinolysin itself, and casein). There are some similar patterns that can be seen across all PICS profiles – most notably, the presence of leucine and other hydrophobic residues at position P2 and P1'. This profile is also similar to that observed previously for the flagellinolysin protease from *C. haemolyticum*, which had a preference for cleaving AVTYY↓LVIA as its consensus peptide[43].

However, one pattern unique to the whole protein-based PICS profiles is the preference for glutamine in several peptide positions. Although the statistical approach used to generate the PICS profiles takes into account the background frequencies of amino acids present in the peptide library, it is possible that it did not completely account for the high glutamine frequency in the proteins being tested. Another possibility is that this reflects the true preference of flagellinolysin protease, and that it was not as well captured in the *E. coli* peptide library. Interestingly, a preference for glutamine residues within the peptide specificity profile has been observed previously for *C. haemolyticum* flagellinolysin when tested on whole protein substrates and also when using purified flagella[43], which suggests that poly-glutamine rich peptides could reflect the true biological substrate preferences of flagellinolysins in general. Further support for this idea comes from the TAILS results (Figure 5.3), which also identified a preference for glutamine-rich peptides, specifically when analyzing differences in protease cleavage among extracellular proteins of wild-type biofilms. In mutant

biofilms, this specificity was not observed, which adds support for the idea that flagellinolysin might be capable of cleaving glutamine-rich (e.g., Q-Q) peptides in extracellular targets.

Interestingly, there are several candidate peptides within the identified VCBS adhesin protein that could be cleavage sites for flagellinolysin depending on which specificity profile is used as a guide. That is, the N-terminal peptide of VCBS contains an intriguing LLAL peptide which matches the profile shown in Figure 5.1a, and also an LQY tripeptide which matches several PICS profiles. N-terminal cleavage of VCBS by flagellinolysin could mediate its extracellular release, similar to that observed for LapG-mediated cleavage of the LapA and LapV adhesins in other species[151,152]. Alternatively, there are also six QL dipeptides and one QQ peptide within the sequence of VCBS, which could be cleavage sites assuming the Q-rich profiles are more reflective of true specificity. Future work that performs protease assays using purified whole VCBS protein or synthesized peptide fragments of VCBS will be critical in order to establish whether VCBS is a direct protease target of flagellinolysin and, if so. determine the precise cleavage site.

Implications of VCBS as a potential flagellinolysin target

Although future work is needed to verify whether VCBS is a direct protease substrate/target of flagellinolysin, numerous independent experiments performed in this Chapter combined with the mutant phenotypic differences observed in previous Chapters suggest that it is a highly plausible target. First, when flagellinolysin's proteolytic activity is inactivated through the E238A mutation, a substantial accumulation of VCBS was detected in biofilms. In fact, VCBS was the top increased protein overall with a nearly 7-fold increase in its abundance. This result suggests that the WT strain in which flagellinolysin is active, reduces the abundance of VCBS in biofilms through proteolytic degradation or by promoting extracellular release of the VCBS adhesin. Indeed, evidence for extracellular release of VCBS was then seen in two independent experiments: one done using shotgun proteomics and the other using an SDS-PAGE analysis to visually identify protein differences in the extracellular (secretome) fraction. Both of these experiments again identified VCBS as the most accumulated protein in the extracellular fraction of WT cells compared to flagellinolysin mutant cells. To confirm that this finding was not unique only to the E238A mutation, experiments were also performed with ΔFlaMP and ΔVCBS strains. As expected, in both of these strains, the extracellular

release of VCBS was abolished as well, indicating that it is a phenomenon that is completely unique to, or dramatically enhanced, in the WT only.

Future studies of VCBS will be needed to determine its precise role in biofilm development of *P. tunicata*, but it appears highly likely that VCBS is a Ca-dependent biofilm adhesin. Bioinformatically, VCBS is related to other biofilm adhesins, particularly in its C-terminal region containing numerous repetitive immunoglobulin-like and cadherin-like repeat domains. Furthermore, in unpublished work, VCBS has been shown to be the top biofilm-associated protein in *P. tunicata* based on comparative proteomics, and its deletion mutant shows major defects in biofilm formation in pellicles as well as in crystal violet assays (Sura Ali, unpublished doctoral work).

If flagellinolysin mediates the proteolytic release of VCBS to avoid premature biofilm formation, this has several implications. It would suggest that flagellinolysin-VCBS may be analogous to the LapG-LapA protease-adhesin systems found in other species. It also raises a key question: how would flagellinolysin, a protease localized to the surface of flagellar filaments that are covered by a membranous sheath, access VCBS? One intriguing hypothesis is that VCBS itself is localized to the flagellar sheath membrane, or that it is broadly present in the outer membrane of *P. tunicata*, but periodically becomes localized within the flagellar sheath when the sheath forms from outer membrane extension. Studies of the flagellar sheath have indeed shown that it is an extension of the outer membrane and may therefore carry outer membrane proteins as well.

Does flagellinolysin generally cleave outer-membrane proteins in the flagellar sheath?

Given the idea that flagellinolysin may cleave outer membrane protein present in the sheath, it is also then worth considering the additional outer membrane proteins that were identified as possible flagellinolysin targets by TAILS. These include peptidoglycan-associated protein (PAL, EAR26516), its gene neighbor YbgF/cpoB (EAR26515), two TonB proteins [TonB (EAR30271), TonB2 (EAR29314)], and an MlaC-like transporter (EAR26697). All of these proteins were identified as high-confidence TAILS targets and are periplasmic or outer membrane associated. It is possible that flagellinolysin therefore does not have a single target (e.g., VCBS), but multiple targets (e.g., VCBS

and multiple outer membrane proteins) that end up in the flagellar sheath and contain cleavage sites matching flagellinolysin's substrate specificity. A schematic depicting this behaviour is shown below in Figure 5.6. In this model, if sheathed flagella are contiguous with the outer membrane, membrane attachment proteins might accumulate in the sheath if not periodically removed. These proteins could produce torsion in the sheath membrane. This could have a range of impacts, ranging from a slowing of the filament rotation due to drag in the sheath to the complete destruction of the sheath from tearing or the paralysis of the filament through membrane tension. Flagellinolysin would then serve the useful purpose of clearing these anchors to allow for free filament rotation. However, no differences were observed in motility under microscope or in soft agar and TEM observations did not show mutant flagella to lack sheaths when attached to cells (see Chapter 4). Future work to analyze this possibility would likely require direct membrane staining and observations of single cells during swimming to look for membrane spinning, similar to methods used by Chen and colleagues to observe sheathed flagellar growth in *V. alginolyticus*[153].

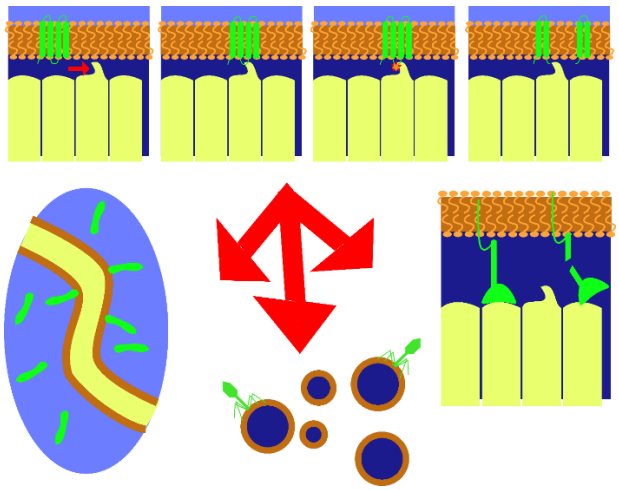


Figure 5.6. Hypothetical model illustrating flagellinolysin's method of action for processing sheath-associated proteins. Sheath-associated proteins drift as a part of normal membrane fluidity and encounter flagellinolysin, whereupon they undergo proteolytic processing. This can result in release of sheath-associated proteins (lower-left), the release of phage decoys (lower middle), or the cleavage of filament-binding proteins (lower-right).

The biology of flagellins makes them ideal for highly localized proteolysis. Flagellins are kept unfolded during transcription and are exported in an unfolded form, only assuming their final conformation once integrated into the filament. This would preclude flagellinolysin in *P. tunicata* from acting on internal proteins. Its sheath would also prevent it from acting on extracellular proteins unless the sheath is first stripped. While sheath is shed during flagellar rotation, the brief window of exposure this would entail makes it more likely flagellinolysin is acting on the interior-facing portions of sheath-associated extracellular proteins.

Outer membrane vesicle shedding

An additional biological feature of sheathed flagella that is worth considering is their ability to aid in the production of outer membrane vesicles through flagellar rotation[39]. The shedding of outer membrane vesicles (OMVs) through flagellar rotation may provide an evolutionary advantage to a sheathed flagellum bearing flagellinolysin: the production of anti-viral decoys. Some bacteriophage, such as bacteriophages T5 and H8, are capable of entering a cell via TonB-dependent receptors. The production of "shut" receptors that are opened by proteolytic processing and then released through OMV shedding would act as a way of creating vesicular traps to absorb free-floating phage from the surrounding medium. This kind of mechanism has been observed in other sheathed bacteria, such as *V. cholerae*[154,155]. Organisms that prematurely enter a dense biofilm state may use this mechanism as a way to clear the surroundings of phage before they enter a state where they engage in increased gene transfer and are highly vulnerable to phage attack.

However, while not incompatible with other functions, phage decoys do not explain the effects flagellinolysin's inactivation has upon the development course of biofilms. The formation of dense microcolonies without undergoing lateral spread is a phenotype resembling those found in mutants with flagellar switching disabled (see Figure 4.8). This effect speaks more of a motility phenotype than one that is phage-related. VCBS, an adhesin-related protein, presents a more plausible candidate for a target that produces a biofilm-related phenotype. By releasing adherent proteins, it could allow

cells to maintain motility in a near-surface state. This function could also be generalized to unsheathed flagella too since flagellinolysin could release bound flagella from organic substrates.

Chapter 6
Discussion

6.1 Brief summary

Flagellinolysin is a novel protein: the first identified enzymatically active flagellin, whose function and biological role prior to this thesis was largely unknown. The overarching objective of the work presented in this thesis was to attempt to characterize flagellinolysin as it exists in the natural world and its biological function.

In Chapter 3, I began with a bioinformatic overview of flagellinolysin's distribution across the bacterial tree of life, and then a focused on analysis of an identified flagellinolysin gene in *Pseudoalteromonas tunicata*, its genomic context, predicted regulation by FliA, and its inferred 3D structure. It was then studied biochemically through recombinant protein expression and purification, and demonstrated to be an active metalloprotease using a zymography assay. Catalytically inactive mutants were generated, including making an E238A mutation within the zinc-binding site.

In Chapter 4, I then investigated two mutant strains in which flagellinolysin was inactivated by point mutation (E238A) or deletion (ΔFlaMP). Although both mutants were motile and showed no observable differences in growth behaviour or flagellar structure, I observed significant impacts on their biofilm structure and development. In particular, flagellinolysin mutant strains showed hyperaggregation and were defective in biofilm spreading on a surface. I investigated this phenotype in greater detail in the E238A, which showed the most exaggerated phenotypic difference and produced highly aggregated biofilms that were more amenable to further characterization. The biofilm hyperaggregation and surface spreading defect phenotype was identified and quantified in confocal microscopic imaging assays of glass surface attached static biofilms as well as in more natural flow-cell conditions. Based on these results, it was hypothesized that flagellinolysin cleaves biofilm adhesins to avoid premature aggregation and biofilm formation during motility.

In Chapter 5, I then used proteomic surveys and other approaches to study flagellinolysin's substrate specificity and identify potential flagellinolysin targets. PICS profiling produced a profile of flagellinolysin's peptide specificity, revealing a preference for peptides with certain characteristics (see Chapter 5 discussion). N-terminomics/TAILS profiling identified differences in proteolysis between the wild type and E238A biofilms, revealing several outer membrane and periplasmic proteins as possible targets of flagellinolysin. Analysis of differential protein abundance, however, provided the strongest possible candidate, VCBS, a protein that appears to be a Ca^{2+}-dependent outer membrane adhesin. It was then demonstrated through proteomics and SDS-PAGE that flagellinolysin activity is associated with the extracellular release of VCBS into the supernatant in planktonic conditions. Therefore, disruption of flagellinolysin activity likely results in an inability to release VCBS, an accumulation of VCBS on cells, and an increased potential for cell-cell aggregation and premature biofilm formation. Together, all these data combine to identify flagellinolysin as a biofilm regulator in *P. tunicata*. It also represents the first characterization of a flagellinolysin in an organism with a sheathed flagellum.

6.2 Flagellinolysin's context within the bacterial world

The development of the bacterial flagellum is a story of gene duplication, drift, and fusion. The flagellar proximal rod protein FlgF's closest homolog is the distal rod protein FlgG, suggesting some ancestral duplication event and subsequent selection leading to divergence and specialization of function between the two[16]. The closest homolog to FlgE, the flagellar hook, is FlgG, which also contains an intragenic duplication of the N-terminal region of FlgG. Meanwhile, in *P. mirabilis,* fusions of the organism's two flagellins have led to the development of a third and antigenically distinct family of flagellins[156,157]. Non-chemotactic *S. typhimurium* strains with inactivated *cheY* genes were able to revert to functional chemotaxis by gene fusion or partial deletion events in the ring and switch proteins FliF and FliG[158]. However, prior to the discovery of flagellinolysin, there were no known instances of naturally-occurring flagellar gene fusions between a flagellin and an enzyme. This makes flagellinolysin an intriguing mystery but also one somewhat without precedent: if there are no other enzymatic flagellins, there are heretofore no known roles for enzymatic flagella.

The bioinformatic explorations of this thesis expanded the list of known flagellinolysins to 731 proteins distributed across the bacterial tree of life. The distribution of flagellinolysin within the tree is wide and patchy, with flagellinolysin lacking any evident clustering or confinement to a given branch of the tree. This patchy distribution is indicative of considerable lateral transfer and/or multiple independent instances of the protein evolving – the strong presence of the protein in the *Gammaproteobacteria* and *Firmicutes*, for example, but not the broad swath of intervening genera make it unlikely that flagellinolysin evolved ancestrally to both strains but was then lost independently in all the intervening taxa. However, in *P. tunicata*, the shared identity with *C. haemolyticum*'s flagellinolysin does suggest that at least here there is a common lineage. This is particularly impressive given their disparate environments of soil and seawater. Indeed, flagellinolysin's environmental distribution includes soil, saltwater, freshwater, deep sea sediments, arctic lakes, in association with plant roots, and in the guts of complex eukaryotes. This wide distribution suggests considerable hopping between organisms. The presence of flagellinolysin in flagellar gene clusters and the presence of the FliA promoter also supports horizontal transfer, as operons or clusters are often transferred together and are driven towards selection for shared function within the cluster.

This diversity of environments and organisms complicates and enhances flagellinolysin's biological context. A protein found in a single environment likely serves a purpose unique to that environment, but one widely distributed likely serves a purpose common to all bacteria, and yet even that is complicated by the diversity of the organisms in question. Flagellinolysin-bearing organisms can be polyflagellate or uniflagellate, can swarm or are confined to swimming motility, can have highly proteinaceous biofilms or biofilms that are primarily polysaccharide. Most intriguingly of all, as in the case of *P. tunicata*, they can even have a flagellar sheath. *P. tunicata* represents the third organism with a formally studied flagellinolysin but the first with a flagellar sheath, a structure which would ostensibly interfere with any flagellar protease attempting to attack an external target. All of this suggests that flagellinolysin's function might be species-specific or that it is able to adapt to provide a range of functions.

Here, I have established that flagellinolysin's biological context is – or perhaps contexts are – diverse. While this thesis does contextualize the protein as it exists in *P. tunicata* and its role in regulating biofilm, future work is going to be required to ascertain whether it plays a similar role in other organisms. Already, I have attempted some preliminary work with *P. peli*, *A. olearius*, and *A. radicis*, which is not presented in this thesis, but which provides additional support that flagellinolysin plays a biofilm-regulatory role in those strains as well. Given their relative distance and lack of sheathed flagella, it seems reasonable to conclude that any biofilm-regulatory role on their part would likely follow a different mechanism to that recognized in *P. tunicata*.

6.3 Defective flagellinolysin produces defective biofilms

The development of the E238A and ΔFlaMP mutant strains provided an opportunity for reverse genetic analysis of flagellinolysin in *P. tunicata*. The gene was successfully disrupted in both strains and the E238A mutation shown to eliminate proteolytic activity in constructs. The result in both strains was the development of defective biofilms, which showed premature aggregation and inability to spread across surfaces. The natural conclusion from this was that some kind of motility defect had been introduced but both strains showed evidence of motility, both directly and indirectly. Flagellar swimming was observed directly with light microscopy and swimming motility measured in soft agar. In addition, both strains produced pellicle biofilms, which do not usually form in organisms incapable of flagellar motility. This means that flagellinolysin's inactivation or removal does not inhibit – at least directly – flagellar rotation or swimming motility.

Proteomic analysis also supports biofilm dysfunction. When *P. tunicata* begins to enter a biofilm state, it produces large quantities of VCBS (unpublished results from Ali et al.), a protein which shares homology with *P. fluorescens*' LapA and *V. fischeri* LapV. LapA is a biofilm-specific adhesin, released by the LapG protease, creating a gradient condition between permanent and transient attachment[61,151,152]. While a rise in biofilm VCBS is detected in all strains, VCBS is only detected in the supernatant via silver stain in the wild type and there is a 10-fold difference between the supernatant content between the wild type and mutants. An inability to release biofilm adhesins would remove the ability to maintain the gradient between "sticky" and "non-sticky" cells and adjust the cell's attachment state. This proteomic detection is consistent with other bacterial phenotypes

deficient in their ability to detach from early biofilm microcolonies and spread on surfaces, whether by inactivation of chemotactic machinery or the disabling of flagellar machinery outright.

VCBS would make a suitable target for flagellinolysin. Strikingly, flagellar sheath-specific proteins with similarity to VCBS have been detected in other genera, such as the sheath-associated protein (SAP) of *Magnetococcus*[48]. Future work to characterize the role and localization of VCBS would require the generation of a suitable antibody to localize VCBS to the flagellar sheath, to confirm its role as a sheath-specific adhesin. As a large (1600 aa) surface- or membrane-associated protein with many repetitive binding domains, VCBS is likely challenging for protein expression/purification studies. A more suitable method of assessing its viability as a target would be self-quenching fluorescent peptides drawn from segments of VCBS. This could be done in tandem with differential proteomics examining extracellular VCBS for native cut sites in the protein. The identification of a unique cut site found exclusively in wild type but not mutant samples would serve as a strong clue to any potential cleavage. It is also possible that flagellinolysin could act indirectly, rather than directly cleaving VCBS, possibly by cleaving a pore protein it uses to translocate or to isolate to the flagellar sheath. It is even possible that some of the identified outer-membrane proteins (e.g., PAL) might act as linker proteins that play a role in attaching VCBS to the surface, and that their cleavage by flagellinolysin is what promotes the extracellular release of VCBS.

The flagellinolysin story laid out here is complicated by differences between the two mutant strains. While both do generate biofilm defects, the phenotypes seem to go in opposite directions: E238A forming thicker and denser biofilms, ΔFlaMP forming weaker and thinner ones. This demands an explanation. This is explored below.

6.4 New model for flagellinolysin in biofilm development

It is well understood that, when bacteria first encounter a surface, the transition from transient to permanent attachment is not necessarily an immediate or irreversible one: during early biofilm development, bacteria need to be able to release from the developing biofilm and spread across the surface. Otherwise, all biofilms would be the singular large lumps seen in the *P. tunicata* mutants.

However, this process is a contest between attachment and spread: if all cells attach, this can result in lumpy aggregates. If all cells spread, this might produce no biofilm at all. Hence there is a need to be able to moderate the ability for cells to attach and release, both to each other and to surfaces. This is enabled by the LapA/G system to which flagellinolysin has been analogized here: regulating the rate of attachment and release.

According to this model, when *P. tunicata* approaches surfaces, it expresses VCBS and also cleaves it through the flagellum-localized protease, flagellinolysin. It is even possible that flagellar rotation itself drives this process. This may not only prevent permanent attachment, but it could play a dual role in pre-coating the surface in a material favourable to adhesion (i.e., VCBS adhesin). As the organism approaches a chemotactically favourable location, it begins to express more VCBS and other adhesins, and slows flagellar rotation. This reduces the cleaving of VCBS and allows for the transition to permanent attachment. This bond is reversible, however, and, when cells receive signals to disperse or rearrange themselves, they may reactivate their flagella and cleave the adhesins to spread or disperse. Cells at the edge of the expanding early microcolonies exist in a dynamic equilibrium between “stay” and “go” that drives the advance of the colony’s frontier.

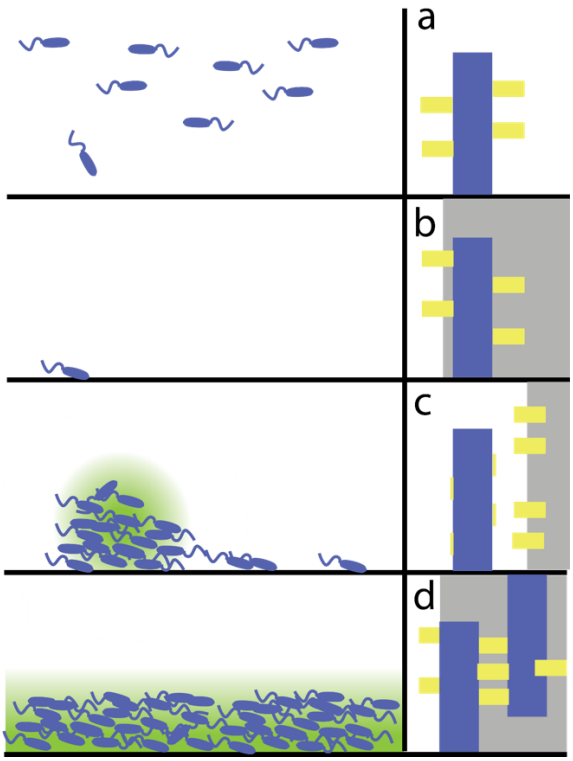


Figure 7.1. Flagellinolysin "stick and release" model for aiding in surface-related spread. (a) A free-swimming cell approaches the surface. (b) The cell makes contact with the surface. Sheath adhesins provide transient attachment. The cell begins to undergo division and cell-cell contact triggers the release of quorum sensing molecules. These molecules downregulate flagellinolysin activity. (c) Cells near the edge of the microcolony release their adhesins and engage in surface-based movement, either by flicking or twitching. (d) As the cell surface layer grows denser, quorum sensing molecules trigger the reduction of cleavage and an end to motion. Biofilm begins to form and thicken.

There are a number of advantages to this model, the first of which being that it accords with a known mode of bacterial surface spread. *Pseudomonas* is known to engage in similar stay-or-go dynamic switching in its flat biofilms, though it uses twitching rather than flagellar motility to drive its spread. However, inactivating flagella switching in these *Pseudomonas* films produces disrupted and large biofilms. *P. tunicata* is not reported to be able to engage in twitching motility but it does contain homologs to *algZ/algR*, the two component regulatory system responsible for mediating both alginate synthesis and twitching motility[95]. It also possesses a number of potential pilin homologs, which may be analogous to Type IV pili (necessary for twitching motility) or curli fibers (facilitating

attachment). These would suggest that *P. tunicata* is likely capable of surface-associated non-flagellar motility in which flagella play an accessory, though not primary, role. Flagella would act less as the main driver in such circumstances and more like an anchor, able to be released or attached as needed.

Another advantage of this model is that it can explain the evident divergence of *P. tunicata*'s mutant strains. An inert enzyme's binding pocket may still retain a strong affinity for its target, which would mean that E238A's flagellinolysin would be able to bind but not cleave VCBS. While both the E238A and deletion mutant would be diminished in their capacity to spread across surfaces, E238A would interfere more severely: binding to anchored VCBS and inhibiting flagellar rotation. The inability of the flagellum to rotate is a common method of mechanosensing, which the cell might interpret as a permanent attachment signal, triggering the rapid activation of late stage biofilm genes. ΔFlaMP mutants would not suffer from this impedance. It is not yet known if or how much the flagellar filament is able to rotate independently of the sheath but *P. tunicata* is not known to possess the O-ring characteristic of many sheathed flagella. Absent some means of isolating the sheath from the rest of the outer membrane, any anchored rotation would induce torsion that would either halt the flagellum or shred the sheath, making it unlikely. Ergo, it is unlikely that the ΔFlaMP mutants are inhibited in rotation the same way E238A mutants are.

Future work will be needed to parse some of the uncertainties in this model. While the presence of biofilm defects in both mutants indicates strongly that flagellinolysin is involved in biofilm regulation, the divergences hint at potential clues and nuances in the mechanism that may complicate a simple explanation. Many of the experiments derived in the previous section would contribute to clearing up this ambiguity. Experiments specific to clearing up this model may involve the development of a strain capable of constitutively expressing VCBS or which expresses it under controlled conditions. Overexpression of VCBS would represent an opportunity to test flagellinolysin's ability to alleviate an overexpression phenotype. This effort will require considerable fine-tuning, as it would be possible that VCBS overexpression could simply overwhelm flagellinolysin's ability to counteract it, or the flagellum's capacity to contain it. For this model to be sufficient, VCBS must also be shown to localize to the flagellar sheath, as flagellinolysin in *P. tunicata* would otherwise be able to directly interact with it.

At the time of writing, additional experiments to investigate this model are underway to use captive monolayers and imaging of cells over time. Thanks to the flagellar sheath, fluorescent membrane-adherent dyes like NanoOrange can be used to identify the direction and motion of the flagellum at the level of a single cell, as has already been demonstrated by Chen and colleagues[153]. Being able to directly observe the degree of flagellar rotation in samples spreading across surfaces, both wild type and mutant, would also provide support for or against this model. It is plausible that rotation may still be observed in ΔFlaMP mutants but its absence or significant reduction in E238A would accord with this model.

6.5 Conclusion

The ambition of this thesis has been to characterize flagellinolysin. In the process, it has answered many questions about flagellinolysin's evolution, ecological niches, regulation, and functional role. As is so often the case, answering one question has the somehow still astonishing capacity to create more: deeper facets of reality's great fractal revealing themselves all the way down. Owing to certain discoveries made during the course of its investigation, it is impossible to say whether the findings within *P. tunicata* are generalizable to a wider variety of species. While the findings here are likely to hold true in other genera with both flagellinolysin and flagellar sheaths, the absence of a sheath adhesin in unsheathed species makes it difficult to conclude that flagellinolysin serves a comparable role in *C. haemolyticum* where it was initially characterized. Even so, the findings here support a model role for flagellinolysin as a regulator of biofilm development, spread, and escape within sheathed flagellar organisms.

It is not impossible or even unlikely that flagellinolysin does also serve a role in a biofilm capacity in other species without flagellar sheaths. Flagella and biofilms are so heavily comingled as processes that a change in one rarely leaves the other unaffected. Flagellinolysin's ability to cleave flagellins may allow it to autocleave bound filaments that undergo coiling or to break down matrix proteins that bind the filament or otherwise inhibit its release during biofilm escape. These roles will be the first to look for in future investigations of the protein. Efforts in that regard are already in their preliminary

stages. HAXXH variants of *P. peli, A. olearius*, and *A. radicis* have all already been generated. Any researchers interested in collaborating on such an investigation or who wish to access the strains may reach out to the Doxey Lab to do so. However, the conclusions of any such investigations are in the future and not within the scope or province of this thesis.

References

1. Berg, H. C. E. coli in Motion. *E. coli Motion* (2004) doi:10.1007/b97370.
2. Münch, R. Robert Koch. *Microbes Infect.* **5**, 69–74 (2003).
3. Rhodes, M. E. The Cytology of Pseudomonas spp. as Revealed by a Silver-Plating Staining Method. *J. Gen. Microbiol.* **18**, 639–648 (1958).
4. West, M., Burdash,', N. M. & Freimuth2, F. Simplified Silver-Plating Stain for Flagella. *J. Clin. Microbiol.* 414–419 (1977).
5. Berg, H. C. & Anderson, R. A. Bacteria Swim by Rotating their Flagellar Filaments. *Nat. 1973 2455425* **245**, 380–382 (1973).
6. Calladine, C. R. Construction and operation of bacterial flagella. *Sci. Prog.* **68**, 365–385 (1983).
7. Cohen-Bazire, G. & London, J. Basal Organelles of Bacterial Flagella. *J. Bacteriol.* **94**, 458 (1967).
8. Blair, D. F. & Berg, H. C. The MotA protein of E. coli is a proton-conducting component of the flagellar motor. *Cell* **60**, 439–449 (1990).
9. Blair, D. F. & Berg, H. C. Restoration of Torque in Defective Flagellar Motors. *Science (80-.).* **242**, 1678–1681 (1988).
10. Dean, G. E., Macnab, R. M., Stader, J., Matsumura, P. & Burks, C. Gene sequence and predicted amino acid sequence of the motA protein, a membrane-associated protein required for flagellar rotation in Escherichia coli. *J. Bacteriol.* **159**, 991–999 (1984).
11. Asai, Y. *et al.* Putative channel components for the fast-rotating sodium-driven flagellar motor of a marine bacterium. *J. Bacteriol.* **179**, 5104–5110 (1997).
12. Chaban, B., Hughes, H. V. & Beeby, M. The flagellum in bacterial pathogens: For motility and a whole lot more. *Semin. Cell Dev. Biol.* **46**, 91–103 (2015).
13. Belas, R. Biofilms, flagella, and mechanosensing of surfaces by bacteria. *Trends in*

Microbiology vol. 22 517–527 (2014).

14. Friedlander, R. S. *et al.* Bacterial flagella explore microscale hummocks and hollows to increase adhesion. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 5624–5629 (2013).

15. Haiko, J. & Westerlund-Wikström, B. The role of the bacterial flagellum in adhesion and virulence. *Biology (Basel).* **2**, 1242–1267 (2013).

16. Liu, R. & Ochman, H. *Stepwise formation of the bacterial flagellar system*. *Proceedings of the National Academy of Sciences of the United States of America* vol. 104 www.pnas.org/cgi/content/full/ (2007).

17. Houry, A. *et al.* Bacterial swimmers that infiltrate and take over the biofilm matrix. *Proc. Natl. Acad. Sci.* **109**, 13088–13093 (2012).

18. Thomson, N. M., Ferreira, J. L., Matthews-Palmer, T. R., Beeby, M. & Pallen, M. J. Giant flagellins form thick flagellar filaments in two species of marine γ-proteobacteria. *PLoS One* **13**, (2018).

19. Macnab, R. M. Type III flagellar protein export and flagellar assembly. *Biochim. Biophys. Acta - Mol. Cell Res.* **1694**, 207–217 (2004).

20. Minamino, T. & Imada, K. The bacterial flagellar motor and its structural diversity. *Trends Microbiol.* **23**, 267–274 (2015).

21. Carroll, B. L. & Liu, J. Structural conservation and adaptation of the bacterial flagella motor. *Biomolecules* vol. 10 1–24 (2020).

22. Beeby, M., Ferreira, J. L., Tripp, P., Albers, S. V. & Mitchell, D. R. Propulsive nanomachines: The convergent evolution of archaella, flagella and cilia. *FEMS Microbiology Reviews* vol. 44 253–304 (2020).

23. Kearns, D. B. A field guide to bacterial swarming motility. *Nat. Rev. Microbiol. 2010 89* **8**, 634–644 (2010).

24. Charon, N. W. & Goldstein, S. F. Genetics of Motility and Chemotaxis of a Fascinating Group of Bacteria: The Spirochetes. *https://doi.org/10.1146/annurev.genet.36.041602.134359* **36**, 47–73 (2003).

25. Kreutzberger, M. A. B. *et al.* Convergent evolution in the supercoiling of prokaryotic flagellar

filaments. *Cell* **185**, 3487-3500.e14 (2022).

26. Song, W. S. & Yoon, S. Il. Crystal structure of FliC flagellin from Pseudomonas aeruginosa and its implication in TLR5 binding and formation of the flagellar filament. *Biochem. Biophys. Res. Commun.* **444**, 109–115 (2014).

27. Takeda, K., Kaisho, T. & Akira, S. Toll-like receptors. *Annual Review of Immunology* vol. 21 335–376 (2003).

28. Steimle, A. *et al.* Flagellin hypervariable region determines symbiotic properties of commensal Escherichia coli strains. *PLoS Biol.* **17**, (2019).

29. Hajam, I. A., Dar, P. A., Shahnawaz, I., Jaume, J. C. & Lee, J. H. Bacterial flagellin—a potent immunomodulatory agent. *Exp. Mol. Med.* **49**, e373 (2017).

30. Smith, K. D. *et al.* Toll-like receptor 5 recognizes a conserved site on flagellin required for protofilament formation and bacterial motility. *Nat. Immunol. 2003 412* **4**, 1247–1253 (2003).

31. Nedeljković, M., Sastre, D. E. & Sundberg, E. J. Bacterial flagellar filament: A supramolecular multifunctional nanostructure. *Int. J. Mol. Sci.* **22**, (2021).

32. Arora, S. K., Ritchings, B. W., Almira, E. C., Lory, S. & Ramphal, R. The Pseudomonas aeruginosa flagellar cap protein, FliD, is responsible for mucin adhesion. *Infect. Immun.* **66**, 1000–1007 (1998).

33. Pallen, M. J. & Matzke, N. J. From the origin of species to the origin of bacterial flagella. *Nat. Rev. Microbiol.* **4**, 784–790 (2006).

34. Diepold, A. & Armitage, J. P. Type III secretion systems: The bacterial flagellum and the injectisome. *Philosophical Transactions of the Royal Society B: Biological Sciences* vol. 370 (2015).

35. Blocker, A., Komoriya, K. & Aizawa, S. I. Type III secretion systems and bacterial flagella: Insights into their function from structural similarities. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 3027 (2003).

36. Gophna, U., Ron, E. Z. & Graur, D. Bacterial type III secretion systems are ancient and evolved by multiple horizontal-transfer events. *Gene* **312**, 151–163 (2003).

37. Zhu, S., Nishikino, T., Kojima, S., Homma, M. & Liu, J. The Vibrio H-Ring facilitates the

outer membrane penetration of the polar sheathed flagellum. *J. Bacteriol.* **200**, 387–405 (2018).

38. Imada, K. Bacterial flagellar axial structure and its construction. *Biophys. Rev.* **10**, 559–570 (2018).

39. Chu, J., Liu, J. & Hoover, T. R. Phylogenetic distribution, ultrastructure, and function of bacterial flagellar sheaths. *Biomolecules* vol. 10 363 (2020).

40. Echazarreta, M. A. & Klose, K. E. Vibrio flagellar synthesis. *Front. Cell. Infect. Microbiol.* **9**, (2019).

41. Subramanian, S. & Kearns, D. B. Functional regulators of bacterial flagella. *Annual Review of Microbiology* vol. 73 225–246 (2019).

42. Wilson, D. R. & Beveridge, T. J. *Bacterial flagellar filaments and their component flagellins*. *Canadian Journal of Microbiology* vol. 39 www.nrcresearchpress.com (1993).

43. Eckhard, U. *et al.* Discovery of a proteolytic flagellin family in diverse bacterial phyla that assembles enzymatically active flagella. *Nat. Commun.* **8**, 521 (2017).

44. Eckhard, U. *et al.* Identification and characterization of the proteolytic flagellin from the common freshwater bacterium Hylemonella gracilis. *Sci. Rep.* **10**, (2020).

45. Zhu, S. *et al*. Molecular architecture of the sheathed polar flagellum in Vibrio alginolyticus. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 10966–10971 (2017).

46. Luket, C. J. & Penn, C. W. Identification of a 29 kDa flagellar sheath protein in Helicobacter pylori using a murine monoclonal anti body. *Microbiology* **141**, 597–604 (1995).

47. Jones, A. C. *et al*. A flagellar sheath protein of Helicobacter pylori is identical to HpaA, a putative N-acetylneuraminyllactose-binding hemagglutinin, but is not an adhesin for AGS cells. *J. Bacteriol.* **179**, 5643–5647 (1997).

48. Lefèvre, C. T. *et al.* Calcium ion-mediated assembly and function of glycosylated flagellar sheath of marine magnetotactic bacterium. *Mol. Microbiol.* **78**, 1304–1312 (2010).

49. Kumar, M. S. & Philominathan, P. The physics of flagellar motion of E. coli during chemotaxis. *Biophys. Rev.* **2**, 13 (2010).

50. Rashid, M. H. & Kornberg, A. Inorganic polyphosphate is needed for swimming, swarming, and twitching motilities of Pseudomonas aeruginosa. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 4885–4890 (2000).

51. Guttenplan, S. B. & Kearns, D. B. Regulation of flagellar motility during biofilm formation. *FEMS Microbiology Reviews* vol. 37 849–871 (2013).

52. Belas, R. When the swimming gets tough, the tough form a biofilm. *Mol. Microbiol.* **90**, 1–5 (2013).

53. Van Dellen, K. L., Houot, L. & Watnick, P. I. Genetic analysis of Vibrio cholerae monolayer formation reveals a key role for ΔΨ in the transition to permanent attachment. *J. Bacteriol.* **190**, 8185–8196 (2008).

54. Flemming, H. C. *et al.* Biofilms: An emergent form of bacterial life. *Nat. Rev. Microbiol.* **14**, 563–575 (2016).

55. Kostakioti, M., Hadjifrangiskou, M. & Hultgren, S. J. Bacterial biofilms: development, dispersal, and therapeutic strategies in the dawn of the postantibiotic era. *Cold Spring Harb. Perspect. Med.* **3**, (2013).

56. Donlan, R. M. *Biofilms: Microbial Life on Surfaces*. *Emerging Infectious Diseases* • vol. 8 http://www.microbelibrary.org/ (2002).

57. Mai-Prochnow, A. *et al.* Hydrogen peroxide linked to lysine oxidase activity facilitates biofilm differentiation and dispersal in several gram-negative bacteria. *J. Bacteriol.* **190**, 5493–5501 (2008).

58. Mai-Prochnow, A. Biofilm development and cell death in the marine bacterium Pseudoalteromonas tunicata. *Appl. Environ. Microbiol.* **70**, 3232–3238 (2004).

59. Petrova, O. E. & Sauer, K. Escaping the biofilm in more than one way: Desorption, detachment or dispersion. *Curr. Opin. Microbiol.* **30**, 67–78 (2016).

60. Davies, D. G. Biofilm Dispersion. *Nat. Rev. Microbiol.* 1–28 (2011) doi:10.1007/978-3-642-19940-0_1.

61. Rumbaugh, K. P. & Sauer, K. Biofilm dispersion. *Nat. Rev. Microbiol.* **18**, 571–586 (2020).

62. Ramasubbu, N., Thomas, L. M., Ragunath, C. & Kaplan, J. B. Structural analysis of dispersin

B, a biofilm-releasing glycoside hydrolase from the periodontopathogen Actinobacillus actinomycetemcomitans. *J. Mol. Biol.* **349**, 475–486 (2005).

63. Monroe, D. Looking for chinks in the armor of bacterial biofilms. *PLoS Biol.* **5**, 2458–2461 (2007).

64. Ghasemi, M., Jenkins, B., Doxey, A. C. & Sivaloganathan, S. A study of nitric oxide dynamics in a growing biofilm using a density dependent reaction-diffusion model. *J. Theor. Biol.* **485**, 110053 (2020).

65. Valentin, J. D. P. *et al.* Role of the flagellar hook in the structural development and antibiotic tolerance of Pseudomonas aeruginosa biofilms. *ISME J. 2021 164* **16**, 1176–1186 (2021).

66. Hobley, L., Harkins, C., MacPhee, C. E. & Stanley-Wall, N. R. Giving structure to the biofilm matrix: an overview of individual strategies and emerging common themes. *FEMS Microbiol. Rev.* **39**, 649–669 (2015).

67. Rawlings, N. D. & Barrett, A. J. *Evolutionary families of metallopeptidases*. *Methods in Enzymology* vol. 248 (1995).

68. Rawlings, N. D., Waller, M., Barrett, A. J. & Bateman, A. MEROPS: the database of proteolytic enzymes, their substrates and inhibitors. *Nucleic Acids Res.* **42**, D503 (2014).

69. Duarte, A. S., Correia, A. & Esteves, A. C. Bacterial collagenases - A review. *Crit. Rev. Microbiol.* **42**, 106–126 (2016).

70. Doxey, A. C., Lynch, M. D. J., Müller, K. M., Meiering, E. M. & McConkey, B. J. Insights into the evolutionary origins of clostridial neurotoxins from analysis of the Clostridium botulinum strain a neurotoxin gene cluster. *BMC Evol. Biol.* **8**, 1–9 (2008).

71. Reid, S. D., Selander, R. K. & Whittam, T. S. Sequence Diversity of Flagellin (fliC) Alleles in Pathogenic Escherichia coli. *J. Bacteriol.* **181**, 153 (1999).

72. Mortimer, C. K. B., Gharbia, S. E., Logan, J. M. J., Peters, T. M. & Arnold, C. Flagellin gene sequence evolution in Salmonella. *Infect. Genet. Evol.* **7**, 411–415 (2007).

73. Deitsch, K. W., Lukehart, S. A. & Stringer, J. R. Common strategies for antigenic variation by bacterial, fungal and protozoan pathogens. *Nat. Rev. Microbiol.* **7**, 493 (2009).

74. Holmström, C., James, S., Neilan, B. A., White, D. C. & Kjelleberg, S. Pseudoalteromonas

tunicata sp. nov., a bacterium that produces antifouling agents. *Int. J. Syst. Bacteriol.* **48**, 1205–1212 (1998).

75. Holmström, C., James, S., Egan, S. & Kjelleberg, S. Inhibition of common fouling organisms by marine bacterial isolates with special reference to the role of pigmented bacteria. *Biofouling* **10**, 251–259 (1996).

76. Rowe-magnus, D. A. Integrase-directed recovery of functional genes from genomic libraries. *Nucleic Acids Res.* **37**, e118 (2009).

77. Bachmann, B. J. Pedigrees of some mutant strains of Escherichia coli K-12. *Bacteriol. Rev.* **36**, 525–557 (1972).

78. Chain, P. S. G. *et al.* Genome of Ochrobactrum anthropi ATCC 49188 T, a versatile opportunistic pathogen and symbiont of several eukaryotic hosts. *J. Bacteriol.* **193**, 4274–4275 (2011).

79. Wood, D. W. *et al.* The genome of the natural genetic engineer Agrobacterium tumefaciens C58. *Science (80-.).* **294**, 2317–2323 (2001).

80. Pearson, G. D. N., Woods, A., Chiang, S. L. & Mekalanos, J. J. CTX genetic element encodes a site-specific recombination system and an intestinal colonization factor. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 3750–3754 (1993).

81. Austin, B. & Zhang, X. H. Vibrio harveyi: a significant pathogen of marine vertebrates and invertebrates. *Lett. Appl. Microbiol.* **43**, 119–124 (2006).

82. The genus Salmonella lingieres, 1900 issued by the salmonella subcommittee of the nomenclature committee of the International Society for Microbiology.

83. Bagdasarian, M. *et al.* Specific-purpose plasmid cloning vectors II. Broad host range, high copy number, RSF 1010-derived vectors, and a host-vector system for gene cloning in Pseudomonas. *Gene* **16**, 237–247 (1981).

84. Schroeter, J. Über einige durch Bacterien gebildete Pigmente. *Beitrage zur Biol. der Pflanz.* 109–126 (1872).

85. Baba, T., Bae, T., Schneewind, O., Takeuchi, F. & Hiramatsu, K. Genome Sequence of Staphylococcus aureus Strain Newman and Comparative Analysis of Staphylococcal

Genomes: Polymorphism and Evolution of Two Major Pathogenicity Islands. *J. Bacteriol.* **190**, 300 (2008).

86. Licitra, G. Etymologia: Staphylococcus. *Emerg. Infect. Dis.* **19**, 1553 (2013).

87. Ehrenberg, C. G. *Physikalische Abhandlungen der Koeniglichen Akademie der Wissenschaften zu Berlin aus den Jahren*. (1835).

88. Muller, R. Eine Diphtherie und eine Streptothrix mit gleichen bisuen Farbstoff, sowie Untersuchungen uber Streptothrixarten in Allgemeinen. *Zentr. Bakt. Parasitenk* **1**, 195–212 (1908).

89. Schäfer, A. *et al.* Small mobilizable multi-purpose cloning vectors derived from the Escherichia coli plasmids pK18 and pK19: selection of defined deletions in the chromosome of Corynebacterium glutamicum. *Gene* **145**, 69–73 (1994).

90. VÄÄTÄNEN, P. Effects of Composition of Substrate and Inoculation Technique on Plate Counts of Bacteria in the Northern Baltic Sea. *J. Appl. Bacteriol.* **42**, 437–443 (1977).

91. Beatson, S. A., Minamino, T. & Pallen, M. J. Variation in bacterial flagellins: from sequence to structure. *Trends Microbiol.* **14**, 151–155 (2006).

92. Steimle, A. *et al.* Flagellin hypervariable region determines symbiotic properties of commensal Escherichia coli strains. (2019) doi:10.1371/journal.pbio.3000334.

93. Evans, F. F., Raftery, M. J., Egan, S. & Kjelleberg, S. Profiling the secretome of the marine bacterium Pseudoalteromonas tunicata using amine-specific isobaric tagging (iTRAQ). *J. Proteome Res.* **6**, 967–975 (2007).

94. Egan, S., James, S., HolmstrÃ¶m, C. & Kjelleberg, S. Inhibition of algal spore germination by the marine bacterium Pseudoalteromonas tunicata. *FEMS Microbiol. Ecol.* **35**, 67–73 (2006).

95. Thomas, T. *et al.* Analysis of the Pseudoalteromonas tunicata genome reveals properties of a surface-associated life style in the marine environment. *PLoS One* **3**, (2008).

96. Franks, A. *et al.* molecules Isolation and Structure Elucidation of a Novel Yellow Pigment from the Marine Bacterium Pseudoalteromonas tunicata. *Molecules* **10**, 1286–1291 (2005).

97. Mirdita, M. *et al.* ColabFold: making protein folding accessible to all. *Nat. Methods 2022 196* **19**, 679–682 (2022).

98. Mezulis, S., Yates, C. M., Wass, M. N., E Sternberg, M. J. & Kelley, L. A. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat. Protoc.* (2015) doi:10.1038/nprot.2015.053.

99. Kelley, L. A. & Sternberg, M. J. E. Protein structure prediction on the Web: a case study using the Phyre server. *Nat. Protoc.* **4**, 363–373 (2009).

100. Eckhard, U., Schönauer, E., Nüss, D. & Brandstetter, H. Structure of collagenase G reveals a chew and digest mechanism of bacterial collagenolysis. *Nat. Struct. Mol. Biol.* **18**, 1109 (2010).

101. Eckhard, U., Schönauer, E. & Brandstetter, H. Structural basis for activity regulation and substrate preference of clostridial collagenases G, H, and T. *J. Biol. Chem.* **288**, 20184–20194 (2013).

102. Yang, Z. *et al.* UCSF Chimera, MODELLER, and IMP: an Integrated Modeling System. *J. Struct. Biol.* **179**, 269 (2012).

103. Delano, W. L. PyMOL: An open-source molecular graphics tool. *CCP4 Newsl. Protein Crystallogr.* 82–92 (2002).

104. Mendler, K. *et al.* AnnoTree: visualization and exploration of a functionally annotated microbial tree of life. *Nucleic Acids Res.* **47**, 4442–4448 (2019).

105. Hickman, J. W. & Harwood, C. S. Identification of FleQ from Pseudomonas aeruginosa as a c-di-GMP-responsive transcription factor. *Mol. Microbiol.* **69**, 376–389 (2008).

106. Arora, S. K., Neely, A. N., Blair, B., Lory, S. & Ramphal, R. Role of motility and flagellin glycosylation in the pathogenesis of Pseudomonas aeruginosa burn wound infections. *Infect. Immun.* **73**, 4395–4398 (2005).

107. Zebian, N. *et al.* Comprehensive analysis of flagellin glycosylation in Campylobacter jejuni NCTC 11168 reveals incorporation of legionaminic acid and its importance for host colonization. *Glycobiology* **26**, 386–397 (2016).

108. Schirm, M. *et al.* Structural, genetic and functional characterization of the flagellin glycosylation process in Helicobacter pylori. *Mol. Microbiol.* **48**, 1579–1592 (2003).

109. Thibault, P. *et al.* Identification of the carbohydrate moieties and glycosylation motifs in

Campylobacter jejuni flagellin. *J. Biol. Chem.* **276**, 34862–34870 (2001).

110. Gomis-Rüth, F. X., Botelho, T. O. & Bode, W. A standard orientation for metallopeptidases. *Biochim. Biophys. Acta - Proteins Proteomics* **1824**, 157–163 (2012).

111. Ali, S. *et al.* Slr4, a newly identified S-layer protein from marine Gammaproteobacteria, is a major biofilm matrix component. *Mol. Microbiol.* **114**, 979 (2020).

112. Lawrence, J. G. & Roth, J. R. Selfish Operons: Horizontal Transfer May Drive the Evolution of Gene Clusters. *Genetics* **143**, 184–1860 (1996).

113. Ballouz, S., Francis, A. R., Lan, R. & Tanaka, M. M. Conditions for the Evolution of Gene Clusters in Bacterial Genomes. *PLOS Comput. Biol.* **6**, e1000672 (2010).

114. Tremblay, B. J. M., Lobb, B. & Doxey, A. C. PhyloCorrelate: inferring bacterial gene–gene functional associations through large-scale phylogenetic profiling. *Bioinformatics* **37**, 17–22 (2021).

115. Wolfe, M. S. Intramembrane Proteolysis. *Chem. Rev.* **109**, 1599 (2009).

116. Dalbey, R. E., Wang, P. & Dijl, J. M. van. Membrane Proteases in the Bacterial Protein Secretion and Quality Control Pathway. *Microbiol. Mol. Biol. Rev.* **76**, 311 (2012).

117. Nakashima, N. & Miyazaki, K. Bacterial Cellular Engineering by Genome Editing and Gene Silencing. *Int. J. Mol. Sci.* **15**, 2773 (2014).

118. Blomfield, I. C., Vaughn, V., Rest, R. F. & Eisenstein, B. I. Allelic exchange in Escherichia coli using the Bacillus subtilis sacB gene and a temperature-sensitive pSC101 replicon. *Mol. Microbiol.* **5**, 1447–1457 (1991).

119. Guerry, P., Alm, R. A., Power, M. E., Logan, S. M. & Trust, T. J. Role of two flagellin genes in Campylobacter motility. *J. Bacteriol.* **173**, 4757 (1991).

120. Kühn, M. J. *et al.* Spatial arrangement of several flagellins within bacterial flagella improves motility in different environments. *Nat. Commun.* **9**, (2018).

121. Maretzky, T., Blobel, C. P. & Guaiquil, V. Characterization of Oxygen-Induced Retinopathy in Mice Carrying an Inactivating Point Mutation in the Catalytic Site of ADAM15. *Invest. Ophthalmol. Vis. Sci.* **55**, 6774 (2014).

122. Houston, S., Hof, R., Honeyman, L., Hassler, J. & Cameron, C. E. Activation and Proteolytic Activity of the Treponema pallidum Metalloprotease, Pallilysin. *PLoS Pathog.* **8**, 17 (2012).

123. Bergström, F. C. *et al.* Scabies Mite Inactivated Serine Protease Paralogs Inhibit the Human Complement System. *J. Immunol.* **182**, 7809–7817 (2009).

124. Altman, M. D., Nalivaika, E. A., Prabu-Jeyabalan, M., Schiffer, C. A. & Tidor, B. Computational design and experimental study of tighter binding peptides to an inactivated mutant of HIV-1 protease. *Proteins* **70**, 678 (2008).

125. Wandersman, C. Secretion, processing and activation of bacterial extracellular proteases. *Mol. Microbiol.* **3**, 1825–1831 (1989).

126. Beaume, M. *et al.* Metabolic pathways of Pseudomonas aeruginosa involved in competition with respiratory bacterial pathogens. *Front. Microbiol.* **6**, 138664 (2015).

127. Menetrey, Q., Dupont, C., Chiron, R., Jumas-Bilak, E. & Marchandin, H. High Occurrence of Bacterial Competition Among Clinically Documented Opportunistic Pathogens Including Achromobacter xylosoxidans in Cystic Fibrosis. *Front. Microbiol.* **11**, 558160 (2020).

128. Jeyaraman, M., Muthu, S., Sarangan, P., Jeyaraman, N. & Packkyarathinam, R. P. Ochrobactrum anthropi – An Emerging Opportunistic Pathogen in Musculoskeletal Disorders – A Case Report and Review of Literature. *J. Orthop. Case Reports* **12**, 85 (2022).

129. Hwang, H.-H., Yu, Manda, Lai & Erh-Min. Agrobacterium-Mediated Plant Transformation: Biology and Applications. *https://doi.org/10.1199/tab.0186* **2017**, e0186 (2017).

130. Pearson, G. D. N., Woods, A., Chiang, S. L. & Mekalanos, J. J. CTX genetic element encodes a site-specific recombination system and an intestinal colonization factor. *Proc. Natl. Acad. Sci. USA* **90**, 3750–3754 (1993).

131. Zhang, X. H., He, X. & Austin, B. Vibrio harveyi: a serious pathogen of fish and invertebrates in mariculture. *Marine Life Science and Technology* vol. 2 231–245 (2020).

132. Won, G. & Lee, J. H. Salmonella Typhimurium, the major causative agent of foodborne illness inactivated by a phage lysis system provides effective protection against lethal challenge by induction of robust cell-mediated immune responses and activation of dendritic cells. *Vet. Res.* **48**, 66 (2017).

133. Revolledo, L. Vaccines and vaccination against fowl typhoid and pullorum disease: An overview and approaches in developing countries Fowl Typhoid and Pullorum Disease Situation in Developing Countries. *J. Appl. Poult. Res* **27**, 279–291 (2018).

134. Diggle, S. P. & Whiteley, M. Microbe Profile: Pseudomonas aeruginosa: opportunistic pathogen and lab rat. *Microbiology* **166**, 30 (2020).

135. Tong, S. Y. C., Davis, J. S., Eichenberger, E., Holland, T. L. & Fowler, V. G. Staphylococcus aureus Infections: Epidemiology, Pathophysiology, Clinical Manifestations, and Management. *Clin. Microbiol. Rev.* **28**, 603 (2015).

136. O'Toole, G. A. Microtiter Dish Biofilm Formation Assay. *J. Vis. Exp.* (2011) doi:10.3791/2437.

137. Klausen, M. *et al.* Biofilm formation by Pseudomonas aeruginosa wild type, flagella and type IV pili mutants. *Mol. Microbiol.* **48**, 1511–1524 (2003).

138. Watnick, P. I. & Kolter, R. Steps in the development of a Vibrio cholerae El Tor biofilm. *Mol. Microbiol.* **34**, 586 (1999).

139. Singh, P. K. Iron sequestration by human lactoferrin stimulates P. aeruginosa surface motility and blocks biofilm formation. *BioMetals* **17**, 267–270 (2004).

140. Serra, D. O., Richter, A. M., Klauck, G., Mika, F. & Hengge, R. Microanatomy at cellular resolution and spatial order of physiological differentiation in a bacterial biofilm. *MBio* **4**, 103–116 (2013).

141. Prigent-Combaret, C. *et al.* Developmental pathway for biofilm formation in curli-producing Escherichia coli strains: Role of flagella, curli and colanic acid. *Environ. Microbiol.* **2**, 450–464 (2000).

142. López-Otín, C. & Overall, C. M. *Protease degradomics: A new challenge for proteomics*. *Nature Reviews Molecular Cell Biology* vol. 3 509–519 (Nature Publishing Group, 2002).

143. López-Otín, C. & Bond, J. S. Proteases: Multifunctional Enzymes in Life and Disease. *J. Biol. Chem.* **283**, 30433 (2008).

144. Wu, J. W. & Chen, X. L. Extracellular metalloproteases from bacteria. *Appl. Microbiol. Biotechnol.* **92**, 253–262 (2011).

145. Bond, J. S. Proteases: History, discovery, and roles in health and disease. *J. Biol. Chem.* **294**, 1643–1651 (2019).

146. Neurath, H. & Walsh, K. A. Role of proteolytic enzymes in biological regulation (A review). *Proc. Natl. Acad. Sci. U. S. A.* **73**, 3825–3832 (1976).

147. Schilling, O. & Overall, C. M. Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. *Nat. Biotechnol. 2008 266* **26**, 685–694 (2008).

148. Schilling, O., Huesgen, P. F., Barré, O., Auf Dem Keller, U. & Overall, C. M. Characterization of the prime and non-prime active site specificities of proteases by proteome-derived peptide libraries and tandem mass spectrometry. *Nat. Protoc. 2010 61* **6**, 111–120 (2011).

149. Kleifeld, O. *et al.* Isotopic labeling of terminal amines in complex samples identifies protein N-termini and protease cleavage products. *Nat. Biotechnol. 2010 283* **28**, 281–288 (2010).

150. Szczepaniak, J., Press, C. & Kleanthous, C. The multifarious roles of Tol-Pal in Gram-negative bacteria. *FEMS Microbiology Reviews* vol. 44 490–506 (2020).

151. Christensen, D. G., Marsden, A. E., Hodge-Hanson, K., Essock-Burns, T. & Visick, K. L. LapG mediates biofilm dispersal in Vibrio fischeri by controlling maintenance of the VCBS-containing adhesin LapV. *Mol. Microbiol.* **114**, 742 (2020).

152. Boyd, C. D., Chatterjee, D., Sondermann, H. & O'Toole, G. A. LapG, Required for Modulating Biofilm Formation by Pseudomonasfluorescens Pf0-1, Is a Calcium-Dependent Protease. *J. Bacteriol.* **194**, 4406 (2012).

153. Chen, M. *et al.* Length-dependent flagellar growth of vibrio alginolyticus revealed by real time fluorescent imaging. *Elife* **6**, (2017).

154. Reyes-Robles, T. *et al.* Vibrio cholerae outer membrane vesicles inhibit bacteriophage infection. *J. Bacteriol.* **200**, (2018).

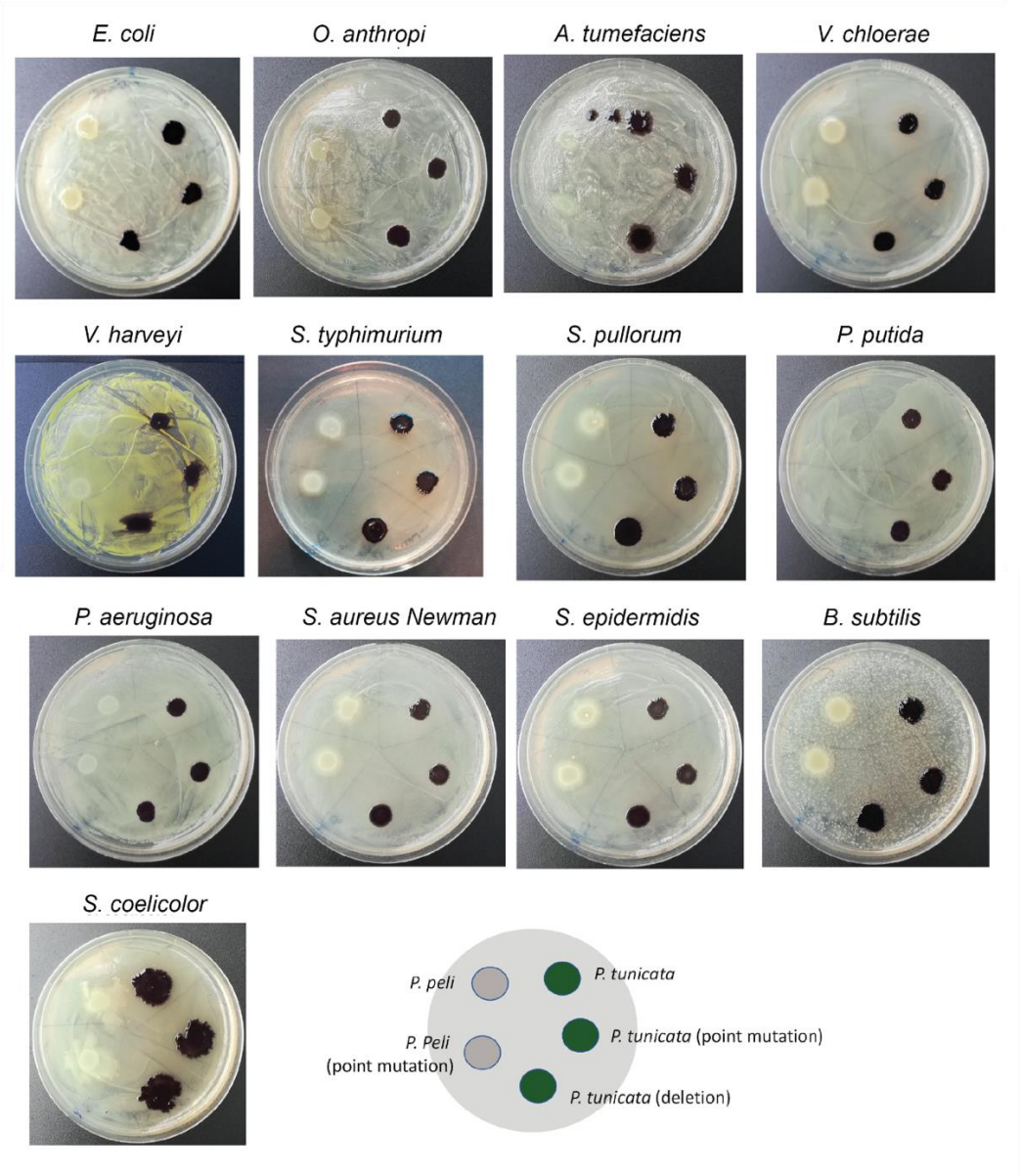
155. MacDonald, I. A. & Kuehn, M. J. Offense and defense: Microbial membrane vesicles play both ways. *Res. Microbiol.* **163**, 607–618 (2012).

156. Murphy, C. A. & Belas, R. Genomic rearrangements in the flagellin genes of Proteus mirabilis. *Mol. Microbiol.* **31**, 679–690 (1999).

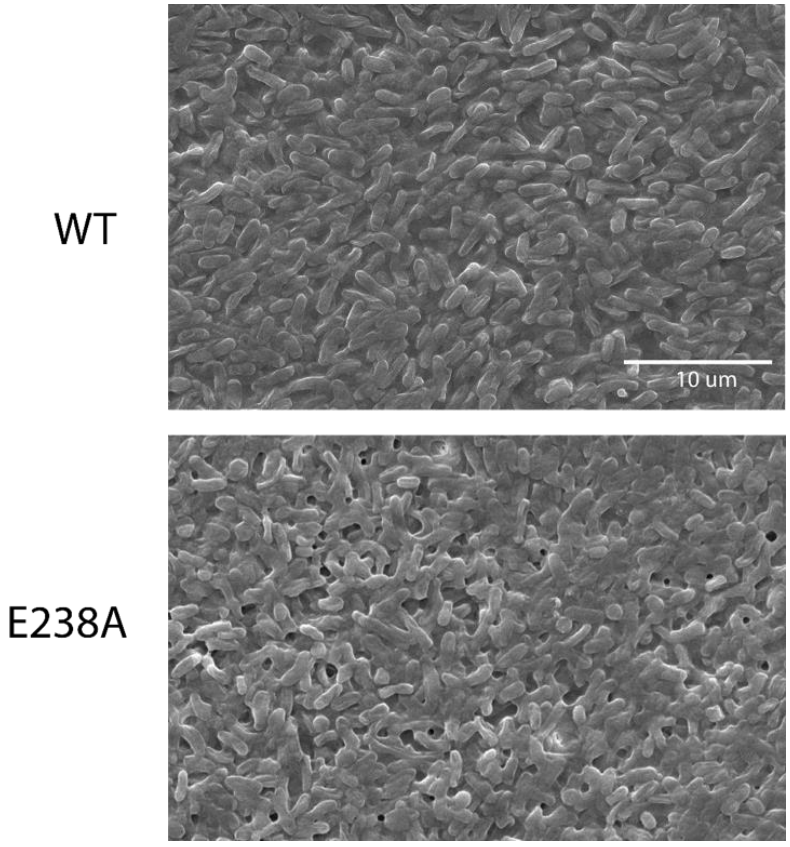
157. Manos, J. & Belas, R. Transcription of Proteus mirabilis flaAB. *Microbiology* **150**, 2857–2863 (2004).

158. Francis, N. R. *et al.* Localization of the Salmonella typhimurium flagellar switch protein FliG to the cytoplasmic M-ring face of the basal body (bacterial flageflum/rotary motor/FliF). *Cell Biol.* **89**, 6304–6308 (1992).

Appendix A
Additional Figures



Appendix Figure A.1. An array of plates showing *P. tunicata* wild type and mutant colonies on lawns of a variety of species. The colonies are of variable size and shape by plate but always of roughly similar size. There are no identified differences in any zones of clearance. *Pseudomonas peli*, which also possesses flagellinolysin, was also included on the plates, along with a point mutant. This figure was produced by Dr. Jiujun Cheng.



Appendix Figure A.2. SEM images of *P. tunicata* wild type and E238A colonies. Both images are at the same scale. SEM imaging revealed no clear visual differences in surface structure of colonies grown on agar surfaces. Both colonies expressed mucoid ECM material and showed a largely undifferentiated lawn of cells. The mucoid material appeared slightly more abundant in the wild type than the mutant but there is no quantitative measure available here. No microstructures, such as channels or pores, were observed, possibly due to the mucoid layer attempting to retain moisture. When the biofilms were coated with the fixing buffers, they lost their ability to maintain a moisture barrier and the colonies "deflated", which might've also collapsed any differences in 3D structure, rendering them undetectable.

Appendix Table 1. High-confidence extracellular TAILS candidates identified in E238A mutant vs WT. These proteins contain signal peptides, are membrane-associated or are secreted. Cytoplasmic proteins have been excluded.

FC vs WT	TAILS log H/L	preTAILS FC over WT	preTAILS H/L	Unique preTAILS peptides	Proteins	Name	log H/L value RepA	log H/L value RepB	Length	P1'	p5-p1	Signal?	p1'-p5'	p5-p5'
3.20	**1.68**	0.62	-0.70	2	EAR26410.1	putative periplasmic peptidyl-prolyl cis-trans isomerase	1.95	1.41	252	166	VVHYK	signal	GTLID	VVHYKGTLID
3.08	**1.62**	0.62	-0.70	2	EAR26410.1	putative periplasmic peptidyl-prolyl cis-trans isomerase	1.69	1.55	252	164	VVVVH	signal	YKGTL	VVVVHYKGTL
3.07	**1.62**	1.30	0.38	2	EAR26688.1	putative TonB-dependent receptor	1.71	1.53	741	36	AQKAS	signal	FEKIE	AQKASFEKIE
2.01	**1.01**	1.02	0.02	4	EAR27555.1	metalloendopeptidase PepO	1.57	0.45	689	443	DYSAL	signal	EITGD	DYSALEITGD
2.09	**1.06**	1.16	0.21	7	EAR27788.1	zinc metallopeptidase family protein	1.08	1.05	613	207	TLANE	signal	GAKEL	TLANEGAKEL
2.17	**1.12**	1.16	0.21	7	EAR27788.1	zinc metallopeptidase family protein	1.23	1.01	613	203	EQQVT	signal	LANEG	EQQVTLANEG
2.64	**1.40**	1.21	0.28	14	EAR28194.1	TonB-dependent receptor	1.49	1.31	751	632	TFQST	signal	EITQS	TFQSTEITQS
2.15	**1.10**	1.21	0.28	14	EAR28194.1	TonB-dependent receptor	1.13	1.08	751	89	GVWVE	signal	SSGGE	GVWVESSGGE
2.28	**1.19**	1.18	0.24	1	EAR28216.1	TonB-dependent receptor	1.35	1.03	986	37	AEETQ	signal	QVEPQ	AEETQQVEPQ
2.19	**1.13**	0.93	-0.11	3	EAR28477.1	putative cytochrome c oxidase, subunit CcoP	1.19	1.08	321	110	SLDAK	membrane	AAGET	SLDAKAAGET
2.39	**1.26**	0.93	-0.11	3	EAR28477.1	putative cytochrome c oxidase, subunit CcoP	1.66	0.85	321	117	GETVQ	membrane	LDTEI	GETVQLDTEI
2.25	**1.17**	0.93	-0.11	3	EAR28477.1	putative cytochrome c oxidase, subunit CcoP	1.28	1.06	321	119	TVQLD	membrane	TEIQK	TVQLDTEIQK
2.25	**1.17**	1.11	0.16	19	EAR28646.1	TonB-dependent receptor	1.24	1.10	957	685	VTRNC	signal	AADGI	VTRNCAADGI
1.97	**0.98**	1.11	0.16	19	EAR28646.1	TonB-dependent receptor	1.08	0.88	957	664	RAESC	signal	AADSI	RAESCAADSI
2.56	**1.36**	1.11	0.16	19	EAR28646.1	TonB-dependent receptor	1.57	1.15	957	99	NNTIY	signal	ANGLN	NNTIYANGLN
2.24	**1.16**	1.11	0.16	19	EAR28646.1	TonB-dependent receptor	1.02	1.31	957	162	VYGAD	signal	AVAGV	VYGADAVAGV
3.34	**1.74**	1.11	0.16	19	EAR28646.1	TonB-dependent receptor	1.70	1.78	957	674	SIAKE	signal	DQYKE	SIAKEDQYKE
2.40	**1.26**	1.11	0.16	19	EAR28646.1	TonB-dependent receptor	1.35	1.18	957	688	NCAAD	signal	GIPAG	NCAADGIPAG
2.79	**1.48**	1.11	0.16	19	EAR28646.1	TonB-dependent receptor	1.61	1.35	957	101	TIYAN	signal	GLNTT	TIYANGLNTT
2.02	**1.02**	1.11	0.16	19	EAR28646.1	TonB-dependent receptor	1.13	0.90	957	362	KNEAF	signal	GQSTP	KNEAFGQSTP
2.41	**1.27**	1.11	0.16	19	EAR28646.1	TonB-dependent receptor	1.10	1.44	957	681	KEQVT	signal	RNCAA	KEQVTRNCAA
3.14	**1.65**	1.11	0.16	19	EAR28646.1	TonB-dependent receptor	1.57	1.73	957	669	SCAAD	signal	SIAKE	SCAADSIAKE
2.17	**1.12**	0.93	-0.11	2	EAR28816.1	Ubiquinol-cytochrome c reductase, cytochrome c1	1.47	0.76	249	35	LMHAN	signal	NDITD	LMHANNDITD

2.25	**1.17**	0.81	-0.31	34	EAR28894.1	hypothetical protein PTD2_07619	1.41	0.94	579	192	SITVA	signal	AAGQA	SITVAAAGQA
3.05	**1.61**	0.81	-0.31	34	EAR28894.1	hypothetical protein PTD2_07619	2.12	1.10	579	214	DSQIR	signal	FTVKA	DSQIRFTVKA
2.08	**1.06**	0.81	-0.31	34	EAR28894.1	hypothetical protein PTD2_07619	1.23	0.88	579	204	GVTFD	signal	ILDIT	GVTFDILDIT
2.43	**1.28**	0.81	-0.31	34	EAR28894.1	hypothetical protein PTD2_07619	1.31	1.26	579	287	EVTTL	signal	LDADI	EVTTLLDADI
3.05	**1.61**	0.63	-0.67	1	EAR29078.1	hypothetical protein PTD2_08539	1.38	1.84	124	35	VASIC	signal	DYVAA	VASICDYVAA
3.61	**1.85**	0.57	-0.81	3	EAR29083.1	hypothetical protein PTD2_08564	1.86	1.84	118	35	SVRIC	signal	EYTQI	SVRICEYTQI
2.08	**1.05**	1.15	0.21	3	EAR29116.1	outer membrane channel precursor protein	1.03	1.08	453	286	TLTLN	signal	ASYGD	TLTLNASYGD
2.15	**1.11**	1.15	0.21	3	EAR29116.1	outer membrane channel precursor protein	1.37	0.84	453	386	VVSAE	signal	SALQA	VVSAESALQA
2.81	**1.49**	1.33	0.41	1	EAR29158.1	putative Methyl-accepting chemotaxis protein	1.67	1.31	671	442	EMSST	membrane	SQAVT	EMSSTSQAVT
2.04	**1.03**	1.40	0.49	12	EAR29317.1	MotA/TolQ/ExbB proton channel family protein	0.98	1.07	454	179	KVATF	signal	DVEVA	KVATFDVEVA
3.11	**1.64**	1.40	0.49	12	EAR29317.1	MotA/TolQ/ExbB proton channel family protein	1.84	1.43	454	222	QVQPL	signal	GRQPD	QVQPLGRQPD
3.27	**1.71**	1.40	0.49	12	EAR29317.1	MotA/TolQ/ExbB proton channel family protein	1.70	1.72	454	313	QLKNA	signal	GTPNT	QLKNAGTPNT
4.06	**2.02**	1.40	0.49	12	EAR29317.1	MotA/TolQ/ExbB proton channel family protein	1.66	2.38	454	338	NQDVE	signal	NLELK	NQDVENLELK
3.94	**1.98**	1.40	0.49	12	EAR29317.1	MotA/TolQ/ExbB proton channel family protein	1.98	1.97	454	184	DVEVA	signal	QLDGA	DVEVAQLDGA
3.29	**1.72**	1.40	0.49	12	EAR29317.1	MotA/TolQ/ExbB proton channel family protein	1.53	1.90	454	97	ELTLA	signal	SKEQE	ELTLASKEQE
3.04	**1.60**	1.40	0.49	12	EAR29317.1	MotA/TolQ/ExbB proton channel family protein	1.55	1.66	454	104	EQELE	signal	TAKGT	EQELETAKGT
2.59	**1.37**	1.40	0.49	12	EAR29317.1	MotA/TolQ/ExbB proton channel family protein	1.31	1.43	454	180	VATFD	signal	VEVAQ	VATFDVEVAQ
3.61	**1.85**	0.80	-0.32	11	EAR29321.1	hypothetical protein PTD2_10914	1.97	1.73	954	96	AFQAN	signal	GASGT	AFQANGASGT
3.93	**1.97**	0.80	-0.32	11	EAR29321.1	hypothetical protein PTD2_10914	1.86	2.09	954	748	ELTVT	signal	QCALT	ELTVTQCALT
2.10	**1.07**	0.90	-0.16	7	EAR29563.1	flagellin	1.15	0.99	586	76	GISMA	secreted	QTAEG	GISMAQTAEG
2.13	**1.09**	1.42	0.51	2	EAR30211.1	putative Zn-dependent aminopeptidase	1.98	0.19	824	125	TYSSD	signal	LQSGP	TYSSDLQSGP

Appendix Table 2. High-confidence extracellular TAILS candidates identified in WT vs E238A mutant. These proteins contain signal peptides, are membrane-associated or are secreted. Cytoplasmic proteins have been excluded.

FC vs Mut	TAILS log H/L	preTAILS FC over Mutant	preTAILS H/L	Unique preTAILS peptides	Proteins	Name	log H/L value RepA	log H/L value RepB	Entry	Accession	Protein names	Gene names	Length	P1'	p5-p1	Signal?	p1'-p5'	p5-p5'
2.47	**-1.31**	1.15	-0.20	6	EAR26516.1	outer membrane protein putative periplasmic protein contains a protein prenyltransferase domain putative ABC-type transport system, auxiliary component hypothetical protein	-1.7	-0.9	A4CFH8	A4CFH8_9GAMM	Peptidoglycan-associated protein	pal PTD2_22582	178	43	TNQAQ	contains signal peptide	QTTDT	TNQAQQTTDT
2.04	**-1.03**	#N/A	#N/A	#N/A	EAR26515.1	putative ABC-type transport system, auxiliary component hypothetical protein	-0.8	-1.2	A4CFH7	A4CFH7_9GAMM	Cell division coordinator CpoB Putative ABC-type transport system, auxiliary component	cpoB PTD2_22577	248	118	PAAEE	contains signal peptide	VSTIS	PAAEEVSTIS
2.19	**-1.13**	1.79	-0.84	2	EAR26697.1	hypothetical protein PTD2_05795 hypothetical protein PTD2_18595 hypothetical protein PTD2_07159 hypothetical protein PTD2_10899 hypothetical protein PTD2_02046	-1.1	-1.1	A4CF22	A4CF22_9GAMM	PTD2_17197	227	46	FNRIS	contains signal peptide	QDQSL	FNRISQDQSL
2.32	**-1.21**	#N/A	#N/A	#N/A	EAR27159.1		-1.2	-1.2	A4CDW7	A4CDW7_9GAMM	Uncharacterized protein	PTD2_05795	840	43 7	NIMLA	contains signal peptide	NTLAS	NIMLANTLAS
3.58	**-1.84**	#N/A	#N/A	#N/A	EAR27859.1		-2	-1.7	A4CBX2	A4CBX2_9GAMM	Uncharacterized protein	PTD2_18595	279	18 3	AISVL	contains signal peptide	GWENL	AISVLGWENL
2.90	**-1.54**	#N/A	#N/A	#N/A	EAR28802.1		-1.4	-1.7	A4C887	A4C887_9GAMM	Uncharacterized protein	PTD2_07159	387	21 1	MMQTQ	contains signal peptide	DTAEP	MMQTQDTAEP
1.85	**-0.89**	0.80	0.32	8	EAR29318.1		-1.1	-0.7	A4C5R4	A4C5R4_9GAMM	Uncharacterized protein	PTD2_10899	256	57	DNTFD	contains signal peptide	QTQEL	DNTFDQTQEL
2.62	**-1.39**	#N/A	#N/A	#N/A	EAR30312.1		-1.2	-1.6	A4C430	A4C430_9GAMM	Uncharacterized protein	PTD2_02046	338	24 1	DNHSN	transmembrane domain	LVEGN	DNHSNLVEGN
2.58	**-1.37**	#N/A	#N/A	#N/A	EAR30271.1	putative TonB2 protein	-1.4	-1.3	A4C3Y9	A4C3Y9_9GAMM	Protein TonB	PTD2_01841	201	10 9	GLSLD	membrane associated	SSDGE	GLSLDSSDGE
2.45	**-1.29**	0.82	0.29	3	EAR29314.1	TonB2 protein	-1.2	-1.4	A4C5R0	A4C5R0_9GAMM	Protein TonB	PTD2_10879	207	97	MPTIN	membrane associated	TGGMS	MPTINTGGMS