The University Of Sheffield.

Structure and Function of Flagellar Cap Complex in Filament Elongation

Natalie Saber Al-Otaibi

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

Doctor of Philosophy

In the

Faculty of Science

Department of Molecular Biology and Biotechnology

08/2021

ACKNOWLEDGEMENTS

First, I would like to thank my supervisor Dr Julien Bergeron for his continued involvement, advice and support with this project over the course of the last 4 years, before and after his move to King's College London. It couldn't have been easy to maintain the level of genuine excitement about science and emotional support that Julien provides to all the members of his lab, past and present, during the lockdown. And for that, as well as many other qualities, I deeply respect him.

In relation to my work, my sincerest thanks go out to Dr Aidan J. Taylor and Professor David J. Kelly for collaborating on the *C. jejuni* cloning, mutagenesis and motility assays. Especially Dr Taylor, who graciously took time out of his busy schedule to walk me through the microbiology part of working with *C. jejuni* (as I have only worked with *E. coli* expression system prior to this) as well as letting me use the facilities. Many thanks go out to the whole of University of Sheffield Electron Microscopy Facility and especially to Dr Svetomir B. Tzokov for his help with negative stain and cryo-EM sample preparation and data collection on the inhouse microscopes, as well as his infinite patience and hard work along with Dr Christopher J. Hill to keep the facility running efficiently.

Thanks go out to the collaborators outside of the University, namely the lab of Frank DiMaio on their assistance in refinement of the $FliD_{cj}$ structure using ROSETTA. Additional thanks go out to Matteo Pizzuto and Davide Corti from HumAbs BioMed for providing the $FliD_{cj}$ antibodies and contacting us about this collaboration. I would also want to thank Professor Per Bullough's lab for fruitful discussions and making me feel welcome, as well as providing a close-knit community with Bergeron lab which was a pleasure to work in.

ABSTRACT

The bacterial flagellum is a molecular motor that allows motility through the rotation of a long filament protruding from the cell. In enteric pathogens, like *Campylobacter jejuni,* it facilitates adhesion to surfaces, infection and evasion of the immune system. FliD is an oligomeric capping protein at the tip of the filament and has been shown to play an essential part in filament assembly and adherence. However, the structure of the intact FliD cap complex, and the molecular basis for its interaction and assembly of the filament, remains elusive.

While the first cryo-EM structure of FliD from *Salmonella enterica* showed a pentameric complex at low resolution, crystal structures of truncated FliD in several other species have revealed a range of crystallographic symmetries. This and a unique oligomerization of *C. jejuni* flagellar filament into 7 protofilaments suggested that the mechanism of filament assembly across bacterial species is not uniform. The aim was to determine the structure and oligomerization of FliD and flagellar filament across different species to conclude if the varied oligomeric states are functional and analyse the interaction of the two complexes to propose a mechanism of elongation.

This thesis reports the first near-atomic cryo-EM structure of the most complete capping complex to date, revealing a pentameric FliD with previously uncharacterised terminal regions essential to its function. FliD in *Serratia marscecens* and *Pseudomonas aeruginosa* forms pentamers of similar dimensions, contrary to the crystal structures. Through construction of FliD mutant strains in *C. jejuni* and motility assays, the function of the capping protein in motility and filament elongation was analysed. There also was an investigation of potential of *C. jejuni* FliD in therapeutic development due to it being an antibody target. Finally, this work demonstrates that the native *C. jejuni* flagellum filament is 11-stranded and proposes a molecular model for the filament-cap interaction.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

EM	Electron Microscopy
ATP	Adenosine Triphosphate
γ-proteobacteria	gamma-proteobacteria
ε-proteobacteria	epsilon-proteobacteria
MA	methyl-accepting domain
CW	Clockwise
CCW	Counter clockwise
HS	Heparan sulphate
GAG	glycosaminoglycan
T3SS	type III secretion system
Cia	*Campylobacter* invasion antigen
IL-8	Interleukin 8
PRR	pattern recognition receptors
PAMP	pathogen-associated molecular patterns
TLR5	toll-like receptor 5
NLR	Nod-like receptor
OTase	oligosaccharyltransferase
GT	glycosyltransferase
SCV	salmonella containing vacuole
LOS	lipooligosaccharide
GBS	Guillain-Barre Syndrome
CDT	Cytolethal Distending Toxin
DNAse	deoxyribonuclease
NF-kB	Nuclear Factor kappa-light-chain-enhancer of activated B cells
kDa	kilodalton
Kbp	kilo-base pairs
IgG	immunoglobulin g
SIgA	secretory immunoglobulin A

$FliD_{cj}$	*Campylobacter jejuni* FliD
RNAP	RNA polymerase
σ^{70}	Primary sigma factor 70
HBB	Hook basal body complex
MS-ring	Membrane and Supramembrane ring
C-ring	Cytoplasmic ring
IM	Inner membrane
L-ring	Lipopolysaccharide ring
P-ring	Peptidoglycan ring
σ^{28}	Sigma factor FliA
TCS	FlgSR two component system
OM	Outer membrane
TM	Transmembrane helices
PMF	Proton motive force
SEC/GSP	General secretion pathway
LPS	lipopolysaccharide layer
IMF	Ion motive force
TAT	Twin arginine translocation system
T5SS	Type 5 secretion system
T8SS	Type 8 secretion system
C-U	chaperone-usher pilus pathway
T2SS	Type 2 secretion system
T9SS	Type 9 secretion system
T4P	Type 4 pilus
T4aP	Type 4 a pilus
T4bP	Type 4 b pilus
T1SS	Type 1 secretion system
T4SS	Type 4 secretion system
OMC	Outer membrane complex
IMC	Inner membrane complex
T6SS	Type 6 secretion system
T7SS	Type 7 secretion system

AAA+ ATPase	ATPases Associated with diverse cellular Activities
NC	Needle complex
EPEC	enteropathogenic *Escherichia coli*
PDB	Protein Data Bank
R-state	right hand inclination of protofilament supercoiling
L-state	left hand inclination of protofilament supercoiling
EMDB	Electronic microscopy data bank
$FliD_{ec}$	*E. coli* FliD
$FliD_{pa}$	*P. aeruginosa* FliD
$FliD_{se}$	*S. enterica* FliD
$FliD_{sm}$	*S. marcescens* FliD
$FliD_{bb}$	*B. bacteriovorus* FliD
$FliD_{hp}$	*H. pylori* FliD
NMR	Nuclear Magnetic Resonance
TEM	Transmission EM microscope
CCD	Charge-coupled device
DED	Direct electron detector
FT	Fourier Transform
CTF	Contrast transfer function
MTF	modulation transfer function
SPA	Single particle analysis
PSF	Point spread function
RCT	random conical tilt
OTR	orthogonal tilt reconstruction
FSC	Fourier schell correlation
LB	Luria-Bertani broth
MH	Mueller Hinton broth
TTC	triphenyltetrazolium chloride
APS	Ammonium Persulfate
SDS	Sodium Dodecyl sulphate
TEMED	Tetramethylethylenediamine
PBS	Phosphate-buffered Saline

Rpm	Rotations per minute
$FlaA_{cj}$	*C. jejuni* Flagellin A
PCR	Polymerase Chain Reaction
DMSO	Dimethyl sulfoxide
dNTP	Deoxynucleoside triphosphate
IPTG	Isopropyl β-D-1-thiogalactpyranoside
OD	Optical Density
FPLC	Fast protein liquid chromatography
Mpa	Megapascal
PAGE	polyacrylamide gel electrophoresis
ITC	Isothermal titration calorimetry
ISA	Isothermal assembly cloning
AV	Amphotericin β and Vancomycin at working concentrations
rSAP	Shrimp Alkaline Phosphatase
pGEM32ZFΔ$fliD$$Kan^R$	$FliD_{cj}$ knockout plasmid
pCmetK0509Chl^R	Complement plasmid
SDM	Site directed mutagenesis
KLD	Kinase, Ligase and *Dpn*I
ΔH	Enthalpy changes of a reaction in a closed thermodynamic system (kcal/mol)
K_a	Association constant, a special case of equilibrium constant associated with binding between two molecules.
K_d	dissociation constant at equilibrium, an inverse value of the association constant ($1/K_a$)
ΔG	Gibbs free energy change in a closed thermodynamic system (kcal/mol)
ΔS	Entropy changes in a closed thermodynamic system ($kgm^2s^{-2}K^{-1}$)
T	Absolute temperature (K)
K	Kelvin measurement of temperature

R	Gas constant 8.315 ($kgm^2s^{-2}K^{-1}mol^{-1}$)
SPR	Surface Plasmon Resonance

DECLARATION

I, the author, confirm that the Thesis is my own work. I am aware of the University's Guidance on the Use of Unfair Means (www.sheffield.ac.uk/ssid/unfair-means). This work has not previously been presented for an award at this, or any other, university. Any collaboration and joint experiments have been acknowledged and referenced.

A large section of this work has been published in Nature Communications as an Article under Open Access license under the following title: "The cryo-EM structure of the bacterial flagellum cap complex suggests a molecular mechanism for filament elongation." (https://doi.org/10.1038/s41467-020-16981-4)

1. INTRODUCTION

1.1 DIVERSITY OF BACTERIAL MOTILITY

The most likely first glimpses of bacteria were observed in the 17th century under a single-lens microscope by Antonie van Leeuwenhoek, due to the fact that they moved around in the water suspension (van Leeuwenhoek, 1677). However, it was only three centuries later that the mechanism by which bacterial cells swim started to become clearer due to developments in structural biology techniques and equipment. While there are many ways a cell can move through liquid and adhere to surfaces, the extracellular helical rotating propeller, called the flagellum, is widespread across different species. The flagellum comes from Latin word meaning “a little whip”, which is somewhat representative of the eukaryotic flagellum that acts through whipping but misleading for the bacterial one, which acts through rotation (Beeby et al., 2020). Through the 20th century it was believed from early electron microscopy (EM) studies that bacterial flagella were a smaller replica of the eukaryotic flagella and used adenosine triphosphate (ATP) as an energy source (Cohen-Bazire and London, 1967). However, in the last 3 decades, with the breakthroughs in microscopy and structural biology techniques, a much clearer picture of the structure and function of the bacterial flagellum started to emerge.

In bacterial species associated with host colonisation, motility is essential. Bacteria use different methods depending on the type and direction of movement. For moving across solid surfaces, called swarming, some bacteria use Type IV pili such as *Myxococcus xanthus*. Others glide through pilus independent motility, sometimes using a thin filament (*Flavobacterium spp.*) or secretion of polysaccharide slime (*M. xanthus*). *Spiroplasma* species swim through contractile cytoskeleton changing the handedness of the helical cell shape and displacing the media around it for propulsion (Jarrell and McBride, 2008). For both swimming and swarming motility many bacterial species employ flagellar rotation, which allows the cells to not only propel themselves through fluids, but also use their flagella for probing and initial interactions with

surfaces, resulting in formation of biofilms (Rossez et al., 2015). As bacterial species come in different shapes and sizes, so do their flagella. Some species, like *Wolinella succinogenes*, have a single monotrichous flagellum at one pole of the cell. Others, like Spirochaetes*,* have internal flagella located in the periplasmic space (Gibson et al., 2020). There are differences within the Gamma-proteobacteria (γ-proteobacteria), with members of the Enterobacteriaceae (enterics), such as *Escherichia coli* and *Salmonella enterica*, having peritrichous randomly positioned flagella across the cell and non-enterics, such as *Vibrio parahaemolyticus* and *Pseudomonas aeruginosa,* having two kinds: high torque sodium driven polar flagella for swimming and secondary lateral flagella driven by proton motive force for surface swarming (Ferreira et al., 2021; Jarrell and McBride, 2008). In most epsilon-proteobacteria (ε-proteobacteria), such as *Campylobacter jejuni* and *Helicobacter pylori*, both the spiral shape of the bacteria and the presence of amphitrichous flagella, meaning a single or multiple flagella at each cell pole, contributes to motility through highly viscous liquids (Beeby, 2015; Beeby et al., 2016; Cohen et al., 2020). The different types of flagellation are summarised in Figure 1.1.1.

1.2 Role of motility in Bacterial Virulence

Bacterial pathogenic cycles can be broken into four stages. First the pathogen must reach its target or host, then colonise, grow and self-maintain, eventually disperse and return to the initial step. Flagella have adapted to assist various bacterial species in their colonisation (Backert and Hofreuter, 2013; Chaban et al., 2015).

To enhance bacterial capability of finding the intended host, flagella enable movement via sensing different attractant and repellent stimuli. This mainly manifests as chemotaxis: sensing of particular compounds to indicate the direction of movement, which was first observed in 1880s by Engelmann and Pfeffer and first investigated in detail around 1960s (Adler, 1966). A single compound can both attract some organisms and repel others, for example Phenol attracts *E. coli* but repels *S. enterica.* These compounds are sensed by chemoreceptors, which are multidomain proteins either

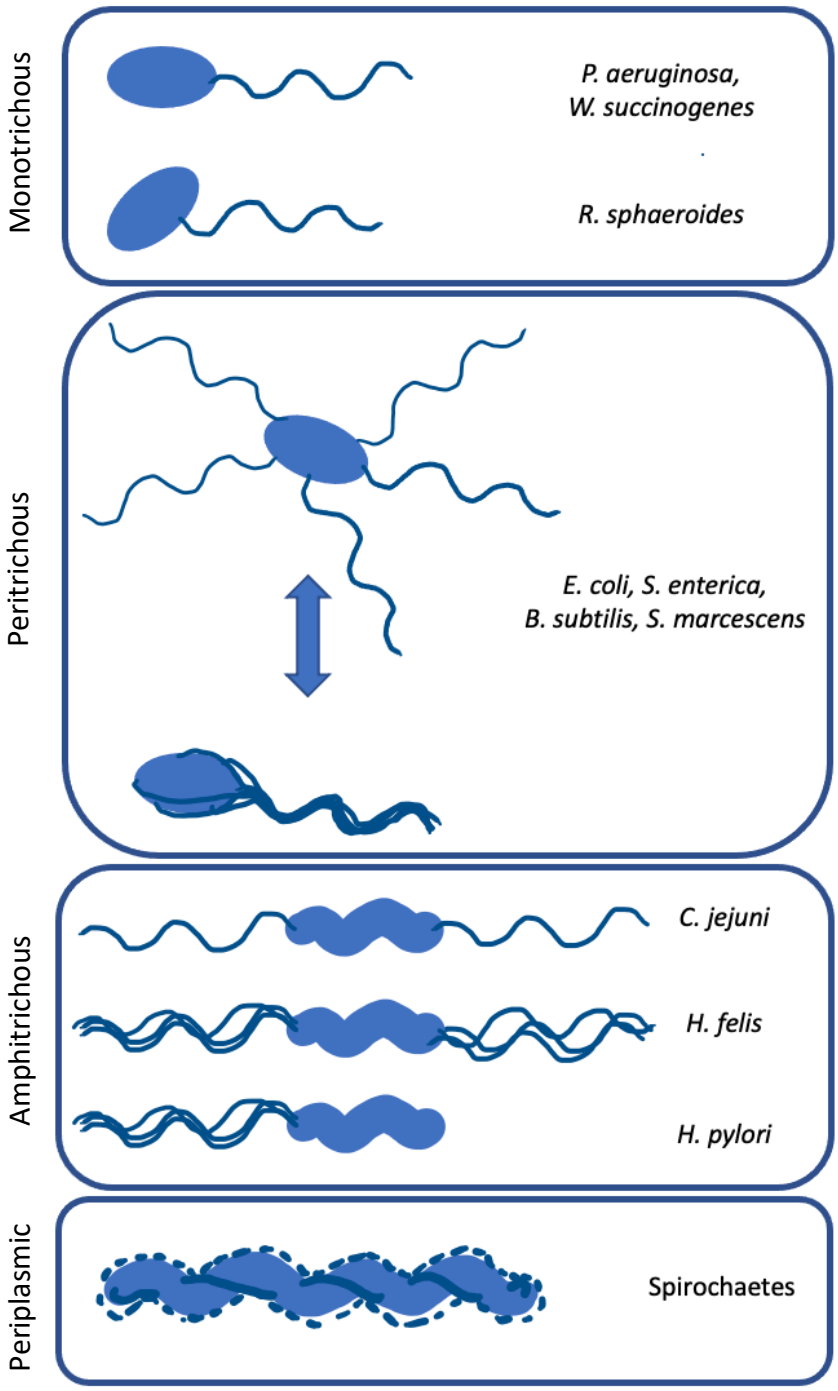


Figure 1.1.1: Diversity of bacterial flagellum arrangements. From top to bottom: A single monotrichous flagellum at the pole propelling the cell forward (*Pseudomonas aeruginosa)*; A single flagellum between the poles (*Rhodobacter sphaeroides)*; 5-10 flagella randomly distributed around the cell, which supercoil into a more powerful propeller (*Escherichia coli, Salmonella enterica, Bacillus subtilis)*; Amphitrichous flagella at cell poles with helical bacterial cell morphology allow for swimming in viscous surfaces (*Campylobacter jejuni)*; Multiple flagella at the two poles (*Helicobacter felis*); Multiple flagella at one cell pole (*Helicobacter pylori)*; Periplasmic flagella cause cells to bend and relax providing movement (Spirochaetes). Adapted from (Beeby, 2015; Eisenbach, 2011).

embedded into the cell membrane or soluble in the cytoplasm. A chemoreceptor would commonly contain a versatile sensory domain capable of direct binding to the ligand or protein-protein interaction with a ligand binding protein and a signalling domain, often called methyl-accepting domain (MA), which is highly conserved and interacts with the rest of the chemosensory cascade.

In a well-studied system such as *E. coli*, the core of the chemotaxis pathway consists of three main components: CheW coupling protein, CheA histidine kinase and CheY response regulator. CheW binds the MA domain of the chemoreceptors and anchors them to the CheA kinase (Lertsethtakarn et al., 2011). CheA kinase phosphorylates CheY, which in turn controls the direction of flagellar rotation by binding the switch proteins in the flagellar C-ring in its phosphorylated state to induce flagellar rotation in a clockwise direction (CW). Upon ligand binding to the receptors, CheA activity is inhibited and non-phosphorylated CheY fails to interact with the switch locking the flagella in the counter-clockwise rotation (CCW) (Chang et al., 2020; Santiveri et al., 2020). In ε-proteobacteria the core of the chemotaxis pathway remains mostly the same. However, their histidine kinases differ from the enteric CheA through the addition of the REC domain contributing to phosphorylation and thus being named CheAY. Several proteins outside of the core set are found in many but not all motile bacteria and augment the function of the main three. CheZ and FliY accelerate the intrinsic dephosphorylation of CheY. CheR and CheB respectively methylate and demethylate chemoreceptor glutamyl residues to control adaptation response to the environment, while CheV is a homolog of CheW with an additional receiver domain. *H. pylori* lacks the CheB and CheR methyl transferases but instead has three CheV variants (Lertsethtakarn et al., 2011). The summary of the chemotaxis signal transduction system is illustrated in Figure 1.2.1.

Other motility stimuli extend to light, temperature and osmolarity. In the case of *C. jejuni* there are seven transmembrane and three soluble receptors responding to a variety of stimuli. CetA transmembrane chemoreceptor is bound to CetB cytoplasmic protein, which is triggered by the response to pyruvate and fumarate. Tlp7 receptor senses formate while Tlp1 transmembrane receptor senses aspartate directly. *H. pylori* mediates low pH response through TlpB, while soluble TlpD senses tactic response to

bacterial energy levels (Lertsethtakarn et al., 2011).

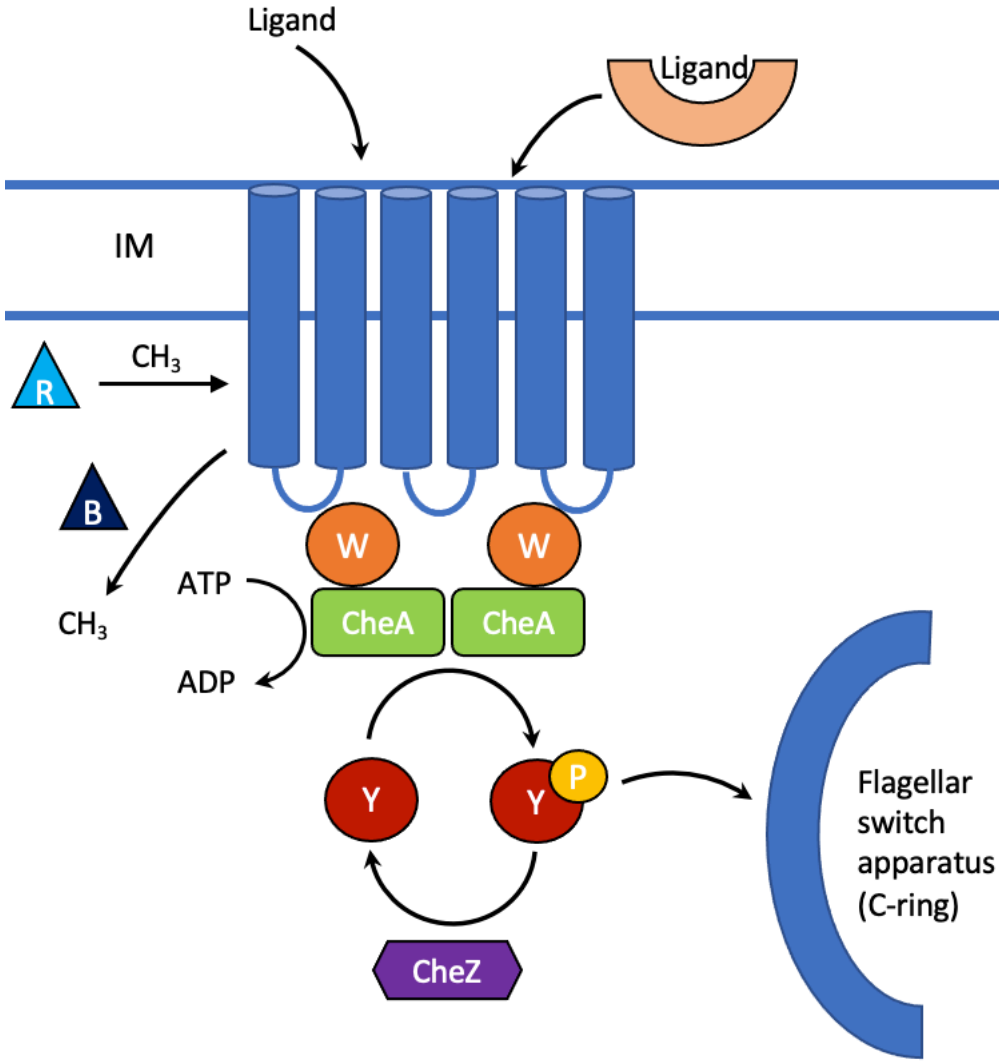


Figure 1.2.1: Diagram of the chemotaxis regulatory pathway in bacteria. Ligands bind directly or through a binding protein (beige) to the receptor, illustrated in this diagram as transmembrane helices (blue bars). The receptor is linked through CheW (W) (and CheV in ε-proteobacteria) to the CheA kinase (CheAY in ε-proteobacteria), which phosphorylates the regulatory protein CheY (Y) that can now interact with flagellar switch apparatus to induce CW rotation. CheZ assists in dephosphorylating CheY and returning the rotational direction to CCW. CheR (R) methylates the receptor and CheB (B) demethylates as an adaptation response. Adapted from Lertsethtakarn et al., 2011.

In enteric pathogens, such as *E. coli*, the trajectory of the cell consists of two main events. A near straight "run" segment, where the peritrichous filaments are bundled together and propel the cell forward through combined CCW rotation and a

brief turning “tumble” event, where one or more of the flagella in the bundle change rotation to CW and break it, thus allowing for a change in direction (Berg, Howard C; Brown, 1972). Some bacteria that colonise the gastrointestinal tract, such as *H. pylori* and *C. jejuni*, enhance their motility as a response to encountering a highly viscous mucus, which needs to be penetrated prior to reaching the infection site (van Putten et al., 2009). In high viscosity media *C. jejuni* wraps one flagellar filament around the cell body leaving a lagging filament with both left-handed filaments having the CCW rotation and the right-handed cell body rotating CW. This wrapped-mode swimming is at a significantly higher speeds than unwrapped. Upon sensing an environment requiring switching of direction, the wrapped filament changes rotation from CCW to CW and within one full revolution the filament unwraps from the cell body. Initially lagging flagellum wraps around the cell body and wrapped filament becomes the lagging. This drastic change in direction allows *C. jejuni* cells to escape narrow crevices in the gastrointestinal tract and reposition for better infection of the host (Cohen et al., 2020).

Upon reaching their target destination pathogens need to recognise they have arrived and stop swimming to activate biofilm formation or swarmer-cell differentiation protocols. The flagellum works as a mechanical sensor for *H. pylori* to find the gastric injury, or for *C. jejuni* to identify and adhere to the abundance of glycoproteins on intestinal mucus crypts. As *C. jejuni* does not have any pili, it adheres through a few potential adhesins such as CadF, which seeks out fibronectin, JlpA, a surface lipoprotein crucial for Hep-2 cell binding, and potentially the FliD flagellar capping protein, which was observed to bind Heparan sulphate (HS) glycosaminoglycan (GAG) on the surface of epithelial cells (Freitag et al., 2017; Young et al., 2007).

Colonisation may differ according to the pathogen. Biofilm formation is the most protected and long-lasting form of multicellular aggregate bound by a matrix of polymers allowing the cells to attach to solid surfaces. As this relies on close packing of the cells, flagellar rotation must be switched off. As illustrated in Figure 1.2.2 a, *P. aeruginosa* first attaches to the surface, turns off the flagellar rotation and forms the biofilm. *S. enterica* tends to use the flagellar type III secretion system (T3SS) to inject

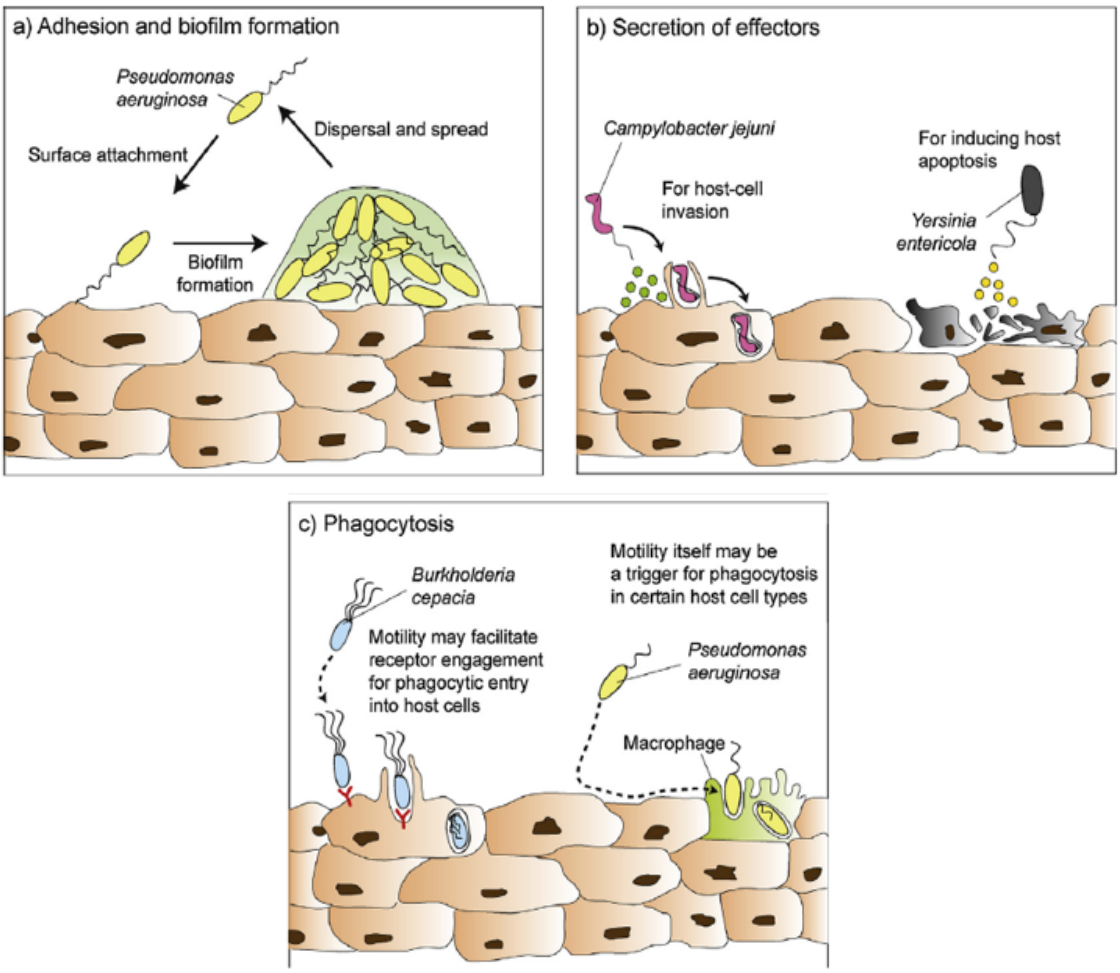


Figure 1.2.2: Summary of bacterial invasion of host cells via flagella. (a) Some organisms like *P. aeruginosa* form biofilms on the surface of the infected cells. (b) *C. jejuni* exports Cia protein to initiate incorporation into a vacuole inside the host cell, while *Y. entericola* induces apoptosis. (c) Some bacteria trigger host receptors to induce phagocytosis into the host cell. Adapted from (Chaban et al., 2015).

virulence factors into the host epithelial cell (Chaban et al., 2015). The T3SS needle is a hollow conduit and has a tip complex containing translocons, which attaches to the host membrane and exports effectors upon contact. In *P. aeruginosa* PopN assembles over the base PcrD ring preventing effector access to the needle, and upon contact with the host PopN dissociates removing the blockage. The effectors induce a range of events, such as invasion and colonization through *S. enterica* SipA and SipC causing actin polymerization within the host, or evasion of the immune system through *E. coli* EspF assisting in evasion of macrophage internalization (Hajra et al., 2021).

Meanwhile, *C. jejuni* lacks a separate T3SS and thus uses its flagellar T3SS to inject the *Campylobacter* invasion antigens (Cia) and FspA proteins to promote internalization of the cells into the host, as illustrated in Figure 1.2.2 b. Multiple exported proteins, including CiaD that activates a pathway for interleukin 8 (IL-8) secretion thus inducing inflammation, contain putative T3SS secretion signal which is recognised by the flagellar T3SS (Backert and Hofreuter, 2013; Samuelson et al., 2013).

Rotation of flagella can drive bacterial penetration into tissues, as in the case of *P. aeruginosa* infection, where motility itself is a trigger for phagocytosis, as illustrated in Figure 1.2.2 c (Rossez et al., 2015). However, flagella do not mechanically bore through cell membranes due to their inability to generate the required force to pierce the lipid bilayer (Chaban et al., 2015). Another unique method of infection is tricking the host cell receptors to induce phagocytosis and remain in a vacuole for growth and dispersion later on. However, as the interior of the vacuole is a hostile environment to the bacterial cell with a low pH, poor nutrient content and abundance of lysosomal enzymes, the pathogen must adapt and resist these conditions or alter the environment around it by affecting host translation and transcription (Cossart and Sansonetti, 2004).

The eukaryotic innate immune system has a set of pattern recognition receptors (PRRs) to recognise the pathogens' highly conserved pathogen-associated molecular patterns (PAMPs). As the flagellum has a 13 amino acid PAMP which is sensed by the toll-like receptor 5 (TLR5) extracellularly and Nod-like receptor (NLR) intracellularly, pathogens must modify or turn off flagellar expression to hide their flagella and evade the immune system (van Putten et al., 2009). Modifications include alternating expressed flagellins within the same population (phase variation), different subsets of cells expressing and lacking flagella (bistability), altering the PAMP sequence to evade TLR5 receptors and adding glycosylation to the filament along with other post-translational modifications (Chaban et al., 2015).

Glycosylation involves addition of glycan to specific residues of proteins as a means of post translational modification. There are three different types of glycosylation: N-linked, C-linked, and O-linked. N-linked glycosylation occurs when a sugar molecule is attached to a nitrogen atom on an asparagine by

oligosaccharyltransferase (Otase) (Gavel and Heijne, 1990). This is common in proteins used for cell differentiation, signalling and pathogenesis, where preassembled blocks of sugars are transferred onto a lipid anchor then flipped from cytosolic side to the outer layer of the bacterial membrane where they are glycosylated (Schwarz and Aebi, 2011). C-linked glycosylation is a rare covalent attachment of mannose to the C2 carbon of a tryptophan residue, usually restricted to mammals (Crine and Acharya, 2021). O-linked glycosylation is a covalent linkage of sugar to the side chain hydroxyl of serine, threonine, or tyrosine. The process usually occurs in the cytoplasm through either an Otase-dependent pathway, with the similar lipid carrier to the N-linked method flipping the substrate into the periplasm or Otase-independent pathway, where monosaccharides are directly transferred onto acceptor proteins by cytoplasmic glycosyltransferases (GTs), which are then exported or secreted through a secretion system (Iwashkiw et al., 2013). This is the process usually used for flagellin glycosylation (Salah Ud-Din and Roujeinikova, 2018).

While it has not yet been characterised in enteric bacteria, flagellin glycosylation loci have been observed throughout the *Enterobacteriaceae*. An example of potential post-translational modifier of flagellin in *S. enterica* is the FliB methylase (De Maayer and Cowan, 2016). Glycosylation is prevalent in polarly flagellated bacteria but not usually essential for motility. However, lack of glycosylation in ε-proteobacteria prevents formation of a functioning flagellar filament. In *C. jejuni,* both minor and major flagellins undergo this process, while *H. pylori* FlaB flagellin is the one to be glycosylated. This process contributes approximately 10% of the protein mass: 17-19 different serine and threonine residues of FlaA are glycosylated in *C. jejuni* with different chains depending on the strain. The localization of these residues is limited to the external domains of flagellin facing away from the lumen towards the surface, with the majority being in the D4 domain (Kreutzberger et al., 2020). While glycosylation occurs independently of the flagellar regulon and absence of many essential structural components does not prevent production of glycosylated flagellin, the genes required for glycan biosynthesis are located adjacent to the structural genes (Lertsethtakarn et al., 2011). Pseudaminic acid is a common glycosylation chain in *C. jejuni*, synthesis of which is a six-step process requiring PseB, PseC, PseH, PseG, PseI

and PseF respectively as catalysts for each step (Salah Ud-Din and Roujeinikova, 2018). PseC is localized to the poles of the cell, indicating that the glycosylation of flagellin occurs closer to the rest of the structural machinery (Ewing et al., 2009).

The final step of the cycle is dispersal, which may require reinitiation of flagellar growth after successful gestation period within the host cell. Another method is export of the effector proteins into the host cytosol through non-flagellar T3SS, inducing an inflammatory response resulting in host cell apoptosis and release of pathogen cells to continue the cycle from the beginning (Cossart and Sansonetti, 2004). *S. enterica* exports SopB to induce salmonella containing vacuole (SCV) formation and maturation to evade host response, while *E. coli* effector EspH induces cytoskeletal rearrangements within the host cell to better persist within the host. *E. coli* EspF is then secreted to induce cell death and mitochondrial dysfunction and disseminate the pathogen through the area of infection (Hajra et al., 2021).

1.3 FLAGELLUM AS THE TARGET OF NOVEL THERAPEUTICS

As motility is one of the essential functions for the bacterium to successfully colonise its host, targeting the proteins that make up the flagellum in development of new therapeutics is a viable option. While many bacteria utilise both their secretion systems and flagella to infect and propagate, *C. jejuni* does not have a T3SS injectisome and instead secretes its effectors through the flagellar T3SS (Samuelson et al., 2013). Which is why this organism is often chosen as a model for development of motility related anti-bacterial therapies. It is also an interesting subject of study as *C. jejuni* flagellum manages to evade stimulating the TLR5 receptor unlike other bacterial flagella. While in *P. aeruginosa* glycosylation promotes TLR5 evasion, in *C. jejuni* it has been observed that glycosylation appears to have no effect on evasion or activation of the receptor. The previously characterised TLR5 binding sequence in the N-terminal domain of *Salmonella* flagellin was not-conserved across *C. jejuni* strains. At least three different sections of *Salmonella* flagellin were required to induce TLR5 response in *C. jejuni* chimeric protein and with the glycosylation having no effect, it appears that the

evasive capability of FlaA is located within the sequence itself (De Zoete et al., 2010). More specifically, the range between 86 and 96 residues is the TLR5 recognition site in *B. subtilis* containing conserved 89R and 93L residues. This is different from *C. jejuni* TLR5 recognition site where these are mutated to 91T and 95K aiding TLR5 evasion (Kreutzberger et al., 2020).

Campylobacter infection in humans is a main cause of gastroenteritis and currently causes about 400 million cases per year. It manifests as diarrhoea, fever and abdominal cramps after a 2-5 day incubation period and the symptoms can last for up to two weeks (Dasti et al., 2010; Young et al., 2007). In the long term, *C. jejuni* lipooligosaccharide (LOS) structure may mimic neuronal gangliosides in the human body thus triggering antibody production and autoimmune response, which after 1-3 weeks of infection can lead to Guillain-Barre Syndrome (GBS), the most common cause of neuro-muscular paralysis world-wide (Epps et al., 2013; Hughes, 2004). In addition, *C. jejuni* strains have developed antibiotic resistance to the majority of antimicrobials such as penicillians, cephalosporins and vancomycin (Epps et al., 2013).

C. jejuni pathogenesis is different to that of many other bacteria in the sense that it does not secrete a large variety of toxins other than the Cytolethal Distending Toxin (CDT) and CiaD, which both induce secretion of IL-8 from host cells (Samuelson et al., 2013). CDT as a part of AB toxin family consists of 3 subunits CdtA, CdtB and CdtC. Subunit B internalises in eukaryotic cells with its deoxyribonuclease (DNAse) I like activity within the nucleus and arrests the cell cycle while stimulating proinflammatory genes in the Nuclear Factor kappa-light-chain-enhancer of activated B cells (NF-kB) pathway (Faïs et al., 2016). One of the other secreted proteins is FspA, an 18 kilodalton (kDa) protein which binds eukaryotic cells and induces apoptosis (van Putten et al., 2009). Within *C. jejuni* small genome of about 1700 kilo-base pairs (Kbp) there are multiple hyper variable sequences causing extensive genetic variation not only among the species but even among the strains (Young et al., 2007). Thus *C. jejuni* strain 11168 is used as a standard laboratory strain while 81-176 is best characterised for pathogenic behaviours and is much more motile and invasive invitro (Backert and Hofreuter, 2013; Poly and Guerry, 2008).

C. jejuni produces 18 different kinds of Cia and only a few are characterised to

date. The current hypothesis of delivery is that *C. jejuni* cell binds the host cell through adhesins such as FspA and exports CiaC and CiaI through the flagellar T3SS either straight into the cytosol of the host cell through a yet unknown mechanism, or into the exocellular space, where the effectors are taken up by the host cell through another carrier. Both *C. jejuni* cell binding to the host and secretion of CiaC and CiaI through a functioning flagellar T3SS are required for maximum uptake by the host cell (Neal-McKinney and Konkel, 2012). Another hypothesis is that *C. jejuni* filament length is controlled by FlaG, which is also secreted, and in many instances when the leading filament is much shorter than the lagging one, it could be used as a Cia injection port (Inoue et al., 2018). However, this poses of question as to how the effectors get secreted past the FliD capping protein at the flagellar tip. One possible mechanism would be that FliD is not tightly situated at the tip and the conformation allows for controlled secretion of effectors through the gaps between the filament and the cap. It has been observed that secretion of Cia does not depend on the presence of FliD or FlaA/FlaB flagellin (Neal-McKinney and Konkel, 2012). A modified flagellar T3SS apparatus lacking FliD was also capable of secreting non-flagellar proteins in *E. coli* (Majander et al., 2005). This, with the combination of the short flagella mechanism mentioned above, could indicate that FliD does not need to be present on top of the secreting flagellum and simply the short filament could suffice as an export machinery. This aspect of flagellar secretion still requires further investigation.

While the flagellin export apparatus secretes multiple infection factors and toxins, the flagellin protein and FliD capping protein are antigens themselves. In *H. pylori* FliD reacts with 97% of infected patient sera and induces strong Immunoglobulin G (IgG) response but not with uninfected, suggesting it as an immune target (Ghasemi et al., 2018). Secretory immunoglobulin A (SlgA) is the main antibody in mucous secretions and a first line of defence against adhesion of enteric pathogens. It was observed to enhance clearance of *C. jejuni* cells at early post-infection stages, making *C. jejuni* FliD ($FliD_{cj}$) an attractive drug target (Perruzza et al., 2020).

1.4. Overall Assembly and Rotation of the Bacterial Flagellum

1.4.1 Gene Expression Regulation

The flagellum is a complex molecular machine consisting of multitude of proteins, which assemble into three main parts: transmembrane basal body, hook and flagellar filament as illustrated in Figure 1.4.1. The order of filament assembly is a highly sequential process requiring complete assembly of each section before the

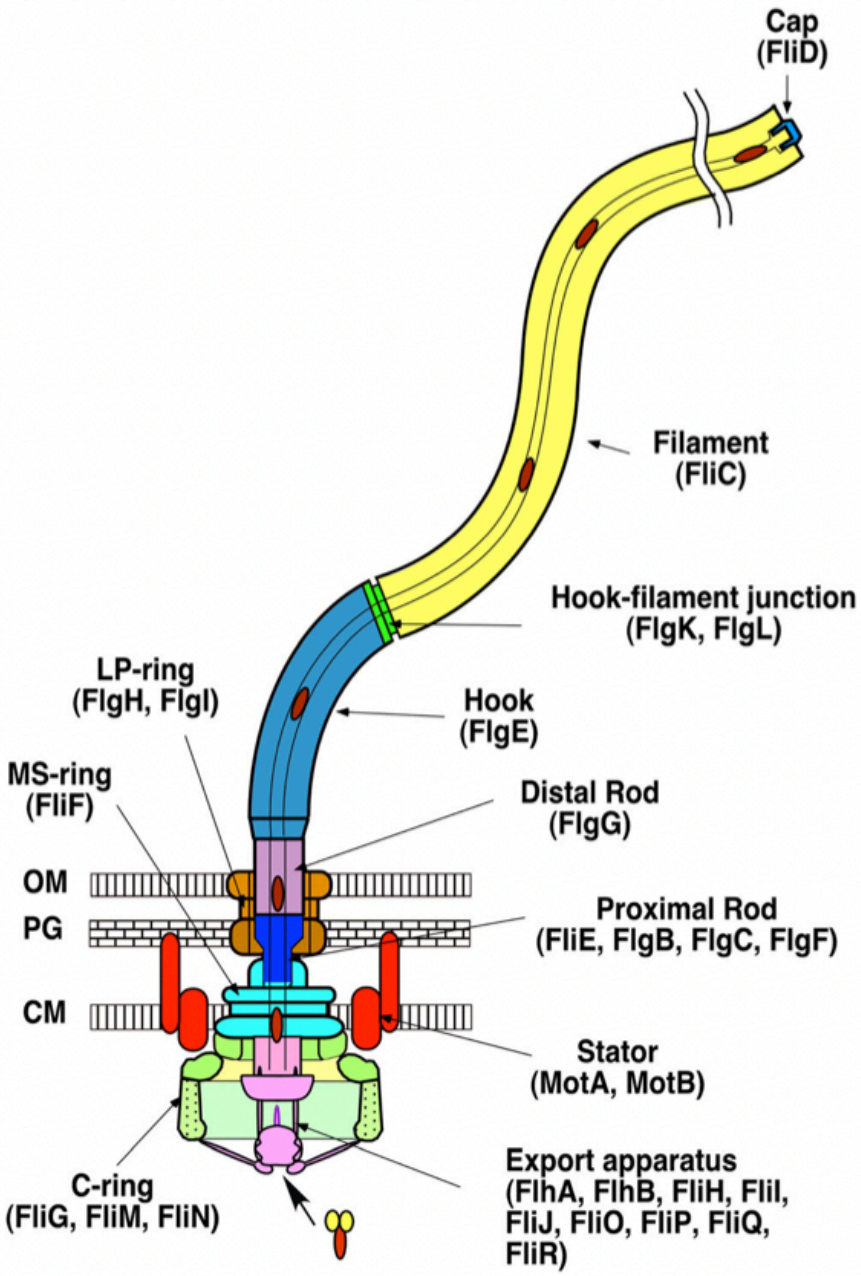


Figure 1.4.1: Diagram of proteins involved in bacterial flagellum assembly. Three main parts of the flagellum are illustrated as follows: **Yellow:** Filament and capping protein, **Blue**: Hook and hook-filament junction, **Multicolour:** Basal body. From (Imada, 2018).

genes coding for the next one are expressed. Over 50 different genes are involved in the assembly of the flagellum, with about half coding for the structural components and half responsible for regulation of the assembly process.

In enteric pathogens, which include *E. coli* and *S. typhimurium*, transcription of these genes is divided into three classes. Class 1 genes *flhD* and *flhC* are transcribed by a single promoter in response to chemotaxis signalling, resulting in the expressed FlhDC interacting with RNA Polymerase (RNAP) and primary sigma factor (σ^{70}) to activate the transcription of Class 2 genes coding for hook-basal body (HBB) components. Class 2 genes first translate the T3SS components (*flhB, fliP, fliR, fliQ and flhA*) followed by the membrane and supramembrane (MS)-ring (*fliF*) and cytoplasmic I-ring (*fliG, fliM, fliN*), which co-assemble in the inner membrane (IM). Genes coding for the ATPase complex are expressed next (*fliH, fliI, fliJ*), followed by axial components of the proximal rod (*fliE, flgB, flgC, flgF*), distal rod (*flgG*) and hook (*flgE*). The hook and filament capping proteins (*flgD* and *fliD*) and the chaperones required for the assembly of the hook-basal body complex are also expressed from Class 2 genes. Some of the Class 2 genes encoding the flagellum are not universal, such as *flgH* and *flgI*, which form the lipopolysaccharide and peptidoglycan (LP)-rings and are not present in Gram-positive bacteria and firmicutes due to the absence of the outer cell membrane (Beeby et al., 2020). In Spirochaetes the *flgH* gene is absent, as the filaments never cross the outer membrane and remain in the periplasmic space, while *flgI* is present in *B. burgdorferi* and absent in *T. primitia*, indicating an inconsistent presence within the phylum (Chen et al., 2011). Class 3 gene expression and Class 2 gene inhibition occurs upon the detection of the hook reaching a determinate length by the molecular ruler FliK, which is intermittently secreted through the formation of the HBB complex. Upon the rod-hook component exceeding FliK length, determined by the contact of the N-terminus of FliK with FlgD at the tip of the hook and C-terminus of FliK with the FlhB subunit of the T3SS, T3SS export specificity switches to late-stage substrates such as FliD and FlgM through a cleavage event in FlhB (Erhardt et al., 2011). Export of FliD frees up the chaperone FliT which binds FlhC to inhibit Class 2 transcription. FliT also binds anti sigma protein FlgM to release the bound FliA (σ^{28} factor/ *fliA*) thus triggering Class 3 gene transcription. FlgM is then exported out of the cytosol. Class 3 genes

include the hook-filament junction (*flgK, flgL*), flagellin (*fliC*), and stators (*motA, motB*). In addition to these structural genes, Class 3 also expresses a variety of chemoreceptors sensing chemotaxic environments (Beeby et al., 2020). A schematic summarising the regulation of flagellar genes in enterics is illustrated in Figure 1.4.2.

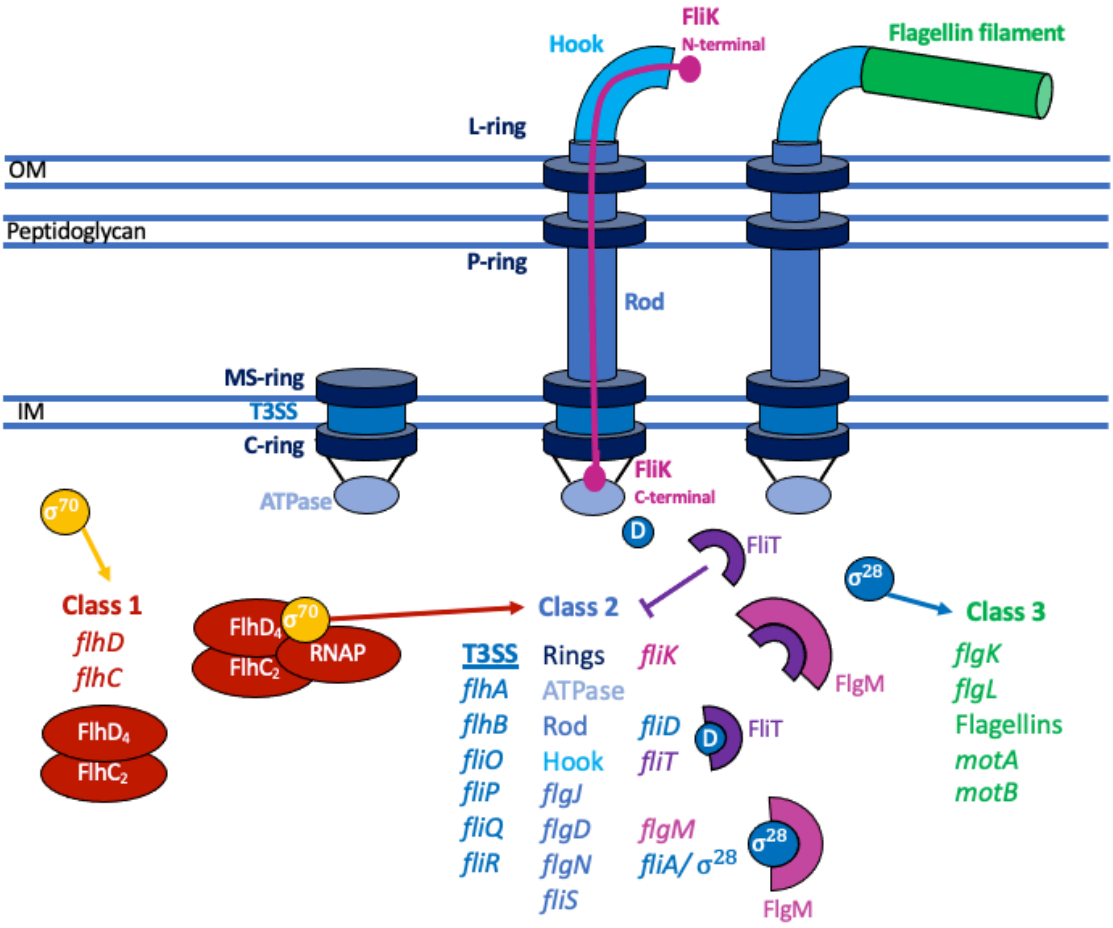


Figure 1.4.2: Diagram of flagellar regulatory cascade for *E. coli* and *S. enterica*. Class 1 genes are in red, class 2 genes are in different shades of blue and purple, class 3 genes are in green. Master regulator FlhDC activates the transcription of T3SS genes, which export other class 2 genes forming the hook-basal body complex. FliK acts as a molecular ruler measuring the distance between the FlgD hook capping protein and FlhB of the T3SS. Upon reaching the desired length, FlhB undergoes a cleavage changing substrate specificity of the T3SS to late substrates. FliD export frees the chaperone FliT which binds FlgM antisigma factor that was inhibiting FliA from activating the Class 3 genes. FlgM factor is then exported out of the cell.

The 3 class system is not universal across all the bacterial species: *P. aeruginosa* has a 4 tiered expression system and *C. crescentus* has its own unique control system (Jarrell and McBride, 2008). While *C. jejuni* also has a 3-tier cascade system, the genes

coding for the flagellar components are spread across 32 separate loci and not clustered as observed in enterics. The global regulator FlhDC is absent in *C. jejuni*, so class 1 genes including the T3SS, FlgSR two component system (TCS), *flhF, flhG*, σ^{28} and σ^{54} are regulated by σ^{70}. Formation of the T3SS signals the TCS causing FlgR to activate σ^{54} factor to initiate class 2 genes transcription, which code for the rest of the basal body, hook complex, FlgM and FlaB minor flagellin (Ren et al., 2018). In *C. jejuni*, FlgM anti-sigma factor does not appear to inhibit σ^{28} activity during HBB formation and their interaction is regulated by temperature. FlgM interacts with FliA and is exported out of the cell only at 42 °C, indicating that at lower temperature there is another regulator for FliA (Wösten et al., 2010). *flaG* gene located in the same operon *as fliD* and *fliS* transcribes a protein secreted into the media which appears to bind and inhibit FliA, thus preventing the transcription of class 3 genes. One of the main class 3 genes is the *flaA* major flagellin, which contributes to the majority of the long filament protruding out of the cell (Inoue et al., 2018). In bacteria with polar flagella a pair of class 1 genes *flhF* and *flhG* express proteins localization of which controls the polarity of the flagella to different ends of the cell (Beeby, 2015). FlhF has also been shown to directly regulate *flgI* gene expression as well as binding to the promoters of *fliA, flgS* and σ^{70} (Li et al., 2020). FlhG ATPase is a mediator of numerical control of polar flagella through multiple mechanisms. In *P. aeruginosa* FlhG/FleN regulates flagellar number by controlling the expression of FlhCD master regulator. In *Bacillus subtilis*, it limits the number of flagella at each cell pole through inactivation of the FlhF GTPase. In *C. jejuni*, which lacks the master regulator genes, FlhG influences FlhF GTPase activity by an alternative mechanism, which involves two regions of the protein that are significantly altered in orthologs instead of the ATPase domain (Gulbronson et al., 2016). ε-proteobacteria motors assemble additional components which allow for increased rotation speeds and torque. Regulation of these components is unknown, but they are a part of clear hierarchical assembly, as observed in *C. jejuni*. *flgP* has been proposed to form the basal disk of 42 nm radius near the outer membrane (OM), *pflA* has a signal for periplasm localisation and proposed to be the main component of the medial disk required for *pflB* forming the proximal disk near the IM (Beeby et al., 2016). A schematic summarising the known regulation steps of flagellar genes in *C. jejuni* is illustrated in Figure 1.4.3.

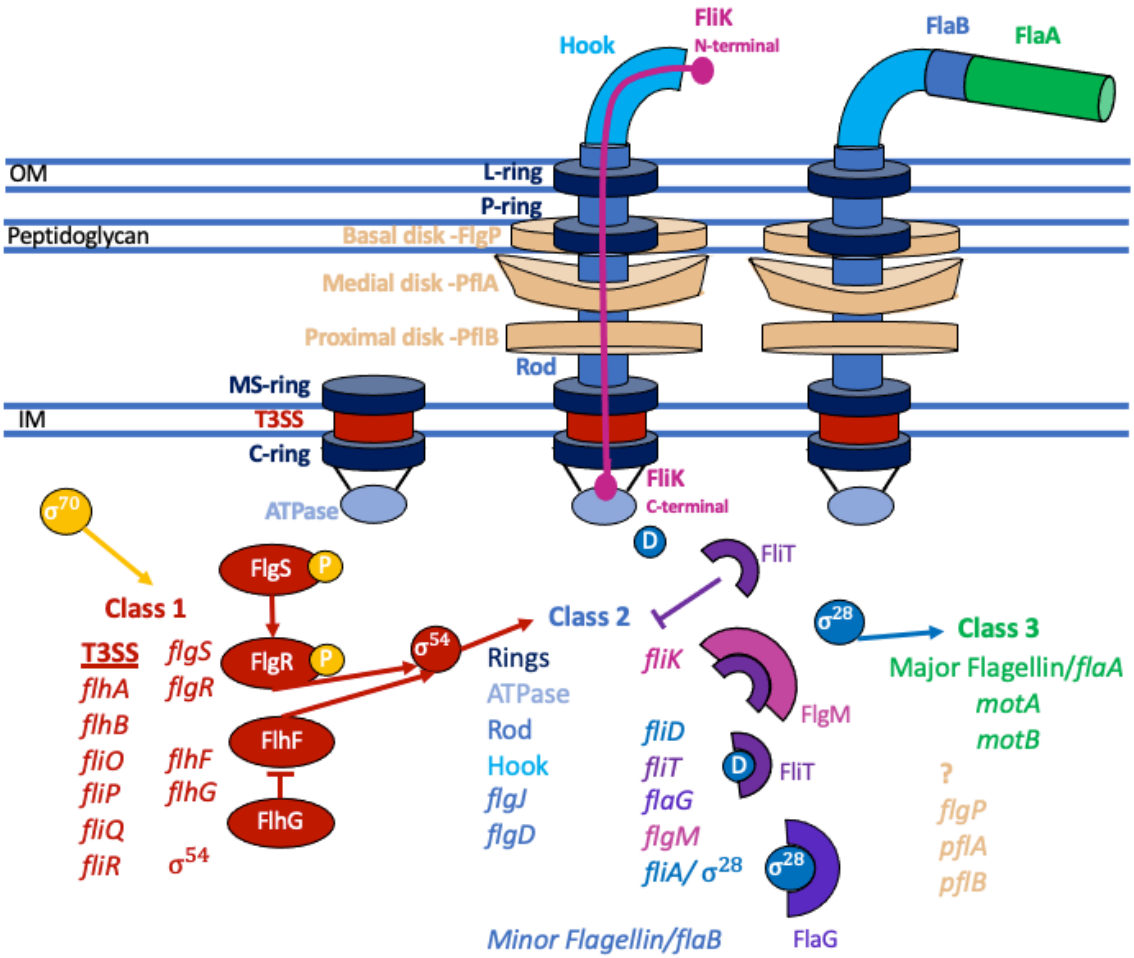


Figure 1.4.3: Diagram of flagellar regulatory cascade for *C. jejuni*. Class 1 genes are in red, class 2 genes are in different shades of blue and purple, class 3 genes are in green. Master regulator FlhDC is absent in *C. jejuni* so class 1 genes include T3SS, TCS and localization regulators FlhF and FlhG. The TCS component FlgS kinase senses the signal from the T3SS and phosphorylates FlgR, activating it and allowing interaction with σ^{54}. FlhF GTPase activates σ^{54} using a different mechanism, potentially through binding directly to promoter gene, and is in turn regulated by FlhG ATPase, which controls the number of polar flagella. σ^{54} in turn activates class 2 genes coding for the rest of the basal body, hook and minor flagellin FlaB. FliK acts as a molecular ruler measuring the distance between the FlgD hook capping protein and FlhB of the T3SS. Upon reaching the desired length, FlhB undergoes a cleavage changing substrate specificity of the T3SS to late substrates. FliD export frees the chaperone FliT which binds FlgM antisigma factor. FlaG, a class 2 late substrate that was inhibiting FliA from activating the Class 3 genes is exported and used for flagellar length control. Additional disks found in *C. jejuni* basal body (beige) are proposed to be formed by *flgP, pflA* and *pflB* and their order in the cascade has yet to be determined.

1.4.2 BASAL BODY ASSEMBLY

The summarised process of a stepwise assembly of Gram-negative *E. coli* flagellum is illustrated in Figure 1.4.4. The first complex to assemble is the T3SS export apparatus consisting of core proteins (FliQ, FliR, FliP), export gate forming proteins (FlhA, FlhB) and ATPase complex components (FliI, FliJ, FliH) which functions to export the rest of the flagellar proteins through the lumen of the rings. The core complex consists of five copies of FliP, which are proposed to oligomerise with the help of membrane chaperone FliO within the membrane, followed by addition of a single FliR. Once this preliminary complex is formed, FliO is through to dissociate from the complex. The $FliP_5R_1$ complex acts as a centre for assembly of the remaining subunits of the export gate $FliQ_4$, FlhB and FlhA nonamer. FlhB substrate specificity switch associates to the $FliP_5R_1Q_4$ by wrapping around using the N-terminal transmembrane I helices and extending the C-terminal domain into the cytoplasm (Halte and Erhardt, 2021). FlhA has a C-terminal region interacting with FlhB and later forming ATPase complex and N-terminal TM regions interacting with the MS-ring, which FlhA assists in assembling (Li and Sourjik, 2011). The C-terminal region forms two cytoplasmic rings with $FlhA_{CD1}$ domain ring close to the membrane interacting with FlhB and a larger $FlhA_{CD2}$ domain ring more distal to the export gate interacting with the ATPase (Halte and Erhardt, 2021).

While flagellar assembly is a highly sequential process, it is also cooperative. FliA assembly and C-terminal protein FliG, which at this step is in its cytoplasmic form, both assist the MS-ring oligomerization (Li and Sourjik, 2011). The MS-ring consists of multiple copies of FliF and is a complex interface connecting many different symmetries between flagellar parts, while in itself having a varied symmetry across bacterial species (Armitage and Berry, 2020; Johnson et al., 2021). FliF was observed to act as an adapter with a 22-fold symmetry in the upper part and 34-fold symmetry in the lower part matching the C-ring symmetry (Johnson et al., 2020). Multiple copies of FliG, one of the three proteins constituting the C-ring, bind the MS-ring followed by the other two proteins FliM and FliN in a 1:1:3 ratio forming the rotor and switch complex of the motor (Eisenbach, 2011). While the sequence is conserved across bacterial species for the C-ring, the diameter has been observed to range from 34 to

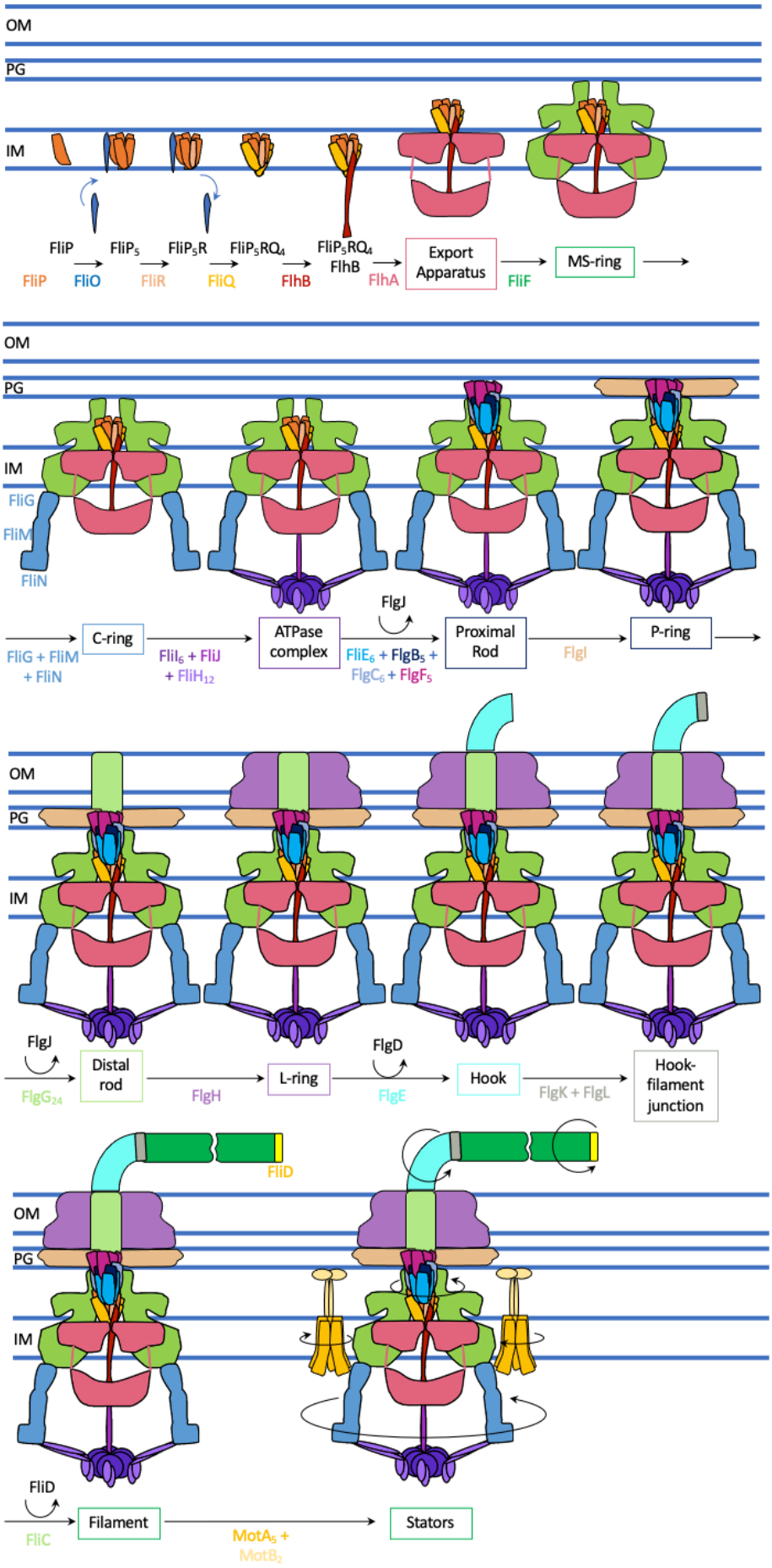

OM
PG
IM
FliP
FliP5
FliP5R
FliP5RQ4
FliP5RQ4 FlhB
Export Apparatus
MS-ring
FliP
FliO
FliR
FliQ
FlhB
FlhA
FliF
OM
PG
IM
FliG
FliM
FliN
C-ring
FliG + FliM + FliN
FliI6 + FliJ + FliH12
ATPase complex
FlgJ
FliE6 + FlgB5 + FlgC6 + FlgF5
Proximal Rod
FlgI
P-ring
OM
PG
IM
FlgJ
FlgG24
Distal rod
FlgH
L-ring
FlgD
FlgE
Hook
FlgK + FlgL
Hook-filament junction
FliD
OM
PG
IM
FliD
FliC
Filament
MotA5 + MotB2
Stators

Figure 1.4.4: Schematic of the stepwise assembly of a Gram-negative enteric flagellum. **OM:** outer membrane, **PL:** peptidoglycan layer, **IM:** inner membrane. Structural proteins are labelled with colours corresponding to their respective schematic. Adaptor proteins binding and assisting the formation of the axial structures are shown with curved arrows, as they dissociate, apart from FliD.

57 nm, indicating additional density which could be an effect of increased torque. In γ-proteobacteria (including enterics) and ß-proteobacterium *Hylemonella gracilis* the diameters of 40 nm are conserved. ε-proteobacteria C-rings are larger with the diameter of 49 nm and Spirochaetes along with firmicutes have the widest observed C-rings (Chen et al., 2011). FliG has three domains which interact with different parts of the motor machinery. $FliG_N$ domain binds MS-ring and anchors it to the cytoplasmic membrane, $FliG_M$ domain binds $FliG_c$ and FliM for switching rotation direction and $FliG_C$ domain interacts with the MotA and MotB stators for torque generation (Armitage and Berry, 2020).

The C-ring is a platform for FliH anchoring the ATPase complex made up of $FliH_{12}I_6J_1$. FliH has three regions: N-terminal domain binding FliN and FliM of the C-ring, central connector domain and C-terminal domain binding FliI hexamer. FliH has also been observed to dimerise in solution with a monomer of FliI. This cytoplasmic carrier binds substrate-chaperone complexes and assists in export of the substrates (Halte and Erhardt, 2021). The flagellar ATPase is a Walker type ATPase structurally similar to the F0F1-type ATP synthase. It takes proteins from chaperones for unfolding and uses both ATP and proton motive force (PMF) to export them through the C-ring for further flagellar assembly. However, it is not essential for flagellar assembly and its main function may lie in expediting the T3SS secretion of proteins (Armitage and Berry, 2020).

Assembled T3SS recruits, unfolds and exports axial components powered by a PMF at 1700 amino acid per second, two orders of magnitude faster than general secretion pathway (SEC) (Beeby et al., 2020). The first protein secreted through the T3SS is a part of the proximal rod called FliE. followed by FlgB, FlgC and FlgF. These proteins bind FlhB without chaperones before export and initial rod assembly is templated by the core T3SS complex of $FliP_5R_1Q_4$. The proximal rod is a collection of six

copies of FlgC and FliE and five copies of FlgF and FlgB each. Rod capping protein FlgJ binds first and helps building the rest of the proteins and has a peptidoglycan hydrolysing ability to pierce through the peptidoglycan and outer membrane (Imada, 2018).

In Gram-negative bacteria such as *E. coli*, the next structure to form is the P-ring made up of FlgI proteins secreted via SEC pathway and polymerised around the proximal rod within the peptidoglycan layer. It recruits 24 copies of FlgG to assemble the distal rod with the assistance of the rod chaperone FlgJ (Beeby et al., 2020). The rod follows a non-integer helical symmetry similar to that of hook and filament and built with C-terminal helices facing into the lumen and N-terminal helices decorating the outside (Johnson et al., 2021). The distal rod in turn acts as an assembly point for L-ring made up of FlgH, also exported through the SEC pathway into the periplasm. The full rod is on average about 22 nm long, spanning the length of the periplasmic space between the OM and IM (Chen et al., 2011). This sequential ring and rod assembly is illustrated in Figure 1.4.5.

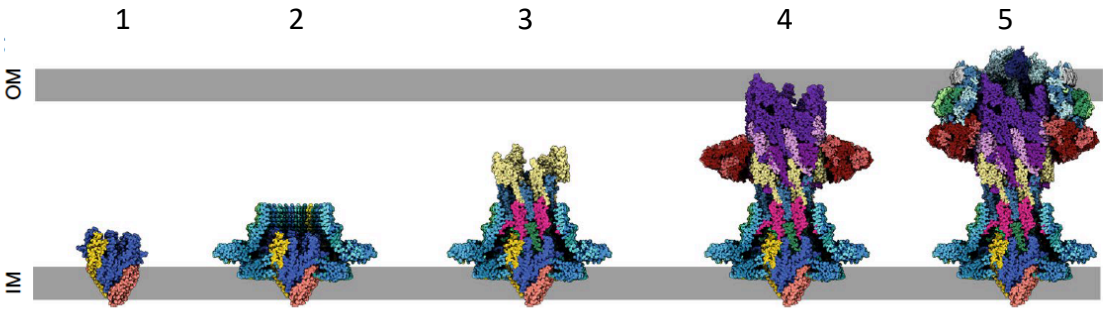


Figure 1.4.5: Step-wise assembly of the flagellar basal body using Cryo-EM maps from (Johnson et al., 2021). (1) Export gate proteins FliQ, FliR, FliP assists the formation of the MS-ring (2) and seeds the protofilaments of the rod. C-ring formation and the rest of the export apparatus follow. (3) Rod is grown with FliE, FlgB, FlgC, FlgF and this enables the formation of the P-ring (4) around the end of the proximal rod which then recruits FlgG to build the distal rod and act as an assembly point for the L-ring (5). From (Johnson et al., 2021).

The LP-ring separates the rotating distal rod from the outer membrane and peptidoglycan and has a 26-fold symmetry in *S. enterica*. This separation is due to the ring and the rod having opposing charges to hold them together, but the smooth

surface of the rod enables free rotation and some degree of sliding. The mismatched P-ring and rod symmetry also minimises friction (Johnson et al., 2021). The rings are not ubiquitous, meaning that while they are both present in their standard state in enterics such as *E. coli, H. gracilis* L-ring may not be embedded in the membrane and exists separately along with P-ring. In spirochaetes the L-ring is absent since the filaments never cross the outer membrane and remain in the periplasm. Even within the spirochaetes the presence of the P-ring is varied (Chen et al., 2011). While the distal rod is in this flexible interaction with the LP-ring, the proximal rod has many contacts with the export apparatus and MS-ring for stable attachment. Top most helix of FlgH interacts with a lipid portion of the lipopolysaccharide layer (LPS) (Johnson et al., 2021).

Unlike the enteric γ-proteobacteria, some species have evolved to contain additional components within their basal body to enable the production of higher torque or for species-specific functions. Sodium driven motors in non-enteric γ-proteobacteria have H-rings composed of FlgO and FlT and T-rings composed of MotX and MotY assembled onto LP-rings. Spirochaetes have a large P-collar made up of FlbB forming a bowl around the rod and stator ring surrounding a wider C-ring. ε-proteobacteria have evolved to output even higher torque through three additional disks within their basal body, as illustrated in Figure 1.4.6 (Chen et al., 2011).

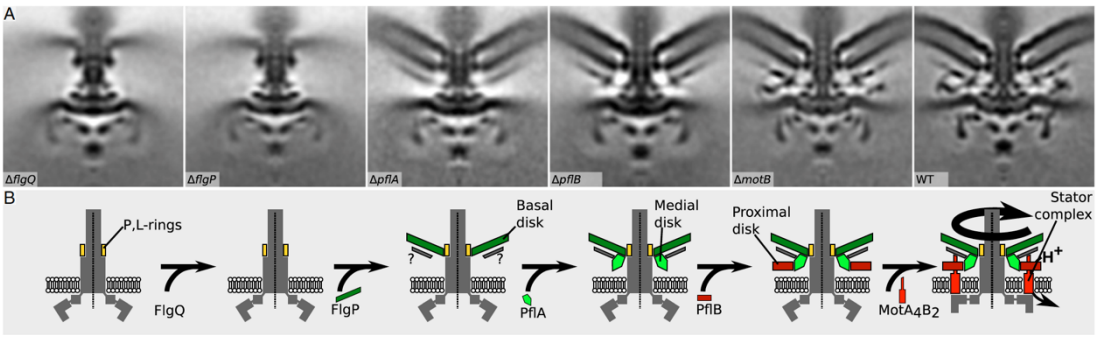


Figure 1.4.6: Stepwise assembly of the *C. jejuni* additional basal body disks. **A.** Sub tomogram averages arranged to illustrate the assembly pathway of the disk complex. **B.** Model of the hierarchical assembly of the complex. From (Beeby et al., 2016).

In *C. jejuni* the basal disk made up of FlgP and FlgQ has a radius of 42 nm and pushes the OM away from the cell body at an increased radius, making a distortion in a shape of a cup. This is not present in *H. pylori*. FlgQ appears to be required first for FlgP association and forming of the disk. The periplasmic median disk made up of PflA

has 17-fold symmetry and requires the presence of the basal disk to assemble. The third IM associated proximal disk made up of PflB is located above the inner membrane and has a 17-fold symmetry as well. It forms underneath the medial disk and creates a wider area for stator binding (Beeby et al., 2016).

1.4.3 HOOK AND FILAMENT ASSEMBLY

As illustrated in the lower half of Figure 1.4.4, after the basal body forms, L-ring displaces FlgJ rod chaperone with FlgD at the tip of the rod. FlgD then uses monomers of FlgE threaded through the rod lumen to build the hook, length of which is controlled by FliK. FlgD is a five-fold complex with helical hairpins which blocks exit points from the rod requiring rearrangement prior to building the hook. Previous models required cap rotation, but most recent findings show that helical hairpins are tucked deep into the rod to sterically prevent this. Along with the information that one subunit sits high not following the helical pitch, the proposed mechanism is that of upward and downward piston-like movement as illustrated in Figure 1.4.7. In the first transition, the lowest copy in the uneven helical pitch of FlgD (illustrated in light blue) becomes the highest copy (red arrow), lifting that part of FlgD up from its previous position. This process repeats for each of the five subunits moving the cap upwards. If viewed from the top, a full revolution of the cap around the rod would sequentially open rod/hook subunit binding sites without rotation on the axis, with every subunit still located on the same axis but translated upwards instead (Johnson et al., 2021).

The hook is made up of multiple repeats of FlgE and can be up to 55 nm in length, which is controlled by FliK. FliK does this through a mechanism dependent on intermittent secretion during the formation of HBB. FliK N-terminal regions binds FlgD, while C-terminal region interacts with FlhB substrate specificity switch, inducing autocleavage and switch to late substrate export such as flagellins. FliK is secreted through the HBB rapidly during a growing hook and slows down as the hook reaches the physiological length. If the hook is too short, the probability that FliK C-terminus will have enough time to interact with FlhB is low, as the secreted N-terminus folds outside of the growing hook and pulls the rest of FliK to itself too quick for interaction. If the hook is of the physiological length or longer (55 nm), FliK spreads

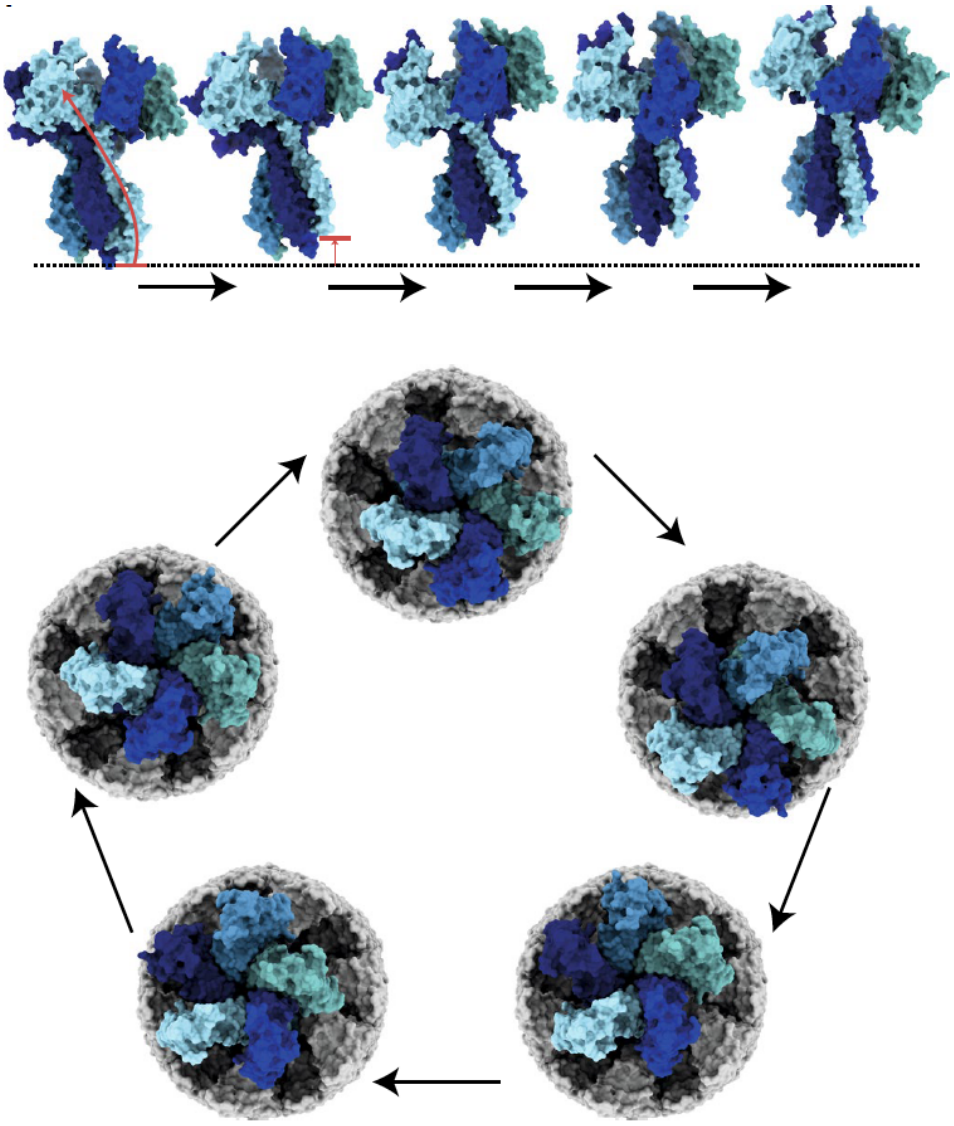


Figure 1.4.7: Proposed motions involved in FlgD-cap catalysed hook assembly mechanism**.** **Top**, a side view of four sequential cap movements with rod removed for clarity. In the first transition the movement of the lowest copy (light blue) to become the highest copy is illustrated by red arrows. **Bottom,** a top-down view of the cap on the rod (grey) showing a full revolution of the cap around the rod with sequential opening of space for hook monomers without rotation of the cap on its axis. From (Johnson et al., 2021).

across the length of the HBB lumen giving ample time for the C-terminal region to switch FlhB to late substrate export (Erhardt et al., 2011). The hook forms an 11-protofilament tube with a diameter of 280 Å in *C. jejuni* as illustrated in Figure 1.4.8 (Matsunami et al., 2016). It is also essential for quick reversal of rotation and uncoiling

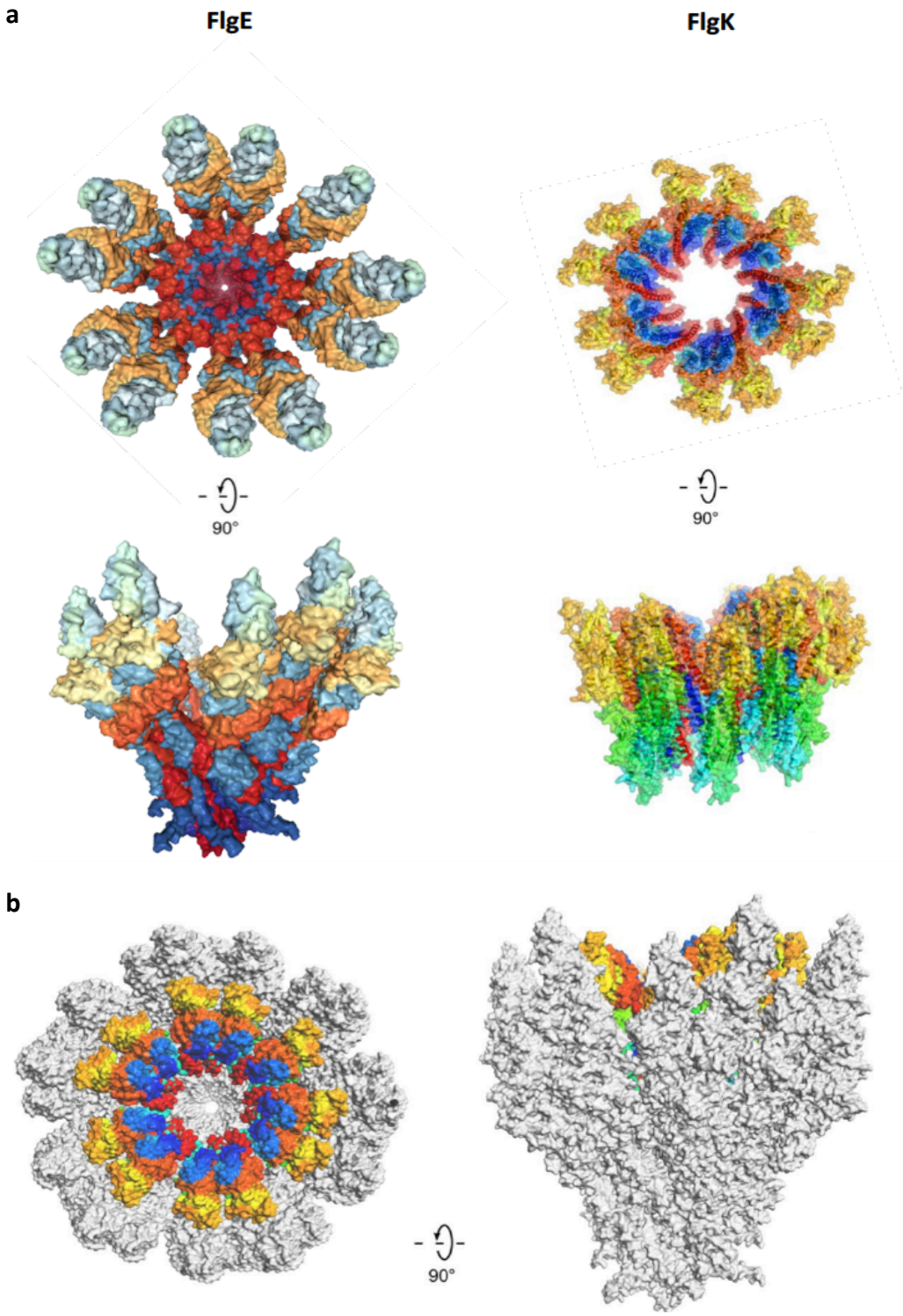


Figure 1.4.8: Structures of FlgE (PDB: 5JXL) and FlgK (PDB: 5XBJ) from *C. jejuni*. (a) Oligomer models of FlgE (left) and FlgK (right) top and side views. There are 11 protofilaments in the hook (FlgE) and FlgK has been modelled according to the same oligomerisation number. (b) Superposition of FlgE and FlgK with FlgK being the coloured structure in the middle. Adapted from (Bulieris et al., 2017; Matsunami et al., 2016).

of the supercoiled flagellar filament bundle (Fujii et al., 2018). While thought to be a rigid structure, has recently been found to be flexible and adapting multiple conformations (Shibata et al., 2019).

After the hook is complete, unfolded FlgK bound to its chaperone FlgN binds the export apparatus (namely FliI and FlhA), dissociates, threads through the lumen of the growing flagellum and binds at the tip of the hook replacing FlgD. Then FlgL bound to FlgN chaperone undergoes the same process and binds FlgK making up the hook-filament junction (Xing et al., 2018). While there is no cryo-EM structure of these complexes, crystal structures have been published of truncated FlgK in *B. pseudomallei, S. enterica* and *C. jejuni*, and truncated FlgL in *B. cereus, X. campestris* and *L. pneumophila* (Bulieris et al., 2017; Gourlay et al., 2015; Hong et al., 2018; Song et al., 2020). Thus, without a structural reference for their oligomerization, the junction proteins were modelled to an assumed 11-fold state onto the hook 11-mer as illustrated in Figure 1.4.8 for FlgK and Figure 1.4.9 for FlgL. However, this hypothesis remains to be confirmed as to how these two proteins oligomerise and interact with the hook and flagellar filaments.

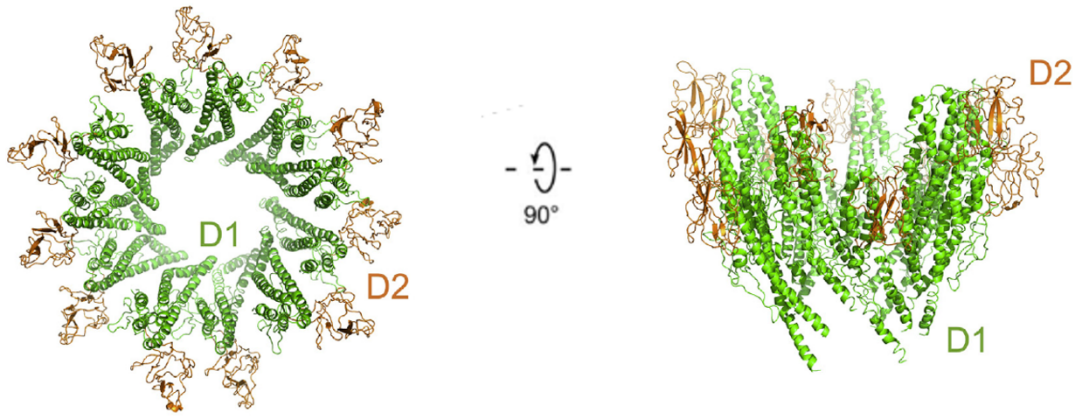


Figure 1.4.9: FlgL from *L. pneumophila* assembly model in the flagellar junction. D1 domain (green) and D2 domain (orange) are labelled. From (Song et al., 2020)

The final steps of flagellar formation involve the filament elongation. Filament capping protein FliD, unfolded and bound in solution to its chaperone FliT, binds the export apparatus, dissociates, and exports through the lumen of the growing flagellum, binding FlgL at the growing end. This is confirmed by biochemical studies and the ability

of FliD and FlgL to form complexes in solution (Imada, 2018; Yonekura et al., 2000). FliT binds FlhC and halts class 2 gene transcription. It also binds FlgM, releasing FliA to express class 3 flagellin genes. FliD oligomerization state has been a subject for debate over the last few years with crystal structures ranging from tetramers in *S. marscecens*, pentamers in *S. enterica* to hexamers in *E. coli* and *P. aeruginosa* (Cho et al., 2017, 2019a; Postel et al., 2016; Song et al., 2017). The only cryo-EM structure of FliD prior to the work reported in this thesis (See result chapter 3.2) showed it to be a pentamer (Maki-Yonekura et al., 2003). While there is evidence of FliD binding to the hook-filament junction in the shape of low-resolution tomography data from *B. burgdorferi*, there are no high resolution structures of interaction of FliD with the growing flagellar filament (Zhang et al., 2019).

After FliD oligomerises on top of FlgL, unfolded flagellin bound to its chaperone FliS attaches to the export apparatus, dissociates, and is exported through the lumen until it reaches FliD, which either rotates or changes conformation to add flagellin monomers to build a nascent filament. While many bacterial species express multiple types of flagellin in their class 3 genes, such as the *S. enterica* expressing FliC or FljB, the situation is more complex in ε-proteobacteria.

The filament then continues to grow to a specific length. Unlike the molecular ruler FliK controlling the length of the hook to a specific physiological size, there are different proposed mechanisms of flagella length control in enterics and ε-proteobacteria. In *S. enterica* the filament forms a 2 nm channel within, through which flagellins and other effectors pass through. The longer the filament becomes, the less efficient the transport of flagellins through the lumen will be, similar to the way that FliK secretion is rapid at first comparing to when the hook is close to the physiological size. This injection-diffusion mechanism explains filament growth dynamics as a function of the total length and diffusion coefficient. Flagellins are partially helical in the channel when exported through the T3SS and move diffusively in one dimension through the length, driven by quick removal at the growing end through polymerization. This length-dependent slowing of export shows why filaments do not grow indefinitely, but sheared filaments can still regrow to the wildtype length (Renault et al., 2017). The rate of flagellar export decreases with the length of filament

structure, so the FlgM that is constantly secreted during flagellar growth will decrease in concentration outside the cell and thus increasing in concentration inside the cell and inhibiting FliA and respectively the flagellin genes, thus controlling the filament length (Paradis et al., 2017).

In *C. jejuni* two possible proteins affect FlaA major flagellin expression. FlgM appeared to bind FliA and limit the class 3 gene expression only at 42 °C, indicating that at lower temperature there is another regulator for FliA and flagellin genes (Wösten et al., 2010). FlaG is secreted into the medium, alike FliK, but cannot be a physical ruler as the filament length is much longer than FlaG in its folded state. FlaG interacts with FliA, so at low temperatures it could act as the anti-sigma factor, thus controlling the expression of the FlaA. The wild type filament length in *C. jejuni* is capped at about 3.5 µm and it is observed that there is a longer and a shorter flagellum on a cell at a time. With a knockout of FlaG, the filament length increases two-fold and through different mutant combinations it has been established that FlaG inhibits the length of FlaA filaments and increases the length of FlaB. The difference is length between long and short flagella is heavily pronounced in the FlaG knockout and in FlaA only filaments, indicating that it is the FlaA component that causes the difference between he long and short flagella on the cell and not FlaB (Inoue et al., 2018). However, due to difficulty of capturing the flagellin elongation process in close snapshots and the fact that these length regulators are diffused in solution, these mechanisms have yet to be confirmed.

1.4.4 Flagellar Rotation

The C-ring acts as the rotor for the flagellum, which interacts with stators made of MotA and MotB and can rotate up to 350 rotations per second (Eisenbach, 2011). MotA to MotB ratio is 5:2, with MotB dimer in the middle surrounded by MotA monomers (Deme et al., 2020; Santiveri et al., 2020). MotB has one transmembrane segment and a peptidoglycan binding domain which anchors it to the rigid cell wall, while MotA has four transmembrane segments and a cytoplasmic domain that interacts with FliG (Chang et al., 2020; Deme et al., 2020). Stators can be inserted into the membrane without collapsing the ion motive force (IMF) and interact in variable numbers with the rotor. Stators are proton channels that use PMF for rotation, and it

is their interaction with the C-ring that induces the whole motor to rotate. This occurs when MotB periplasmic domain, which in an inert state, folds over the entry region like a cap, extends upon encountering the motor ring and interfaces with FliG (Armitage and Berry, 2020). If MotB is removed it causes uncontrolled proton flow across the membrane. Upon activation MotB tethers to cell wall and MotA rotates around it in a clockwise direction causing unidirectional proton flow (Deme et al., 2020; Santiveri et al., 2020). While the first stator binding is slow, it induces the ring rotation which in turn induces rotation and activation of additional stator units (Ito et al., 2021).

The mechanism of flagellar rotation switching involves conformational change in FliG upon FliM and FliN interacting with chemosensory CheY-P kinases. When C-ring is bound to CheY-P the direction of flagellar rotation is CW and when it is in its inert state the direction is CCW (Chang et al., 2020). In the CW rotation $FliG_C$ domain interacts with the inner part of the stator circle and in the CCW rotation it interacts with the outer part as illustrated in Figure 1.4.10.

Some bacterial genera, such as *Bacilli* and *Vibrio*, have sodium ion driven motors which rotate even faster than their enteric counterparts. While the proton driven motors have MotA and MotB stators, sodium ion driven motors have PomA and PomB. The number of stators associated with the motor determines the load. Enterics, such as *E. coli*, have motors with 11 stators that can produce 2000 pN.nm in mechanical output, ε-proteobacterium *H. pylori* motor produces 3600 pN.nm force with 17 stators and sodium motors in *Vibrio* species can produce up to 4000 pN.nm with only 13 stators bound. Stators exist in a highly dynamic system with rapid turnover, associating and dissociating depending on the need (Beeby et al., 2016; Chen et al., 2011).

The tendency to produce a higher mechanical input is linked to species evolving additional structures and larger diameter rings to allow engagement of more stators and higher torque, as illustrated in Figure 1.4.11. Enteric motor has the simplest structure without additional rings when comparing to high torque motors. In *S. enterica* it uses PMF to rotate at around 1300 pN.nm. Non-enteric γ-proteobacteria motor of *V. fischeri* produces torques of more than 2200 pN.nm using sodium driven stators with additional H-ring in dark green (FlgP) and aquamarine (FlgT), and T-ring in fuchsia (MotXY) interacting with the stators directly. *ε*-proteobacteria motor of

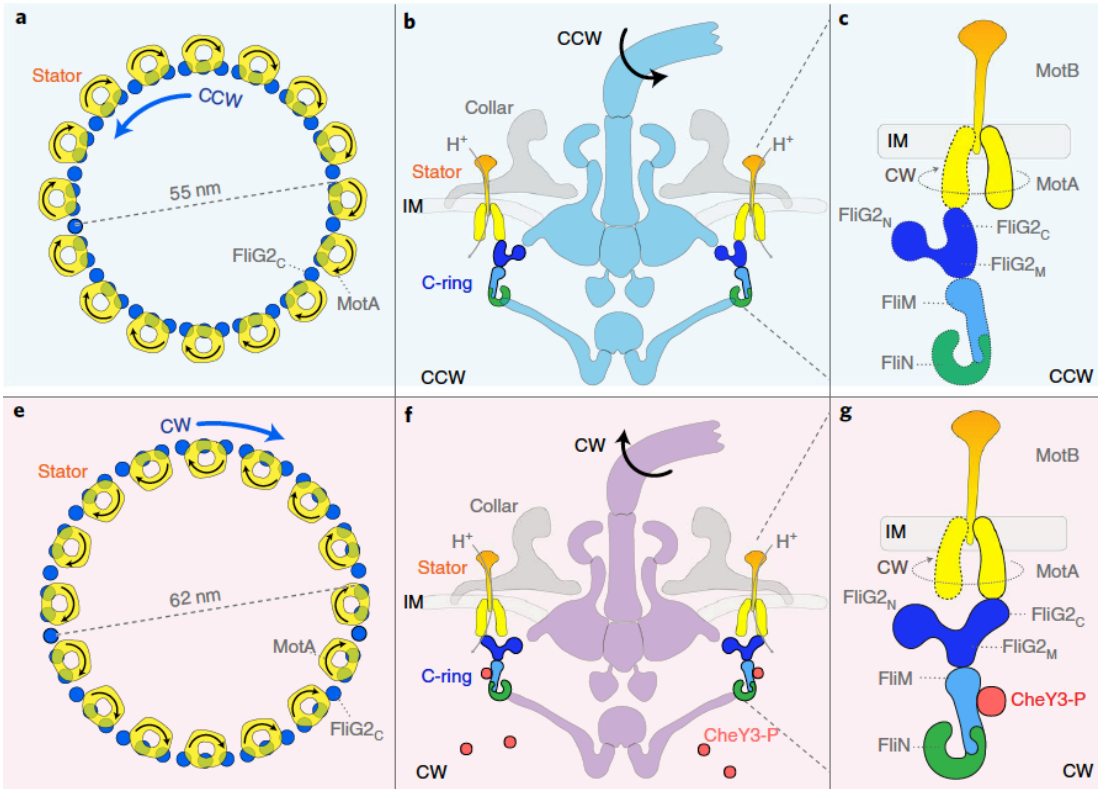


Figure 1.4.10: Model for flagellar rotational switching in spirochaetes. **Top line:** CCW rotation **Bottom line:** CW rotation. (a)(e) Top view of the C-ring FliG (blue) interacting with stator subunit MotA (yellow). (b)(f) Side view of the whole basal body with C-ring (green/blue) interacting with stators (yellow/orange) with and without CheY-P. (c)(g) Side view of stator with MotA rotating CW regardless of flagellar rotation direction. Adapted from (Chang et al., 2020).

C. jejuni produces torques of around 3600 pN.nm using PMF driven stators with additional basal disk in dark green (FlgP), medial disk in light green (PflA) and proximal disk in dark red (PflB), with the medial and proximal disks interacting with extra density (FlgQ) and stators directly (Beeby et al., 2016). The widest C-ring diameter and highest torque belongs to a member of spirochaete family *B. burgdorferi* with 46-fold symmetry and 16 stators forming a 62 nm ring (Chang et al., 2020). The P-collar of spirochaetes (pink) shares no homologue with the other species listed above, but does increase the surface area with which PMF based stators can interact to reach speeds of up to 4000 pN.nm (Beeby et al., 2016).

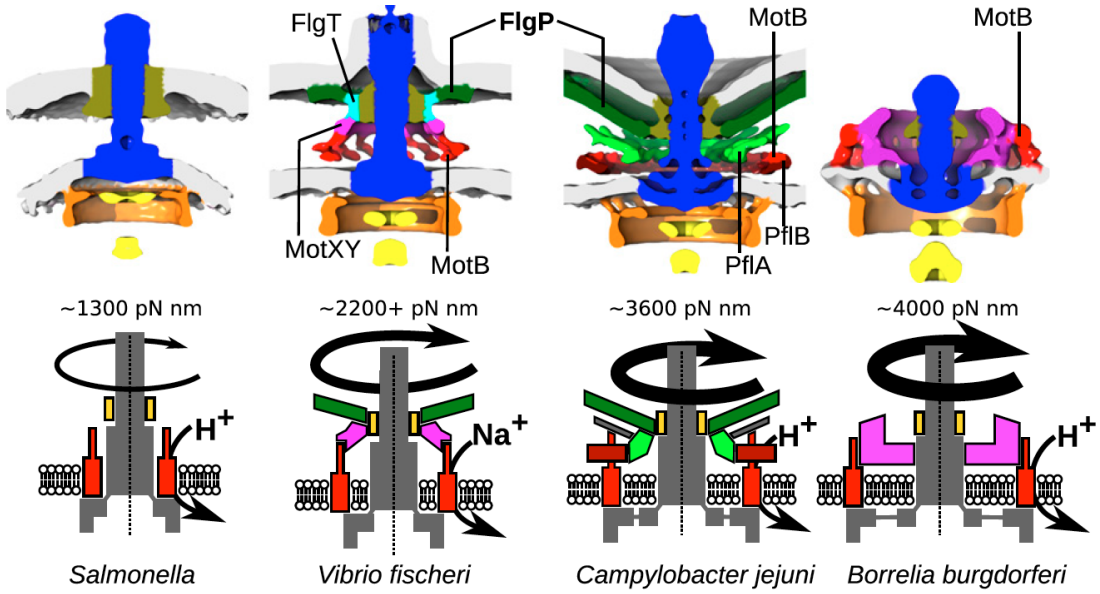


Figure 1.4.11: Examples of wider stators accounting for torque diversity. **Top** Sub tomogram average cross sections of bacterial basal bodies with increasing level of torque and number of stators from left to right. Rod in blue, LP-rings in dark green, C-ring in orange, export apparatus in yellow and the additional rings labelled in other colours respective of their location and species. **Bottom** Schematic representation of the motor complexes with rod in grey, stators in red, basal disk in dark green, median disk in light green, proximal disk in dark red, spirochaetes collar in fuchsia. From Beeby et al., 2016.

1.5 Flagellar and Injectisome T3SS

1.5.1 Secretion Systems

Bacterial secretion systems are large membrane-spanning complexes allowing protein translocation from the cytoplasm across the bacterial membranes out of the cell or into the periplasm. Gram-negative bacteria have more than nine specialized translocation systems and Gram-positive bacteria appropriated several of those systems and evolved their own. There are many threads tying systems to each other either through features inherent among all systems or their mechanistic properties. In a review by Christie, the 11 known secretion systems were classified into 4 groups, as summarised in Figure 1.5.1. Class 1-3 systems assemble in Gram-negative bacteria, while Class 4 are observed across Gram-positives. Class 1 and 2 systems are broad categories of two-step translocation systems, meaning that the substrates are

delivered across the cell membrane in two steps: across the IM via SEC or twin arginine translocation (TAT) systems and across OM via the class secretion system. Class 3 and 4 systems contain examples of one-step translocation from cytoplasm to cell surface. While both two-step translocator systems mediate OM transfer, class 1 systems assemble only in the OM, while class 2 systems assemble across the entire envelope. Class 3 and 4 systems both assemble across the entire envelope with the difference that class 3 span two membranes and peptidoglycan in Gram-negative bacteria, while class 4 pass through the cell membrane in Gram-positive bacteria with specialized functions (Christie, 2019).

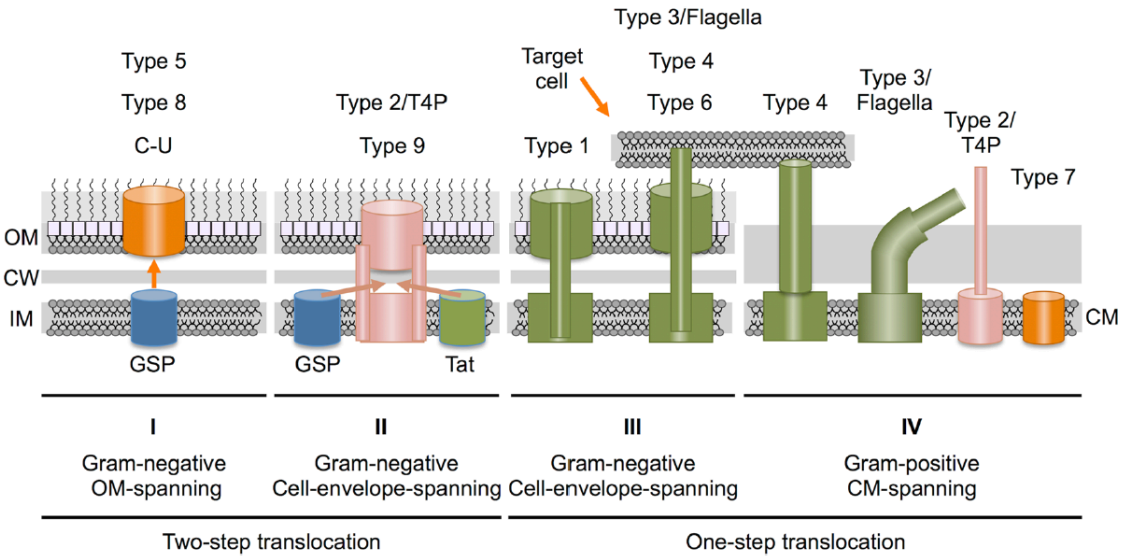


Figure 1.5.1: Overview of the different classes of bacterial secretion systems. **OM:** outer membrane, **CM:** cytoplasmic membrane, **IM:** inner membrane. From (Christie, 2019).

Class 1 systems contain type 5 secretion system (T5SS), type 8 secretion system (T8SS) and chaperone – usher pilus pathway (C-U) best characterised in proteobacteria. Members of this class rely on sources other than ATP hydrolysis or PMF to drive them (Christie, 2019). GSP, or SEC, secretes unfolded proteins powered by ATPase SecA through translocase made of SecY, SecE and SecG from cytoplasm into the periplasm (Halte and Erhardt, 2021). T5SS substrates then bind periplasmic chaperones and bind to Bam complex initiating the insertion of anchor domain into the OM, threading the substrate through the passenger domain. Pilus subunits of the C-U system also bind periplasmic chaperones, which deliver them to the OM usher

domain, extruded through the transport system and fit into the ordered assembly of the pilus. T8SS mediates assembly of amyloid fibres named curli found in proteobacteria and Bacteroidetes responsible for aggregation and biofilm formation. Protein export in this class is driven by charge based interactions promoting insertion and folding of the substrate on the other side of the OM membrane (Christie, 2019).

Class 2 systems contain type 2 secretion system (T2SS), type 9 secretion system (T9SS) and Type 4 pilus (T4P), which shares ancestry with T2SS through their Type IV filament superfamily (Denise et al., 2019). These systems assemble across the entire envelope and transport periplasmic substrates, delivered there by SEC or TAT systems, through the OM (Christie, 2019). The T2SS is prevalent in proteobacteria for the purpose of releasing proteins into extracellular space. It contains cytoplasmic ATPase (GspE) bound to the IM platform (GspC) with pseudopilus made up of GspG extending and retracting through a secretin pore made of GspD secretin and GspS lipoprotein. Substrates are delivered to the periplasm through TAT, which secretes folded proteins and is driven by the PMF. Upon delivery, substrates are tethered to the periplasmic face of the IM. They fold, bind pseudopilus and then are extruded through a secretin pore via piston-like movement (Naskar et al., 2021). Type 4 pili consist of 2 groups: type 4a pili (T4aP) and type 4b pili (T4bP) based on minor differences in assembly. T4aP is responsible for twitching motility and DNA uptake, with genes spread through the genome. T4bP are responsible for biofilm formation, colonization and adhesion, with genes clustered in single operon (Jacobsen et al., 2020). T9SS in Bacteroidetes has a secretory and gliding motility functions. Motility is paired to secretion, involving rapid movement of cell surface adhesins. The two IM subunits PorL and PorM couple IM PMF energy to secrete substrates from the periplasm (Christie, 2019).

Class 3 systems can transport substrates from cytoplasm to the outside of the cell in one-step translocation mechanism and a subset of these can also inject the effectors into the host cells. Type 1 secretion system (T1SS) is related to ABC transporter superfamily and is composed of an ABC transporter ATPase, membrane fusion protein interacting with the ATPase and TolC-like protein, which is recruited upon binding of substrate to the complex. TolC inserts into the OM and extends a helical tunnel through the periplasm connecting the exocellular space to periplasmic

space. Type 4 secretion system (T4SS) is unique among translocation systems in that a large subfamily functions to deliver DNA between cells for conjugation. It is present in nearly all bacterial systems and some archaea. It contains two large subassemblies named outer membrane complex (OMC) and inner membrane complex (IMC) connected by thin flexible stalk. The OMC assembles into a barrel shaped structure in the OM, while the IMC has a platform and two or three ATPases attached. Type 6 secretion systems (T6SS) inject effectors into host cells through a spring-like mechanism, reminiscent of a reversed contractile bacteriophage. Membrane complex spans the entire cell envelope and, while inner needle Hcp and sheath complex made of TssBC assemble in the cell envelope and dock to the baseplate. This results in the expulsion of the tube through the baseplate and through the membrane of the host cell. Class 4 systems assemble across the cellular membrane of Gram-positive bacteria and include SEC and TAT pathways, flagellar T3SS as well as type 7 secretion system (T7SS). T7SS is unique to Gram-positive bacteria and secretes small proteins of about 100 amino acids, homodimers and heterodimers, requiring a wide channel with a tight gate system (Christie, 2019).

The T3SS system family contains a non-flagellar T3SS injectisome and flagellar T3SS, both attributed to either class 4 for Gram-positive bacteria or class 3 for Gram-negative. As illustrated in Figure 1.5.2 and Table 1.5.1, the flagellum and T3SS injectisome have a common core: the IM export apparatus, MS-ring with limited homology and C-ring with ATPase, which have close homology but strongly differ in function. Contrary to other secretion systems, the flagellum basal body has unique LP-ring and stators which pertain to the motility and rotation function. While the flagellum has evolved a complicated axial machinery for rotation and motility, the secretion systems often have an extendable needle/pilus which often can be retracted through disassembly (Milne-Davies et al., 2021). These non-flagellar appendages are dynamic and assist with other types of motility. T4P pilus generates mechanical force to pull bacteria for twitching motility, while T9SS secretes surface adhesins contributing to gliding motility (Christie, 2019).

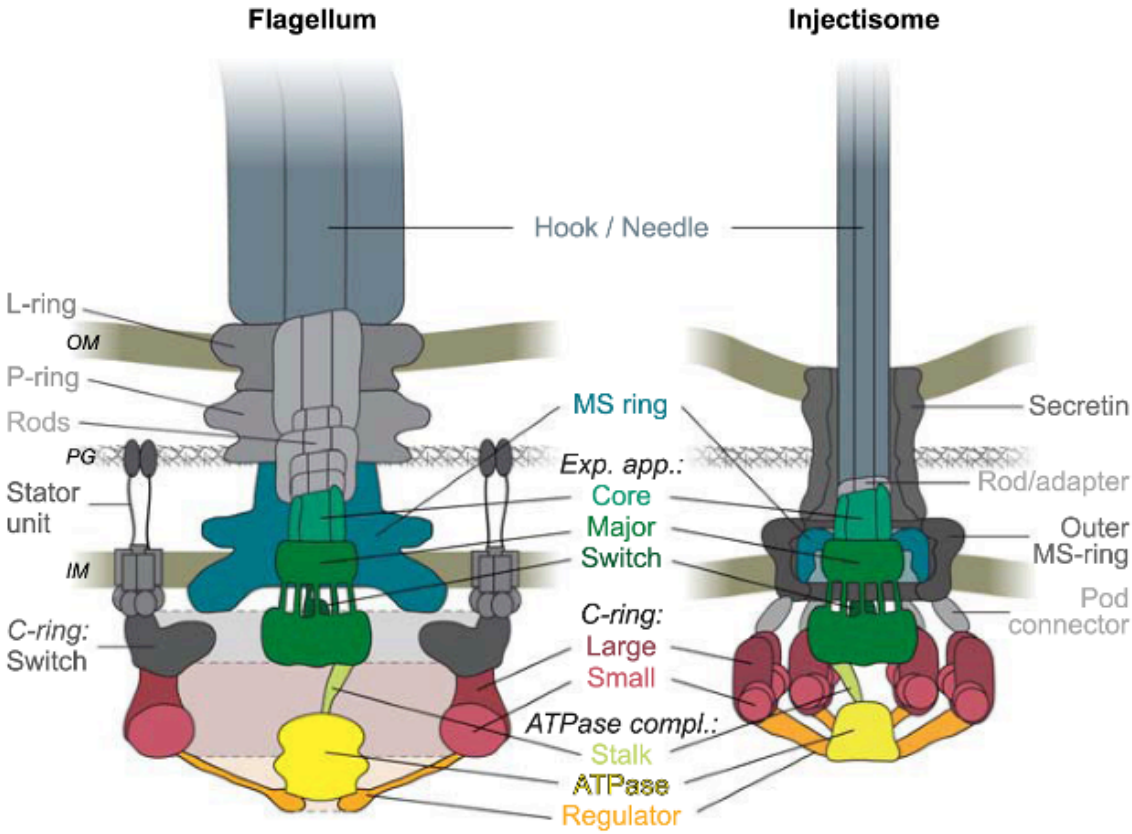


Figure 1.5.2: Overview of similarities between flagellar T3SS and T3SS injectisome. Parts conserved are marked in colours and system specific subunits are shown in grey. **OM:** outer membrane, **PG**: peptidoglycan, **IM:** inner membrane. From (Milne-Davies et al., 2021)

Table 1.5.1: Summary of Flagellar and injectisome T3SS components, their relative stoichiometries (columns marked with St) and functions. The proteins are listed from distal end to cytoplasm and proteins of similar functions have been placed in the same rows. Proteins conserved in both systems are coloured **Blue.** Proteins available in only one system are uncoloured**.** References to stoichiometries are mentioned in text. Information summarized from (Hu et al., 2019; Milne-Davies et al., 2021).

Function	Flagellum		Injectisome	
	Protein	St	Protein	St
Flagellar cap protein	FliD	5		
Flagellar filament	FliC	11		
Hook-filament junction	FlgL	11		
	FlgK	11		

Pore-forming translocators			SctB	8
			SctE	8
Hydrophilic translocator			SctA	5
Hook/Needle	FlgE	~120	SctF	~140
Hook/Needle ruler	FliK		InvJ	
L-ring	FlgH	~26		
P-ring	FlgI	~26		
Rod proteins	FlgG	24		
	FlgF	5		
	FlgC	6		
	FlgB	5		
	FliE	6	SctI	6
Stator	MotA	5*n		
	MotB	2*n		
Export Apparatus (EA) Core	FliQ	4	SctS	4
	FliR	1	SctT	1
	FliP	5	SctR	5
EA gate forming protein	FlhB	1	SctU	1
	FlhA	9	SctV	9
MS (IM) ring	FliF	~26	SctJ	24
			SctD	24
Sorting platform			SctK	6
C-ring/pod large protein	FliG	34		
	FliM	34	SctQ	12/24
C-ring/pod small protein	FliN	3*FliM	$SctQ_c$	2*SctQ
ATPase regulator	FliH	12	SctL	12
ATPase	FliI	6	SctN	6
ATPase stalk	FliJ	1	SctO	1

1.5.2 Common Core within Flagellar and Injectisome T3SS

The ATPases are present in both the flagellum and many of the secretion systems. Like the flagellar ATPase hexamer FliI surrounding the monomer stalk of FliJ, the T3SS injectisome SctN hexamer assembles with SctO. This is attached to the C-ring through 12 copies of SctL, similar to that of flagellar ATPase cage FliH (Milne-Davies et al., 2021). While T3SS in both flagella and injectisome have ATPases structurally similar to that of F0F1-type ATP synthase, T2SS and T4P have ATPases structurally similar to ATPases Associated with diverse cellular Activities (AAA+). T1SS, T6SS, T4SS, T7SS also utilize ATPases in their IM complexes. (Chang et al., 2016; Christie, 2019; McCallum et al., 2019). PMF is used by T9SS instead of ATP hydrolysis for secretion, while other systems use PMF for sheath/tube contraction, such as T6SS. Class 1 secretion systems use peptide diffusion and charge interactions instead to drive the translocation of their respective effectors (Christie, 2019). The C-ring-like structure in the cytoplasm side of the IM is the second feature present in flagellum and the injectisome. While FliG subunit of the C-ring is flagellum specific, FliM and FliN form a similar structure to SctQ in the injectisome. $SctQ_c$ fragment additionally expresses and binds at 1:2 ratio to the main SctQ ring (similar to the 1:3 FliM:FliN ratio). The MS-ring of the flagellum, while being an IM protein of 22-fold symmetry in the upper part and 34-fold symmetry in the lower part matching the C-ring, has a similar injectisome component formed by 24-fold SctD and SctJ (Milne-Davies et al., 2021).

The T3SS has evolved a complex export apparatus with a cytoplasmic gate for chaperone binding and export of cytoplasmic proteins, assembly of which is summarised in Figure 1.5.3. Five units of SctR, one unit of SctT and four units of SctS make up the core of the export apparatus, alike flagellar FliPRQ complex. It forms a right-handed pseudo helical assembly closed at the base with central atrium of 15-20 Å. SctTRS presents a hydrophobic interface to the IM and hydrophilic to the periplasm. SctD assembles around this complex to create an amphipathic environment (Hu et al., 2019). One copy of SctU is then surrounded by nine copies of SctV to form the export gate (FlhBA in flagellum) (Milne-Davies et al., 2021).

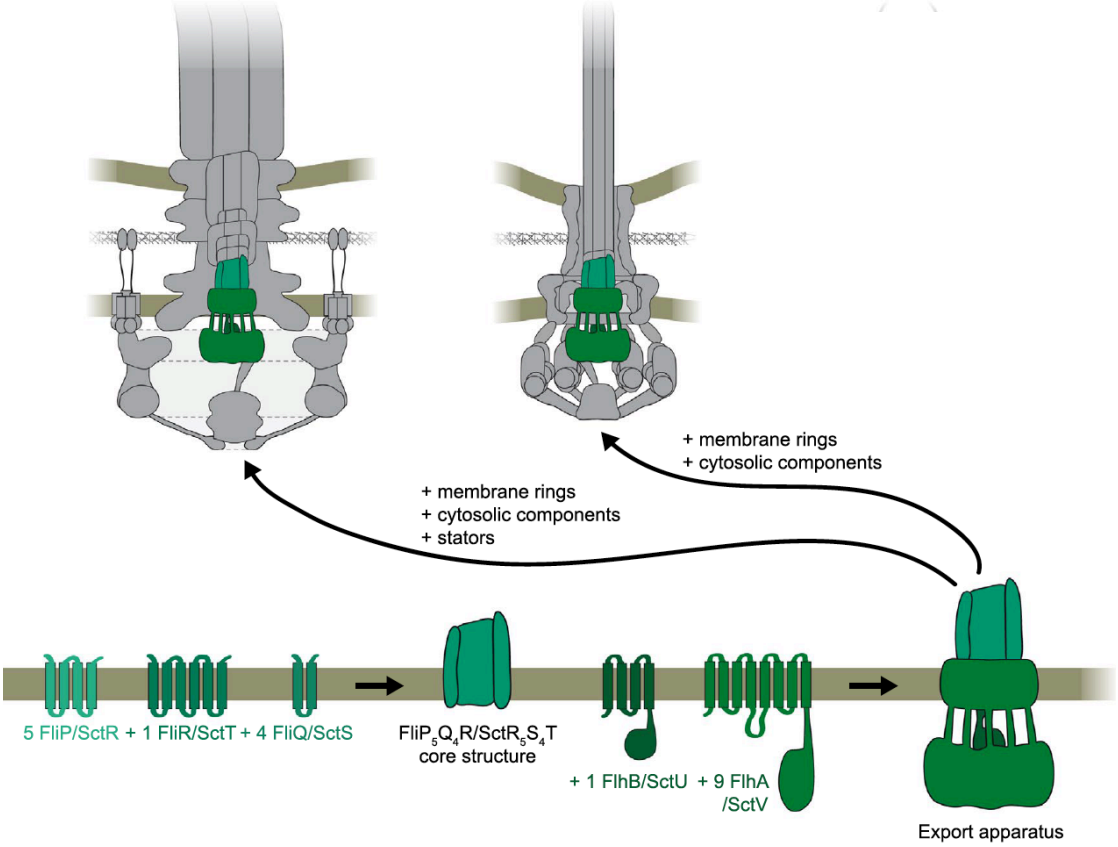


Figure 1.5.3: Assembly of the T3SS export apparatus in flagella and injectisome. **Bottom:** stepwise assembly of the export apparatus in the flagellum (left nomenclature) and injectisome (right nomenclature). **Top:** context of the export apparatus in flagella (left) and injectisome (right). Adapted from (Milne-Davies et al., 2021).

One difference between the flagellar and injectisome T3SS is the architecture of the needle complex (NC) and flagellar hook. Both require a rod passing through the membrane complexes to connect the export apparatus, with flagellar rod being a complex helical assembly of multiple proteins and injectisome rod consisting of hexamer of protein SctI that binds SctTS. The NC forms with 5 copies of SctF protomers binding SctI and completing the first turn of the 11-start needle helix (Hu et al., 2019). Alike FliK controlling the hook length, InvJ is a molecule that control the length of the T3SS injectisome needle (Bergeron et al., 2016). SctF is a small ~9 kDa protein that has a 2 helix coiled coil bundle motif with N-terminal helix facing into the lumen and C-terminal helix buried in the needle wall (Worrall et al., 2011). The NC as well as the

whole T3SS injectisome has a ~25 Å channel running through the whole complex. As many flagellar proteins that form filaments and their respective caps have a terminal coiled coil domain (FliC, FliD, FlgD, FliE) similar to that of the SctF, the mechanism of assembly and polymerization might be similar (Blocker et al., 2008). Stepwise formation and elongation of the needle is illustrated in Figure 1.5.4.

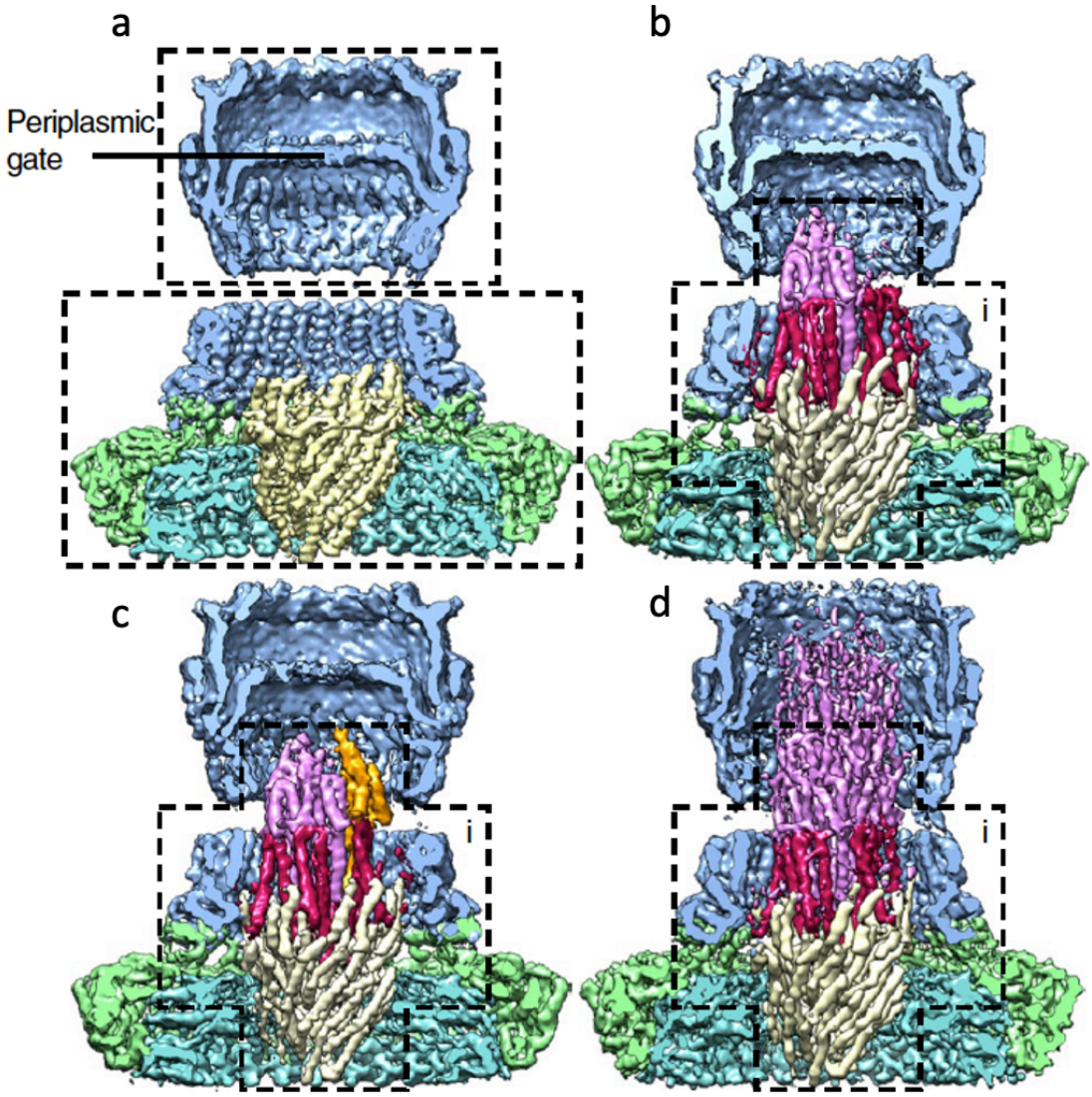


Figure 1.5.4: Reconstructions of NC showing different states of needle assembly. (a) NC base SctI, no needle and secretin periplasmic gate closed, (b) partially unlocked with SctI rod formed and SctF beginning oligomerization, (c) unlocked with additional SctF subunit added (orange) and (d) fully opened and needle elongated. SctD (green), SctV (cyan), $SctR_5S_4T$ (yellow), SctI (cherry), SctF (pink), SctC (blue). Boxed regions correspond to focus refinements. Adapted from (Hu et al., 2019).

However, while the mechanism of assembly of the flagellar hook and filament relies on their respective capping proteins attaching to a platform and elongating the filaments, the injectisome needle polymerizes without such a cap. On the top of the injectisome a pore forming complex forms made up of SctB and SctE octamers and hydrophilic translocator SctA pentamer linking the pore to the needle tip. While this NC tip forms after the needle has polymerized into a filament it also has a coiled coil terminal domain topology similar to that of the needle and flagellar axial proteins (Worrall et al., 2011).

One of the main differences between the flagellum and secretion systems is the formation of complex LP-rings spanning the OM through the lumen of which the rotating rod passes the torque onto the hook, while secretion systems have a secretin ring instead, with different function and assembly. In T3SS injectisome SctC forms a secretin pore with the periplasmic domain. Pilotins are proteins required for stability, localization, and assembly of secretin pores. However, they are not conserved across secretion systems and are specific to their secretin protein. Some systems, like enteropathogenic *Escherichia coli* (EPEC) T3SS do not have pilotins at all. Secretin pores are often the only relatively rigid structures in secretion systems while the rest of the components are flexible (Hu et al., 2019). Overall, while the core of the T3SS apparatus in the flagellum and injectisome are similar in assembly and function, the rest of the structural components are distinctly suited for motility in flagella and substrate export in the injectisome.

1.6 Structure of the Flagellar Axial Region

The flagellar axial regions have evolved to efficiently transfer the motor rotational force into motility as well as play an integral part in pathogenesis of many bacterial species. While these regions have proved difficult to study *in vivo* using structural biology, the improvement in cryo-EM techniques over the last few years has led to an increase in intact structures across the bacterial species, which in turn helped in understanding the mechanism of their assembly.

The proximal rod protein FliE is the first to be exported through an assembled flagellar T3SS complex surrounded by MS-ring, C-ring and connected to the ATPase complex. It is followed by the proteins making up the proximal (FlgB, FlgC and FlgF) and distal (FlgG) rod, assisted into folding with rod capping protein FlgJ. The rod is made up of six copies of FliE and FlgC, five copies of FlgB and FlgF, and 24 copies of FlgG in *S. enterica*, as illustrated in Figure 1.6.1. The rod subunits have a similar packing motif with the C-terminal helix packing into the centre of the growing filament with the N-terminal helix decorating the outside. The sequence between the core terminal helices is variable and changes between different adjacent rod proteins. Due to this fact, FlgG was observed to conform into three main states building up in sequential layers. The packing of the rod components follow the helical symmetry of hook and filament (discussed below) using alternating 5:6 stoichiometries (Johnson et al., 2021).

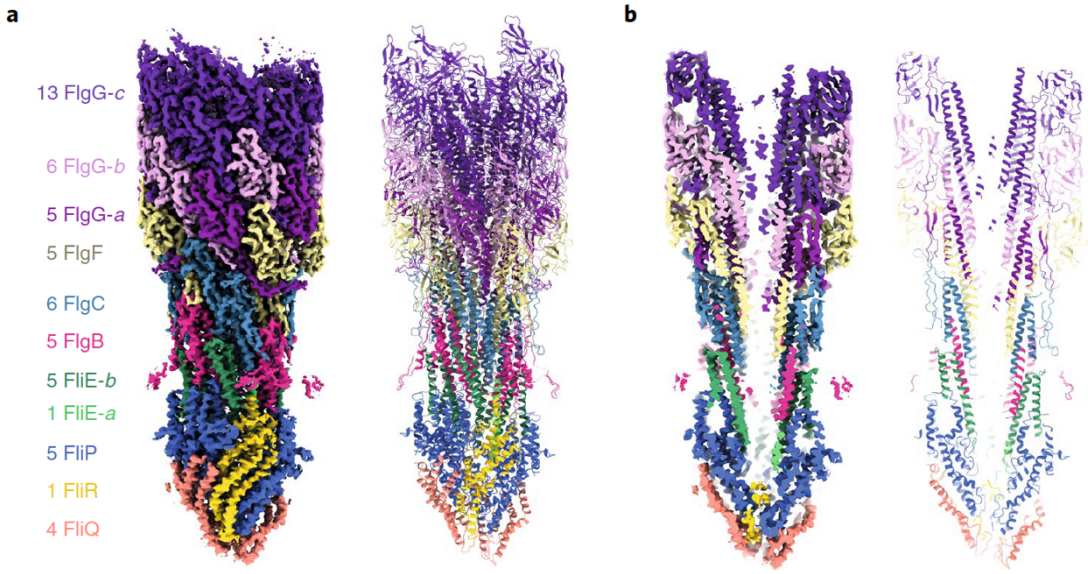


Figure 1.6.1: Structure of flagellar axial rod from *S. enterica*. (a) Cryo-EM volume (left) and cartoon representation of the atomic model (right) with the colour coded protein subunits. (b) A cross-section through structures in (a) to show a hollow lumen. Postscript indicates different conformational states of the same protein chain. From (Johnson et al., 2021).

The flagellar hook is made up of multiple repeats of FlgE. The protein has multiple main domains with D0, D1 and D2 conserved across multiple species while D3 and D4 are found specifically in ε-proteobacteria like *C. jejuni* and *H. pylori* (Matsunami et al., 2016). The X-ray crystallography structures of this protein have been determined

from incomplete constructs of different length, as shown in Table 1.6.1. There are multiple cryo-EM structures of the hook in *S. enterica* and only one other *C. jejuni* structure. However, unlike the crystallography structures, these contain the full length FlgE for their respective bacterial species.

Table 1.6.1: List of FlgE structures from Protein data bank (PDB) determined via X-ray crystallography or Cryo-EM.

Bacterial species	**Construct**	**X-ray or Cryo-EM**	**Resolution (Å)**	**PDB (EMDB)**	**Reference**
C. jejuni	D1-D4	X-ray	2.45	5AZ4	(Yoon et al., 2016).
C. crescentus	D2-D3	X-ray	1.84	5AY6	
T. denticola	D2	X-ray	3.04	6NDW	(Lynch et al., 2019)
H. pylori	D1-D3	X-ray	2.29	5NPY	(Loconte et al., 2017)
C. jejuni	D0-D4	Cryo-EM	3.5	5JXL (EMD-8179)	(Matsunami et al., 2016)
S. enterica	D1-D2	X-ray	1.8	1WLG	(Samatey et al., 2004)
S. enterica	D0-D2	Cryo-EM	4.1	6KFK (EMD-9974)	(Horvath et al., 2019)
S. enterica	D0-D2	Cryo-EM	3.1	6K9Q (EMD-9952)	(Kato et al., 2019)
S. enterica	D0-D2	Cryo-EM	3.4	7CGB (EMD-30354)	(Tan et al., 2021)
S. enterica	D0-D2	Cryo-EM	7.1	6JZT (EMD-1674)	(Saijo-Hamano et al., 2019)
S. enterica	D0-D2	Cryo-EM	2.86	6K3I (EM-9909)	(Shibata et al., 2019)

Core FlgE domains conserved across many bacterial species include the D0 alpha helical coiled coil domain made up of both N and C-terminal regions, D1 globular domain and D2 globular domain found to be essential for curvature in the filament (Shibata et al., 2019). *C. jejuni, C. crescentus* and *H. pylori* have an additional D3 globular domain, with *C. jejuni* and *H. pylori* both having another additional D4 domain, making *C. jejuni* 92 kDa FlgE the longest flagellar hook protein (Yoon et al., 2016). FlgE has an additional small domain localized between D0 and D1 termed Dc, deletion of which does not prevent hook assembly. It is thought to be important in strengthening

the structure of the hook filament (Shibata et al., 2019). Domains behave as rigid bodies connected by flexible hinge regions with D2 observed to move as much as 27 Å (Shibata et al., 2019). This is contrary to previous hypothesis that a hook is a rigid structure, and it can transfer the torque from the motor to the filament while being relatively mobile.

FlgE forms an 11 protofilament tube with a diameter of 280 Å in *C. jejuni* and 180 Å in *S. enterica* (Matsunami et al., 2016). The curvature of the filament occurs due to the shortest protofilament having a 6.3 Å distance between top and bottom subunits due to closer packing, while the longest having 12.5 Å. This difference causes one protofilament to be shorter than the other thus curving the hook into the correct form as shown in Figure 1.6.2 (Kato et al., 2019). FlgE and FlgG distal rod protein have high sequence similarity and domain distribution as shown in Figure 1.6.3. The assembled polyrod has similar helical parameters to the hook filament allowing FlgD to build the hook straight onto the rod without any adapter proteins (Horvath et al., 2019).

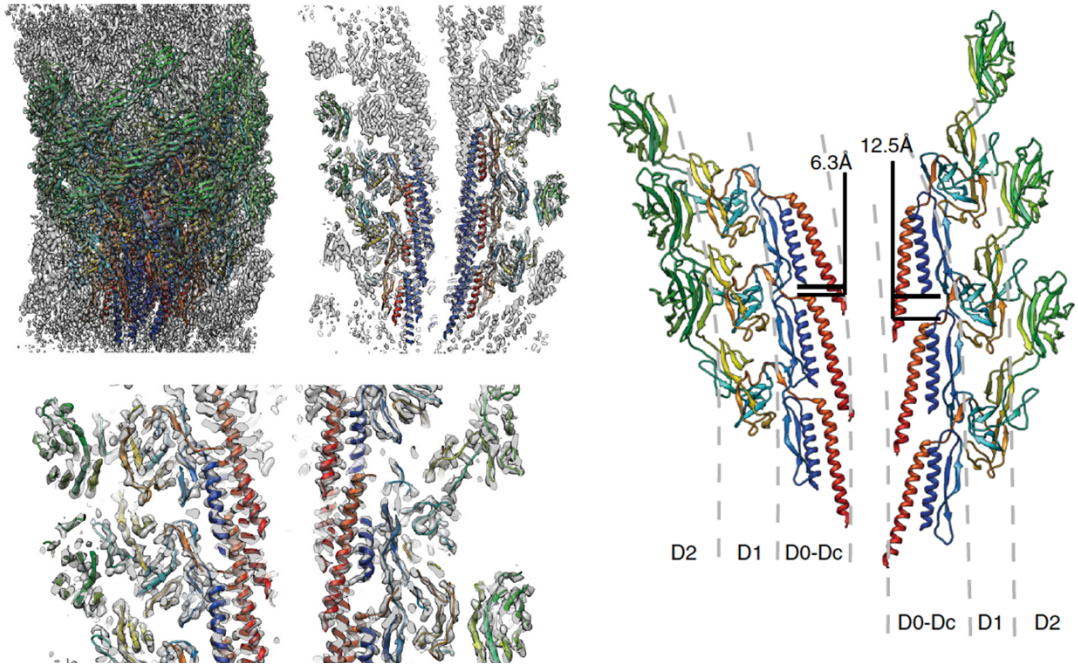


Figure 1.6.2: Structure of supercoiled hook from *S. enterica*. **Left:** Reconstructed density map and models of FlgE subunits coloured in rainbow from N to C-terminus. Domains are labelled underneath the structure. **Right:** Cross section from the model with the difference in the gap distance between the D0 domains along the filament for the protofilament for the shortest and the longest, which gives FlgE hook the curvature. From (Kato et al., 2019).

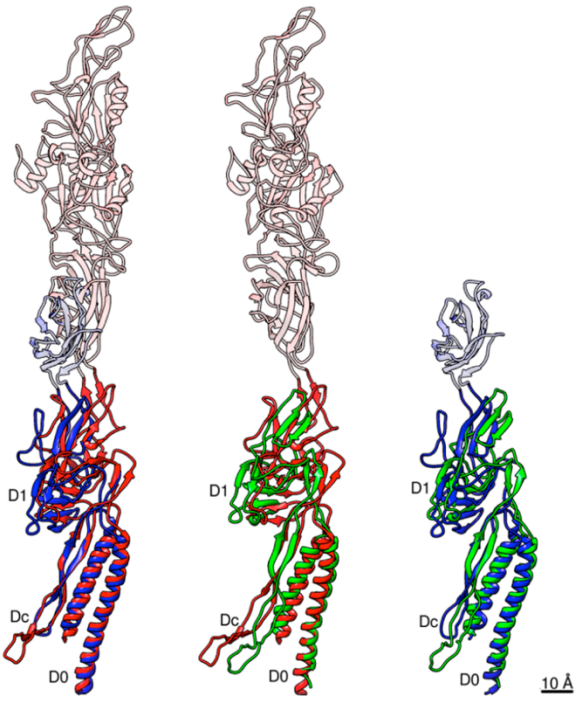


Figure 1.6.3: Comparison of *S. enterica* FlgE, FlgG and *C. jejuni* FlgE through superposition on the D0 domain. Colours only applied to D0-D1 domains. **Left**: *S. enterica* FlgE (blue) and *C. jejuni* FlgE (red) are compared with the rest of the domains coloured in pink for *C. jejuni* and light blue for *S. enterica.* **Centre:** *S. enterica* FlgG (green) and *C. jejuni* FlgE (red) are compared. **Right:** *S. enterica* FlgG (green) and FlgE (red) are compared. From (Horvath et al., 2019).

After the hook reaches a certain length, which is controlled by FliK, the hook-filament junction begins to form to connect it to the filament. FliK does this through being periodically secreted during the formation of HBB and binding both FlgD and FlhB to measure the distance between them. If it is too short (hook incomplete), FliK secrets rapidly through the lumen not having enough time to induce autocleavage in FlhB. If the hook reaches the desired length of 55 nm, then FliK takes longer to exit the lumen and gives time for FlhB to switch to late substrate expression (Erhardt et al., 2011). Hook-filament junction consists of two proteins that localize at the tip of the hook FlgK and FlgL. There is limited structural information on this junction and all currently published structures are listed in Table 1.6.2. There have been no cryo-EM structures of either of these proteins to date, so their assembled oligomerization state has been assumed from modelling the crystal structures to that of the 11-protofilament hook. As observed with hook monomers, the terminal regions appear to prevent

crystallization in solution, so in all crystal structures the D0 domain is not present. FlgK in *C. jejuni* consists of three main domains D0, D1 and D2 domains with D1 split into two subdomains D1a and D1b. Comparing to FlgK in other species, D1 domain is highly conserved while D2 has a different fold. While other structures have an eight strand β-barrel*, C. jejuni* FlgK D3 has a V-shaped six strand barrel complemented by helices as illustrated in Figure 1.6.4. *B. pseudomallei* FlgK has an extra D3 globular domain (Bulieris et al., 2017). FlgL consists of D1 domain helical bundle and a D2 β-sandwich, with overall domain structure similar to that of FlgK, even though the D2 domain is located at a different angle to the D1 helical bundle. While D1 is stable on its own, D2 is essential for proper folding. D1 domain is structurally conserved between FlgL and flagellin as well as illustrated in Figure 1.6.5 (Song et al., 2020).

Table 1.6.2: List of FlgK and FlgL structures from PDB determined via X-ray crystallography.

Bacterial species	**Construct**	**X-ray or Cryo-EM**	**Resolution (Å)**	**PDB (EMDB)**	**Reference**
FlgK					
B. pseudomallei	D1-D3	X-ray	1.8	4UT1	(Gourlay et al., 2015)
C. jejuni	D1-D2	X-ray	2.45	5XBJ	(Bulieris et al., 2017)
S. enterica	D1-D2	X-ray	2.1	2D4Y	unpublished
FlgL					
X. campestris	D1-D2	X-ray	1.9	5ZJ0	(Hong et al., 2018)
B. cereus	D1	X-ray	2.2	5ZIY	
L. pneumophila	D1-D2	X-ray	3.06	7C7Z	(Song et al., 2020)

After the hook-filament junction assembles on top of the hook filament, flagellar capping protein FliD attaches and begins elongating the flagellar filament made up of thousands of copies of flagellin. Comparing to other proteins in the axial region of the flagellum, the filament has been extensively studied for over 20 years using both X-ray crystallography and Cryo-EM as shown in Table 1.6.3.

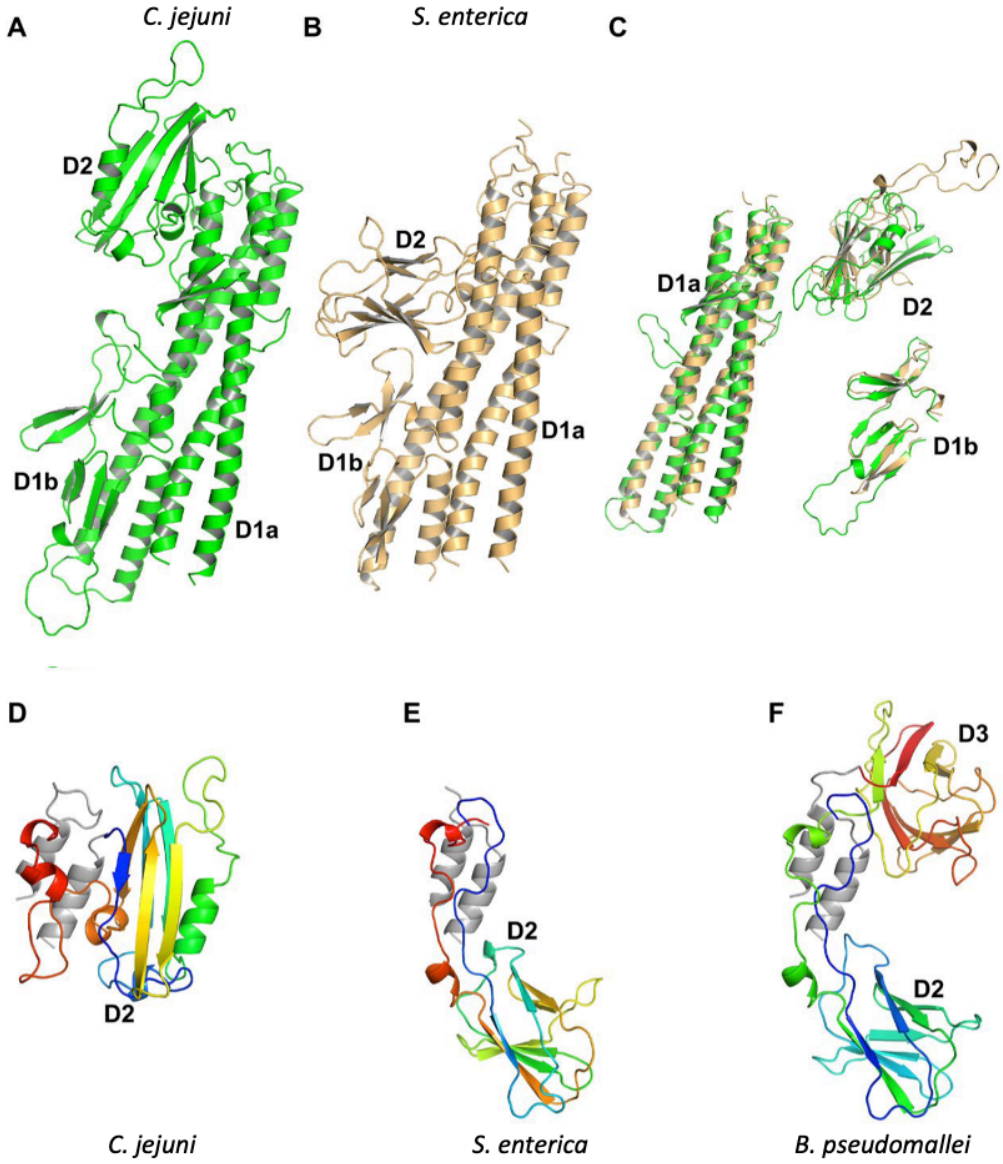


Figure 1.6.4: Comparison of FlgK from different bacterial species. (A) FlgK from *C. jejuni* (B) *S. enterica*. (C) Superpositions of the structures in (A) and (B) by isolated domains. (D) D2 domain of FlgK from *C. jejuni* rainbow colour from N to C terminus. (E) D2 domain of FlgK from *S. enterica* (F) and extra D3 domain from FlgK form *B. pseudomallei.* From (Bulieris et al., 2017).

While flagellin usually has the common denomination of FliC as observed in *E. coli*, the filament is often comprised of more than one type of flagellin. For example, in *S. enterica* filaments can form from FliC or FljB, while *C. crescentus* has FljK as one of six flagellins. However, only one flagellin can be expressed at a time, with the switch between them being called phase variation (Montemayor et al., 2021).

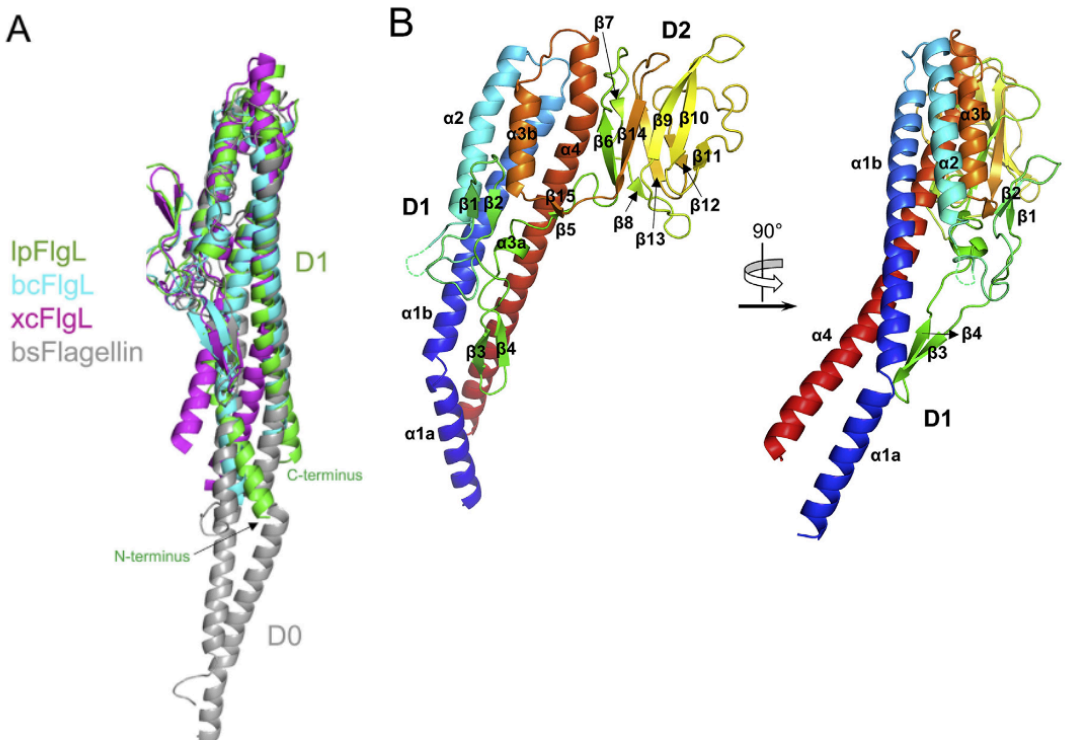


Figure 1.6.5: Comparison of FlgL from different bacterial species. (A) Similarity of FlgL from *L. pneumophila* D1 domain (green) with D1 domain of *B. cereus* (blue) and *X. campestris* (magenta) D1 domains as well as *B.* subtilis flagellin. (B) Structure of FlgL from *L. pneumophila* depicted as a rainbow ribbon (N-terminus blue, C-terminus red). From (Song et al., 2020).

Spirochaetes have even more complicated flagella compared to the exoflagella listed above. While FlaA and FlaB deletions affects curvature and motility and FlaB deletion results in the removal of the filament, the extra FcpA and FcpB flagellins are essential for supercoiling and virulence and form an extra sheath responsible for *Spirochete* specific functions and bacterial internalized environment (Gibson et al., 2020). *C. jejuni* minor flagellin FlaB is a part of the class 2 genes, expressed prior to FliK molecular ruler switching the specificity of FlhB. It was observed that flagellar filaments consist of a short length of FlaB near the hook-filament junction, followed by FlaA major flagellin. In *C. jejuni* the wild type filament is restricted to the growth of about 3.5 µm, with flagellum at one end being slightly longer than the other. This indicates that while class 3 major flagellin had not been expressed yet, FliD assists in building a short segment of FlaB minor flagellin, prior to the switch in FlhB. Once the switch to late substrates occurs, FlaA is expressed and assembles at the end of the FlaB filament,

still assisted by FliD. FlaA and FlaB flagellins in *C jejuni* appear to pack differently as a FlaA only mutant produces straight filaments and FlaB only results in highly curved segments. This could mean that the helical shape observed through EM is due to the FlaB part transferring the helicity to the straight FlaA section (Inoue et al., 2018).

Table 1.6.3: List of flagellin structures from PDB determined via X-ray crystallography or Cryo-EM.

Bacterial species	**Construct**	**X-ray or Cryo-EM**	**Resolution (Å)**	**PDB (EMDB)**	**Reference**
S. enterica	D1-D3	X-ray	2.00	1IO1	(Samatey et al., 2001)
S. enterica	D0-D3	Cryo-EM	4.00	1UCU	(Yonekura et al., 2003)
S. enterica	D0-D3	Cryo-EM	4.00	3A5X (EMD-1641)	(Maki-Yonekura et al., 2010)
P. aeruginosa	D1-D2	X-ray	2.10	4NX9	(Song and Yoon, 2014)
B. pseudomallei	D1-D2	X-ray	1.3	4CFI	(Nithichanon et al., 2015)
B. subtilis	D0-D1	Cryo-EM	4.5	5WJY (EMD-8852)	(Wang et al., 2017)
B. subtilis	D0-D1	Cryo-EM	6.7	5WJX (EMD-8851)	
B. subtilis	D0-D1	Cryo-EM	5.7	5WJZ (EMD-8853)	
B. subtilis	D0-D1	Cryo-EM	4.3	5WJU (EMD-8848)	
P. aeruginosa	D0-D3	Cryo-EM	4.3	5WK6 (EMD-8856)	
B. subtilis	D0-D1	Cryo-EM	3.8	5WJT (EMD-8847)	
P. aeruginosa	D0-D3	Cryo-EM	4.2	5WK5 (EMD-8855)	
B. subtilis	D0-D1	Cryo-EM	4.4	5WJW (EMD-8850)	
B. subtilis	D0-D1	Cryo-EM	5.5	5WJV (EMD-8849)	
Kurthia	D0-D1	Cryo-EM	2.8	6T17 (EMD-10362)	(Blum et al., 2019)
S. enterica	D0-D3	Cryo-EM	3.56	6JYO (EMD-9896)	(Yamaguchi et al., 2020)
C. jejuni	D0-D4	Cryo-EM	3.5	6X80 (EMD-22088)	(Kreutzberger et al., 2020)
C. crescentus	D0-D1	Cryo-EM	3.4	6XL0 (EMD-22231)	(Montemayor et al., 2021)
C. crescentus	D0-D1	Cryo-EM	3.2	6XKY (EMD-22229)	

According to the structure of *S. enterica* FliC, the flagellin monomer has four domains and folds on itself like a hairpin starting with N-terminal coiled coil D0, followed by a D1 domain which is a bundle of 3 α-helices, two β-turns, a β-hairpin and a long chain stretched over three helices. D2 and D3 domains are rich in antiparallel β-sheets and are variable in both size and sequence. After forming the D3 domain the sequence goes back into D2, D1 and ends in the D0 C-terminal coiled coil which is illustrated in Figure 1.6.6. The architecture of the D0 and D1 domains is highly conserved across bacterial species. The structure of *P. aeruginosa* flagellin has D0-D1 domains solved, but the protein sequence indicates the presence of the D2-D3 domains, similar to the domain distribution of *S. enterica*, which could not be resolved due to low resolution (Maki-Yonekura et al., 2010; Montemayor et al., 2021).

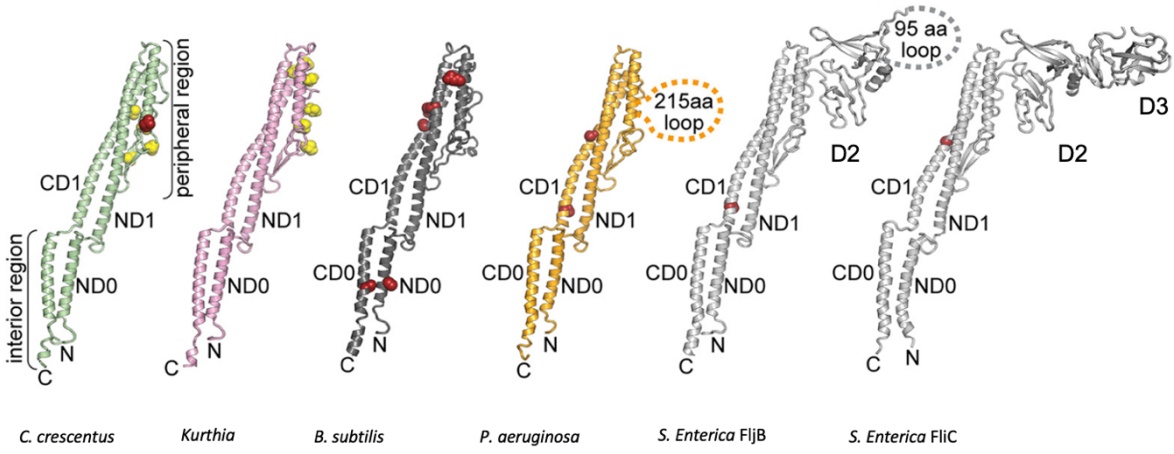


Figure 1.6.6: Comparison of bacterial flagellins. Known straightening substitutions are shown in red spheres and residues undergoing post translational modifications are in yellow spheres. Adapted from (Montemayor et al., 2021).

C. jejuni flagellar filament is made up of two separate flagellins FlaA and FlaB, and both have a standard flagellar domain distribution (D0-D3) with an additional head domain protruding from the D3 domain termed D4. However, while D0-D1 domains are conserved, but D2-D3 domains are not homologous to the *S. enterica* D2-D3, as illustrated in Figure 1.6.7 (Kreutzberger et al., 2020).

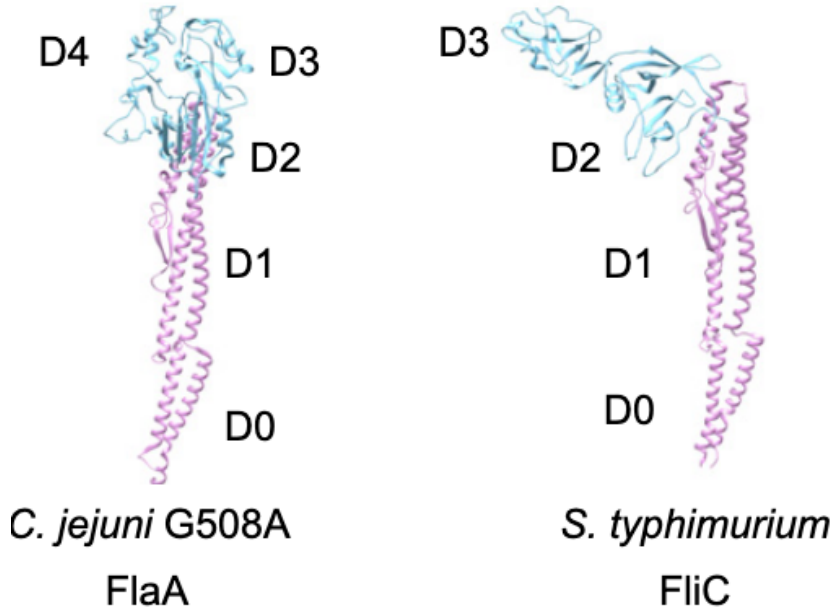


Figure 1.6.7: Comparison of *C. jejuni* and *S. enterica* major flagellins. Conserved homologous domains in pink and non-homologous head domains in blue. Adapted from (Kreutzberger et al., 2020).

Using the hook-filament junction as a base, flagellins polymerize into protofilaments which in turn form the whole flagellum with the assistance of FliD capping protein. However, spontaneous polymerization of flagellin was observed in supersaturating conditions without the presence of other flagellar proteins (Wakabayashi et al., 1969). This indicates that the regions driving initial polymerization are located within the flagellins themselves and most likely is the D0 terminal domain. While this spontaneous polymerization theory has been disproven, the terminal domains still may be involved in the polymerization process. This is one of the reasons the flagellar axial proteins such as FlgE, FlgG, FlgK and FlgL as well as flagellin could not be crystallized without removal of terminal coiled coil sequences. Another issue during structure determination of filaments is that they supercoil and create difficulty during helical reconstruction. Similar to the short and long protofilaments in the hook giving it the curved shape, in *Salmonella* protofilaments can either be short or long, which also reflects in the Right-hand inclination (R-state) and Left-hand inclination (L-state) respectively, depending on the ratio of different types of minor and major flagellins. This variety leads to supercoiling (Wang et al., 2017). This is why straightening the

filaments through mutation of A449V in *S. enterica* and G508A in *C. jejuni* allowed for production of straight filaments amenable for structure determination by cryo-EM (Galkin et al., 2008).

The number of protofilaments from the structures of six different organisms, including *B. subtilis, P. aeruginosa* and *S. enterica*, was determined to be 11. When comparing the D0-D1 domains, the width of the inner diameter (25 Å) and outer diameter (125 Å) as well as the conformation is highly conserved. However, while *B. subtilis* flagellin ends with the D0-D1 domains, *P. aeruginosa* and *S. enterica* contain extra domains D2-D3, which widen the outer diameter to 170 Å and 230 Å respectively while maintaining the inner diameter width, as illustrated in Figure 1.6.8 (Wang et al., 2017). The helical parameters, namely the rise and the twist, are relatively conserved across bacterial species with the rise being in the range of 4.61 – 4.90 Å and twist in the range of 65.1 – 65.8 ° (Montemayor et al., 2021). However, the first structure of *C. jejuni* flagellum solved to 15 Å in 2008 indicated the presence of 7 protofilaments. The diameter of the flagella was also smaller than that of *Salmonella* (180 Å) (Galkin et al., 2008).

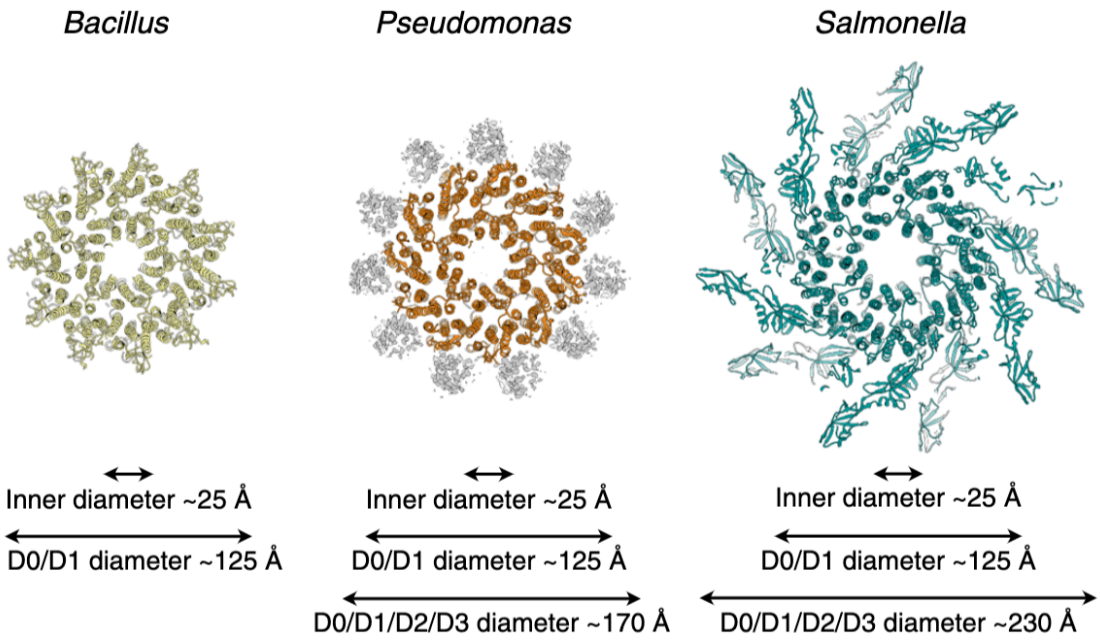


Figure 1.6.8: Comparison of bacterial flagellar filaments. Compared maps of *B. subtilis, P. aeruginosa* and *S. enterica* (3A5X (EMD-1641)). Unresolved density is shown in grey. From (Wang et al., 2017).

The irregular structure of *C. jejuni* flagella sparked a discussion about its unique properties, mainly the discrepancy between the hook oligomerization state (11 subunits) and the protofilaments (7 subunits), as well as the connection between the number of protofilaments and oligomerization state of the FliD capping protein. However, in 2020 an updated structure was published, showing that *C. jejuni* forms an 11-start helical filament as well, with helical parameters of 4.8 Å rise and 65.4 ° twist within the range of other structures and a 200 Å width, which is slightly shorter than that of *Salmonella* (230 Å) due to the difference in head domain packing. *C. jejuni* FlaA also contains an additional D4 domain to the D2-D3 domains present in *Salmonella* (Kreutzberger et al., 2020).

Many bacteria, including *C. jejuni*, have flagella that are glycosylated. Glycosylation of flagella involves O-link attachment of pseudaminic or legionaminic acids to Serine and Threonine residues on the surface of the flagellum (see section 1.2). It prevents exposure of a sensitive peptide region which triggers an immune response (Rossez et al., 2015; Young et al., 2007). In *C. jejuni* filament the glycosylation congregates around the D4 domain of FlaA major flagellin, which is important in evading the TLR5 receptors. It is also important in structural integrity of the filament, as shown by the fact that if two serine residues S455 and S461 are mutated to alanine and glycosylation is removed, the full filament cannot assemble and motility is greatly reduced (Kreutzberger et al., 2020). Similarly, in *Aeromonas* spp. lack of pseudaminic acid glycosylation of flagella prevents filament assembly (Mendoza-Barberá et al., 2021).

1.7 Flagellar Capping Protein and Mechanism of Filament Elongation

Similar to how FlgD binds the flagellar distal rod and assists in building the hook, flagellar capping protein FliD binds the FlgL section of the hook-filament junction and then proceeds to elongate the flagellar filament. Unlike the filament itself, which has been studied extensively through cryo-EM, FliD structures from different bacterial species were determined via X-ray crystallography in a truncated state without

terminal domains. While there have been attempts at observing FliD using cryo-EM in both *in vitro* and *in vivo*, the resolution of the maps obtained was too low to build a comprehensive model (Maki-Yonekura et al., 2003; Zhang et al., 2019). The known structures of FliD are listed in Table 1.7.1.

Table 1.7.1: List of FliD structures from PDB determined via X-ray crystallography or Cryo-EM. The last two entries are maps available in the Electron Microscopy Data Bank (EMDB) but do not have a corresponding PDB model.

Bacterial species	**Construct**	**X-ray or Cryo-EM**	**Resolution (Å)**	**PDB (EMDB)**	**Oligomeric state**	**Reference**
P. aeruginosa	D2-D3	X-ray	2.2	5FHY	6	(Postel et al., 2016)
S. enterica	D2-D3	X-ray	2.5	5H5T	5	(Song et al., 2017)
E. coli	D2-D3	X-ray	2.15	5H5W	6	
E. coli	D1-D3	X-ray	3.0	5H5V	6	
S. marcescens	D2-D3	X-ray	1.9	5XLJ	4	(Cho et al., 2017)
S. marcescens	D2-D3	X-ray	3.05	5XLK	4	
H. pylori	D2-D3	X-ray	2.6	6IWY	-	(Cho et al., 2019a)
B. bacteriovorus	D2-D3	X-ray	1.99	6KTY	4	(Cho et al., 2019b)
S. enterica	D0-D3	Cryo-EM	26	EMD-1873	5	(Maki-Yonekura et al., 2003)
B. burgdorferi	D0-D4	Cryo-EM	32	EMD-0525	5	(Zhang et al., 2019)

FliD self-oligomerises as a capping protein at the tip of the flagellum. In *S. enterica* it forms a five-legged stool with flexible leg domains, where the plate consists of the core regions and the flexible regions make up the legs (Maki-Yonekura et al., 2003). While in solution it is bound as a monomer to the chaperone FliT, preventing premature oligomerization (Song et al., 2017). However, when no chaperone is present it forms a decameric complex consisting of two pentamers to stabilize the flexible regions (Maki-Yonekura et al., 2003).

While only one structure obtained through cryo-EM to 26 Å exists in *S. enterica*,

FliD structures from *E. coli, P. aeruginosa, S. enterica, S. marcescens, H. pylori* and *B. bacteriovorus* were obtained via X-ray crystallography. They share flagellin-like domain distribution with D1 helical bundle forming the leg, which resembles flagellin and FlgL structurally but not according to sequence identity. Connected to the D1 leg domains are D2-D3 domains rich in anti-parallel β-sheets forming the head. While their protomer structures are quite similar, the main difference between species is the oligomerization state as well as extra domains in some species as illustrated in Figure 1.7.1 (Postel et al., 2016).

E. coli FliD ($FliD_{ec}$) structure was obtained to 3 Å resolution with terminal deletions to allow crystallization. As with the flagellin, the protein folded in a γ-shaped hairpin with D0 missing, and the remaining leg domain D1 formed an α-helix bundle with two β-strands forming a hairpin loop. D2 domain has two antiparallel β-sheets forming a cup-like domain with hydrophobic top exposed to solvent as a monomer but covered in the hexamer. D3 domain also has two antiparallel β-sheets but has a different overall fold as it is surrounded by two α-helices and two extra β-strands (Song et al., 2017).

P. aeruginosa FliD ($FliD_{pa}$) structure was solved to 2.2 Å resolution on a construct with a large portion of C and N termini removed. It contains a complete D3 domain, partial D2 and D1 domains caused with the deletion of N-terminal 77 residues and missing D0 domain from an additional C-terminal 69 residue deletion. The complete D3 domain has a similar fold to that of previously mentioned FliD structures as does the partially complete D2 (Postel et al., 2016). *S. enterica* FliD ($FliD_{se}$) structure was solved to 2.5 Å resolution but only containing D2-D3 domains, which are structurally similar to that of $FliD_{ec}$ (Song et al., 2017).

S. marcescens FliD ($FliD_{sm}$) is yet another structure solved by X-ray crystallography to 1.9 Å resolution which also contained only the D2-D3 domains as well as the deleted N and C termini. D3 domain has slightly more variation than D2 domain. When aligned by their D2 domains the structures of monomers are almost identical. However, when aligned by their D3 domains there are some discrepancies in flexible loops connecting the helices around the antiparallel β-sheets (Cho et al., 2017).

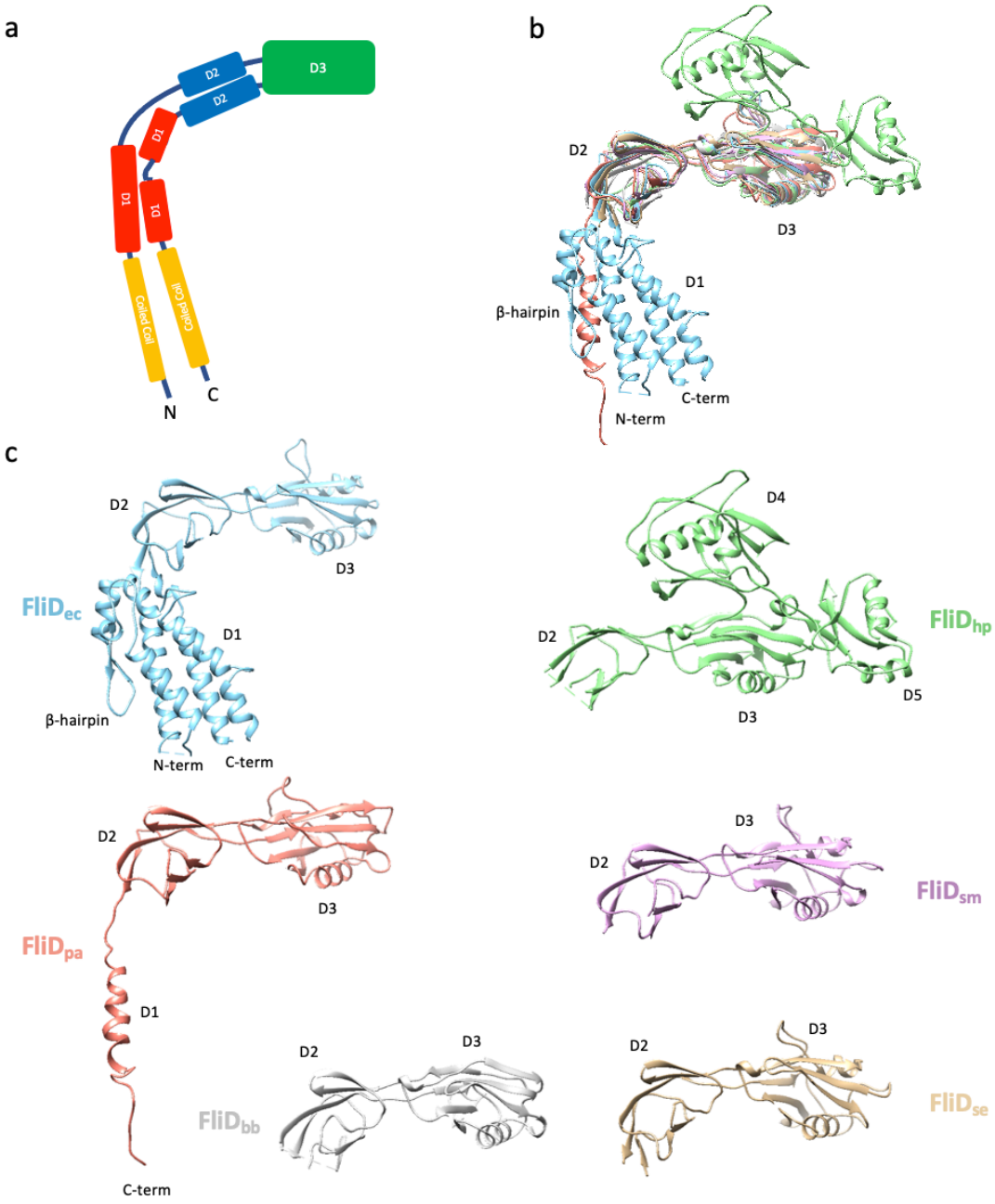


Figure 1.7.1: Structures of FliD monomers across bacterial species. (a) Schematic representation of the common domains of FliD monomer from the studied bacterial species. (b) Alignment of the structures in (c) using $FliD_{ec}$ as a reference via D2 domain. (c) Domain allocation and structures of FliD monomers from the structures listed in Table 1.7.1

B. bacteriovorus FliD ($FliD_{bb}$) crystal structure was solved to 2 Å resolution with the D2-D3 construct. The domains are similar to other FliD structures but instead of an α-helix there is a six residue loop (Cho et al., 2019b). *H. pylori* FliD ($FliD_{hp}$) is longer in sequence than all other FliD analysed in this work to date. To determine the crystal

structure, $FliD_{hp}$ domains D0 and D1 were cut leaving D2, D3, D4 and D5 with the D2-D3 head domains conserved across bacterial species and the *H. pylori* specific D4 and D5 domains having an antigenic effect (Cho et al., 2019a). There is no known atomic structure of *C. jejuni* FliD ($FliD_{cj}$) to date, but the amino acid sequence also has the conserved terminal coiled coil regions and the predicted secondary structure similar to that of FliD monomer. It also has multiple inserts that could potentially be a D4 additional domain as observed in $FliD_{hp}$ (see results chapter 3.1).

Oligomerisation is assumed to be regulated by the head and not the leg domains, as shown through crystallographic studies of a shorter D2-D3 construct in $FliD_{ec}$ (Postel et al., 2016). The subunits are connected through D2-D3' interactions between adjacent monomers to make a circular construct with a lumen (Song et al., 2017). While the protomer domain structure is highly conserved between species, they appear to differ in their oligomerization according to the crystal symmetry. $FliD_{ec}$ and $FliD_{pa}$ appear to form hexamers, $FliD_{se}$ a pentamer, $FliD_{bb}$ and $FliD_{sm}$ tetramers as illustrated in Figure 1.7.2. Formation of $FliD_{ec}$ hexamer was demonstrated by a crystal structure of a D2-D3 construct (Song et al., 2017). $FliD_{pa}$ was proven to form hexamers through negative stain EM and complementation of a knockout mutant with the purified hexamer. At the same time complementation with $FliD_{se}$ did not restore motility to the *P. aeruginosa* mutants. However, even though they are both hexamers, the packing of the domains represented by the structures indicates a larger lumen in $FliD_{pa}$ as well as a wider overall diameter (Postel et al., 2016).

$FliD_{se}$ is a pentamer in the crystal, which is also shown via cryo-EM. The lumen is smaller than that of the hexamers due to the denser packing of the domains, causing the pentamer to have smaller diameter than the hexamer. In the low resolution cryo-EM data the full legs are visible (Maki-Yonekura et al., 2003). $FliD_{sm}$ appears to be a tetramer not within the asymmetric unit, but due to crystallographic symmetry. It has a very small lumen of about 11 Å which could possibly indicate that the tetrameric state is a non-native oligomerization state but a crystallographic artefact (Cho et al., 2017).

The main function of FliD capping protein is the assembly of a fully functional

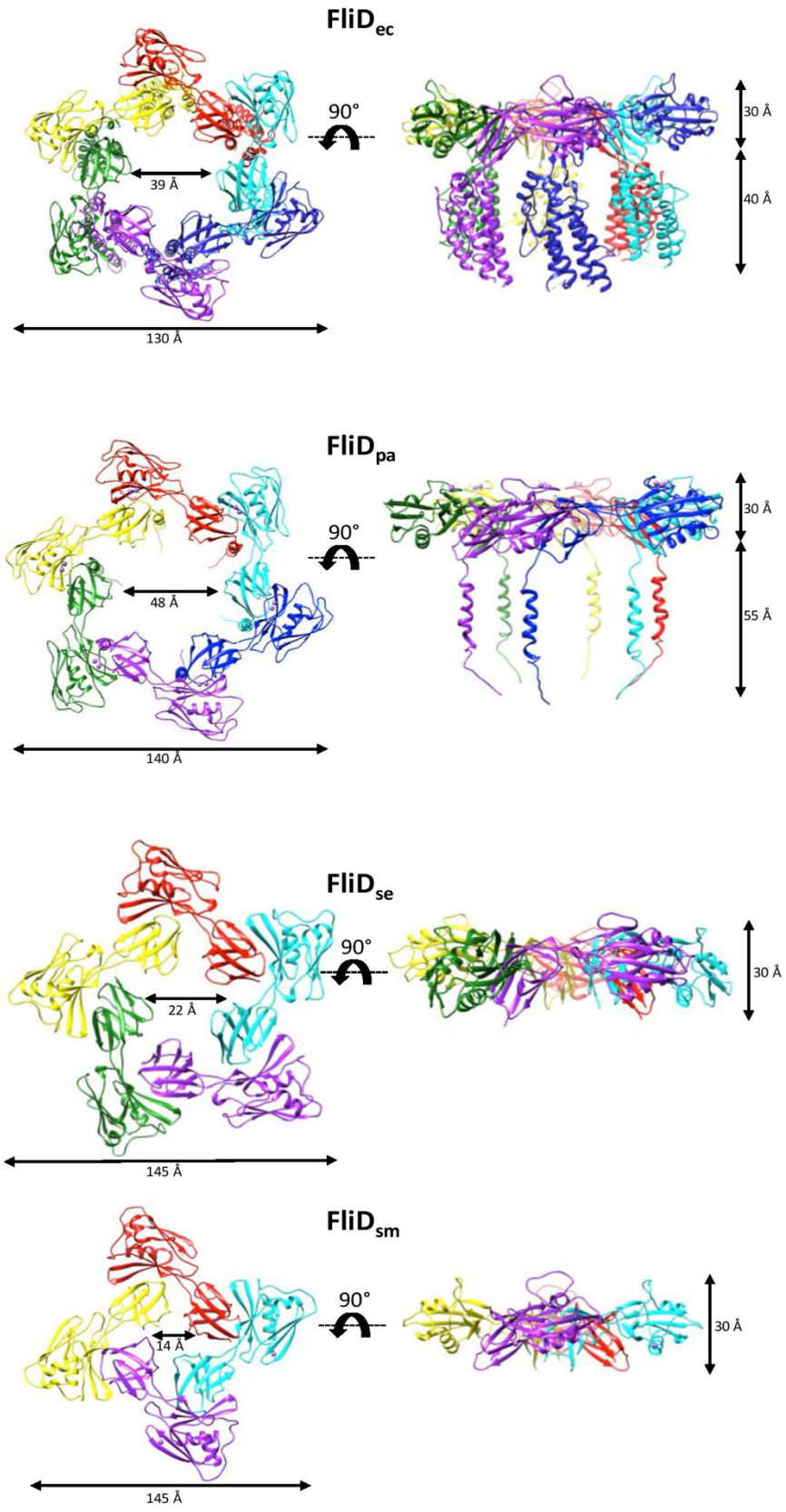


Figure 1.7.2: Known oligomers of FliD capping protein. Crystal structures of $FliD_{ec}$ (5H5V), $FliD_{pa}$ (5FHY), $FliD_{se}$ (5H5T) and $FliD_{sm}$ (5XLJ) as bio assembly units. The distances were calculated in Chimera or provided by the reference paper in Table 1.7.1.

flagellum. If FliD is not present, the cells lose their motility completely (Diószeghy et al., 2004; Imada et al., 1998). However, it can be complemented by adding purified recombinant FliD which is also interchangeable in between certain species. $FliD_{ec}$ can reconstitute salmonella flagellum but $FliD_{pa}$ cannot (Inaba et al., 2013). In solution, when FliD is not bound to its chaperone, it exists in an equilibrium between monomers and large oligomeric assemblies. As the stool-like structure has the flexible terminal regions exposed, it strives to stabilize through binding the terminal regions of another oligomer. Thus, $FliD_{se}$ pentamer forms a decamer (Maki-Yonekura et al., 2003). In alkaline pH the equilibrium is shifted towards monomeric assemblies, while the acidic pH shifts it towards the biological dimers. The flagellar lumen is not wide enough to fit the head domains, so it is more likely that the stool-like structure is the native state when bound to the flagellum (Maki et al., 1998)

While the legs of FliD are assumed to bind flagellin, the mechanism is still unknown. FliD shows low sequence similarity to the flagellar hook proteins and FliC with the exception of the highly conserved terminal coiled coil regions (Postel et al., 2016). According to nuclear magnetic resonance (NMR) studies of FliT-$FliD_{se}$ fusion protein, the C-terminal regions missing from current structures form alpha helices or coiled coils (Khanra et al., 2016). Upon deletion of these regions, the oligomerization of the biological assemblies is abolished (Furukawa et al., 2002).

As there is information present on the *S. enterica* flagellar filament as well as FliD, a mechanism of elongation for enteric flagellar filaments has been proposed which is outlined in Figure 1.7.3 (Yonekura et al., 2000). The mechanism relies on both rotation and axial translation of the cap plate accompanied by rearrangement of the legs through the flexible loop between D1 and D2 domains. This mechanism is based on a 3D density map where a cavity in place of one of the flagellar legs was observed. This led to a hypothesis that the leg domain shown in position 1 cannot fill the indentation in the growing filament as it is too deep and FliD leg cannot reach and bind flagellin subunits below it properly. The other four legs in positions 2-5 can interact with flagellins in a stable manner. The position illustrated is hypothetical but the idea being that the D1-D0 leg domains of FliD are not within that cavity. In the top of the figure 1.7.3, the red pentagon signifies the strict 5-fold symmetry of the head domains

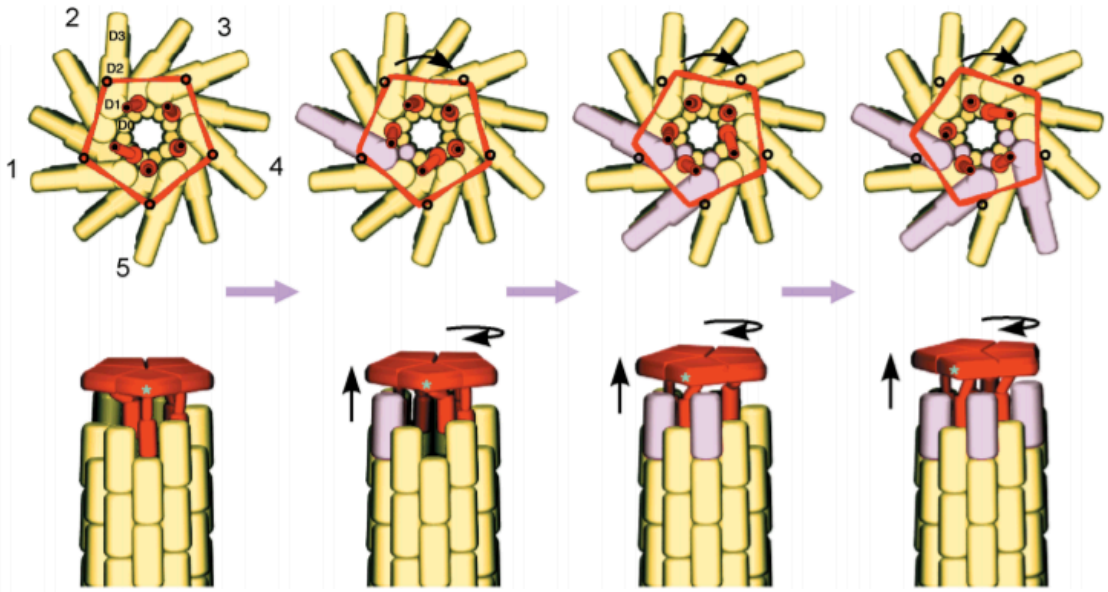


Figure 1.7.3: Proposed mechanism of *Salmonella* flagellar assembly. The pentamer head of FliD is represented by the flat red pentagon. The units in violet are newly incorporated flagellin molecules. Five circles in the upper panel indicated the initial positions of the cap domain. As the cap rotates clockwise, the leg domains shift slightly to allow a new flagellin molecule to incorporate in its place. From (Yonekura et al., 2000)

and the red tubes on the inside of the pentagon show how flexible the position of each leg can be in response to the rigid cap. According to this model, the flagellin assembly proceeds along a 1-start helix in a CCW direction, approximately at every 65.5°. This value was calculated from assuming that to build a full ring of 11 flagellins in the filament, the cap would need to rotate 360° twice (360 x 2) and then divided by the number of monomers to be added (360 x 2 / 11). 65.5° is also the angle of rotation after which the next binding site appears. However, because the legs of the cap are located every 72° (360/5), a 6.5° CW rotation with change in the leg conformations is sufficient to make appropriate interactions. In summary, as the filament is growing in a CCW direction, adding a flagellin monomer every 65.5°, FliD is rotating CW with the legs shifting 6.5° to open up a new binding site. FliD would rotate completely by the assembly of 55 flagellin subunits. The energy driving the cap rotation is supplied by newly polymerizing flagellin subunits due to the injection-diffusion mechanism (Yonekura et al., 2000).

According to this mechanism the number of subunits of FliD requires a

particular number of protofilaments to work. A 2(N)+1=Y equation has been proposed, where N is the subunit number in FliD and Y is the protofilament number (Cho et al., 2017). Thus a pentameric $FliD_{se}$ requires 11 protofilaments (which has been confirmed), hexameric $FliD_{ec}$ and $FliD_{pa}$ require 13 protofilaments and $FliD_{sm}$ needs only 9 protofilaments (Cho et al., 2017). However, as there is no structural information on these bacterial filaments, this equation cannot be proven yet. *C. jejuni* has seven protofilaments as per the first cryo-EM data, thus the number of subunits in $FliD_{cj}$ should be three (Galkin et al., 2008). However, if the lumen of $FliD_{sm}$ is already quite small, a trimeric FliD would not be able to efficiently export all the required extracellular proteins. This model has now largely been disproved by the work reported in this thesis.

Another model based on the *Salmonella* structures involves a similar 5-to-11 model of FliD monomer to protofilament number but differs in the shape of FliD. While the previous model relied on the head domains staying symmetrical and stationary while only the legs moved along with the full oligomer rotation, this model is based on a slight asymmetry in the FliD pentamer observed in the pentamer crystal structure with one pair of monomers at a wider distance from each other, as illustrated in Figure 1.7.4 (Song et al., 2017). The FliD legs fill the indented space coloured in green and labelled with $\alpha, \beta, \gamma, \delta, \varepsilon$. Due to the symmetry mismatch between FliD and the filament the distance between two of the legs would be longer than the rest: in (b) this distance would be between the $\beta - \gamma$ legs. This would shape the FliD into an asymmetrical pentagon. Structural similarity between the D0-D1 leg domains of FliD and flagellin allows γ-leg to preoccupy position 9. Position 9 is where an upcoming flagellin molecule will be incorporated (shown in yellow). Once it is polymerized into the filament, the γ-leg moves to the next position 8 and the entire FliD oligomer is hoisted up. Leg δ that used to occupy position 11 is now replaced by a new flagellin and shifted to position 10. Each FliD leg moves on to a new position for every five repeats of flagellin insertion (b-f) until it returns to its original position in (b) (Song et al., 2017).

None of these models explain how FliD builds the first few flagellins on top of the hook-filament junction and propose elongation mechanisms of already pre-

existing flagellar starter, limited to enteric flagellar filaments. FliD is exported and binds FlgL, which is thought to oligomerise as an 11-mer on top of FlgK and hook, containing the same helical symmetry due to the D0-D1 domains of axial proteins forming highly conserved coiled coils. As FliD oligomerises as a pentamer on top of an 11-mer of FlgL there will be a symmetry mismatch similar to the one that exists between the growing flagellar filament and FliD. Based on this, the first flagellin monomer will travel through the HBB complex and the Hook-filament junction lumen and insert into the cavity opened by symmetry mismatch between FliD and FlgL. This will act as a primer to build the filament further. Presence of the minor flagellin FlaB patch of subunits prior to FlaA major flagellin expression and polymerization complicates the situation significantly. It is yet unknown if $FliD_{cj}$ binds FlgL prior to FlaB polymerization and packs it in a different manner to FlaA, or FlaB polymerises on its own into a short differently packed segment prior to FlaA secretion (Inoue et al., 2018). Until a high enough resolution EM dataset is available capturing the tip of the filament at different stages of growth (bound to FlgL with first flagellin monomers, elongating the filament, at the end of a fully grown filament), the actual elongation mechanism will remain elusive.

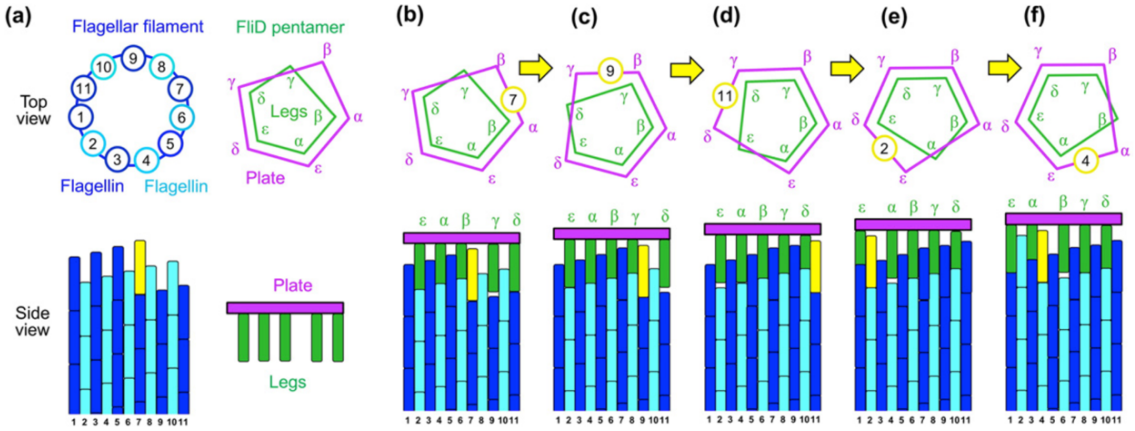


Figure 1.7.4: Model of $FliD_{se}$ mediated flagellar elongation. (a) Schematic diagrams of flagellar filament (left) in blue and cyan which corresponds to the side view below, and $FliD_{se}$ (right) with the plate (D2-D3 domain) in pink and legs in green. (b-f) Sequential cycles of the flagellin insertions and asymmetrical cap shifting. Last flagellin protein added is in yellow. Flagellin positions labelled with numbers as per (a). From (Song et al., 2017).

1.8 Theory and Methodology of Single Particle Cryo-EM

1.8.1 Architecture of the TEM

The first transmission EM microscope (TEM) was built in 1931, and though modernised the design remained essentially the same to this day (Knoll and Ruska, 1932). A basic TEM consists of an electron source, accelerator stack, condenser lenses, specimen holder, objective lens, projector lens and screen/camera detector as illustrated in figure 1.8.1 (Franken et al., 2020). The emission source consists of a metal wire sharpened to a point with a current passing through it, allowing for electrons to escape from their positions at the tip through receiving enough energy from collisions. It needs to be bright, stable, and coherent. While the first two are relatively self-explanatory the coherence requirement is comprised of spatial and temporal coherence. Spatial coherence is a measure of how close the source is to a sharp point. Temporal coherence is a measure of how similar the wavelengths of the escaping electrons are. A perfectly coherent beam is made up of electrons with the same wavelength, phase, and point of origin (Franken et al., 2020). The current at the emission gun is usually low (-300 to -80 kV) and the bulk of the microscope is grounded causing a strong electric field accelerating the electrons down through the accelerator stacks (Cheng, 2015). To prevent excessive electron scattering, the system is at high vacuum (Carroni and Saibil, 2016).

The accelerated electrons pass through lenses made up of coiled metal wire with a current going through them creating an electron field that deflects incoming electrons and can be used to focus them. The lens system consists of two or more condenser lenses, which condense and shape the electron beam from the gun to focus on the sample and control the spot size (C1) and intensity (C2). Next are the objective lenses, which produce a magnified image of the sample and there is usually two of them as well. The diffraction lens is just under the back focal plane so in diffraction mode the image of the diffraction pattern can be viewed. The intermediate and projector lenses amplify the signal from the magnified image to be read using a detector. The lenses suffer from three types of aberration: spherical, chromatic and astigmatism. Spherical aberration is a consequence of the fact that the electron that

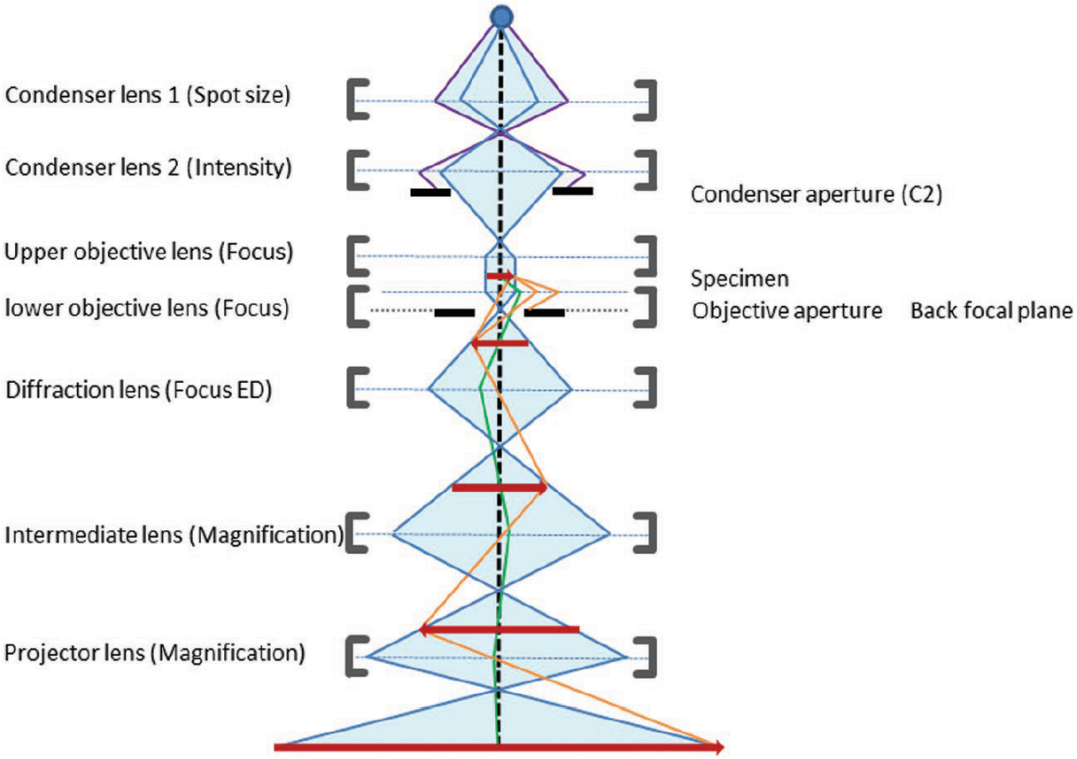


Figure 1.8.1: Schematic of a basic electron microscope. Emission source is shown as a blue dot, the electron beam is shown in blue; grey brackets represent lenses; black bars are apertures; red arrow is the specimen and intermediate images; paths of scattered electrons are drawn in orange and green. From (Franken et al., 2020).

passes through the centre of the magnetic field will be exposed to a weaker force compared to the one closer to the magnetic coils. Chromatic aberration arises when not all electrons passing the lens have the same speed. Astigmatism is the result of a difference in the strength of the lens in x and y directions due to the coil conducting the current unevenly. Apertures, both C2 and the objective, can be inserted or retracted to assist the lens in focusing the electron beam and block electrons scattering too far (Franken et al., 2020).

There are two types of detectors: charge-coupled device (CCD) and direct electron detector (DED). When electrons hit the CCD scintillator, they are scattered and produce photons, which are captured within the fibreoptic and brought down to specific pixels on the CCD layer (Bammes et al., 2011). This used to be the preferred option over film, until the appearance of DED. Electrons hitting the direct detector are converted to charge without the intermediate photon state, which significantly

improves quality and speed of image collection. This discovery contributed significantly to the rapid improvement in EM over the last decade (Kühlbrandt, 2014). The speed of data collection can be used to record movies in combination with electron counting. Processing movies frame by frame allows for the use of lower dose per frame and correction of beam-induced movement (Scheres, 2014).

1.8.2 IMAGE FORMATION AND DIFFRACTION

The basis of how EM images, or micrographs, are taken is rooted in the fact that electrons exist both as particles and as waves. They are optimal for studying biological samples at high magnifications as their low wavelength increases the resolving power. They also interact significantly and specifically with matter through scattering and can be focused, unlike X-rays, allowing for use in imaging mode rather than only diffraction mode. (Nogales and Scheres, 2015).

As primary electrons from the beam pass through the sample, they may not interact with any atoms in the protein and remain unscattered, maintaining their initial energy, frequency, and phase. They may also interact with the sample atoms and lose energy through inelastic scattering or be deflected without losing energy through elastic scattering. High resolution information is present in the elastically scattered electron, while inelastic scattering occurs when energy is transferred to the sample and results in radiation damage (Franken et al., 2020).

TEM image is formed by wave interference. Interference is constructive when waves are in phase leading to increase in signal and amplitude. It is destructive when the waves are out of phase leading to reduction in amplitude as illustrated in Figure 1.8.2. In both schematics the amplitude of the parallel light before the grating is the same. The only difference is that in b) the red wave has a smaller scattering angle, leading to destructive interference and loss of amplitude of the original wave. The resulting image on the diffraction plane is called a diffraction pattern. The relationship between these three parameters can be described in a formula where λ is the beam wavelength, (θ) is the scattering angle relative to the primary electron and d is the spacing between scattering centres (Franken et al., 2020).

$$d \sin(\theta) = n\lambda$$

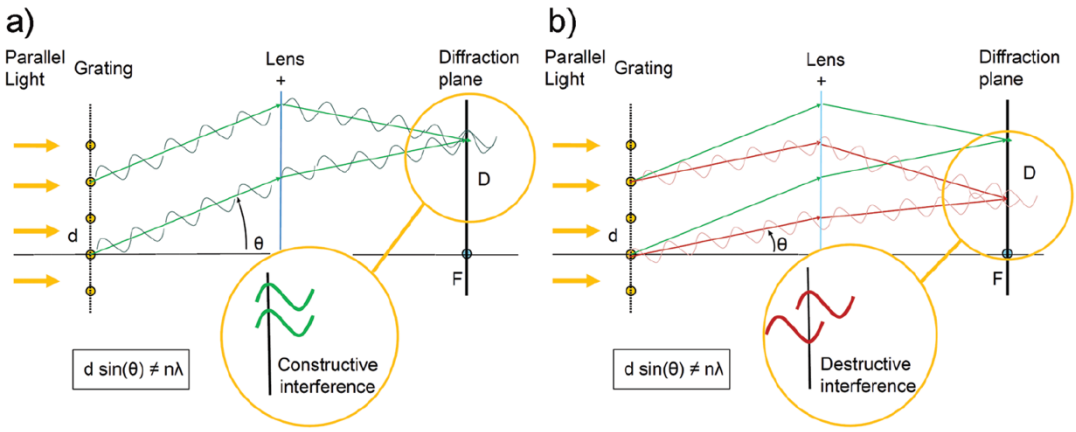


Figure 1.8.2: Illustration of a) constructive and b) destructive interference. Parallel scattered waves interact in the back focal plane of the lens. Amplitudes are either added or subtracted depending on the phase of each wave. D is the magnified reciprocal distance between events of wave interfaces. As smaller details need larger scattering angles and require larger D values, on the back focal plane the smallest details are localized to the edges (Franken et al., 2020).

In a protein sample, there are different values of d with wide centres representing molecular subunits, medium centres alpha helices and beta sheets, and small centres representing individual atoms. The primary electron amplitude and phase is the same and no matter how scattered the parallel waves are, the lens focuses them to meet at the back focal plane of the microscope, where the diffraction pattern can be imaged from (TEM diffraction mode). In this pattern every point carries the information about how many scattering centres were there originally. The drawback of diffraction patterns is that they do not carry phase information, only amplitudes squared (intensities) (Franken et al., 2020).

This is what causes the "phase problem" in X-ray crystallography, as without the phases a reverse Fourier transform (FT) cannot be done without experimentally derived phase data. FT is a mathematical concept of decomposing a complex wave into simplified functions depending on their spatial or temporal frequency, where each wave has a distinct frequency and amplitude. The FT deciphers the interfering scattered rays that just passed through the sample into their respective simplest functions to quantify how much of each original component there is as a function of

spatial frequency (Franken et al., 2020). In EM though, if the waves are allowed to propagate past the diffraction plane to the image plane, they interfere with each other once more to produce the complex wavefront and undergo Fourier synthesis (inverse FT). This recreates the complex wave function and the scattering centre localization on the image plane identical to the sample but at much higher magnification and with the phase information intact (Cheng, 2015). This would be the case in an ideal microscope.

1.8.3 CONTRAST AND CTF

There are two types of contrast in the TEM: scattering and phase. To generate scattering contrast, the electrons that pass through the sample and scatter at large angles are blocked from reaching the image plane by inserting of the objecting lens. Thus, the scattering from a heavy atom will be blocked from reaching the image plane, making the atom look darker in comparison to the rest. This is observed in the case of metals, gold fiducials, etcetera. Biological samples with low atomic number elements result in small angle scattering only, making this method of contrast generation have little effect. Therefore, a focused image has little to no contrast as illustrated in Figure 1.8.3 a-b. Phase contrast is generated by going out of focus. Changing the focus means making the current in the objective lens weaker or stronger thus affecting the scattering of the electrons passing through the electron field. The original unscattered electron will arrive with the same phase and frequency, but the scattered electrons will have their phases delayed. Thus imaging out of focus gives a per-frequency phase shift (Franken et al., 2020).

Imaging by the detector does not physically happen on the image plane. Below the image plane are the intermediate and projector lens followed by the detector. There are multiple possible image planes, but they are all conjugate. So, changing the focus of the lens theoretically shifts the image plane up or down and whatever state hits the image plane is detected. When the focus is changed, the phases of the scattered electrons increase with increasing frequency, creating frequency-dependent interference and contrast, as illustrated in figure 1.8.3 c-e. Images can be either defocused or overfocused. A defocused image has the scattered electrons meet below

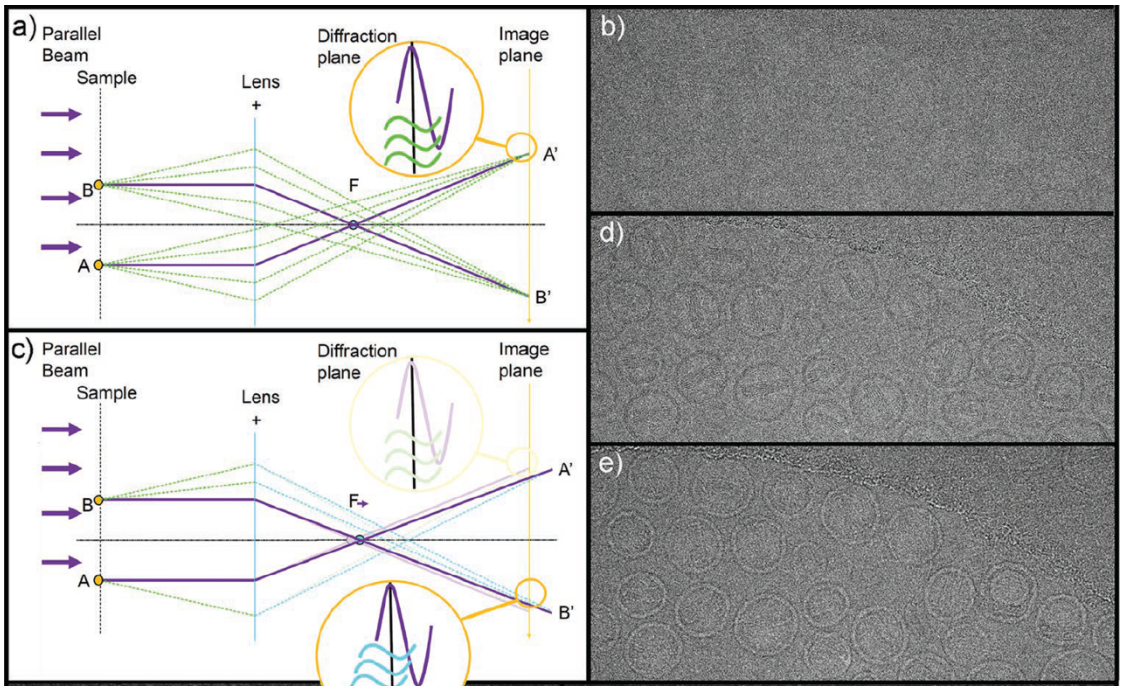


Figure 1.8.3: Illustration of phase contrast in the electron microscope. a) small-scattering angles are too small to be removed by the aperture without resolution lost and reconstitute on the image plane in focus, b) in focus image does not have contrast as scattered information recombines into a complex wave at the image plane, c) Out of focus scattered waves have delayed phases and reconstitute with the primary electron wave below or above the image plane creating defocus or overfocus. d) defocused and e) overfocused images. From (Franken et al., 2020).

the recorded image plane (1.8.3. d) and an overfocused image has the image reconstituted above the recorded plane with electrons continuing past it onto the recorded conjugate image plane (1.8.3. e) (Franken et al., 2020).

FT in other words is a mathematical conversion of the diffraction pattern from an electron microscopy image into a frequency power spectrum of the image. The difference between diffraction and FT is the second round of wave interference form the image providing the phases in the image and (phase) contrast transfer function (CTF). This function links the total phase shift due to spherical aberration and defocus to a respective frequency. In other words, it is a measure of how much contrast is transferred to the image as a function of spatial frequency. Each spatial frequency shows how much of that Fourier component is transferred into the final image. Some are transferred fully (designated value of 1), some have not been transferred

(designated value of 0) some have a Fourier component fully present but contributing in the opposite direction, or have reversed contrast (designated value -1) (Carroni and Saibil, 2016).

CTF of a perfect microscope would show all scattered wavelengths at the same amplitude: at full contrast of 1. But due to enveloping it doesn't as CTF is damped at high spatial frequencies. This is called CTF enveloping, which occurs from spatial and temporal coherence of the electron beam, specimen motion and modulation transfer function (MTF) of the image recording device. Small defocus values maximise the enveloping at high resolution but minimises the CTF at low resolution frequencies. The inverse is also true (Cheng, 2015). An image showing an example of a CTF plotted against 1/amplitude (d) overlayed on top of a corresponding Fourier Transform is illustrated in Figure 1.8.4.

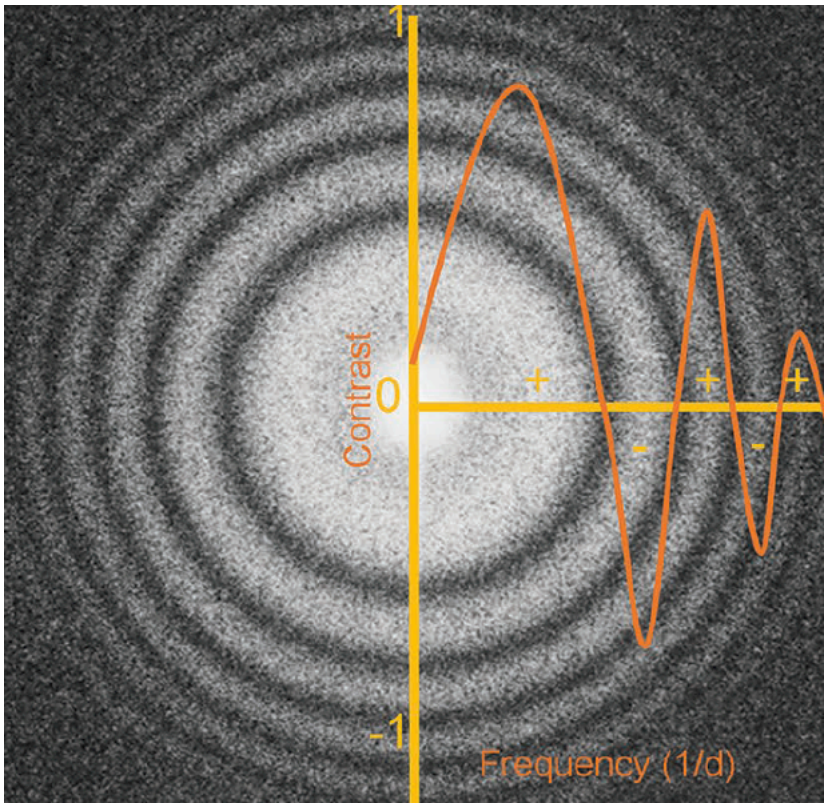


Figure 1.8.4: FT of an image showing signal intensity (contrast) sorted by frequency of the image, demonstrating the effect of the CTF shown in orange and the enveloping effect. From (Franken et al., 2020).

1.8.4 Cryo-EM Sample preparation

Until recently, the major limitation of EM was the need for sample staining and dehydration to provide contrast and withstand the vacuum inside the column. The staining often resulted in elution of molecular constituents and dehydration caused shrinkage artifact and loss of cell turgidity. While this is still used as a negative stain method to produce high-contrast images of samples, it is limited in resolution to about 15 Å due to the grain size of the uranyl formate/ uranyl acetate stain (Nogales and Scheres, 2015). This sample preparation limitation was overcome by Dubochet and colleagues in 1987 based on earlier work by Taylor and Glaeser in 1979 through a method of rapid freezing to produce vitrified samples (Dubochet et al., 1988; Taylor and Glaeser, 1974). Vitrification occurs when a thin layer of aqueous solution is solidified by rapid cooling preventing ice crystallization. Vitrified samples are stable in the column vacuum and low temperature slows down radiation damage from the electron beam, which has been a major drawback in working with biological samples (Baker and Rubinstein, 2010). Now this technique of using vitrified samples is called Cryoelectron microscopy (Cryo-EM).

For standard Cryo-EM sample preparation a small aliquot of sample in solution is applied to an electron microscopy grid, usually copper with thin carbon coating, blotted to a thin layer and frozen in liquid ethane at -180 °C. Grids are about 3 mm across with carbon mesh containing holes to trap sample particles in thin vitreous ice (Cheng, 2015). Thicker samples such as cells or tissues can be used, but they would need to be prepared either by chemical fixation or high pressure freezing followed by sectioning or ion beam milling prior to being imaged (Carroni and Saibil, 2016).

1.8.5 SPA data collection and processing

Single particle analysis (SPA) is most often used for soluble purified homogenous macromolecular complexes. Thousands of identical particles are imaged and averaged using multiple 2D projections containing many particles with different angular orientations. Unlike crystallography, where the protein crystal must have a

highly ordered protein lattice for efficient data collection, some heterogeneity in the SPA samples can be allowed. This is because minor heterogeneity stemming from flexible complexes or multiple conformations can be reduced through processing stages (Thompson et al., 2016).

As electrons are highly damaging to the biological sample, even the cryo-fixed one, the exposure time and amount of electrons that can be used to obtain high resolution information prior to the sample degrading is limited. Most commonly the amount of electrons per image, in other words dose, is limited to 20 – 40 electrons/ $Å^2$ (Nogales and Scheres, 2015). This creates a problem, as lower dose imaging is better for the sample but creates less contrast. Recent developments in direct electron detectors resulted in the capability of the camera to take movies, a collection of micrographs, or frames, taken in the same area with a low dose, such as 1 e/$Å^2$, thus spreading the total damaging dose of 50 or above over as many frames. This allows for two major corrections that can be accounted for during data processing. First one is motion correction, where beam-induced movement can be corrected by registering identical features in the subframes to produce a motion corrected image (Scheres, 2014). Beam induced sample motion can be decomposed into two components: uniform whole-frame motion and nonuniform local motions varying across the image (Zheng et al., 2017). These motions are thought to be caused by the reaction of the specimen to the high energy electron beam, resulting in the build-up of charge and radioactive decay of the sample and vitrified medium. The charge build-up causes weak electron beam deflection and blurs the image through interference with the scattering of electrons. Radiolysis leads to a build-up of internal pressure and "bubbling" of the molecules (Brilot et al., 2012).

While aligning fractions for the micrograph improves contrast slightly and prevents motion-caused blurring, the problem of sample decay persists. To rectify this issue dose weighting can be used, which is using the weighted average of aligned frames with elimination of initial 2-3 and final frames. This is performed as initial frames have the least radiation damage and contain high frequency information but suffer from beam-induced movement the most. By frame 3 the beam-induced movement Is still present but not as harsh and the frames still contain high resolution

information. By the last frame, the sample has been exposed to the full dose of electrons and the frames have greater loss of high spatial frequencies (Carroni and Saibil, 2016).

When collecting a dataset, the micrographs should have been taken at varying defocus values. High resolution micrographs would be taken closer to focus and low-resolution data further from focus to cover the full range of frequencies and make up for the ones lost in each image CTF. After aligning and correcting for electron damage, defocus is determined by fitting a CTF model to the observed oscillations of the image (Rohou and Grigorieff, 2015).

While the CTF is a representation of intensities as a function of spatial frequency in Fourier space, point spread function (PSF) is its equivalent in real space. Meaning that every density in the original sample is degraded by the same PSF in the image (Franken et al., 2020). Thus, we can determine the relationship between image and PSF as per the following equation:

$$I = O \times PSF$$

Where (I) is the image, (O) is the object and PSF is point spread function. The above equation is true in real space. When this is represented in reciprocal space:

$$F\{I\} = F\{O \times PSF\} = F\{O\} \times F\{PSF\}$$

As it was previously established, the relationship between CTF and PSF is as such that one could be transformed into the other through FT.

$$F\{I\} = F\{O\} \times CTF$$

To determine the image, an inverse FT needs to be performed as follows:

$$I = F^{-1}[F\{O\} \times CTF]$$

The above equation shows that FT of the image could be divided in the reciprocal space by the CTF and inverse FT performed so that a better version of the object can be portrayed.

To correct for the CTF, an FT of the image needs to be taken resulting in the FT (I), which is in turn divided by the thresholded CTF in reciprocal space resulting in a FT (I) that is CTF corrected in a process called Wiener filtering. A final FT is performed to convert the power spectrum to a real space image, which after CTF correction should be identical to the initial object on the specimen plane (Frank, 2006). This process is summarised in Figure 1.8.5

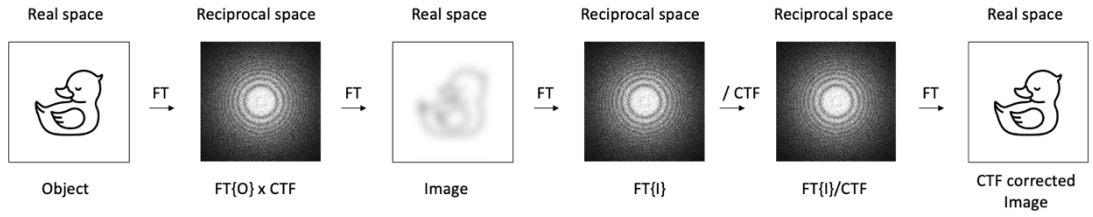


Figure 1.8.5: Process of CTF correction. As the image is a factor of the signal from the object degraded by the PSF, it appears blurry when a reverse FT is taken from the object FT. To correct for it, an FT of the degraded image is divided by the CTF in reciprocal space. The image recovered from that FT will be CTF corrected and resemble the object pre PSF degradation.

As the above process requires division by the CTF, this becomes an issue when the value of CTF at a particular frequency is 0. Instead of dividing by value of CTF at this step, the CTF is thresholded prior to division through all contrast being made unidirectional, or in other words the negative amplitudes are multiplied by -1 or 180 phases to get positive components. However, there is no recovery of the information with no contrast, which is why a range of defocus must be present across the dataset to account for the missing frequencies (Carroni and Saibil, 2016)..

After an image has been motion corrected, dose weighted and CTF corrected, the desired particles are ready to be picked. This could be done manually or computationally using a reference. Particles are picked, centered, and sorted into similar classes to generate 2D class averages through a comparison of all images using multivariate statistical analysis. These components are used to classify images according to their features. Resulting class averages can serve as reference images and refine the dataset (Carroni and Saibil, 2016). More recent approach is Bayesian analysis, in which particles are not assigned to a class but given probability distribution.

Relion, a software package containing the algorithms required for cryo-EM image processing, includes a model for the noise, automates filtering and weighting judgements (Scheres and Chen, 2012).

Once suitable 2D classes are obtained, preferably with different orientations of the particle in vitreous ice, they can be used to determine an *ab initio* 3D structure. This is done due to projection-slice-theorem, which states that FT of a 2D projection is

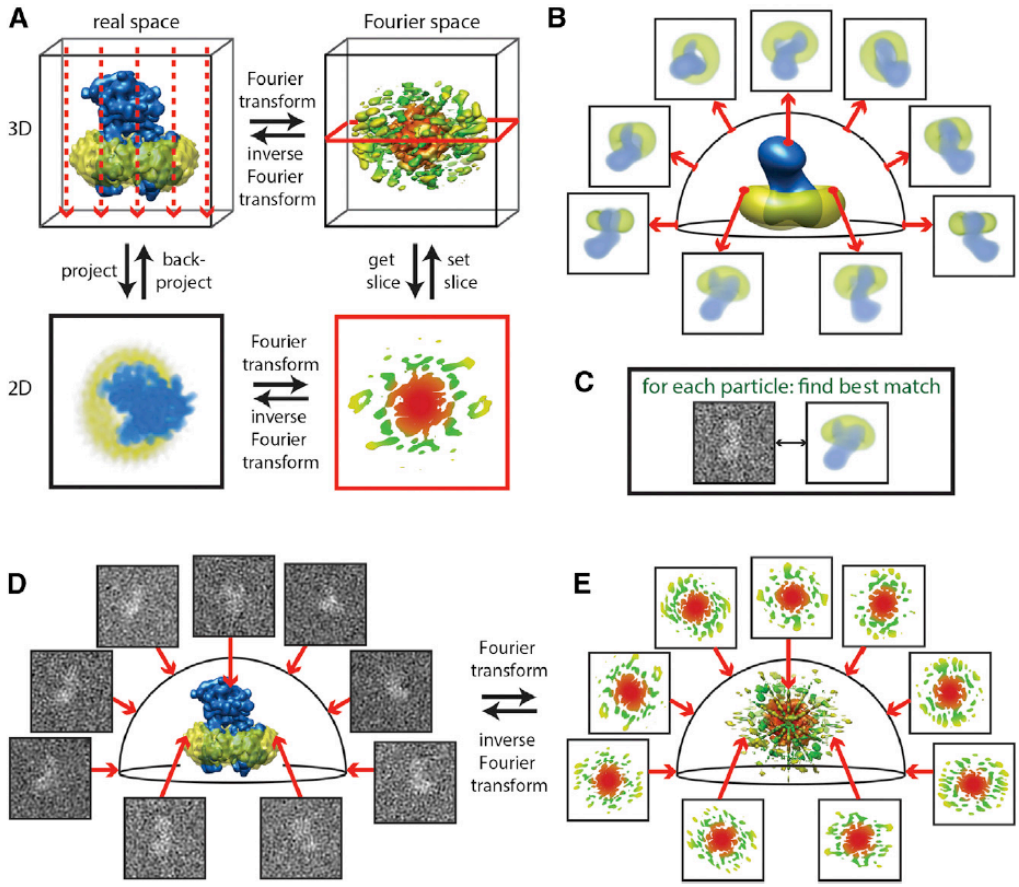


Figure 1.8.6: Determination of 3D structures from 2D projections. A) The projection-slice-theorem stating that 2D projection of a 3D volume in real space is equivalent to a 2D slice of the 3D FT in reciprocal space. B) Projection matching process used to combine experimental 2D projections into a 3D reconstruction. To determine relative orientations of the experimental projections, projections from a reference 3D object are calculated in all directions. C) experimental projections are compared with reference projections to find the best match. D) the projection-slice-theorem implies that 3D reconstruction (D) can be calculated by positioning many 2D FT slices into the 3D transform (E) and calculating inverse FT. Iterative B-E gradually improves orientation and resolution of the reconstruction. From (Nogales and Scheres, 2015).

a central slice through the 3D FT of the underlying structure, illustrated in Figure 1.8.6 A. Therefore, if one knows the directions of 2D projections, one can position the slices within the transform and calculate the original structure by computing the inverse FT (Nogales and Scheres, 2015). In theory with only three projections the relative orientations of the particles could be established. This process is unreliable in noisy images, but presence of symmetry in the particle greatly facilitates this. If there is a known homologous structure, then projection matching can be done where reprojections of the known model are used as references for alignment of the dataset by cross-correlation. Each particle is assigned the best matching orientation and a first 3D map can be reconstructed, refined through multiple iteration (Carroni and Saibil, 2016). The simplified pathway for Cryo-EM data processing is illustrated in Figure 1.8.7.

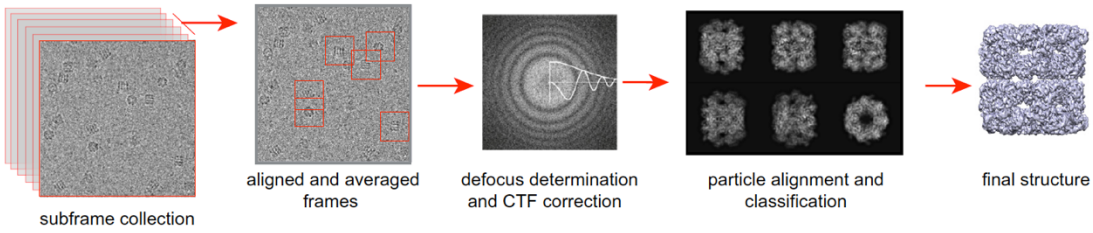


Figure 1.8.7: Illustration of an SPA reconstruction pathway. Left to right: Movie frames are collected, dose weighted and aligned via motion correction. Defocus is determined by fitting a CTF curve to the FT of the image and CTF correction is performed to remove the PSF effect on the outcome image. After particle picking, alignment, classification and cleaning of the dataset, particles are assigned orientation by projection matching to the initial model. Iterative repeats of refinement improve particle orientation and result in their projection onto a final 3D map. From (Carroni and Saibil, 2016).

In some cases when there is no reference model, or there aren't enough different particle views due to preferred orientation of the sample in vitreous ice, there could be a problem with angle assignment. This can be overcome through random conical tilt (RCT) and orthogonal tilt reconstruction (OTR) (Carroni and Saibil, 2016). In RCT, pairs of images are collected through taking an image with stage tilting at an angle between 45° and 60° while the other image is untilted. Since the tilt angle has been set experimentally, a reliable model can be obtained. However, for particles with preferred orientation this method still results in a missing cone of data (Radermacher

et al., 1987). OTR relies on having good distribution of particles in the vitreous ice to fill in the missing data from RCT using -45°, 45° tilt pairs (Rosenthal, 2015).

The final 3D reconstruction is a Coulomb potential density map interpreted the same way as electron density maps obtained from X-ray crystallography (Cheng, 2015). This means that results of the years of development in X-ray crystallography model building software could be utilised in EM as well. A *de novo* model can be built in Coot, the fit of the model to the map and model parameters can be refined through PHENIX and so on (Adams et al., 2010; Emsley et al., 2010). Resolution of EM 3D volumes is estimated by calculating Fourier Schell Correlation (FSC) between two independently refined halves of the dataset. FSC is a measure of reproducibility as a function of spatial frequency and the resolution value is determined through an agreed upon threshold (Scheres and Chen, 2012).

1.9 Aims and Objectives

The main aim of this thesis is to investigate the role of FliD capping protein in bacterial motility and flagellar filament elongation through determining its structure using SPA Cryo-EM. While previous attempts at characterizing FliD from different bacterial species focused on the use of X-ray crystallography, when investigating this particular protein there was a significant drawback (Cho et al., 2017, 2019b, 2019a; Postel et al., 2016; Song et al., 2017). The N- and C-terminal regions of FliD contain many hydrophobic residues, which prevent successful crystallization and so must be removed. These termini make up the D0 domain, observed in many flagellar axial proteins and thought to be essential for correct folding and integration into the flagellum. While X-ray crystallography provides high resolution data capable of reaching sub-atomic resolution, in the case of FliD the regions of interest cannot be observed.

An alternative structural biophysical technique would be NMR, where atomic nuclei are charged and therefore have different resonance frequencies. Structural

information about proteins is obtained from comparison of the resonance frequencies of interacting adjacent atoms. While this is a high precision technique, the limitation is that unless the protein of study is smaller than 40-50 kDa in solution the NMR spectrum is very complicated and difficult to interpret. This technique also requires relatively large amounts of pure samples to achieve acceptable signal to noise ratio.

The most optimal technique to study FliD appears to be SPA cryo-EM. As observed in the cryo-EM structure of *S. enterica* FliD, the untruncated D0 domains are readily visible and do not interfere with FliD oligomerization. Due to the hydrophobic residues stabilizing through interacting with each other in solution, $FliD_{se}$ oligomerises as a decamer (Maki-Yonekura et al., 2003). Although the resolution of this structure is 26 Å, with the current developments in cryo-EM it is possible to reach resolutions similar to that of X-ray crystallography derived structures and characterise the D0 domain.

Another advantage of using SPA cryo-EM, is that unlike NMR and X-ray crystallography, which require a pure homogenous sample at high concentration, SPA image analysis can deal with a certain level of heterogeneity and data collection can be done on a small amount of sample in buffer solution (about 0.1 mg). In addition to these significant advantages in sample preparation, cryo-EM does not have a "phase problem": the reciprocal space diffraction pattern of the sample carries the intensities information but not the phases, which must be calculated through other means. In EM, micrographs are real space images with both amplitude and phase data, although slightly less accurate due to the PSF but can be corrected for it (Cheng, 2015).

X-ray crystallography structures of FliD appear to have a different oligomerization state depending on the species, which has sparked a debate about the functional state of FliD when bound to the filament. SPA cryo-EM often requires purification of a recombinant protein to achieve the minimum required sample purity and concentration similar to crystallography. However, in cryo-EM the protein is vitrified in its surrounding medium, freezing it in its natural state in solution, and not forced to form a crystal lattice. There is a strong argument that the oligomeric state observed via cryo-EM is closer to the native state of the protein than the connections

it makes in a crystal.

To investigate the role of FliD in motility and flagella assembly, this thesis aims to use molecular cloning techniques to determine whether any of the highly conserved residues in the uncharacterised D0 domain significantly affect motility through potentially altering the packing of flagellins in the flagellar filament. *C. jejuni* is the organism of choice to study FliD function, partially due to having limited number of flagella at polar ends of the cell and potentially exhibiting a stronger effect of the mutations on the phenotype of FliD, partially due to the now disproven observation that *C. jejuni* flagellar filament consists of 7 protofilaments, contrary to the 11 of the enteric filaments. Current models of flagellar filament elongation, while proposing different conformational changes in FliD on the growing filament end, are based on the asymmetry in a pentamer cap interacting with the 11 protofilament. To observe if the same asymmetry is present in the *C. jejuni* flagellum, this work aims to solve the structure of a native flagellar filament and propose a mechanism of elongation based on the cryo-EM structural information.

2.METHODS

2.1 GENERAL REAGENTS

All chemicals were purchased from Sigma Aldrich unless otherwise stated. All media were sterilized by autoclaving and all buffers filtered and degassed unless stated otherwise.

2.1.1 GROWTH MEDIA

All prepared with MilliQ water filtered from a Neptune (Purite) System.

Media	**Ingredients**
Luria-Bertani (LB) medium	25 g/l LB-broth, high salt Fluka Analytical
LB agar	LB broth supplemented with 1.5% (w/v) Difco Bacto agar
Mueller Hinton (MH) broth	21 g/l MH-broth
Columbia blood agar	39 g/l Columbia-broth with 5% Horse blood
BHI broth	37 g/l BHI-broth
Motility agar plates with TTC	37 g/l BHI-broth, 37 g/l Bacto agar, 0.005% triphenyltetrazolium chloride (TTC)
BHI glycerol buffer	37 g/l BHI-broth and 15% v/v glycerol
BTS buffer	28 g Brusella broth base, 10 g Tryptone, 2.1 g Serine
ZYM-5052 media	ZY media: 1% w/v Tryptone, 0.5% w/v Yeast 50x M stock: 1.25 M Na_2HPO_4, 1.25 M KH_2PO_4, 0.25 M Na_2SO_4, 2.5 M NH_4Cl 50x 5052 stock: 25% w/v glycerol, 2.5% w/v glucose, 10% w/v lactose monohydrate 1 M $MgSO_4$ stock (all autoclaved separately and then combined)

2.1.2 Core Bacterial Strains

Strain	**Genotype/ Phenotype**	**Source/note**
BL21-CodonPlus(DE3) – RIL *E. coli* strain	Cam^r extra tRNA *argU* (AGA, AGG), *ileY* (AUA), *leuW* (CUA) lacUV5	Agilent Technologies
Subcloning Efficiency™ DH5α™ Competent Cells	*argU* (AGA, AGG), *ileY* (AUA), *leuW* (CUA)	ThermoFisher
Campylobacter jejuni subsp. Jejuni 81116 strain	WT	Professor Dave Kelly's Lab at The Univesity of Sheffield MBB
Campylobacter jejuni subsp. Jejuni 11168 strain	WT	Professor Dave Kelly's Lab at The Univesity of Sheffield MBB
Campylobacter jejuni subsp. Jejuni 81-176 strain	WT	Professor Dave Kelly's Lab at The Univesity of Sheffield MBB

2.1.3 Gels Composition

Gel buffer name	**Ingredients**	**Comments**
1% Agarose gel buffer	1 g Agarose 100 ml MilliQ	Melted in the microwave oven and poured
SDS-PAGE loading buffer	2 ml 1.5 M TRIS pH 8.8 2.4 ml Acrylamide 40% Severn Biotech 80 µl Ammonium Persulfate (APS) 10% 80 µl Sodium Dodecyl sulphate (SDS) 10% 8 µl Tetramethylethylenediamine (TEMED) 3.4 ml MilliQ	In 8 ml of buffer Enough for 1 gel in 1.5 mm or 1.0 mm cast
SDS-PAGE stacking buffer	1.25 ml 0.5 M TRIS pH 6.8 0.5 ml Acrylamide 40% Severn Biotech 50 µl APS 10% 50 µl SDS 10% 5 µl TEMED 3.1 ml MilliQ	In 5 ml of buffer enough for 1 gel in 1.5 mm or 1.0 mm cast

2.1.4 BUFFERS COMPOSITION

Buffer name	**Ingredients**
HEPES buffer	50 mM HEPES, 150 mM NaCl, pH 7
1M Imidazole buffer	50 mM HEPES, 150 mM NaCl, 1M Imidazole, pH 7
C. jejuni Wash buffer	15% v/v glycerol, 9% w/v sucrose
CaCl buffer	0.1 M CaCl
CaCl glycerol buffer	0.1 M CaCl, 15% v/v Glycerol
Colloidal fixation solution	40% v/v ethanol, 10% v/v Acetic Acid
Colloidal staining solution	70 mg Coomassie brilliant blue G-250 in 1.5 ml HCl in 1 litre
Coomassie staining solution	0.1% Coomassie R250, 10% v/v acetic acid, 40% v/v methanol
Coomassie destaining solution	20% v/v methanol, 10% v/v Acetic Acid.
Tris buffer	50 mM Tris, 100 mM NaCl, pH 7.4
Phosphate-buffered Saline (PBS) buffer	137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4 , 1.8 mM KH_2PO_4
Low pH no salt conformation buffers	50 mM HEPES pH 4, 50 mM Tris pH 5
Medium pH no salt conformation buffers	50 mM HEPES pH 6, 50 mM HEPES pH 7, 50 mM Tris pH 8
High pH no salt conformation buffers	50 mM CAPS pH 9, 50 mM CAPS pH 10, 50 mM CAPS pH 11.
Low salt conformation buffers	50 mM HEPES 150 mM NaCl pH 6, 50 mM HEPES 150 mM NaCl pH 7, 50 mM HEPES 150 mM NaCl pH 8
High salt conformation buffers	50 mM HEPES 500 mM NaCl pH 7, 50 mM HEPES 1 M NaCl pH 7
TSE lysis buffer	100 mM TriHCl pH 8, 500 mM sucrose, 1 mM EDTA
$NaPO_4$ buffer	20 mM $NaPO_4$ pH 7.4

Buffer A1	50 mM HEPES pH 6.98
Buffer B2	50 mM HEPES 1M NaCl pH 7.02
Buffer A2	20 mM Tris pH 7.5
Buffer B2	20 mM Tris 1M NaCl pH 7.5

2.1.5 ANTIBIOTICS

Antibiotic	**Stock concentration**
Kanamycin	50 mg/ml in MilliQ
Chloramphenicol	25 mg/ml in ethanol
Vancomycin	10 mg/ml in MilliQ
Amphotericin ß	10 mg/ml in MilliQ
Carbenicillin	100 mg/ml in MilliQ

2.1.6 COMPETENT CELLS

E. coli BL21 and DH5α cells were made competent through growing a single colony overnight culture at 37 °C, using 1 ml to inoculate 100 ml of LB broth in 250 ml conical flask and growing for 1.5-3 hours until mid-log phase at 37 °C. Cells were collected by centrifugation at 6000 rotations per minute (rpm) for 3 minutes at 4°C. Cells were resuspended in 10 ml CaCl buffer and incubated on ice for 20 minutes. Cells were centrifuged using the same parameters again and this time resuspended in 5 ml CaCl glycerol buffer and frozen in 100 µl aliquots at -80°C.

C. jejuni 81116 cells were made competent through being harvested from a 1-day old Columbia blood agar plate, grown in a microaerobic cabinet at 42 ^{o}C with a controlled atmosphere of 10% v/v O_2, 5% v/v CO_2 and 85% v/v N_2, into 800 µl of *C. jejuni* wash buffer and centrifuged at 14000 rpm at 4 ^{o}C for 5 minutes. Cells were resuspended in 800 µl of wash buffer, centrifuged at the same parameters and this

wash repeated 3 more times. Pellet was resuspended in 300 µl and split into 3x 100 µl aliquots in sterile, chilled 1.5 ml aliquots and frozen at -80°C.

2.2 Protein Expression and Purification

2.2.1 Bioinformatics

FliD sequences from different species used in this work were sourced from EMBL-EBI database and were aligned using T-Coffee server and visualized with ESPript 3.0 (Notredame et al., 2000; Robert and Gouet, 2014). The same was done for the flagellin sequences. The co-evolution analysis between $FliD_{cj}$ and *C. jejuni* flagellin (FlaAcj) was performed using RaptorX Complex Contact prediction server (Zeng et al., 2018). Protein parameters were determined using ProtParam ExPASy tool (Gasteiger et al., 2005).

2.2.2 Plasmid Construction and Cloning in E. coli

Campylobacter jejuni 81116 strain *fliD* gene *C8J_0509* was codon optimised and inserted into pET28a(+) plasmid through cloning sites *Ndel/EcoRI* by gene synthesis courtesy of Bio Basic Inc. The resulting construct contained an N-terminal His-tag, Thrombin cleavage site and kanamycin resistance cassette as illustrated in Figures 2.2.1 and 2.2.2. This plasmid design was repeated for $FliC_{cj}$ (NCBI PRK13589), $FliD_{pa}$ (EMBL-EBI PF07195), $FliD_{sm}$ (NCBI PRK08032), $FliD_{ec}$ (NCBI PRK08032), $FliD_{hp}$ (NCBI PRK08453) and $FliD_{cj}$noD4 domain constructs. The full codon optimised sequences are listed in Appendix 1.

Transformation of plasmids into *E. coli* consisted of addition of 1 µl of 100 ng/µl DNA to 100 µl thawing cells, incubating on ice for 30 minutes and subjecting to heat shock in a 42 °C water bath for 45 seconds. The cells were then incubated for 1 hour in 1 ml pre-warmed LB at 37 °C and plated on respective antibiotic selective plates.

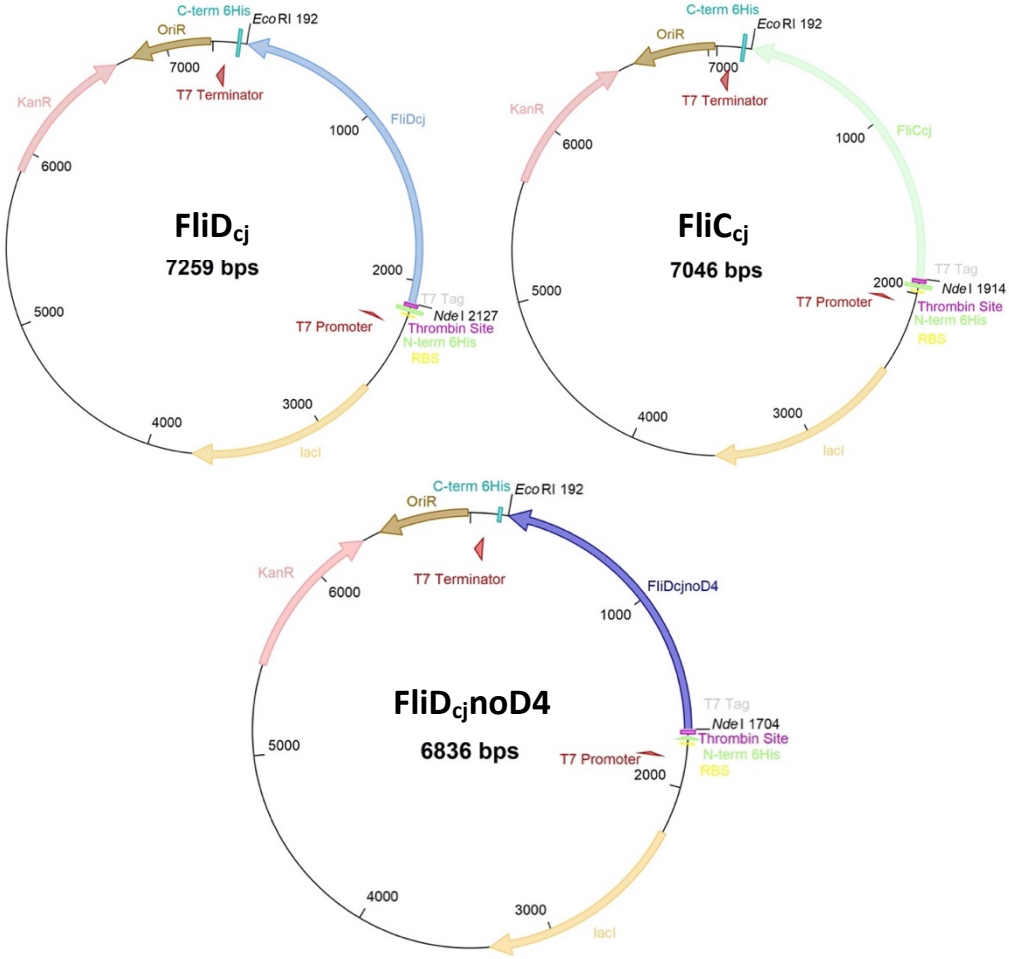


Figure 2.2.1: Constructed plasmids with *C. jejuni* FliD, FliC and FliD construct without D4 domain. Made via Clone Manager 9.

$FliD_{cj}$ domain constructs D2-D3 and D1-D4 were designed using QuickChange Primer design service from Agilent and are listed in Appendix 2. Primers in powdered desalted form were reconstituted to 10 µM concentration in MilliQ water. Polymerase Chain Reaction (PCR) samples were prepared by first making up a master mix, which had per 50 µl reaction: 5 µl PfuUltra II reaction buffer, 3.5 µl Dimethyl sulfoxide (DMSO), 1.25 µl Deoxynucleoside triphosphate (dNTP) solution, 1 µl PfuUltra II Fusion HS DNA Polymerase and 0.25 µl of 100 ng/µl template DNA ($FliD_{cj}$ plasmid). Master mix (total of 11 µl) was added to 36 µl MilliQ and 1.5 µl of respective forward and reverse primers and PCR programme initiated using a PCR thermocycler as detailed in Table 2.2.1.

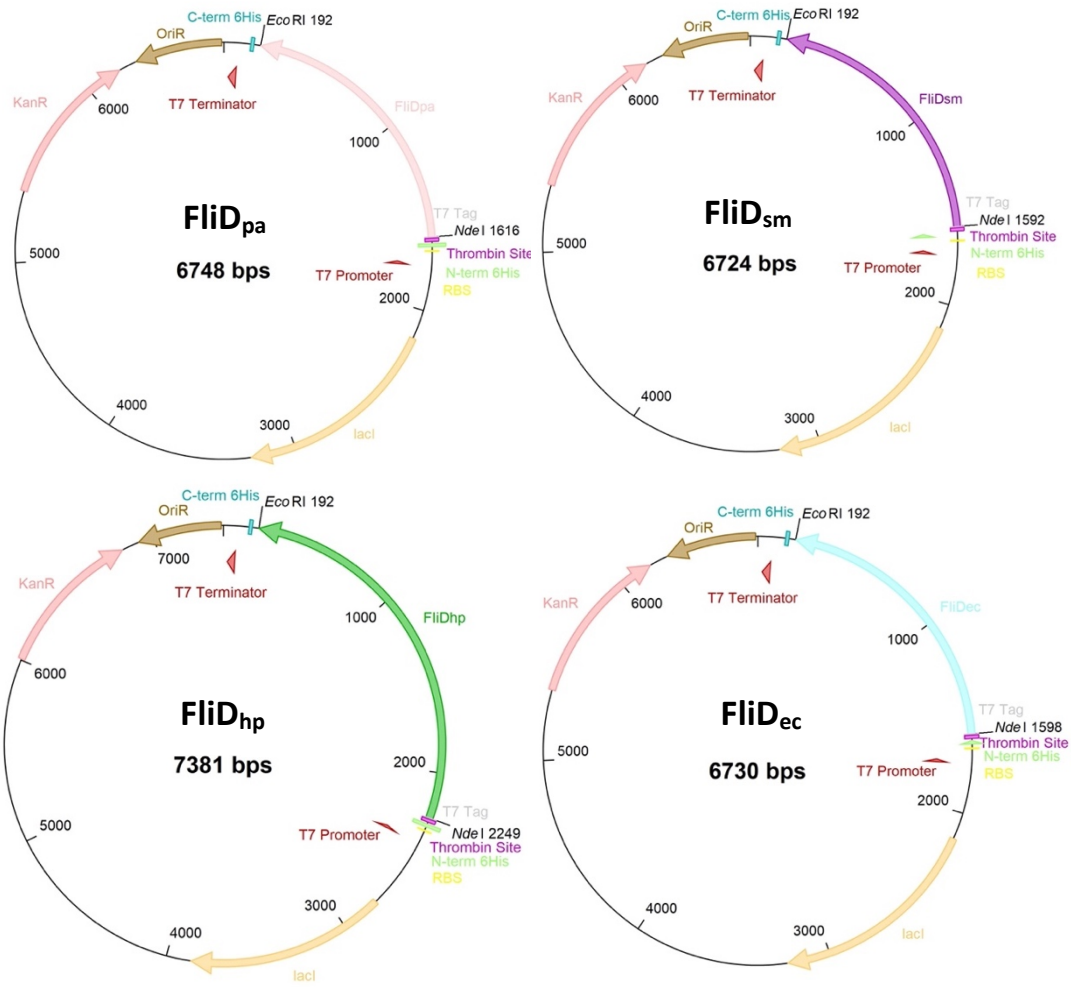


Figure 2.2.2: Constructed plasmids with different species FliD capping protein insert. Made via Clone Manager 9.

Table 2.2.1: PCR thermocycler programme settings.

Step	Cycles	Time	T (°C)
Initial denaturation	1	2 min	95
Denaturation Annealing Extension	18	30 s 1 min 12 min	95 55 68
Final Extension	1	10 min	68
Hold	1	∞	4

1 µl of *DpnI* restriction enzyme was added to digest methylated original DNA and samples incubated for 8 hours at 37 °C. QIAquick PCR purification kit (QIAGEN) was used to extract DNA per protocol provided. The DNA was eluted in 10 µl MilliQ and transformed into DH5α competent cells through addition of 5 µl of resulting DNA to 100 µl thawing cells followed by the standard transformation protocol. Separate colonies were selected from plates, grown overnight in 5 ml LB with antibiotic and plasmids extracted using QIAprep Spin Miniprep Kit (QIAGEN) standard protocol with final volume of 20 µl.

2.2.3 Protein Expression and Purification

Recombinant proteins were expressed in *E. coli* BL21-CodonPlus(DE3)-RIL cells and pLysS cells containing the corresponding plasmids and grown in Kanamycin and Chloramphenicol containing media. A standard Isopropyl β-D-1-thiogalactpyranoside (IPTG) protocol was used for $FliD_{cj}$, $FliC_{cj}$ and $FliD_{cj}D1D4$, where transformants were grown in LB medium with antibiotics at 37 °C until late exponential phase (Optical density $(OD)_{600}$ ~0.6). Overexpression was achieved through induction by 1 mM IPTG overnight at 20 °C. For $FliD_{pa}$, $FliD_{cj}noD4$, $FliD_{hp}$, FliDec and $FliD_{sm}$ expression was auto-induced in ZYM-5052 media using BL21 overnight culture that was added to the premixed media and left to overexpress at 20 °C overnight (Studier, 2005).

For all proteins, cells were collected by centrifugation, resuspended in 50 mM HEPES 150 mM NaCl pH 7 and sonicated. The lysate was centrifuged at 14,000 g at 4 °C for 45 min. The supernatants (except $FliC_{cj}$) were applied onto a 5 ml HisPure™ Ni-NTA resin (ThermoScientific) gravity-based column equilibrated with 50 mM HEPES 150 mM NaCl pH 7 and eluted using a linear 20–500 mM Imidazole gradient in the 500 mM 10 ml fractions (for $FliD_{cj}$) and 5 ml fractions for the rest. Fractions containing proteins were pooled, incubated with Thrombin overnight at 4 °C and concentrated in 10 kDa Sartorius concentrator. The protein was eluted using Amersham Biosciences Akta Fast protein liquid chromatography (FPLC) machine with gel filtration column HiLoad Superdex 200pg 16/600 column (GE Healthcare) equilibrated with 50 mM HEPES 150 mM NaCl pH 7 with 4 ml fraction collection at 0.5 megapascal (Mpa)

pressure and 0.5 ml/min flow unless otherwise specified. 12% SDS-polyacrylamide gel electrophoresis (PAGE) gels were used to analyse the protein purification results and were prepared and assembled as indicated in 2.1.3. They were ran using the standard protocol, stained using either Colloidal (fixation then staining) or Coomassie (staining then destaining) method and imaged using a gel imager.

FliD oligomerization in different buffer conditions was tested by resuspending $FliD_{cj}$, $FliD_{pa}$ and $FliD_{hp}$ in conformation buffers (see 2.1.4) and negative stain grids were made as per the standard protocol (see 2.3.1) within 4 hours of dilution to prevent re-equilibration.

2.2.4 FLAGELLIN PURIFICATION ATTEMPTS

C. jejuni strains 11168 and 81-176 were grown overnight on Columbia blood agar, a pellet of colonies from each was collected and resuspended in 30 ml BTS buffer with Amphotericin ß and Vancomycin and left to grow overnight in a microaerobic cabinet at 42 ^{o}C with a controlled atmosphere of 10% v/v O_2, 5% v/v CO_2 and 85% v/v N_2. OD_{600} was measured and cultures diluted to 0.1, 500 µl of which was used to inoculate 50 ml of broth for each strain and left to incubate overnight under the same conditions. This was repeated for another set of 50 ml flasks for 8-hour incubation. OD_{600} 0.1 of these cultures was used to inoculate 300 ml cultures and left to grow overnight. These 300 ml cultures were then centrifuged at 6000 g for 20 minutes at 20 ^{o}C, supernatant removed, and pellets resuspended in 1 ml of Tris buffer. Cells were then passed 20 times through modified Avanti Polar Lipids apparatus, which consisted of two 1ml glass Hamilton 22 gage Syringes connected by a 0.7 mm tube as shown in Figure 2.2.3. The syringe was first washed with buffer and then loaded with cells. The resulting solution was centrifuged at 15000 g for 20 minutes, supernatants isolated and concentrated through a 100K Sartorius filter at 3000 g for 5 minutes. The resulting solution was resuspended gently on ice prior to ultracentrifugation at 80000 g for 1 hour. The supernatant was poured off and pellet resuspended in 100 µl of Tris buffer.

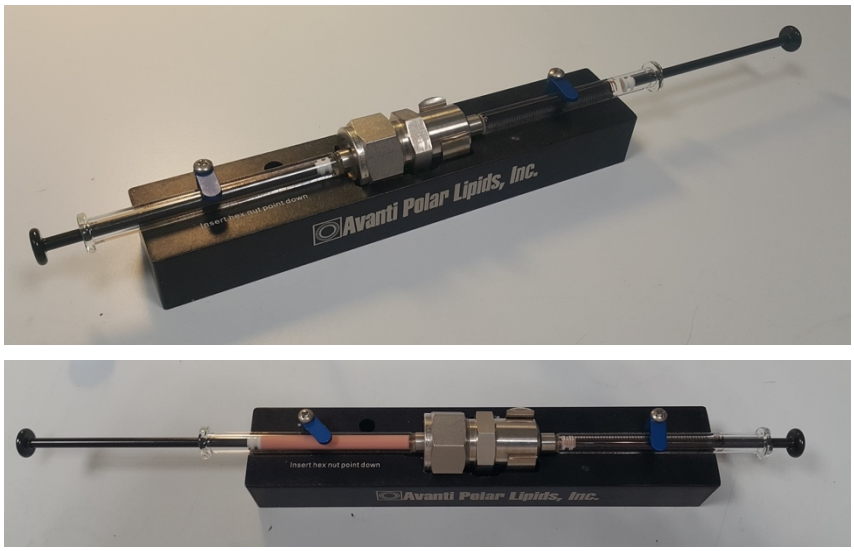


Figure 2.2.3: Customised polar lipids apparatus used for flagellar shearing experiments. Photo on top shows empty state. Photo on bottom shows *C. jejuni* cells on the left to be passed through to the right syringe.

The 81-176 ultracentrifuged sample was eluted using Amersham Biosciences Akta FPLC machine with MonoQ 1ml 5/50 GL anion exchange column (GE Healthcare) equilibrated with Buffer A1 and method set to introduce Buffer B1 gradient over 20 ml (10 x column volume) with 0.5 ml fraction collection at 4 Mpa pressure limit and 1 ml/min flow.

*81116*Δ$fliD$Kan^R knockout strain, cloned as per section 2.4.1, was grown from glycerol stock in blood agar plates with Amphoteracin ß, Vancomycin and Kanamycin. Cells were resuspended in MH broth with the above antibiotics and centrifuged for 10 minutes at 14000 rpm to isolate the media. It was passed through 0.2 µm filter and ultracentrifuged at 55000 rpm for 1 hour at 4 ^oC. 0.3 M $NaSO_4$ was added to polymerise the flagellin in the supernatant and left over 2 days at room temperature. Samples were ultracentrifuged once under the same conditions and both supernatant and pellet tested for flagellin. Pellet sample was further eluted using Amersham Biosciences Akta FPLC machine with MonoQ 1ml 5/50 GL anion exchange column (GE Healthcare) equilibrated with Buffer A2 and method set to introduce Buffer B2 gradient over 40 ml (20 x column volume) with 1 ml fraction collection at 4 Mpa pressure limit and 1 ml/min flow. Supernatant was eluted under the same conditions but with 60 ml (20 x column volume) wash with passing the sample through a

superloop. Supernatant fractions A1-A4 were pooled together and concentrated prior to further purification via gel filtration column HiLoad Superdex 200pg 16/600 (GE Healthcare) using the protocol outlined in section 2.2.3.

2.3 Structural Analysis

2.3.1 Negative-stain grid preparation and Data Collection

For negative-stain TEM experiments, 5 µl of purified protein, or of cell culture in log phase, was applied onto glow-discharged, carbon-coated copper grids (Agar Scientific). After incubating the sample for 2 min at room temperature, the grids were rapidly washed in three successive drops of deionized water and then exposed to three successive drops of 0.75% uranyl formate solution. Images were recorded on a CM100 TEM (Phillips) equipped with a MSC 794 camera (Gatan) or a Technai T12 Spirit TEM (Thermo Fisher) equipped with an Orius SC-1000 camera (Gatan). Datasets were manually acquired with 80 kV emission, 30000x magnification, a pixel size of 2.46 Å pix^{-1}, and a defocus range from −0.8 µm to −2.0 µm. The micrographs were processed using cisTEM package, with CTF parameters determined by CTFFIND4 (Grant et al., 2018; Rohou and Grigorieff, 2015). The number of particles and mask size is listed in Table 2.3.1.

Table 2.3.1: Number of manually picked particles and mask size used for processing via cisTEM for the negative stain datasets. Gel filtration purification fractions from traces in figures 3.1.10 and 3.1.11 are shown next to each protein and antibody samples were all tested with FliDcj.

Sample	Number of particles picked	Mask diameter (Å)
$FliD_{sm}$ (A10)	3500	330
$FliD_{pa}$ (B12)	2700	330
CCG4 Full G1	2322	360
CCG4 rFab	4360	360
CAA Full G1	2865	360
CAA rFab	2540	360
CAA SlgA1	5118	360
CCG4 SlgA2	2686	360

2.3.2 Single Particle Cryo-EM Grid Preparation and Data Collection

For the structural characterization of $FliD_{cj}$, aliquots of (5 µl) of purified protein at a concentration of 1 mg ml^{-1} was deposited onto glow-discharged C-flat holey carbon films 1.2/1.3 200 mesh (EMS). A Vitrobot Mark III (FEI) plunge-freezing device was used for freeze plunging, using double-blotting with a final blotting time of 6.5 s (Snijder et al., 2017). Cryo-EM data was collected with a Diamond Light Source Titan Krios TEM proposal EM19709-1 operated at 300 kV and equipped with an energy filter (Gatan GIF Quantum) and recorded on a K2 Summit direct electron detector (Gatan) operated in counting mode. 1223 micrographs were automatically acquired with the EPU software (Thermo Fisher), at a pixel size of 1.38 Å pix^{-1}, using a total dose of 41 e^{-} $Å^{-2}$ and with 40 frames per micrograph. The defocus range used for data collection was −1.0 µm to −2.6 µm.

For $FliD_{cj}$ and heparin interaction 5 µl protein aliquot at 1 mg ml^{-1} was premixed with heparin at 0.5 mM and deposited onto 20 s glow-discharged C-flat holey carbon films 1.2/1.3 200 mesh (EMS). A Vitrobot Mark III (FEI) plunge-freezing device was used

for freeze plunging, using double-blotting with a final blotting time of 4 s (Snijder et al., 2017). Cryo-EM data was collected with a Diamond Light Source Titan Krios TEM proposal EM20970-1 operated at 300 kV and equipped with an energy filter (Gatan GIF Quantum) and recorded on a K2 Summit direct electron detector (Gatan) operated in counting mode. 1060 micrographs were automatically acquired with the EPU software (Thermo Fisher), at a pixel size of 1.39 Å pix^{-1}, using a total dose of 50 e^{-} $Å^{-2}$ and with 40 frames per micrograph. The defocus range used for data collection was −1.0 µm to −2.2 µm.

For the structural characterization of the native *C. jejuni* filament, wild-type *81116* strain cell culture grown to OD_{600} = 5 was applied onto glow-discharged Cflat holey carbon films 2/2 200 mesh (EMS). A Leica EM GP (Leica) plunge freezing device was used for freezing, with a 6 s blotting time. Cryo-EM data were collected on a Technai Arctica TEM (Thermo Fisher) operated at 200 kV and equipped with a Falcon III camera. 100 micrographs were manually collected using the EPU software (Thermo Fisher) in linear mode, with a pixel size of 2.03 Å pix^{-1}, with a total dose of 45 e^{-} $Å^{-2}$ and 1 frame per micrograph. The defocus range used for data collection was approximately −0.8 µm to −2.0 µm. A second dataset was collected on a Diamond Light Source Titan Krios TEM proposal EM19832-23 operated at 300 kV and equipped with a K3 camera. 2955 micrographs were automatically collected using the EPU software (Thermo Fisher), with a pixel size of 1.386 Å pix^{-1}, with a total dose of 47 e^{-} $Å^{-2}$ and 40 frames per micrograph. The defocus range used for data collection was approximately −1.0 µm to −2.2 µm.

2.3.3 Single Particle Cryo-EM Image Processing and Reconstruction

For $FliD_{cj}$, processing was done first in RELION 2.0 followed by some reprocessing in RELION 3.0 (Scheres, 2012). Motion correction was performed with MotionCor2, with dose-weighting (Zheng et al., 2017). CTF parameters were determined by CTFFIND4 software (Rohou and Grigorieff, 2015). Approximately 2000 particles were manually picked from selected micrographs to generate representative 2D class averages. These classes were used as templates for automated particle picking

for the entire dataset. A total of 130000 particles were picked and extracted using a 280×280 pixels box. After multiple rounds of 2D classification, 55967 particles from the best 2D classes were obtained and used to generate an initial model. Following further 3D classification and refinement with D5 symmetry, a final map to 4.7 Å resolution was generated, which was sharpened using PHENIX 1.16 (Adams et al., 2010; Terwilliger et al., 2018). A mask centering on the head domain was used for further refinement with C5 symmetry, leading to a map of the head domain to 5.0 Å resolution. Further 3D classification of the masked head domain was used to identify 4 different conformations of the D4 domain not resolved in the full map. Local resolution maps were generated using Relion 3.1 Local resolution function.

For the native *C. jejuni* filament manual dataset, processing was done in RELION 3.0. Motion correction was performed with MotionCor2, with dose-weighting. CTF parameters were determined by CTFFIND4. Filaments were manually picked, and particles were extracted using a 4.7 Å rise and 300 pixel box leading to a set of 254041 segments. Multiple rounds of 2D classification gave a final dataset of 71828 good particles which were used for 3D refinement, both with and without imposed helical symmetry. 70 Å low pass filtered map of *P. aeruginosa* flagellar filament (EMDB:8855), in a form of a smooth cylinder with the correct dimensions, was used as a reference for all 3D refinements.

For the native *C. jejuni* filament eBIC collected automatic dataset, processing was done in RELION 3.0 and CryoSPARC (Punjani et al., 2017). Motion correction was performed with MotionCor2, with dose-weighting in RELION 3.0. The rest of the steps were performed in CryoSPARC. Filaments were manually picked, and particles were extracted using a 400 pixel box leading to a set of 31097 segments. Multiple rounds of 2D classification gave a final dataset of 20605 good particles which were used for 3D refinement. The parameters of known *C. jejuni* flagellar filament were imposed (4.8 Å rise and 65.3° twist), and a search initiated with rise range of 1-12 Å. The resulting best structures were used as references for heterogenous refinement classification and best classes further refined with a close-fit mask, until final 13202 particles used to refine 2 nearly identical structures.

2.3.4 Model Building and Refinement for $FliD_{CJ}$

For the D1–D3 domains, a homology model was generated with PHYRE2, using the $FliD_{ec}$ crystal structure (PDB:5H5V) as a template (Kelley et al., 2015; Song et al., 2017). Domains D2 and D3 were fitted into the sharpened head domain map in Chimera (Pettersen et al., 2004). Domain D1 was fitted into the sharpened full $FliD_{cj}$ map in Chimera. Domain D0 was built into the sharpened full $FliD_{cj}$ map using Coot (Emsley et al., 2010). This model was subjected to iterative rounds of real-space refinement and building using PHENIX 1.16 and Coot respectively. The N-terminal stretch was modelled with RosettaES, and then the remaining missing loops were modelled using RosettaCM guided by the electron density (Frenz et al., 2017; Song et al., 2013). The output model was refined once more in Coot to improve the geometry and delete any modelled residues in areas without electron density.

2.3.5 Generation of FliD-Filament Complex Model

The low-resolution tomography map of FliD bound to a hook in FlaB knockout mutant of *Borrelia burgdorferi* was used as a volume for docking atomic models (EMDB: 0525)(Zhang et al., 2019). The pentamer $FliD_{cj}$ model (this study) was used for the region corresponding to the cap, and the *P. aeruginosa* flagellum model (PDB: 5WK6) was used for the filament, as our filament map was not high enough resolution to obtain an atomic model.

2.3.6 Biochemical characterization of $FliD_{CJ}$ binding to heparin and antibodies

Isothermal Titration Calorimetry (ITC) machine (TEA) loading syringe and cell was cleaned with Decon 90 at 2% followed by MilliQ with the assistance of a vacuum pump. The reference cell was filled with 300 µl of MillQ to clean and followed by equilibration with 300 µl of HEPES buffer. 1.89 mg ml^{-1} $FliD_{cj}$ (0.027 mM) was diluted to 0.02 mM in HEPES buffer, degassed and loaded into the cell. 1g Heparin sodium salt was diluted in 4 ml MilliQ, but as the MW was between 8 and 25 kDa, it was assumed to be 15000 g mol^{-1} with final concentration of 16.6 mM. It was loaded into the loading

syringe, placed into the burette and into the cell. The run parameters used are listed in Table 2.3.2.

Antibodies from HumAbs BioMed were diluted to 1 mg ml^{-1} stock and samples prepared in 50 µl with 10 µl $FliD_{cj}$ at 1 mg ml^{-1}, 2 µl antibodies and made up with 50 mM HEPES 150 mM NaCL pH7 buffer (Perruzza et al., 2020). Negative stain grids were made up as in 2.3.1 and screened on Technai T12 Spirit TEM (Thermo Fisher) under the same parameters. Samples were further diluted as follows: 0.01 mg ml^{-1} for CCG4 rFab and 0.001 mg ml^{-1} for CCG4 Full G1, CAA Full G1 and CAA rFab, 0.0014 mg ml^{-1} for CCG4 SlgA2, 0.0037 mg ml^{-1} for CAA SlgA1 and datasets were manually acquired.

Table 2.3.2: ITC set up parameters.

Run number	**Run 1**	**Run 2**	**Run 3**	**Run 4**	**Run 5**
Spin speed (s)	300	300	300	300	300
$FliD_{cj}$ concentration in cell (mM)	0.02	0.01	0.01	0.01	0.01
Heparin concentration in syringe (mM)	16.6	16.6	16.6	16.6	16.6
Injection length (s)	200	200	250	250	250
Injection volume (µl)	1.5	3	1.5	2	2.5
Injection number	25	16	31	25	20

2.4 Effect of $FliD_{cj}$ on Cell Motility

2.4.1 Construction of FliD Knockout Mutant (ΔfliD) in C. jejuni

Isothermal assembly (ISA) cloning was used to create mutant vectors based on pGEM3ZF plasmid with an antibiotic resistance cassette KanR between the two flanking regions of $FliD_{cj}$. Kanamycin cassette primers used were standard ISA primers amplified from pJMK30. These are listed in Appendix 3. The final mutation vector was designed such that spontaneous double crossover with the *C. jejuni* 81116 genome would result in the replacement of the majority of the open reading frame of *fliD* with the kanamycin resistance cassette, allowing a means of selection.

Amplifying the fragments, source DNA and plasmid vector

Working concentrations of Amphotericin β and Vancomycin (AV) was added to 30 ml of MH broth in 100 ml flask and inoculated with *C. jejuni* 81116 strain grown overnight from glycerol stock on Columbia blood agar. It was then left to grow overnight in the 42 °C microaerophilic cabinet. 1 ml of vells was centrifuged at 4 °C 12000 g for 10 minutes to obtain a pellet from which genomic DNA was isolated as per standard protocol in GenElute TM Bacterial Genomic DNA kit (QIAGEN).

Fragments to make up the plasmid were amplified using a Phusion High-Fidelity PCR Master Mix (E2621, New England Biolabs). The 3 fragments were made up in 25 µl with 12.5 µl FlashPhusion mix, 1 µl Forward primer for ($FliD_{cj}$ top, $FliD_{cj}$ bottom, Kanamycin cassette), 1 µl Reverse primer for (same as above), 100 ng genomic DNA template (1µl) and 9.5 µl MilliQ water. The solutions were placed into PCR thermocycler and set up with parameters detailed in Table 2.4.1.

Table 2.4.1: PCR thermocycler programme settings.

Step	Cycles	Time	T (°C)
Initial denaturation	1	5 min	98
Denaturation Annealing Extension	30	10 s 10 s 30 s	98 55 72
Final Extension	1	5 min	72
Hold	1	∞	4

1% Agarose gels were made up and samples tested at 120 V for 30 minutes. The gel was imaged using UV and fragments extracted using QIAquick Gel Extraction Kit (QIAGEN) per product specification.

pGEM3ZF plasmid was digested by *HincII* restriction enzyme in the following mix of 20 µl: Shrimp Alkaline Phosphatase (rSAP) (NEB) 1 µl, *HincII* 1 µl (NEB), 2 µl CutSmart NEB buffer, 3 µl (2000ng) plasmid DNA, 13 µl of MilliQ. The mix was incubated for 1 hour at 37 °C and 10 minutes at 65 °C then linearised plasmid DNA

isolated using QIAprep Spin PCR cleanup Kit (QIAGEN) per product specification. The concentration of DNA was determined by nanodrop.

Ligation and selection in *E. coli*

ISA method was used to ligate the fragments together into a knockout plasmid using 2.5 µl of HiFi mix (NEB), 1.25 µl of pGEM3ZF linear plasmid, 0.5 µl Kanamycin Resistance cassette, 0.5 µl $FliD_{cj}$ top, 0.5 µl $FliD_{cj}$ bottom. It was incubated at 50 °C for 1 hour. 2 µl of the mixture was added to 50 µl of competent DH5α *E. coli* cells and left on ice for 30 minutes. The cells were heat shocked at 42 °C for 45 seconds and returned to ice for 2 minutes, then left to grow at 37 °C in LB for 1 hour. The cells were plated on Kanamycin and Carbenicillin plates and left to grow overnight.

12 transformants in *E. coli* were selected and tested via colony PCR by mixing 6.25 µl of 2X MyTaqRed mix (Bioline), 0.5 µl for Forward screening primer, 0.5 µl of Reverse screening primer and 5.25 µl of MilliQ water was added to 12 samples. Same was done for the negative (pGEM32ZF) and positive (known good knockout plasmid) control samples. The solutions were placed into PCR machine and set up with parameters detailed in Table 2.4.2.

Table 2.4.2: PCR thermocycler programme settings.

Step	Cycles	Time	T (°C)
Initial denaturation	1	3 min	95
Denaturation Annealing Extension	30	10 s 10 s 60 s	95 55 72
Final Extension	1	10 min	72
Hold	1	∞	4

Samples were tested on 1% Agarose gel and imaged. A successful colony was selected and resuspended in 5 ml of LB with Carbenicillin and Kanamycin in a sterile universal tube and left to incubate overnight. Cells were centrifuged down at 8000 rpm for 2 minutes and plasmid DNA isolated through QIAprep Spin Miniprep Kit (QIAGEN) as per product specification. DNA concentration was obtained to be 90 ng/µl.

The pGEM32ZFΔ$fliD$Kan^R (FliD$_{cj}$ knockout) plasmid was tested for correct construction via PCR screening as specified above in Table 2.4.2. It was diluted to 10 ng/µl and 1 µl of plasmid added to each of the reaction tubes containing a combination of the following: 6.25 µl MyTaq Red mix (Bioline), 0.5 µl Forward primer, 0.5 µl Reverse primer, 4.25 µl MilliQ. The primers were set up in following combinations, detailed in Table 2.4.3.

Table 2.4.3: PCR plasmid test setup.

DNA source	Forward Primer	Reverse Primer
pGEM32ZF	32ZF	32ZF
pGEM32ZF	FliDcj top	FliDcj bottom
pGEM32ZFΔ$fliD$Kan^R	32ZF	32ZF
pGEM32ZFΔ$fliD$Kan^R	FliDcj top	FliDcj bottom
pGEM32ZFΔ$fliD$Kan^R	32ZF	FliDcj bottom
pGEM32ZFΔ$fliD$Kan^R	FliDcj top	32ZF

Transforming pGEM32ZFΔ$fliD$Kan^R back into *C. jejuni*

C. jejuni 81116 wild type overnight cells were harvested and resuspended into 800 µl of *C. jejuni* Wash buffer (15% glycerol 9% sucrose). They were centrifuged and resuspended in wash buffer three times at 14000 rpm at 4 °C for 5 minutes, and final pellet resuspended and split into 3 100 µl aliquots. 1000 ng of pGEM32ZFΔ$fliD$Kan^R was added to an aliquot of competent cells and transferred into a pre-chilled electroporation cuvette, shocked at 2.5 kV for 5 ms and plated on a non-selective AV agar plate which was incubated overnight in the 42 °C microaerophilic cabinet. Harvested cells from the plate were resuspended in MH broth and spread on Kanamycin selective AV agar plates left to incubate over 2 days. Colonies were selected and spread onto fresh plates in patches and incubated overnight. Cells were harvested and resuspended in BHI glycerol buffer and stored at -80 °C.

2.4.2 Construction of the complement strain (ΔfliDφmetK) in C. jejuni

For complementation of the mutant, *fliD* gene was amplified from *C. jejuni* 81116 genomic DNA using primers fliDcompF/R, listed in Appendix 3. The amplified

fragment was incorporated into pCmetK complementation vector containing flanking regions of the pseudogene region corresponding to *cj0046* in *C. jejuni* 81116 to allow insertion into the genome. Plasmid was also designed to contain a constitutive promoter from the *C. jejuni metK* gene to drive expression of *fliD*, and a chloramphenicol resistance cassette.

Amplifying the plasmid vector and fragments

pCmetK plasmid was digested by *BsmBI* restriction enzyme in the following mix of 25 µl: *BsmBI* 1 ul (NEB), 1 µl rSAP, 2.5 µl 10x CutSmart NEB buffer, 2 µl plasmid DNA, 19.5 µl of MilliQ. $FliD_{cj}$ amplified insert (see Appendix 3, section 2.4.1) was digested in a similar manner. The solutions were digested at 55 °C overnight and at 80°C for 20 minutes to denature the enzyme. 1% Agarose gels were made up and samples tested at 120 V for 30 minutes. The gel was imaged using UV and fragments extracted using QIAquick Gel Extraction Kit (QIAGEN) per product specification. The DNA concentrations were determined with Nanodrop.

Ligation and selection in *E. coli*

The amplified *fliD* gene was ligated into pCmetK vector using 0.5 µl of T4 ligase, 1 µl of 10x T4 DNA ligase buffer, 2 µl digested pCmetK (20 ng), 7 µl $FliD_{cj}$ (50 ng). Control was made up same as above but with water instead of $FliD_{cj}$. It was incubated at room temperature for 20 minutes and 5 µl of each sample including the control was transformed into competent DH5α *E. coli* cells and plated on Chloramphenicol LB plates and left to grow overnight.

6 transformants in *E. coli* were selected and tested via PCR by mixing 6.25 µl of 2X MyTaqRed mix (Bioline), 0.5 µl for Forward screening primer, 0.5 µl of Reverse screening primer and 5.25 µl of MilliQ water was added to 6 samples. Same was done for the negative (pCmetK) and positive ($FliD_{cj}$ amplified cassette) control samples. The solutions were placed into PCR machine and set up with parameters detailed in Table 2.4.2. Samples were tested on 1% Agarose gel and imaged.

Samples were checked for the correct insertion through PCR. 1 µl of cell lysate added to each of the reaction tubes containing a combination of the following: 6.35 µl MyTaq Red mix (Bioline), 0.5 µl Forward primer, 0.5 µl Reverse primer, 4.25 µl MilliQ

and set up according to the PCR parameters in Table 2.4.2. The primers were set up for all lysates and negative control (pCmetK) in following combinations, detailed in Table 2.4.4.

Table 2.4.4: PCR lysate test setup.

Forward Primer	Reverse Primer	Purpose
meKF	metKR	Checks for fragment insert
$FliD_{cj}$ top	$FliD_{cj}$ bottom	
metKF	$FliD_{cj}$ bottom	Check for correct orientation
metKF	$FliD_{cj}$ top	Checks for incorrect orientation

A successful colony was selected and resuspended in 5 ml of LB broth with Chloramphenicol in a sterile universal tube and left to incubate overnight. Cells were centrifuged down at 8000 rpm for 2 minutes and plasmid DNA isolated through QIAprep Spin Miniprep Kit (QIAGEN) as per product specification. DNA concentration was obtained to be 250 µg/ml.

Transforming pCmetK0509Chl^R back into *C. jejuni*

C. jejuni 81116 FliD knockout strain cells were harvested and resuspended into 800 µl of *C. jejuni* Wash buffer. They were centrifuged and resuspended in wash buffer three times at 14000 rpm at 4 °C for 5 minutes, and final pellet resuspended and split into 3 100 µl aliquots. 2500 ng of pCmetK0509Chl^R was added to an aliquot of competent cells and transferred into a pre-chilled electroporation cuvette, shocked at 2.5 kV for 5 ms and plated on a non-selective AV and Kanamycin agar plate which was incubated overnight in the 42 °C microaerophilic cabinet. Harvested cells from the plate were resuspended in MH broth, pelleted down and resuspended in MH broth with working concentration of AV and spread on Kanamycin and Chloramphenicol selective AV agar plates left to incubate over 2 days. Colonies were selected and replated into fresh plates in patches and incubated overnight. Cells were harvested and resuspended in BHI glycerol buffer and stored at -80 °C.

Growth curves and viability of *C. jejuni* strains

Overnight cells grown on agar with respective antibiotics (WT only on AV, knockout (Δ*fliD*) on Kanamycin and complement (Δ*fliD*φmetK) on Kanamycin and Chloramphenicol), were grown in MH media with AV for 14 hours at 37 °C. Culture was diluted to OD_{600} 0.1 and 30 ml inoculated overnight in same conditions. OD_{600} values were measured every 2 hours and plotted.

RNA isolation was obtained through centrifugation of 5 ml of each culture at OD_{600} 0.8-1.0 at 14000 rpm for 5 minutes, resuspension of the pellet in 1 ml $NaPO_4$ buffer 5 µl Phenol and 50 µl 100% ethanol, and further centrifugation under the above stated conditions. Supernatant was decanted and pellet stored at -20 °C. RNA purification and qRT-PCR was done by Dr. Aidan Taylor at Professor Dave Kelly's Lab at the University of Sheffield.

2.4.3 Construction of FliD Point Mutants in C. jejuni

Point mutations in *fliD* were constructed by Q5 site directed mutagenesis (SDM) of the complementation vector using the Kinase, Ligase and *DpnI* (KLD) method (M0554, New England Biolabs). The template DNA plasmid (pCmetK0509Chl^R) 1 µl at 20 ng/µl was mixed with 12.5 µl master mix, 1.25 µl Forward, 1.25 µl Reverse primers and 9 µl MilliQ for each point mutant and amplified through PCR reactions as detailed in Table 2.4.5. The divergent primers containing targeted nucleotide substitutions in the forward primer are listed in Appendix 4.

Table 2.4.5: PCR thermocycler programme settings.

Step	Cycles	Time	T (°C)
Initial denaturation	1	30 s	98
Denaturation Annealing Extension	25	10 s 30 s 3 min	98 56 72
Final Extension	1	2 min	72
Hold	1	∞	4

An 1µl aliquot of the linear PCR product was treated with 1 µl KLD enzyme mix, 5 µl KLD reaction buffer and 3 µl of MilliQ for 20 minutes at room temperature to circularize the mutated plasmid while degrading any residual template. 5 µl of treated plasmids were transformed into *E. coli* DH5α and transformants selected by chloramphenicol resistance. Plasmid was purified from multiple transformants through QIAprep Spin Miniprep Kit (QIAGEN) as per product specification. *fliD* open reading frame was sequenced, through premixing the plasmids with pCmetK primer, to ensure the correct substitution had been introduced without secondary mutations (LightRun sequencing, Eurofins EU). Point mutated complementation vectors were then transformed into the *C. jejuni* Δ*fliD* strain as in sections 2.4.1 and 2.4.2 to generate the collection of point mutant strains with both Chloramphenicol and Kanamycin resistance.

2.4.4 Motility Assays

Overnight growth of *C. jejuni* on blood agar plates was harvested and resuspended in PBS to an OD_{600} nm of 1.0. 0.5 µl aliquots were then injected into semi-solid agar plates (0.4% w/v agar, 3.7% w/v brain heart infusion) containing 0.005% TTC, a redox dye which allows clear visual assessment of growth. The diameter of growth was measured after 16 h of incubation in *C. jejuni* standard growth conditions listed above. The data was collected in triplicate for each sample in three different experimental sessions.

2.4.5 Flagella Attachment Analysis via Negative Stain TEM

The WT, knockout, complement and all point mutant strains were taken from the same cultures used to inject the agar plates as in section 2.4.4. The point mutant flagella attachment was determined through imaging grids at ×700 magnification using CM100 TEM (Phillips) at about 20 micrographs per mutant containing cell count from 30 to 100 cells. The filament was counted as a "intact flagellum" when it had 2 or more inflection points (whether it was attached or on its own in solution). Filaments with less inflection points were considered fragments and not included in the calculations.

The percentage of attachment was calculated as a proportion of the total flagella observed per mutant. Flagella length was calculated via ImageJ.

3.RESULTS

3.1: PURIFICATION AND CHARACTERIZATION OF FLIID ORTHOLOGUES

3.1.1 INTRODUCTION

Low-resolution cryo-EM studies of the cap complex in *S. enterica* have suggested that it consists of five copies of FliD (also known as HAP2), forming a "stool"-shaped complex with a core "head" domain and five flexible "leg" domains, that interact with the growing end of the filament (Maki-Yonekura et al., 2003; Yonekura et al., 2000). Crystal structures of the truncated FliD head domain have been reported for several species, and revealed a range of crystallographic symmetries, from tetramers in *Serratia marscecens* ($FliD_{sm}$), pentamers in *S. enterica* ($FliD_{se}$) to hexamers in *E. coli* ($FliD_{ec}$) and *Pseudomonas aeruginosa* ($FliD_{pa}$) (Cho et al., 2017, 2019a; Postel et al., 2016; Song et al., 2017). This observation led to the hypothesis that the cap complex can have different, species-specific oligomeric states. This hypothesis was investigated through purification of complete recombinant FliD from the above-mentioned species and observe their oligomerization in a non-crystalline state through EM, as well as determine a high-resolution structure of the leg-domains truncated from the crystal structures.

3.1.2 BIOINFORMATICS

FliD protein sequences from ε-proteobacteria *(C. jejuni* and *H. pylori)*, γ-proteobacteria (*P. aeruginosa* and *S. marcescens*) and the enteric subfamily *(S. enterica* and *E. coli*), which were consequently used to generate plasmids for recombinant expression as per section 2.2.2, were aligned to observe their domain distribution. This alignment, shown in Figure 3.1.1, reveals highly conserved residues at the N and C termini, the D2 head domain and D1 helical bundle. There is one large insert present in *H. pylori* sequence located within the D3 domain, which was shown to form a globular domain in the crystal structure (Cho et al., 2019a). Another large insert was observed in *C. jejuni* sequence located within the D1 leg domain which was named D4. Both of these inserts are reflected in the molecular weight of the proteins: ~70 kDa for

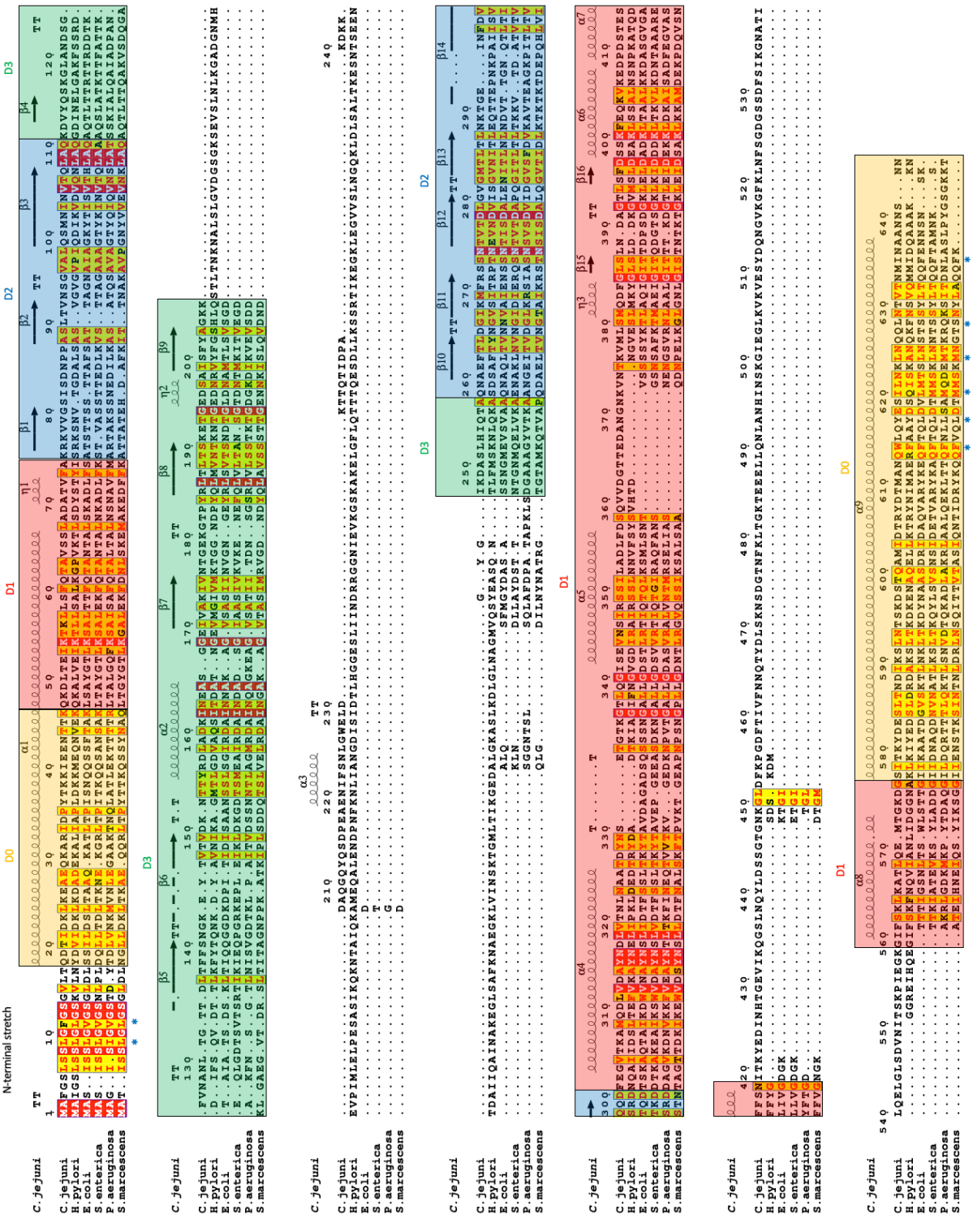


Figure 3.1.1 Sequence alignment for FliD from different proteobacteria species studied in this work. The secondary structure elements for $FliD_{cj}$ are shown at the top. The domains are colour coded as follows: *Yellow*: D0 terminal domains. *Red*: D1 domain. *Blue*: D2 domain. *Green*: D3 domain. *Uncoloured*: extra species-specific domains. Stars indicate important residues in terminal interactions, for which mutagenesis experiments were performed.

$FliD_{cj}$,~75 kDa for $FliD_{hp}$, ~50 kDa for $FliD_{ec}$, $FliD_{pa}$ and $FliD_{sm}$.

The tendency to have extended or extra domain was also observed in flagellin proteins. As illustrated in Figure 3.1.2, *C. jejuni* major flagellin FlaA and minor flagellin FlaB had inserts after the D3 head domain in comparison to some flagellin sequences from other members of proteobacteria. There is also a similarity in conservation of residues at the terminal regions. Due to interaction of FliD and flagellin in the assembly of the flagellar filament, a co-evolution analysis was performed to see any co-evolving residue pairs important for binding listed in Table 3.1.1. Overall, there were many pairs of $FliD_{cj}$ C-terminus mutants with $FlaA_{cj}$ N-terminal residues that appeared to co-evolve. That is to be expected as FliD terminal regions should have residues that interact with that of flagellin to be able to fit and interact with the growing filament end. However, as the point-mutations (see sections 2.4.3 and 3.3.1) were designed prior to conducting of this analysis, only three residues (Tyr617, Leu624 and Trp614) were subject to both co-variation test and point-mutagenesis.

3.1.3 Expression and Purification of full-length FliDcj

While both $FliD_{cj}$ and $FliC_{cj}$ expressed without issue and a solubility assay was conducted, $FliD_{cj}$ is observed to be partly soluble whereas $FliC_{cj}$ is in inclusion bodies thus preventing the purification of a soluble fraction. $FliD_{cj}$ was overexpressed, as detailed in section 2.2.3, and the soluble portion purified with His-trap column as illustrated in Figure 3.1.3. Due to the nature of the construct, the protein was expressed with an N-terminal His-tag and thus would attach to the Ni-NTA beads when in buffer lacking imidazole. Once the Imidazole was introduced, first at 20 mM, some $FliD_{cj}$ managed to elute from the column along with most of the contaminants. At 500 mM imidazole most of the protein eluted in Fraction 6 and the remainder slowly came off in further fractions, as illustrated in Figure 3.1.4. However, it was not pure enough for structural work and thus a second purification step was required.

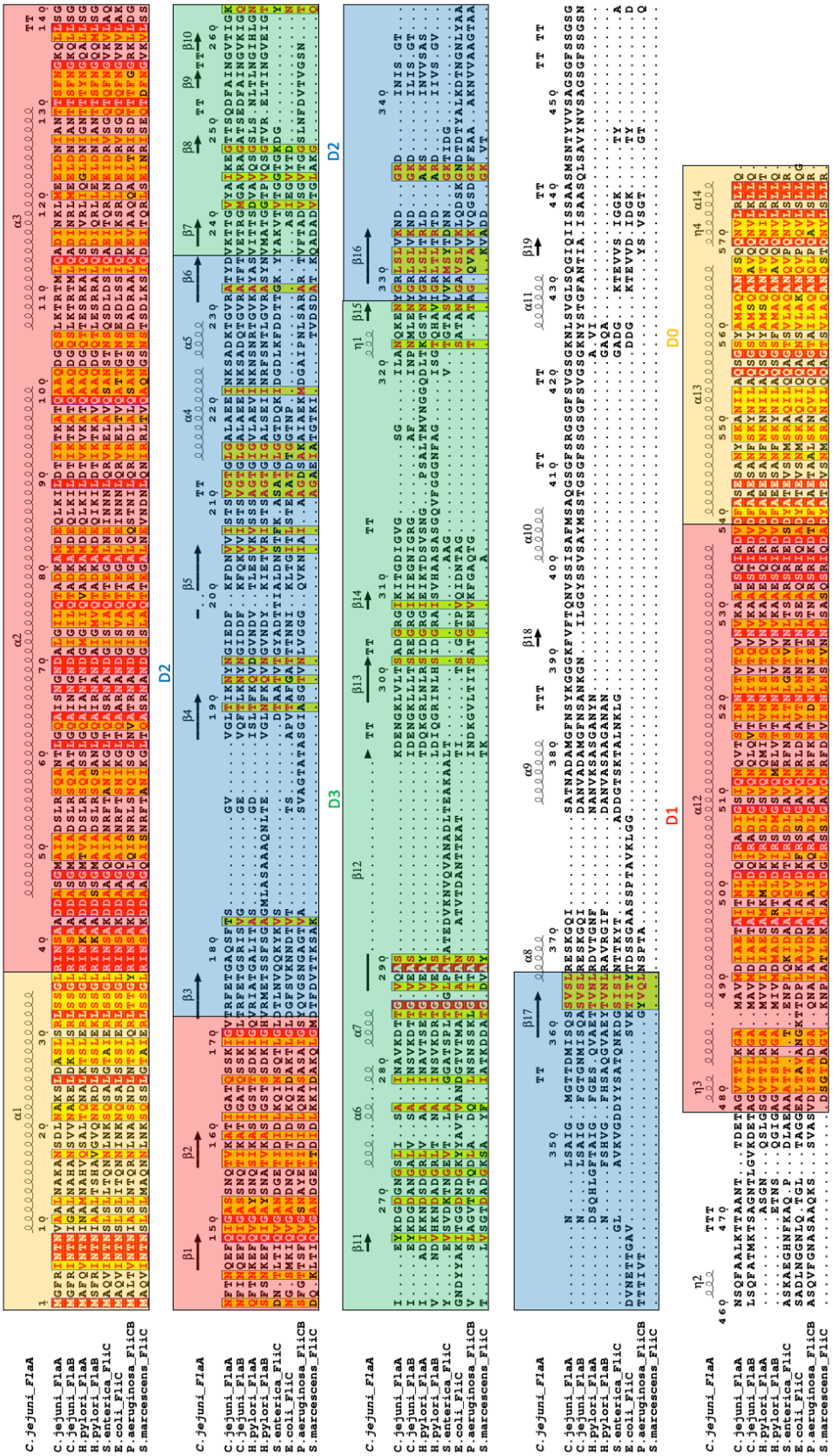


Figure 3.1.2: Sequence alignment for Flagellin from different proteobacteria species studied in this work. The domains are colour coded as follows: *Yellow:* D0 terminal domains. *Red*: D1 domain. *Blue:* D2 domain. *Green*: D3 domain.

Table 3.1.1: Co-evolution analysis between aligned FliD and FlaA (flagellin) sequences across bacterial species performed using RaptorX ComplexContact server. Conserved residues are underlined, mutated $FliD_{cj}$ residues (see sections 2.4.3 and 3.3.1) are in red.

$FliD_{cj}$	**$FlaA_{cj}$**	**Probability**
Val631	Ala14	0.610818
Val631	Ser18	0.537091
Asn629	Ser18	0.484225
Ile635	Ala11	0.459525
Leu621	Leu25	0.439240
Ile635	Ala14	0.425490
Met634	Ala14	0.413679
Ala638	Phe3	0.400955
Leu621	Leu29	0.387116
Tyr617	Leu32	0.377176
Asn629	Ala22	0.375285

$FliD_{cj}$	**$FlaA_{cj}$**	**Probability**
Gly340	Val9	0.371620
Ile635	Lys15	0.368281
Leu624	Asn21	0.352808
Trp614	Leu32	0.349173
Leu621	Leu25	0.348483
Leu24	Ile471	0.335947
Val631	Lys15	0.322094
Ser417	Gln487	0.321698
Asn643	Met1	0.321136
Asn629	Asn21	0.317865
Glu28	Ala45	0.317778
Asp390	Lys87	0.314706

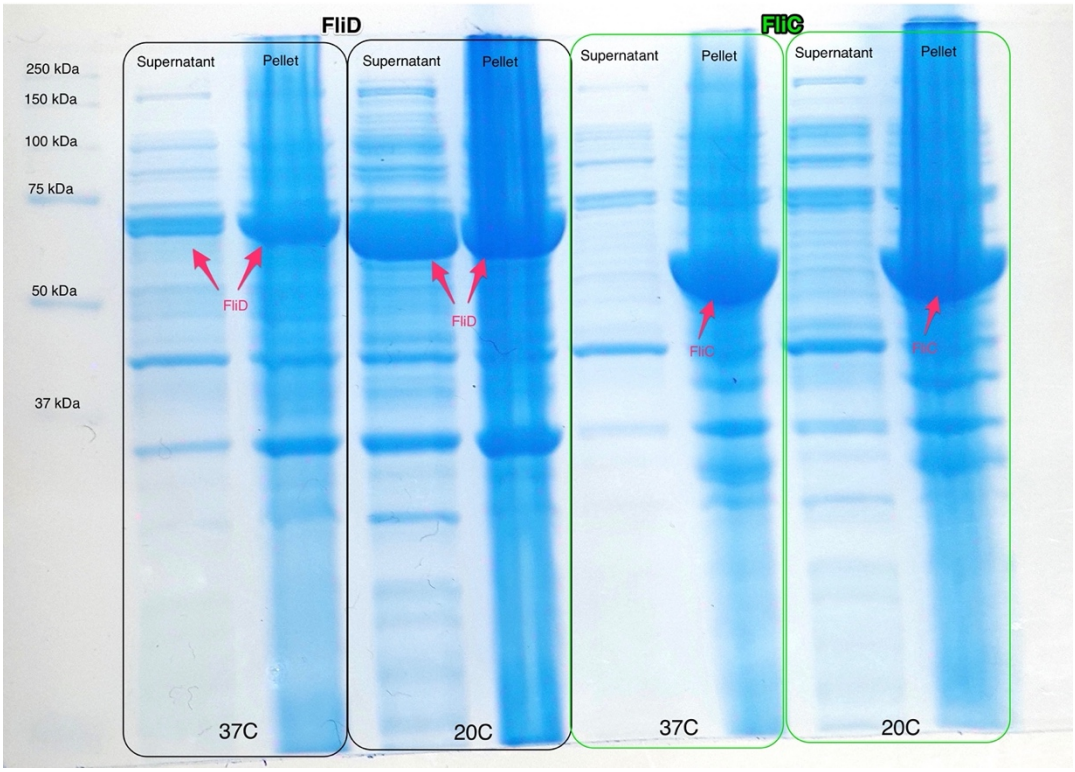


Figure 3.1.3: SDS-PAGE gel of soluble (Supernatant) and insoluble (Pellet) $FliD_{cj}$ and $FliC_{cj}$ post-expression at different temperatures. $FliD_{cj}$ (~70 kDa) is present in both while FliCcj (~60 kDa) is only present in the pellet.

The purified sample was separated into two different stocks: with His-tag and without His-tag after Thrombin digestion. Both were purified through gel filtration as illustrated in Figure 3.1.5 and 3.1.6 respectively. His-tagged protein eluted in a similar fashion to the cleaved one, indicating that Thrombin incubation does not interfere with the native conformation of $FliD_{cj}$. In gel filtration the proteins were separated according to their molecular weight into two separate peaks, suggesting two populations with two different oligomeric states, with the first peak being the largest oligomeric construct present in the sample. Molecular weight calibrations were not performed as we did not run the gel filtration column with the protein standards prior to this elution. Thus, the molecular weight of the components inside the peaks or what the oligomerisation status of FliD is in each peak could not be determined.

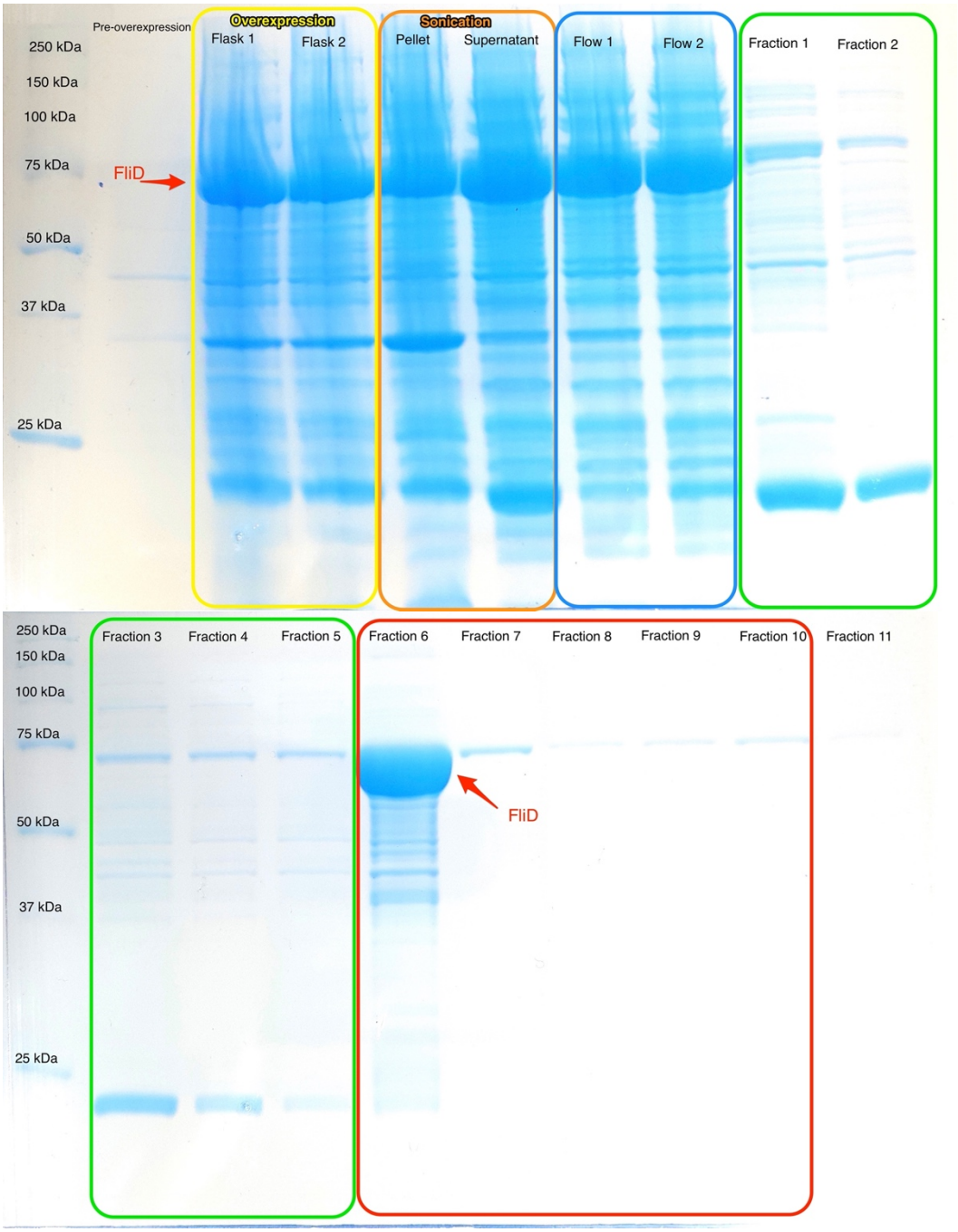


Figure 3.1.4: SDS-PAGE gel of Ni-NTA column output for $FliD_{cj}$ purification. *Yellow:* Pre-sonication, *Orange:* Sonication, *Blue:* no Imidazole, *Green:* 20 mM Imidazole, *Red:* 500 mM Imidazole.

According to the results of Ni-NTA column purification, the largest $FliD_{cj}$ oligomer should elute at the first few peaks of the chromatogram. This is confirmed through the SDS-PAGE gels, where $FliD_{cj}$ elution is split across two peaks in fractions A9-B11. According to previous data on $FliD_{se}$, the protein exists in equilibrium between

monomers and oligomers in solution (Maki et al., 1998; Postel et al., 2016). Thus, the first peak was assigned to the largest oligomer and the second peak could be either an intermediate oligomeric state or monomeric protein.

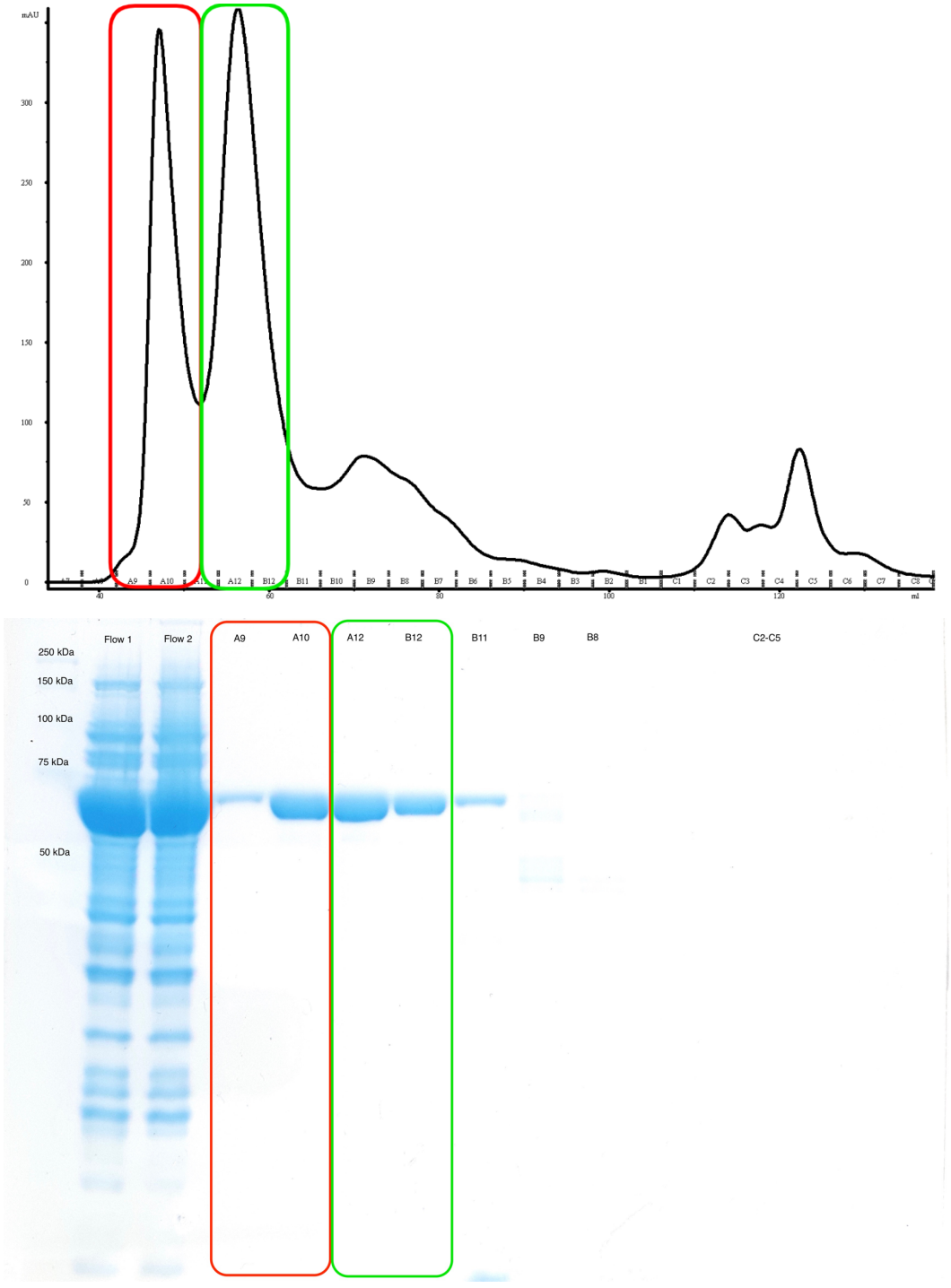


Figure 3.1.5: Chromatogram and SDS-PAGE gel of gel filtration column output for His-tagged $FliD_{cj}$. Left to Right SDS-gel: 1: Protein ladder, 2: Ni-NTA output flow through after passing the supernatant through the column once, 3: twice, 4: A9 (36 ml), 5: A10 (40 ml), 6: A12 (48 ml), 7: B12 (52 ml), 8: B12 (56 ml), 9: B11 (60 ml), 10: B9 (64 ml), 11: B8 (68 ml), 12: C2 (104 ml), 13: C3 (108 ml), 14: C4 (112 ml) , 15: C5 (114 ml). *Red:* larger oligomer fractions, *Green:* smaller oligomer fractions.

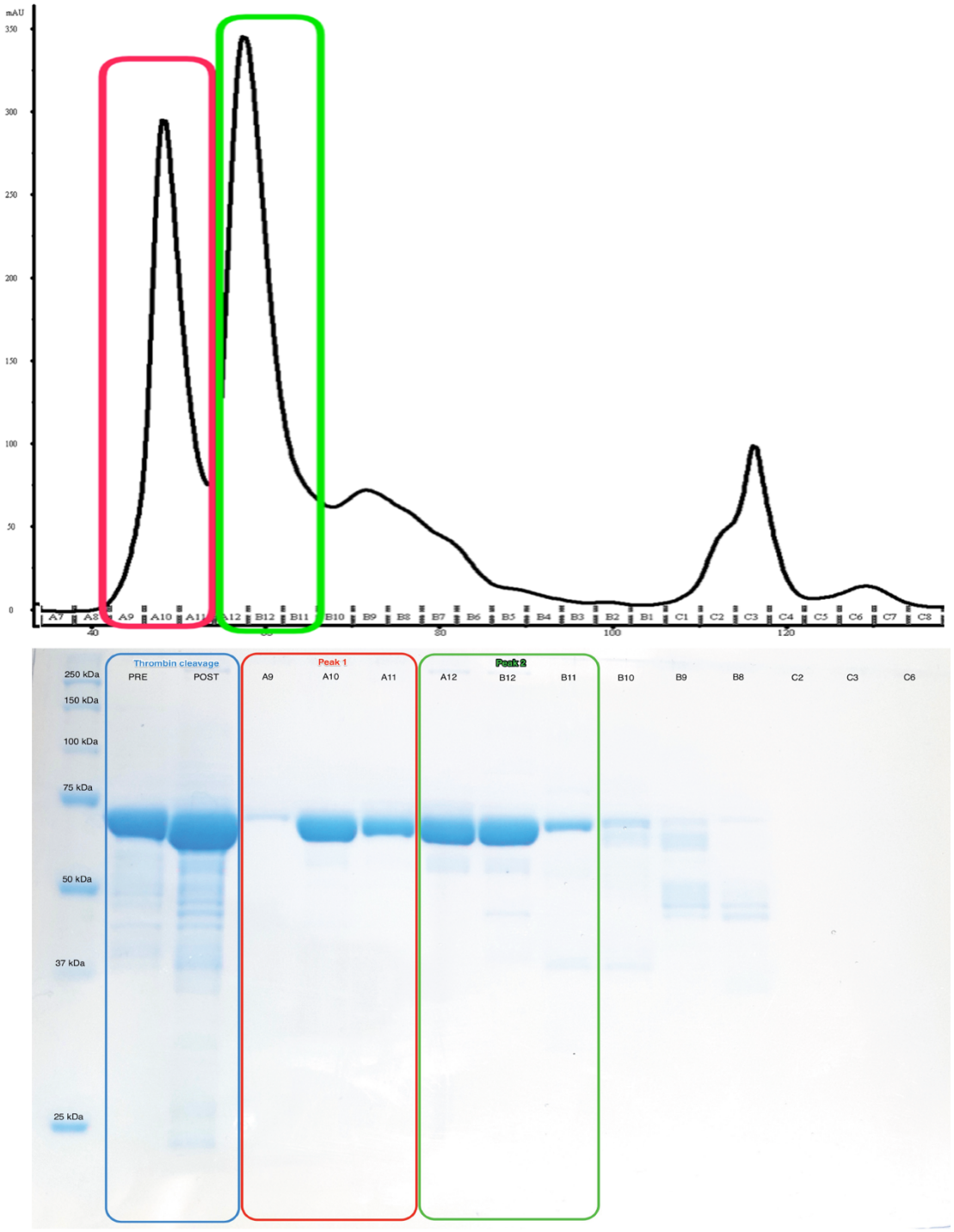


Figure 3.1.6: Chromatogram and SDS-PAGE gel of gel filtration column output for His-cleaved $FliD_{cj.}$ Left to Right SDS-gel: 1: Protein gradient ladder, 2: Pre-thrombin incubation sample with His-tag, 3: Post-thrombin incubation sample with reduced molecular weight due to His-tag cleavage, 4: A9 (36 ml), 5: A10 (40 ml), 6: A11 (44 ml), 7: A12 (48 ml), 8: B12 (52 ml), 9: B11 (56 ml), 10: B10 (60 ml), 11: B9 (64 ml), 12: B8 (68 ml), 13: C2 (104 ml), 14: C3 (108 ml), 15: C6 (118 ml). *Blue*: Thrombin cleavage test, *Red:* Larger oligomer fractions, *Green:* smaller oligomer fractions.

The cleaved samples were observed using negative stain EM. The second peak (2) had no visible large protein formations, which would be common for a monomer, while the first peak (1) had uniform dumbbell-like constructs similar to that observed for $FliD_{se}$ decamers (Maki-Yonekura et al., 2003). This is illustrated in Figure 3.1.7. Neither of the non-cleaved samples contained uniform oligomeric structures, with the A10 fraction having some sort of formation which lacked uniformity and a distinguishable shape (shown in Appendix 5).

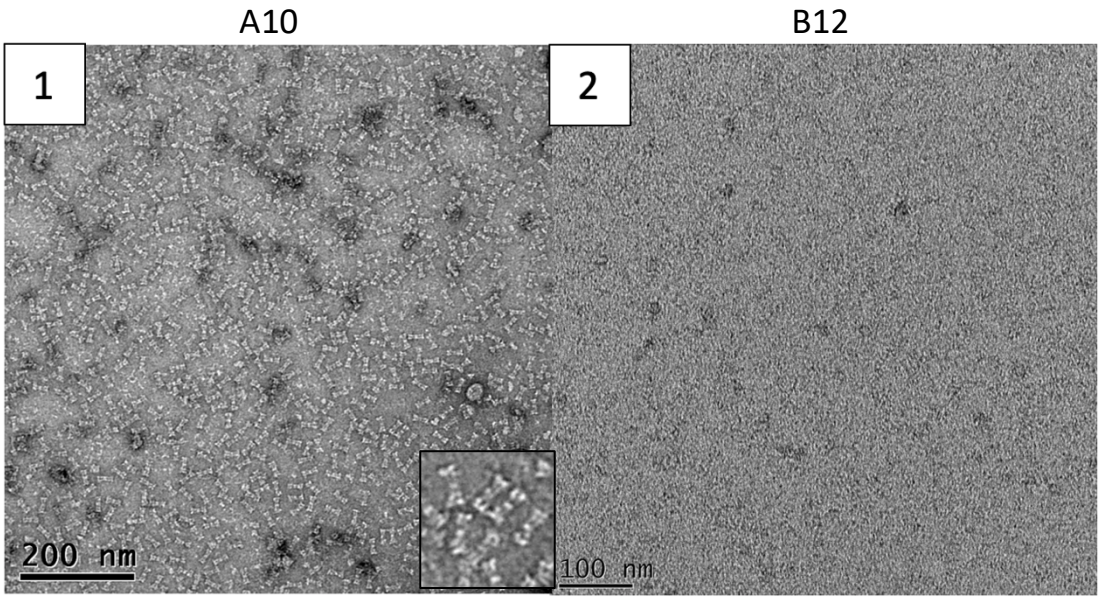


Figure 3.1.7: Negative stain images of $FliD_{cj}$ from 1) Non-tagged fraction A10, example of decamer particles is shown in the right bottom corner, 2) Non-tagged fraction B12.

$FliD_{se}$ was observed to change its oligomerization in different pH and salt buffer conditions (Maki et al., 1998; Vonderviszt et al., 1998). So, to observe if a similar conformational change occurs in recombinant $FliD_{cj}$, it was resuspended in conformation buffers listed in section 2.1.4 and observed via negative stain EM. At low pH there was a lot of decamer structures, the number of which significantly dropped with increase in pH as expected. Salt concentration does not appear to have an effect on the oligomer populations unlike that observed in $FliD_{se}$ (Imada et al., 1998). This change in populations of decamers and monomers is illustrated in Figure 3.1.8 with extended micrographs in Appendix 5.

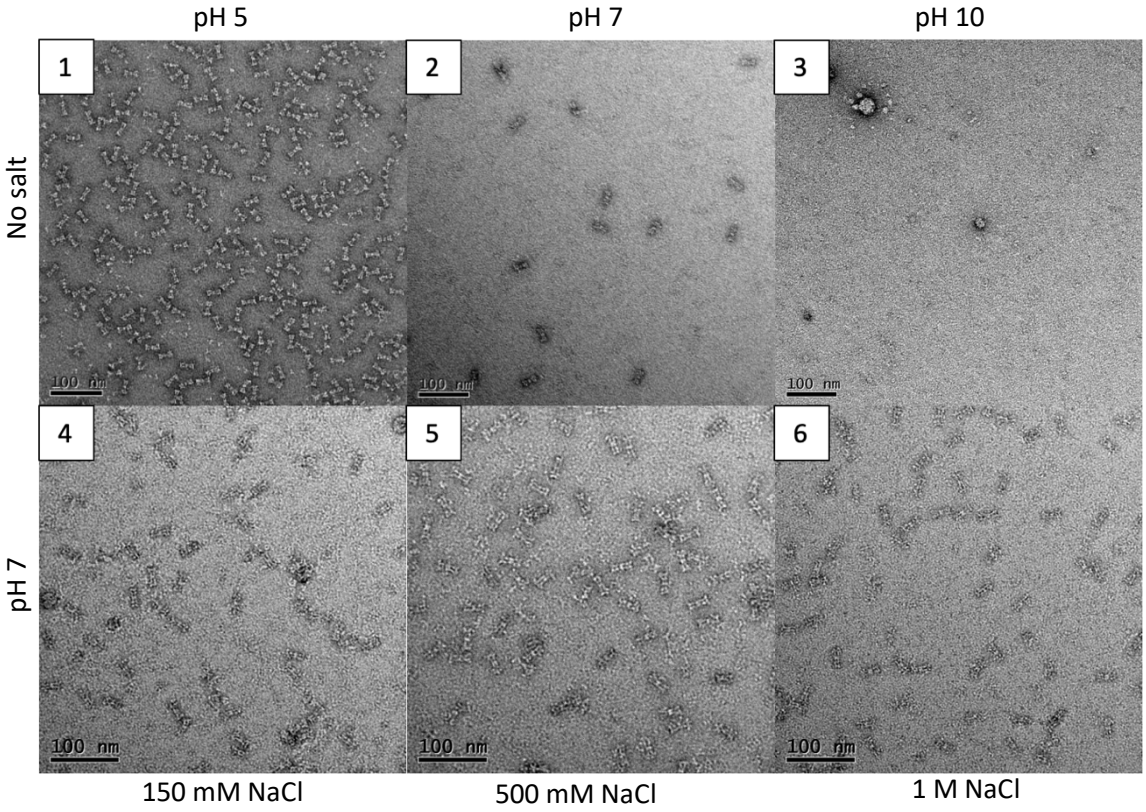


Figure 3.1.8: Negative stain images of $FliD_{cj}$ in 1) 50 mM Tris pH 5, 2) 50 mM HEPES pH 7, 3) 50 mM CAPS pH 10, 4) 50 mM HEPES 150 mM NaCl pH 7, 5) 50 mM HEPES 500 mM NaCl pH7, 6) 50 mM HEPES 1M NaCl pH7. Dumbbell-like shapes are decamers.

3.1.4 Expression and Purification of FliD from orthologues

$FliD_{pa}$ and $FliD_{sm}$ were first both overexpressed using the IPTG based method in BL21, followed by both being overexpressed in pLysS cells, but the expression was not high enough to purify. In instances when it was, the purified fractions did not contain the expected FliD decamer formations (see Appendix 6). So, an auto-induction method was utilised (see section 2.2.3) and both 50 kDa proteins were expressed sufficiently in the soluble fractions as shown on Figure 3.1.9.

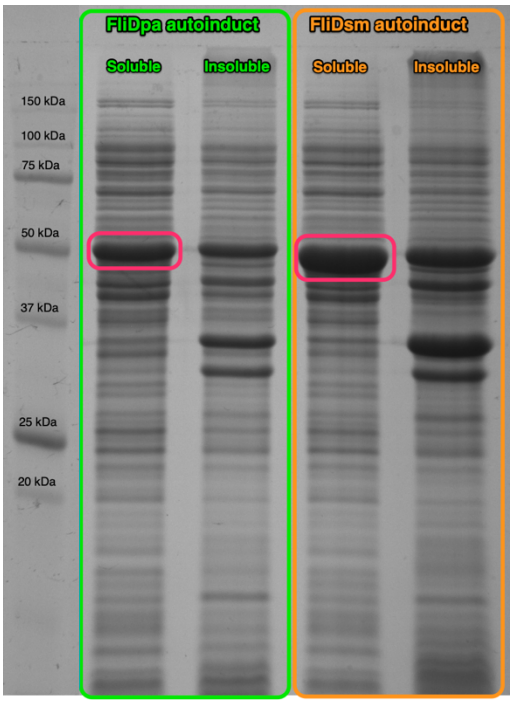


Figure 3.1.9: SDS-PAGE gel of $FliD_{pa}$ and $FliD_{sm}$ purified via auto-induction post cell lysis with soluble and insoluble fractions labelled. The relevant 50 kDa band is marked in magenta.

Both soluble fractions were purified via Ni-Nickel column and behaved similarly to $FliD_{cj}$ (Appendix 7). Gel filtration was attempted using a 75pg HiLoad 16/600 Superdex column instead of 200pg used for $FliD_{cj}$ with all the other parameters remaining the same. This was due to potential better separation of the smaller $FliD_{pa}$ and $FliD_{sm}$ proteins than with the larger bead size. Proteins separated in more than 2 peaks probably owing to more contaminants in the auto-induction sample, as illustrated in Figures 3.1.10 and 3.1.11.

Negative stain analysis of the fractions however, greatly varied from that of $FliD_{cj}$ purification. The first peak fractions for $FliD_{pa}$ A10-A11 (figure 3.1.12 top left) did not include any clear decamer formations. However, the second peak B12 (figure 3.1.12 top right) while lacking decamers had a very clear fraction of oligomer with 5-fold symmetry that could be top views of said decamers. For $FliD_{sm}$, while the first peak A9-A11 (figure 3.1.12 bottom left) contained a variety of different protein oligomers, the remaining peaks had either monomeric or denatured protein as shown in Figure 3.1.12 bottom right.

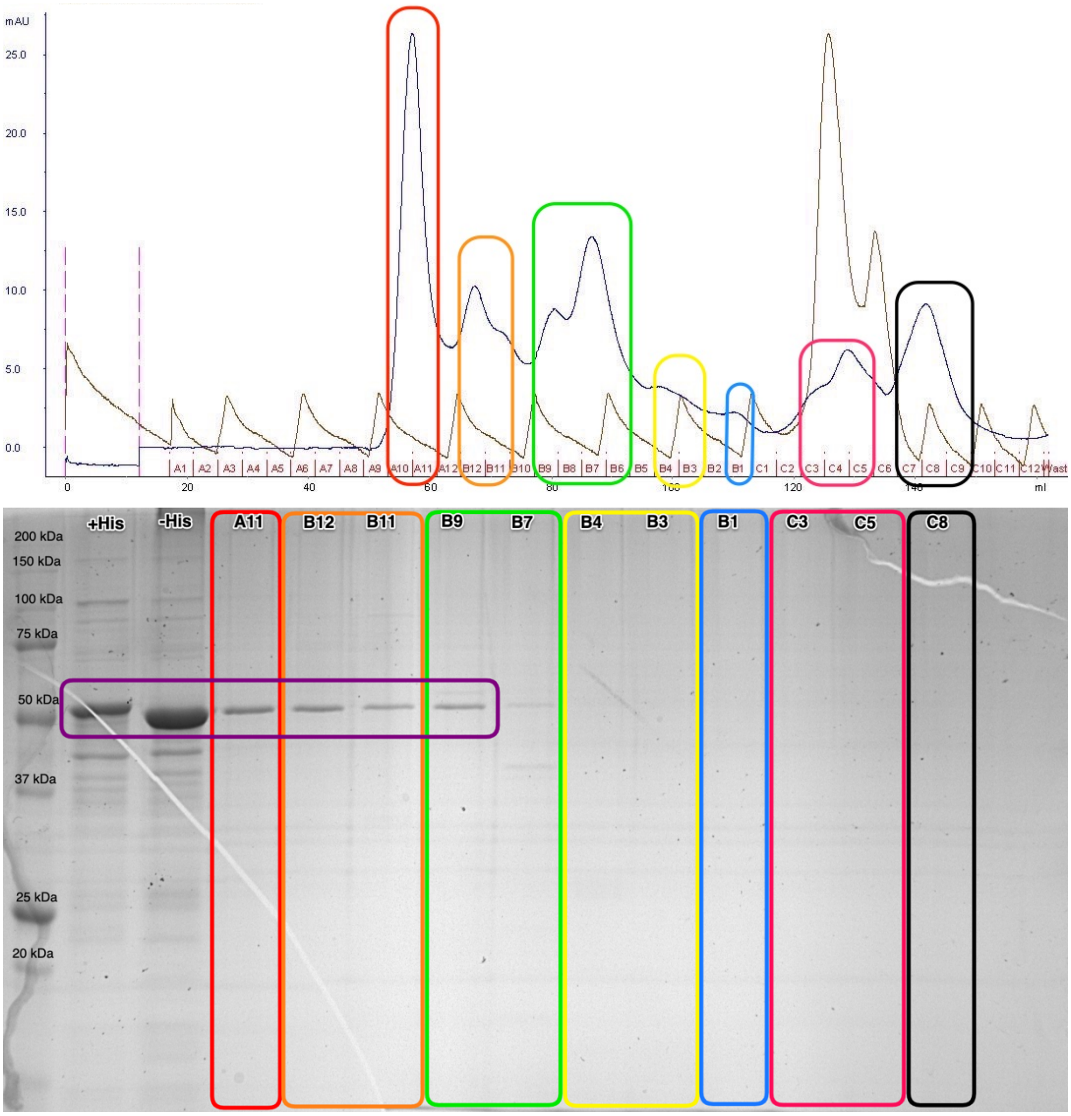


Figure 3.1.10: Chromatogram and SDS-PAGE gel of gel filtration column output for His-cleaved $FliD_{pa}$. Left to Right SDS-gel: 1: Protein gradient ladder, 2: Pre-thrombin incubation sample with His-tag, 3: Post-thrombin incubation sample with reduced molecular weight due to His-tag cleavage, 4-14: Elution fractions. Peaks colours correspond to the fractions of the gel.

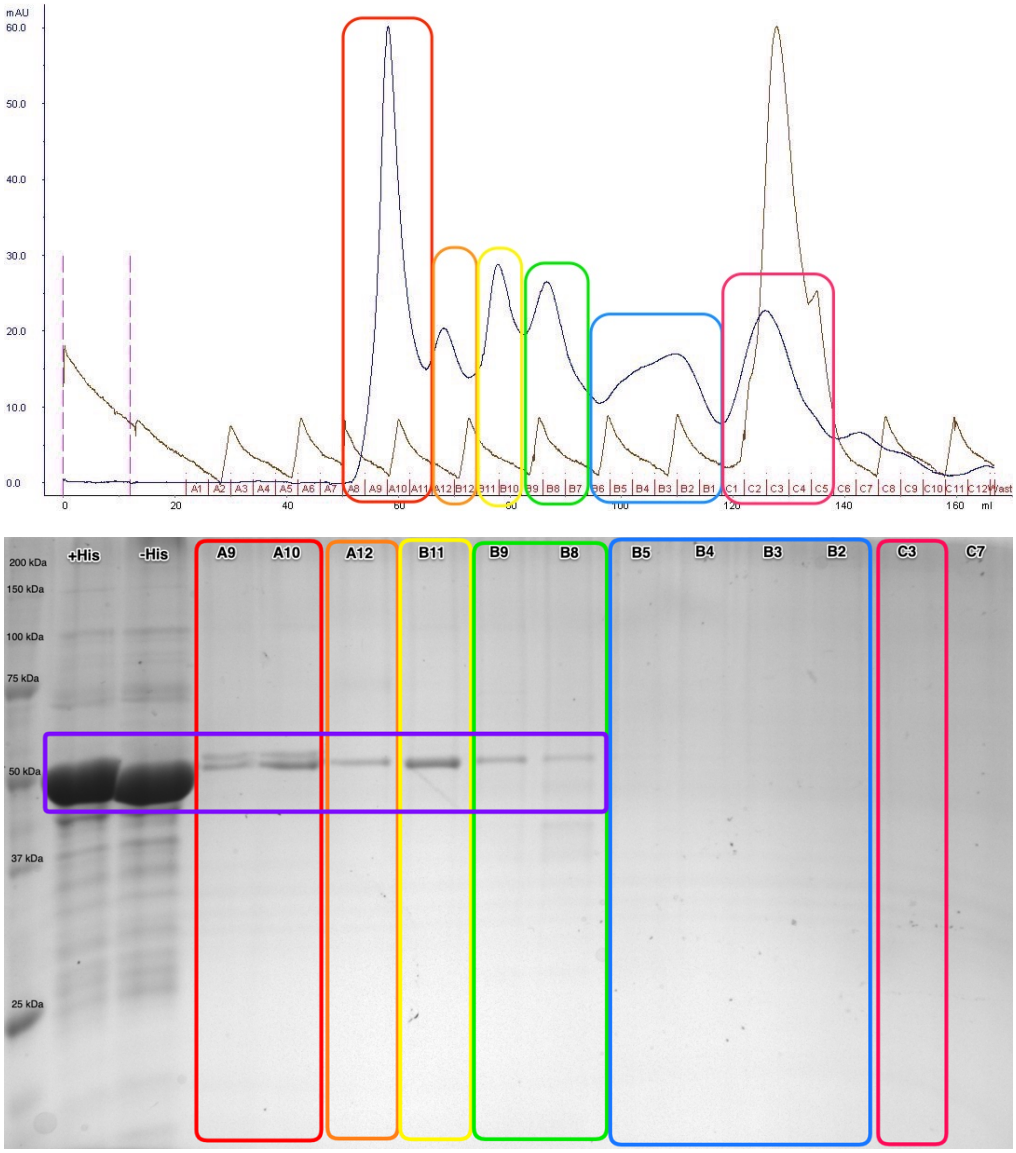


Figure 3.1.11: Chromatogram and SDS-PAGE gel of gel filtration column output for His-cleaved $FliD_{sm}$. Left to Right SDS-gel: 1: Protein gradient ladder, 2: Pre-thrombin incubation sample with His-tag, 3: Post-thrombin incubation sample with reduced molecular weight due to His-tag cleavage, 4-15: Elution fractions. Peaks colours correspond to the fractions of the gel.

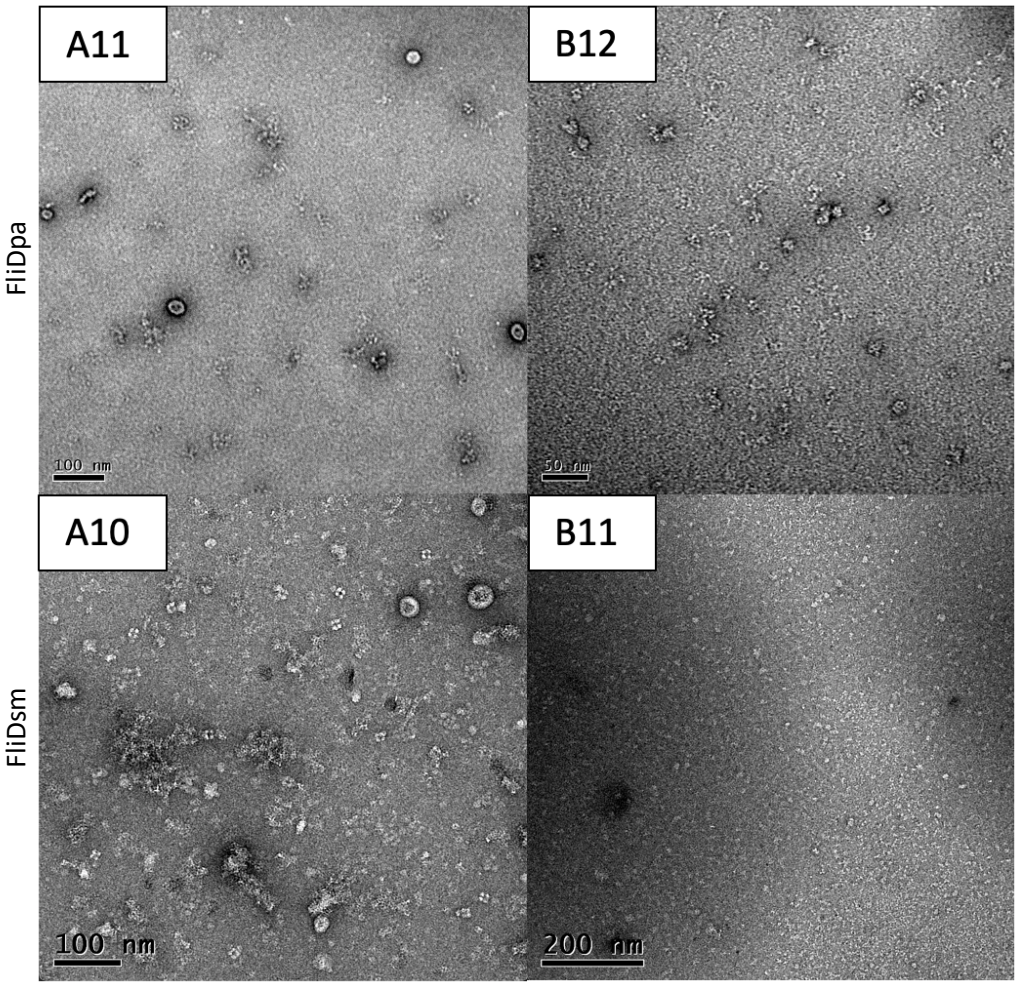


Figure 3.1.12: Negative stain images of $FliD_{pa}$ and $FliD_{sm}$ from gel filtration fractions. FliDpa A11 and $FliD_{sm}$ A10 are from the first peak. $FliD_{pa}$ B12 and FliDsm B11 are from the second peak.

$FliD_{hp}$ and $FliD_{ec}$ samples were overexpressed using the auto-induction method and soluble fractions purified through a Ni-Nickel column in a similar fashion to that of the above-mentioned samples, shown in Appendix 8. However, gel filtration using the similar protocol to $FliD_{cj}$, while successfully separating the sample into more than two peaks akin to $FliD_{pa}$ and $FliD_{sm}$ samples, did not yield any fractions containing either decamer or pentamer formations. The yield was also much less than that of all the previously purified proteins. While there are multiple reasons why the purification of these two proteins was not of equal quality, due to time constraints the optimisation could not be pursued within this work.

3.1.5 Negative Stain Data Collection

While both $FliD_{pa}$ B12 and $FliD_{sm}$ A10 samples were much more heterogenous than $FliD_{cj}$, there was enough of repeating oligomer formations to warrant an in-depth analysis. This was conducted via negative stain with parameters detailed in section 2.3.1. Particles were observed that resembled top views in both fractions as well as in the original $FliD_{cj}$ prep, which allowed the preliminary 2D classification to determine their lateral symmetry. This revealed that for both orthologues, particles with 5-fold symmetry and similar dimensions to that of $FliD_{cj}$ were present, as illustrated in Figure 3.1.13.

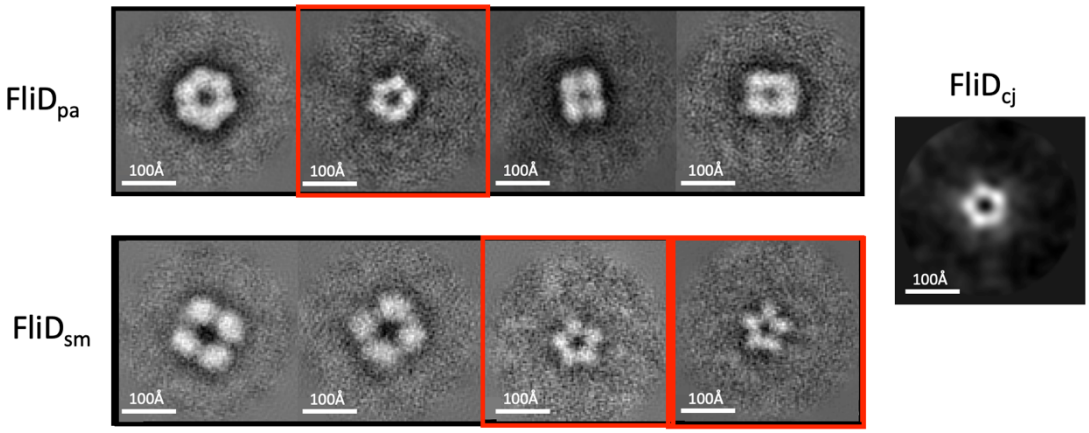


Figure 3.1.13: 2D classes obtained from negative-stain data of recombinant $FliD_{pa}$ and $FliD_{sm}$. $FliD_{pa}$ shows three different sets of particles: a class with particles showing 6-fold symmetry (~120 Å in diameter), a class showing smaller particles with 5-fold symmetry (~100 Å diameter), and classes of particles with 4-fold symmetry (~80 Å and ~110 Å in classes 3 and 4, respectively). $FliD_{sm}$ shows two different sets of particles: large particles with 4-fold symmetry (~150 Å in diameter), and smaller particles with 5-fold symmetry (~100 Å diameter). The red squares indicate classes with particles of similar shape and dimensions to the $FliD_{cj}$ structure, with distinctive five-fold symmetry. $FliD_{cj}$ negative stain top view is shown as a reference to the right. Adapted from (Al-Otaibi et al., 2020).

However, in the $FliD_{pa}$ sample additional particles were observed with 6-fold (16.4%) and 4-fold symmetry (58.5%), while in the $FliD_{sm}$ sample there was a large percentage of particles with 4-fold symmetry (52%). The dimensions of the particles in those 2D classes are larger than the FliD pentamer, and therefore it could not be concluded if these correspond to alternative oligomeric species, or to other negative

stain artifacts and/or non-specific aggregates. However, the presence of pentamers with similar dimensions to that of FliD (25.5% in $FliD_{pa}$ and 48% in $FliD_{sm}$) supports the hypothesis that FliD from these bacterial species is capable of forming a pentamer in solution. However, the percentages of pentamers are low in comparison to other fractions indicating that in solution FliD may adapt multiple stable oligomeric shapes (hexamer in $FliD_{pa}$) which may not necessarily be the native state.

3.1.6 DISCUSSION

Purification of FliD from a variety of species showed that the same purification methodology does not completely work for each sample. While $FliD_{cj}$ had a relatively simple recombinant protein prep, $FliD_{pa}$ and $FliD_{sm}$ required a change in inducing media to be overexpressed and even then, not at the same purity as $FliD_{cj}$. $FliD_{ec}$ and $FliD_{hp}$ did not overexpress even with the change of media, indicating that a separate purification strategy might be necessary for each FliD homologue. This could occur due to various reasons.

The most obvious difference between these FliD homologues is the presence of a species-specific extra domain: $FliD_{cj}$ D4 localized in between the D1 helical bundle domain and $FliD_{hp}$ D4 localized in the D2-D3 head domains. The presence of this domain significantly changes the molecular weight of the protein which could in turn affect its properties, create additional contacts in the oligomer that would improve or detriment their stability in solution. Predicted isoelectric points (pI) of FliD from different species compared in this section vary with the lowest being $FliD_{cj}$ at 4.76 and highest being $FliD_{pa}$ at 6.52. This might affect the protein behaviour during purification dependent on the choice of buffer as oligomerization appears to be affected by the pH in $FliD_{cj}$.

In previous studies, evidence suggesting different stoichiometries of the native state for the cap complex in different species was based on crystallographic symmetries of truncated proteins. In this work an arguably conflicting evidence is provided against this, by demonstrating that FliD is capable of adopting a pentameric stoichiometry in at least three different proteobacteria families, in addition to other

stable oligomeric states. $FliD_{pa}$ is the only other FliD for which a stable hexameric oligomerization state was observed *in vitro* (Postel et al., 2016). This work shows evidence to support that observation, however the presence of a pentamer fraction shows that this hexameric oligomerization state is not limiting. This finding contributed to our flagellar elongation model discussed in later chapters. However, these are preliminary results and further investigation is required to understand FliD oligomerization mechanisms *in vitro* and *in vivo*.

3.2 $FliD_{CJ}$ Structure Determination

3.2.1 Introduction

Single particle Cryo-EM is a rapidly developing method of determining protein structure in a vitrified state without the need for crystallization. Due to the difficulty of crystallization of the full length FliD protein and thus all available crystal structures having truncated terminal domains, the purified uniform sample of $FliD_{cj}$ was used to determine the full-length structure. As there is only one low-resolution cryo-EM studies of the cap complex in *S. enterica*, the structure determined in this section is the first near-atomic EM structure of FliD (Maki-Yonekura et al., 2003; Yonekura et al., 2000). After solving the structure, the observed domains were analysed in isolation. There are many alternative structural biology techniques to determine the structure of protein domains constructed through molecular cloning. The aim was to generate a $FliD_{cj}$ construct of just head domains to test if it behaves similarly to other FliD crystal structures and forms hexamers or tetramers in a crystalline lattice. Isolating the *C. jejuni* specific D4 domain would allow us to determine its structure via X-ray crystallography or NMR and removing it would show if it plays an essential role in $FliD_{cj}$ function and stability.

3.2.2 Cryo-EM Data Collection and Processing

The data was collected and processed as detailed in sections 2.3.2 – 2.3.4 and the processing pipeline in figure 3.2.1. The protein formed discrete particles in vitreous ice, and 2D classification confirmed that it adopted the dumbbell shaped structure previously reported for $FliD_{se}$ (Maki-Yonekura et al., 2003; Yonekura et al., 2000). While the preferred orientation was that of the side-view, a significant subset of particles adopted top-view orientations, with clear 5-fold symmetry. The structure of the full complex was solved to 4.7 Å resolution, as illustrated in Figure 3.2.2. However, the map shows a wide range of local resolution, with the leg domain well defined and with visible density for the large sidechains, while the head domain is more poorly defined as shown in Figure 3.2.3.

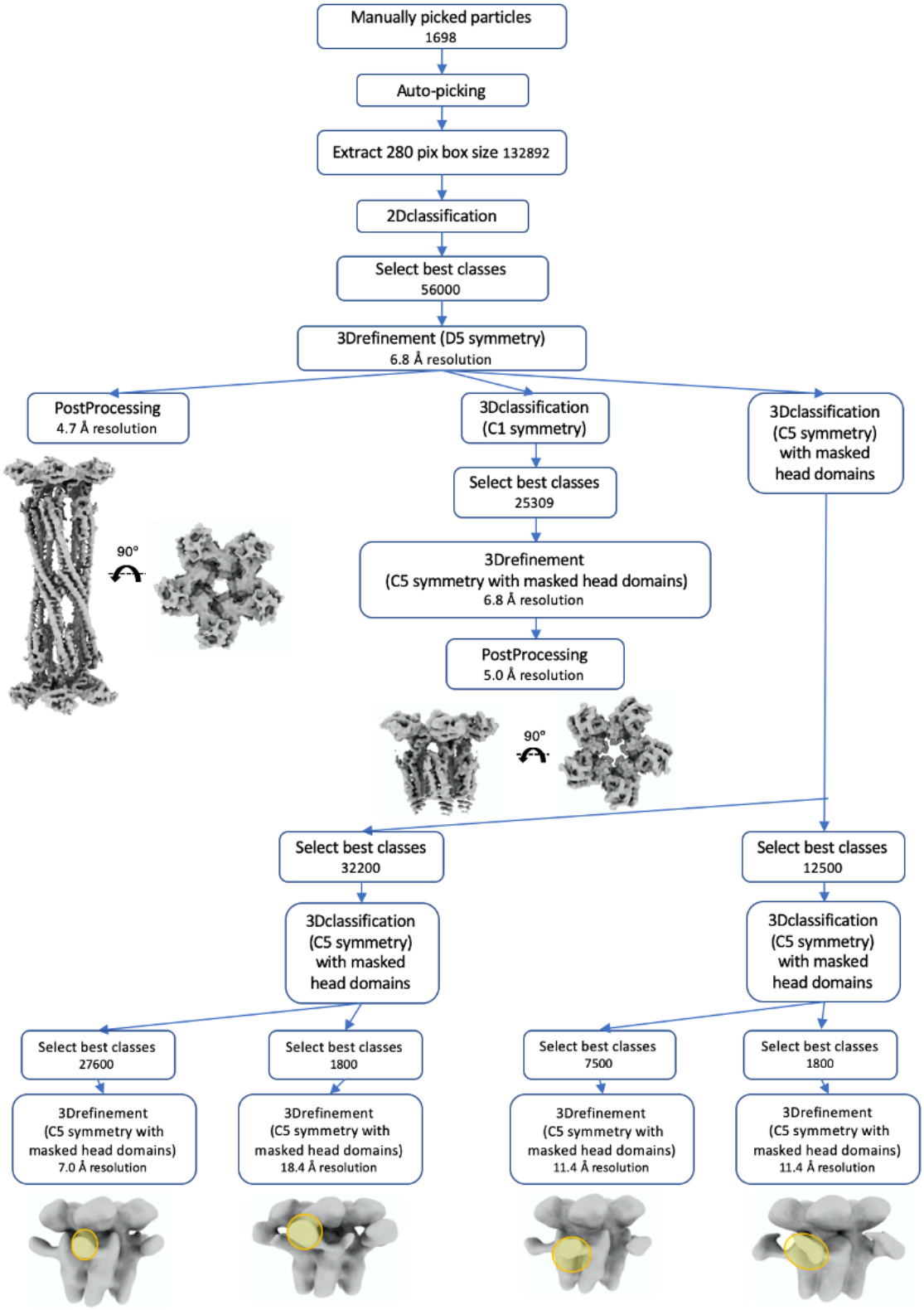


Figure 3.2.1: Workflow of $FliD_{cj}$ processing in Relion.

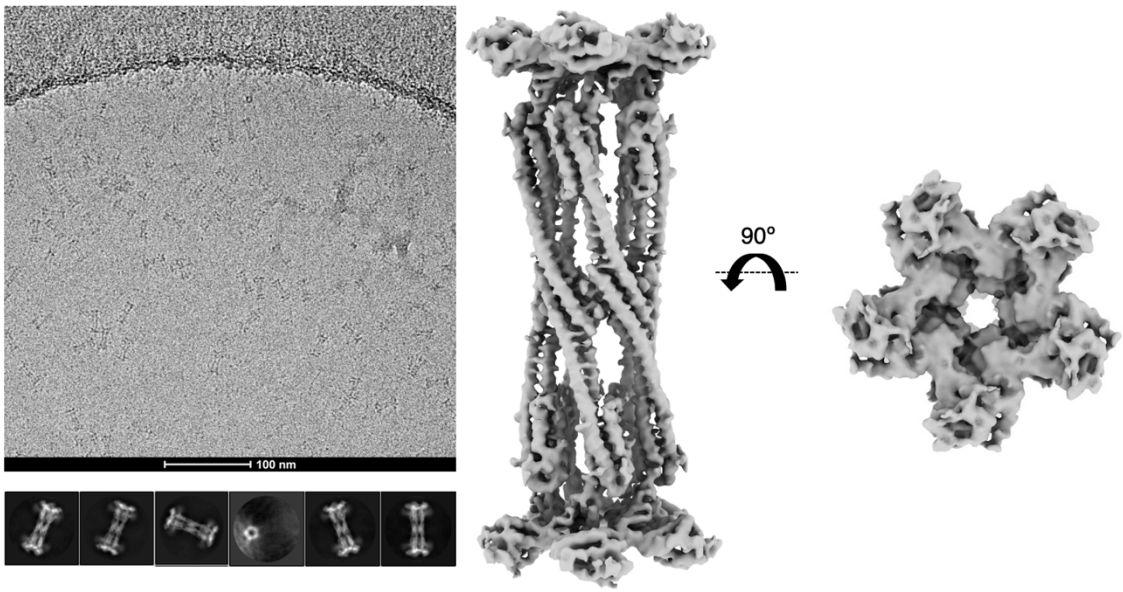


Figure 3.2.2: **Left:** Cryo-electron micrograph of the $FliD_{cj}$ complex. Large particles (~ 30 nm x 5 nm) are visible. Below are 2D classes generated from ~56000 particles. **Right:** Cryo-EM map of the full complex. Side view (left) and top view (right).

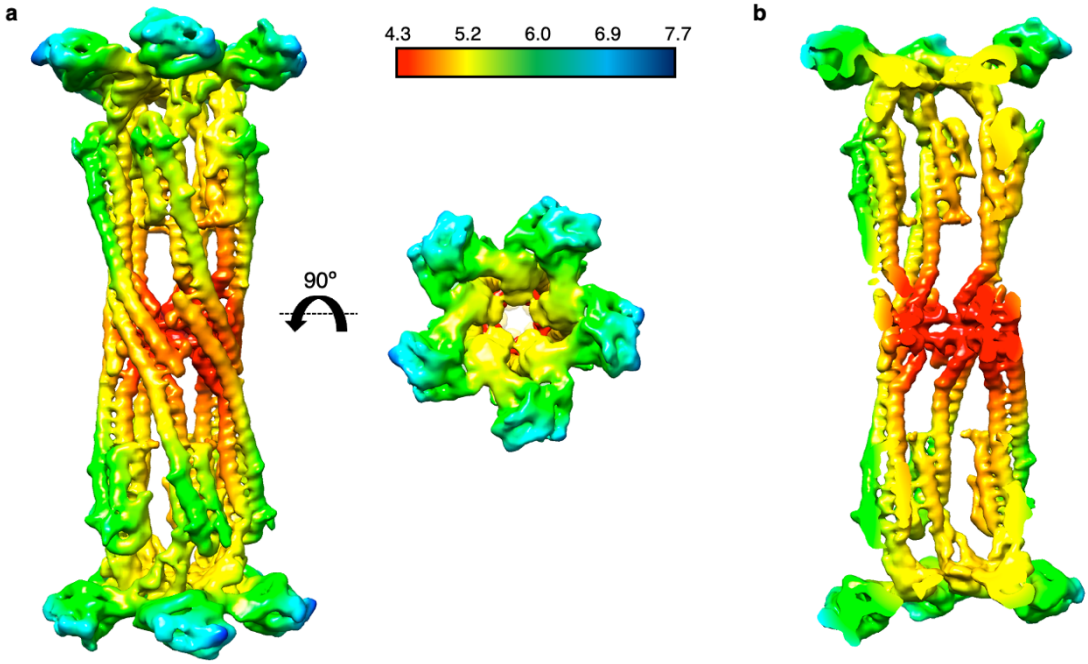


Figure 3.2.3: (a) Local resolution map for decamer $FliD_{cj}$ showing the uneven resolution distribution. Side view (left) and top view (right). The colour to resolution scale is shown above. (b) Cross section view of map in (a) with the colour to resolution scale as shown in (a). The colour to resolution scale is shown above.

The FSC plot of both maps, illustrated in Figure 3.2.4, shows a bump in what supposed to be a smooth curve, reflecting the uneven local resolution. This could suggest that there is some flexibility in the head domains, or that each subunit adopts slightly different conformations, which were averaged out in the map because of the applied 5-fold symmetry.

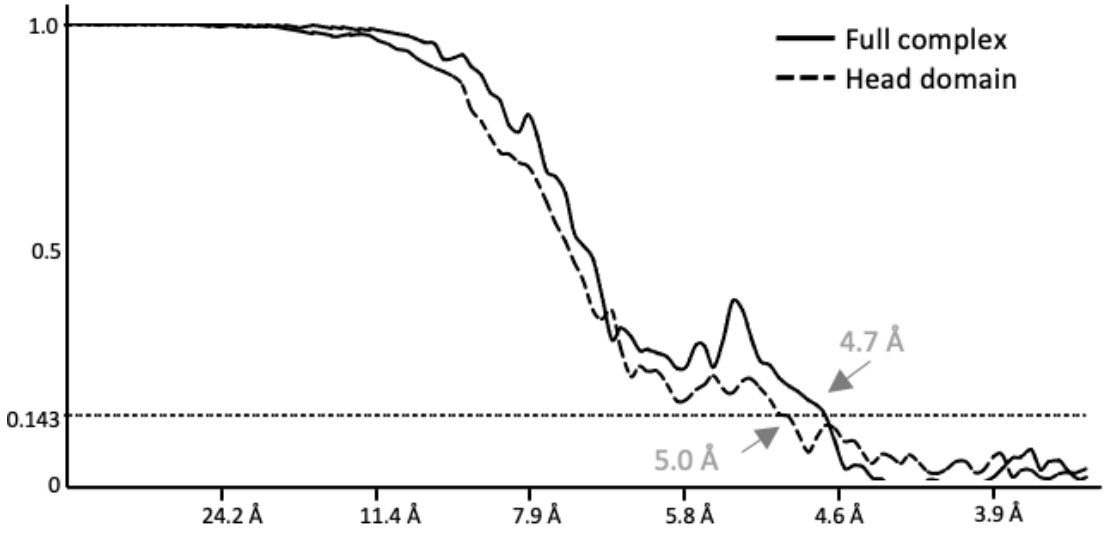


Figure 3.2.4: FSC plots for the maps in 3.2.2 and 3.2.5 showing the resolution to be 4.7 Å and 5.0 Å respectively.

To obtain a better map for this region of the complex, a focused refinement was performed on the head domain only, leading to a map at 5.0 Å resolution, illustrated in Figure 3.2.5. When the contouring was highly reduced, an extra domain bound to D1 was visible and assumed to be the extra *C. jejuni* specific $FliD_{cj}$ D4 domain in consensus with the alignment data in section 3.1.1.

3.2.3 $FLID_{CJ}$ MODEL BUILDING AND REFINEMENT

$FliD_{cj}$ model was built as detailed in section 2.3.4 and with parameters detailed in Table 3.2.1. The $FliD_{cj}$ complex possesses an overall architecture similar to $FliD_{se}$, consisting of ten subunits, with two pentamers interacting in a "tail-to-tail" orientation, through the leg domains as illustrated in Figure 3.2.6. A pentamer is about 170 Å in height (the decamer is ~300 Å) and 130 Å in width with a 20 Å lumen. Using

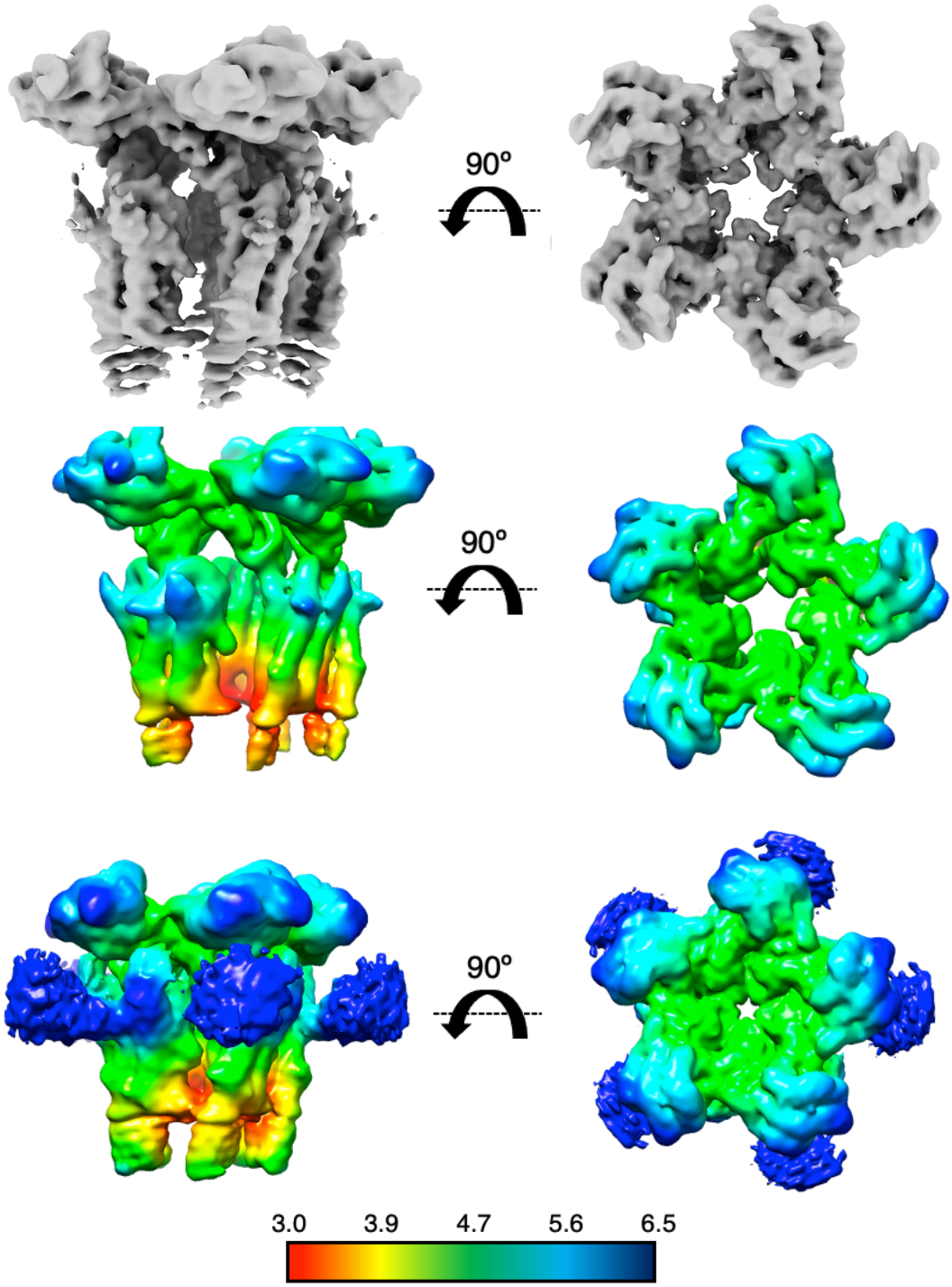


Figure 3.2.5: Cryo-EM map of the complex obtained from a masked refinement of full $FliD_{cj}$ map in Figure 3.2.2 allowing for better resolution in the head domains. Below are the local resolution maps of the same contour and low-contour rendering of the head domain-focused cryo-EM map with the D4 domain at a low resolution (6.5 Å). Side view (left) and top view (right). The colour to resolution scale is shown below.

the focused head domains map, atomic model was generated for this region of $FliD_{cj}$ (Domains D1, D2, and D3), based on the PHYRE2 server modelling of crystal structure of $FliD_{ec}$ (PDB ID: 5H5V) (Postel et al., 2016).

Table 3.2.1: Cryo-EM data collection, refinement, and validation statistics for $FliD_{cj}$.

	$FliD_{cj}$ (EMDB-10210) (PDB: 6SIH)
Data collection and processing	
Microscope	Titan Krios
Magnification	36232
Voltage (kV)	300
Camera	K2 Summit
Pixel size (Å)	1.38
Symmetry Imposed	D5 (full map), C5 (head map)
Defocus range (μm)	−1.0 to −2.6
Total dose ($eÅ^{-2}$)	41
Number of micrographs	1223
Total particles used	55967
Map Resolution (Å)	4.7 (full map) 5.0 (head domain)
Refinement	
Map sharpening *B* factor ($Å^2$)	−285
Model composition	
Non-hydrogen atoms	37320
Protein Residues	4880
B factors ($Å^{-1}$)	
Protein	148
R.m.s. deviations	
Bond lengths (Å)	0.005
Bond angles (°)	0.543
Validation	
MolProbity score	1.85
Clashscore	8.36
Poor rotamers (%)	0.00
Ramachandran plot	
Favored (%)	94.13
Allowed (%)	5.87
Disallowed (%)	0.00

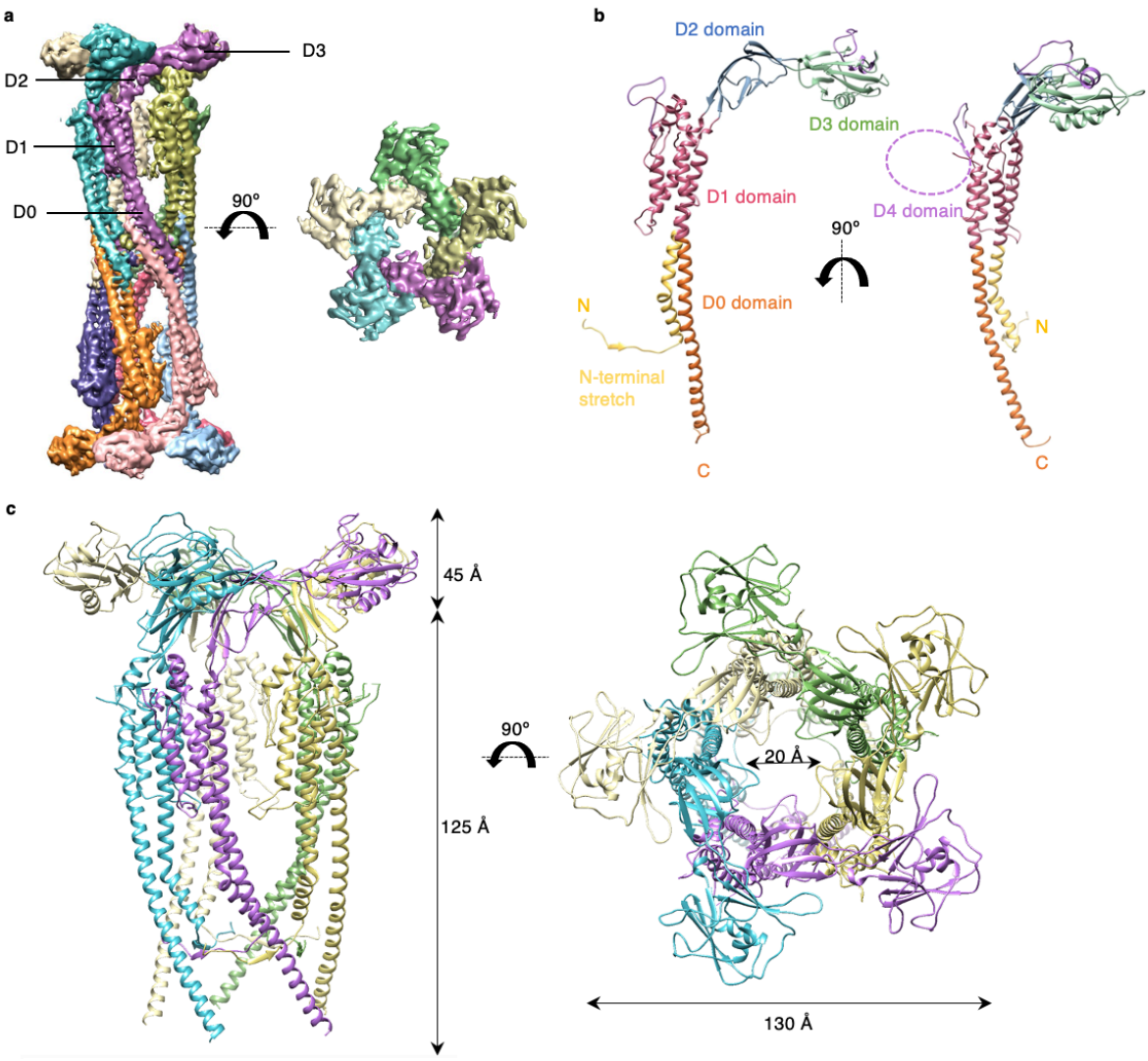


Figure 3.2.6: The cryo-EM map of $FliD_{cj}$, segmented and coloured by subunit. (a) A side view of the full complex map is shown on the left, and a top view of the head domain map is shown on the right. Domains are indicated with labelling lines. (b) Cartoon representation of the full-length $FliD_{cj}$ monomer, coloured according to domain organization. The purple dotted circle indicates the position of the D4 domain. (c) Cartoon representation of the $FliD_{cj}$ pentamer, corresponding to the intact cap complex, with respective measurements. The lumen was measured at the bottom as the width between the N-terminal stretches. In our model the D4 domain is not visible as it was not built in. For reference it should be localized as indicated in panel (b). Side view (left) and top view (right) are shown, and colour coded as in (a). Adapted from (Al-Otaibi et al., 2020).

The N and C termini were manually built into the density in Coot by inserting a helix into the density and mutating the residues to the ones in the sequence. The

helical register was determined by fitting large residues to their corresponding densities and examples of such are shown with labels in Figure 3.2.7. N-terminal stretch specifically was refined to the EM density in Rosetta by Dan Farrell from DiMaio Lab, revealing the short ß-strand.

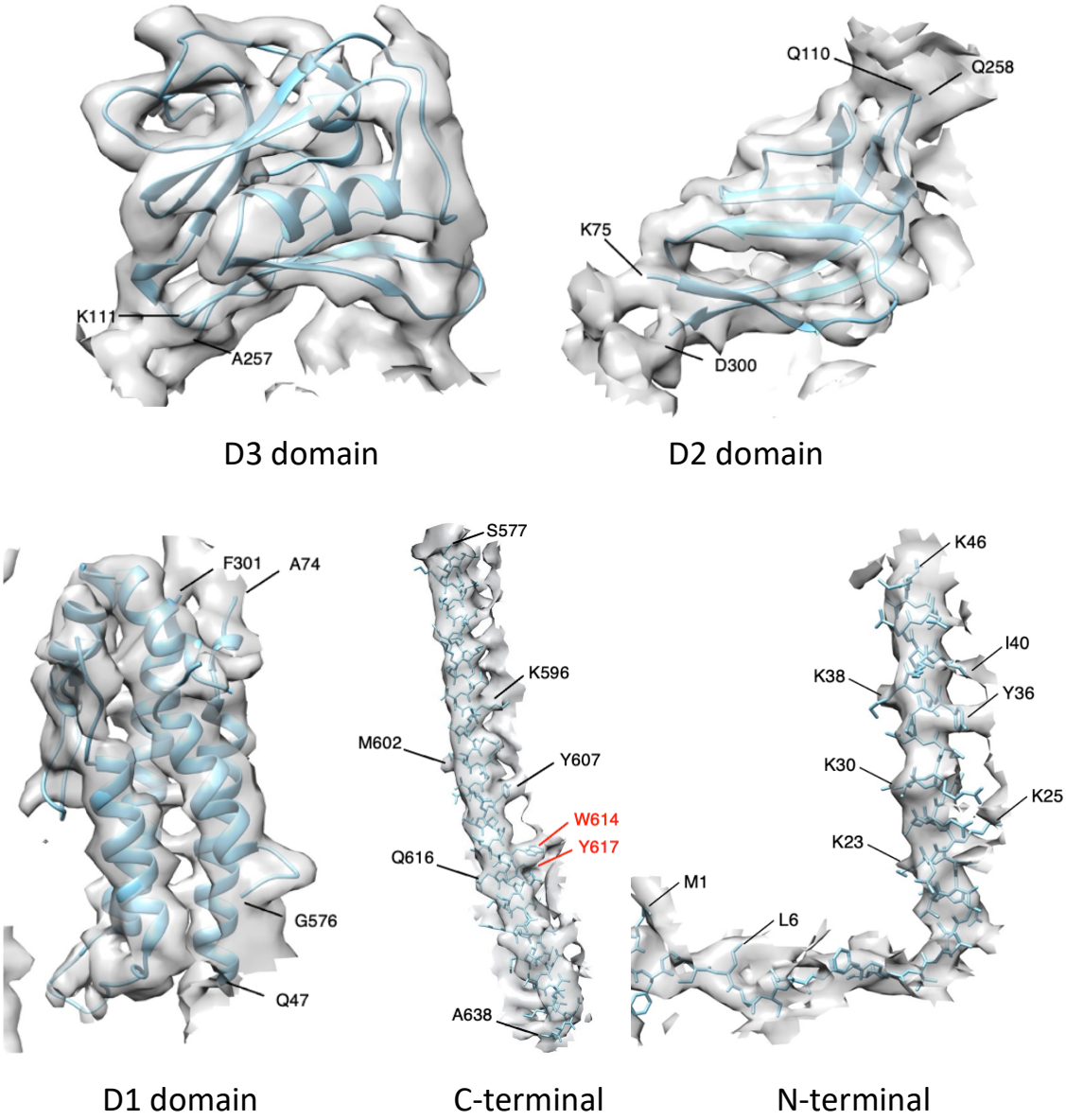


Figure 3.2.7: Map to Model Fit of the $FliD_{cj}$ model. The N and C termini were manually built into the density in Coot by inserting a helix into the density and mutating the residues to the ones in the sequence. The helical register was determined by fitting large residues to their corresponding densities and examples of such are shown with black labels. The red labels correspond to co-varying residues shown in Table 3.1.1. D1, D2 and D3 domains were modelled in PHYRE2 into the density of map in figure 3.2.2(for D1) and figure 3.2.5 (for D2 and D3). The start and end residues for each domain as well as their locations are labelled.

The model-map FSC determines the fit of the model to the EM map and while the accuracy of the fit of the head domains is lower than that of the leg domains, they were modelled and fitted as globular domains while the legs were built into the density manually. The FSC plots are shown in Figure 3.2.8.

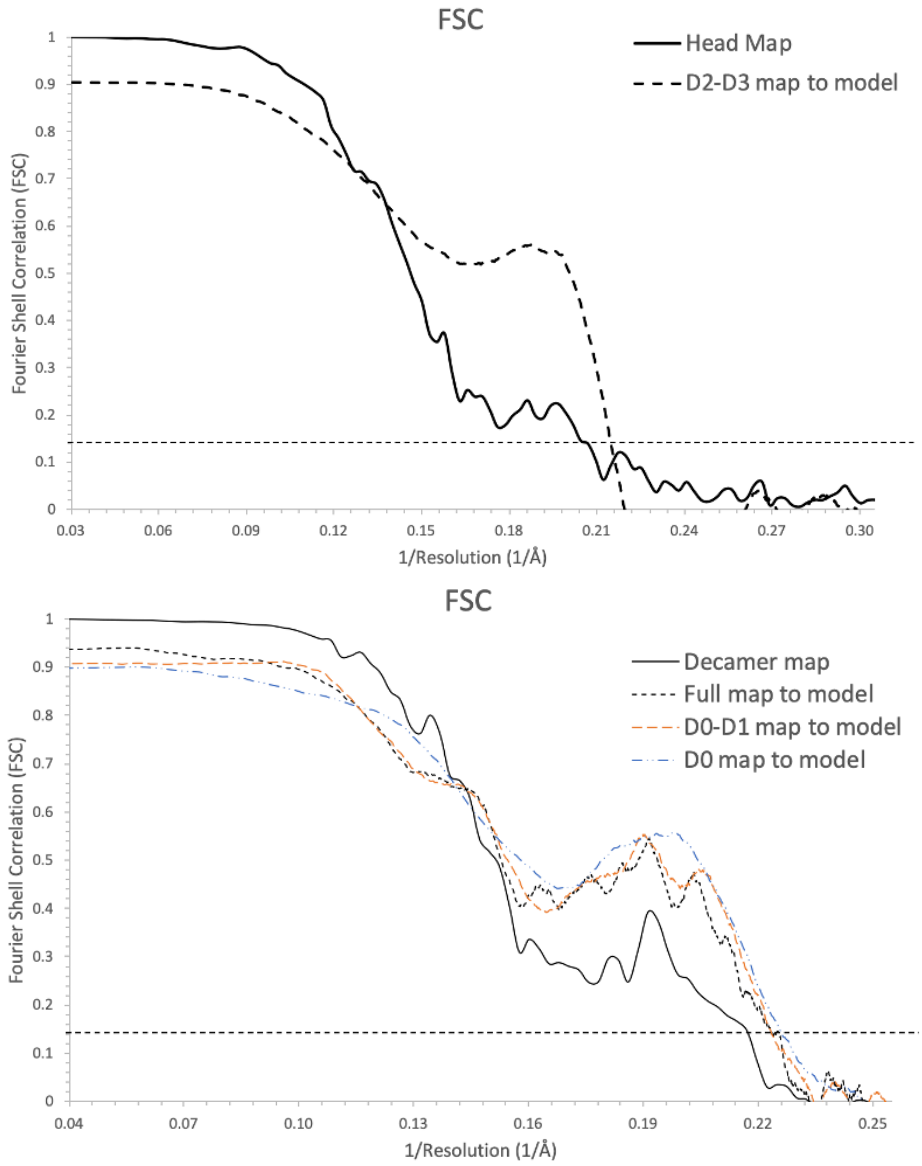


Figure 3.2.8: Map to model FSC maps for different domains of $FliD_{cj}$. On the left graph the resolution of the map of the head domain (5.0 Å) is compared to the modelled density of D2-D3 domains at the 0.143 FSC cut off point (4.8 Å). On the right graph the resolution of the map of the full decamer (4.7 Å) is compared to the full model (4.6 Å), D0-D1 leg domains (4.6 Å) and D0 terminal regions (4.5 Å).

The $FliD_{cj}$ structure shows that the FliD protomer folds in on itself in a v-shape, which results in N and C termini next to each other in the leg domain, as illustrated in the topology diagram in Figure 3.2.9.

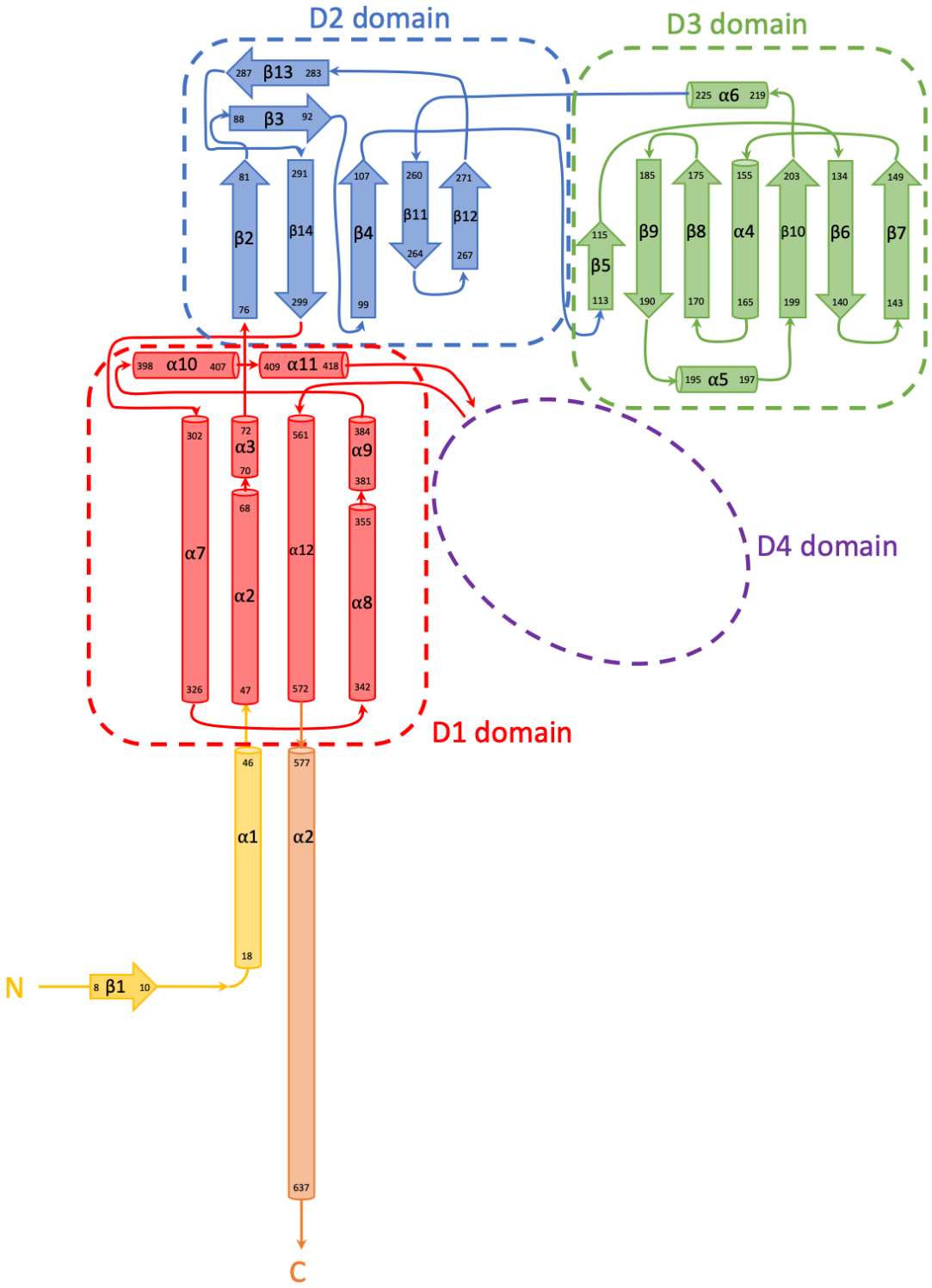


Figure 3.2.9: Topology map of FliDcj drawn from the structure built in 3.2.6.

The overall architecture, as proposed previously, consists of a D0 domain formed by a long-coiled coil, consisting of two helices located at the termini. A four-helix bundle forms the D1 domain. Connected to the D0–D1 leg domains are D2–D3 domains, rich in anti-parallel β-sheets, forming the head (Cho et al., 2017, 2019a; Postel et al., 2016; Song et al., 2017). This overall architecture is similar to that of the flagellin and hook, and in agreement with the previously reported structures of the FliD head domain (Imada, 2018). In particular, the D0 and D1 domains in the flagellin are largely structurally similar to the D0 and D1 domains in FliD in their function. However, while it was predicted that the D0 domain consists of a two-helix coiled-coil, as present in the flagellin and hook, this structure reveals that the N-terminal 17 residues are extended into a stretch that folds under and behind the monomer, interacting with the preceding subunit via a short β-strand. As a consequence, the C-terminal helix of the coiled-coil is not partnered with the N-terminus, but instead interacts with that of another molecule through hydrophobic interactions, forming the pentamer-to-pentamer interface. This intriguing architecture likely explains the strong tendency of FliD to form tail-to-tail complexes during isolation, as observed in $FliD_{se}$ and $FliD_{cj}$.

$FliD_{cj}$ possesses a long insert within the D1 helix bundle, not present in other orthologues. Secondary structure prediction indicates that this insert is likely globular, leading to the hypothesis that it forms an additional domain, termed D4. This type of domain insertion is not unusual, and has been observed in other FliD orthologues, as well as in flagellin and hook proteins (Cho et al., 2019a; Imada, 2018; Matsunami et al., 2016; Shibata et al., 2019; Wang et al., 2017). As per the workflow in Figure 3.2.1, the dataset was classified focused on the head domains and obtained the different conformations of the D4 extra domain, illustrated in Figure 3.2.10. However, due to the low resolution of the domain map an atomic model could not be built. This domain was modelled using Phyre2 server, which predicted a mostly ß-sheet globular conformation with two alpha helices, like that of FliD head domains D2 and D3. The role of this D4 domain is not known, but it could be related to $FliD_{cj}$'s capacity to bind to heparin, a feature involved in *C. jejuni* adherence (Freitag et al., 2017).

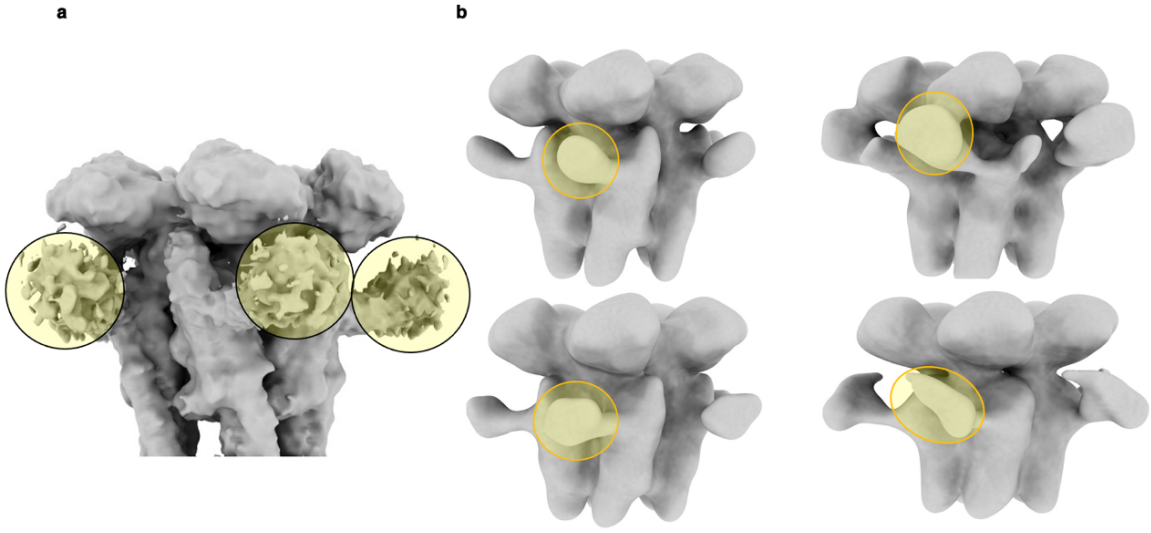


Figure 3.2.10: (a) Low-contour rendering of the head domain-focused cryo-EM map. Density attributed to the D4 domain is indicated with yellow circles. (b) Further 3D classification revealed distinct conformations of the D4 domain but remained at low resolution because of the low number of particles in each class.

3.2.4 Cloning and Analysis of FliD$_{CJ}$ domains

Prior to solving the structure of FliD$_{cj}$ domains, the protein sequence was modelled to existing FliD structures and primers designed to create three separate constructs: D2-D3 head domains, D1-D4 helical bundle and D4 domain in isolation as detailed in Figure 3.2.11, section 2.2.2 and Appendix 2. D2-D3 construct was designed for crystallization to observe if FliD$_{cj}$ forms tetramers or hexamers in a crystal lattice, unlike its cryo-EM form, similar to FliD from other species. The position of D1 relative

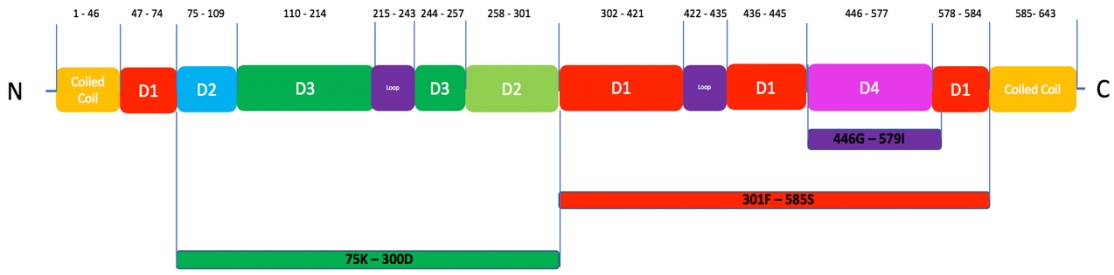


Figure 3.2.11: Primer design lead by modelled domain distribution in FliD$_{cj}$. D2-D3 head domain construct in Green (75K-300D), D1-D4 construct in Red (301F-585S) and D4 construct in purple (446G-579I).

to D2-D3 domains is different in $FliD_{cj}$ from other cap structures potentially due to the flexible hinge or the additional D4 domain protruding out of D1. The D2-D3 construct would also show if this flattening is a common occurrence in truncated FliD or a unique structural feature. D1D4 helical bundle and D4 domain in isolation were designed to also determine their structures through crystallography or NMR in the case of D4 domain alone, as it is below the maximum sample size threshold.

The first cloning step aiming to remove the N-terminal unwanted sequence from the three constructs succeeded for D2-D3 and D1-D4 but failed for D4 as observed through transformation or PCR products in *E. coli*. This could occur due to many reasons, some of which could be low acceptance of plasmid or low plasmid count. The successful constructs were then digested with *DpnI* to remove methylated DNA and the second cloning step removing the C-terminal unwanted sequence was conducted resulting in the success of D1-D4 but failure of D2-D3.

After solving the structure of $FliD_{cj}$ the leg domain assignment appeared to be slightly different from that of the model. Therefore, while the primers for the D2-D3 construct were correct, primers for D1-D4 incorporated a bit of the C-terminal coiled coil and primers for D4 were off by a significant margin, which could have contributed to issues of cloning it as illustrated in Figure 3.2.12.

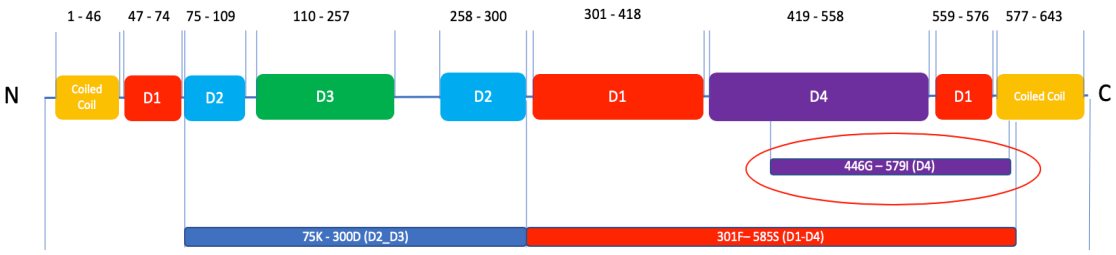


Figure 3.2.12: Post-structure solving primer locations. While D2-D3 and D1-D4 constructs are relatively correct, the D4 domain primers (circled in red) are in a wrong location.

Due to time constrains, constructing the correct primers for D4 on its own and testing them was not possible. Cloning the second set of primers for D2-D3 construct was attempted a few more times, but no viable colonies were observed. This could be due to the primer being incorrect or interfering with the kanamycin resistance cassette on the original plasmid and would require more experimentation to ascertain the

issue. Thus, out of the three constructs, D1-D4 was successfully cloned and used for further structural studies.

30 kDa $FliD_{cj}$D1D4 construct (301F-585S) was expressed and purified as detailed in section 2.2.3. behaving similarly to the full length $FliD_{cj}$. The soluble fraction was purified via Ni-Nickel column (Appendix 9). Gel filtration via Hi-Load Superdex 200 pg column was attempted using the same conditions as $FliD_{cj}$ and it appears that the construct did not separate properly and a smaller bead column such as 75 pg is required as illustrated in Figure 3.2.13.

To observe the effect of the D4 domain on stability and folding of $FliD_{cj}$, 50 kDa $FliD_{cj}$noD4 construct (1M-417S, 559G-643N) was designed and synthesized from Bio Basic as per section 2.2.2 and expressed and purified as detailed in section 2.2.3. The soluble fraction was purified via Ni-Nickel column (Appendix 10) and gel filtration via Hi-Load Superdex 200 pg column was attempted using the same conditions as $FliD_{cj}$ as illustrated in Figure 3.2.14.

The protein eluted in a large peak fractions B12-B10 and a smaller peak of B6-B4. While the smaller peak had some formations, they were not the expected $FliD_{cj}$ decamers. However, the large peak fractions contained nonuniform formations usually observed in void fraction and circular clumps of protein illustrated in Figure 3.2.15. These clumps look like a group of decamers stuck together tightly as shown by red circles. As the only difference between this construct and full length $FliD_{cj}$ is the lack of D4 domain, it might be the factor shielding a hydrophobic surface which became exposed upon deletion of the domain. Resuspending the sample in detergent and observing its behaviour further using EM could clarify this observation.

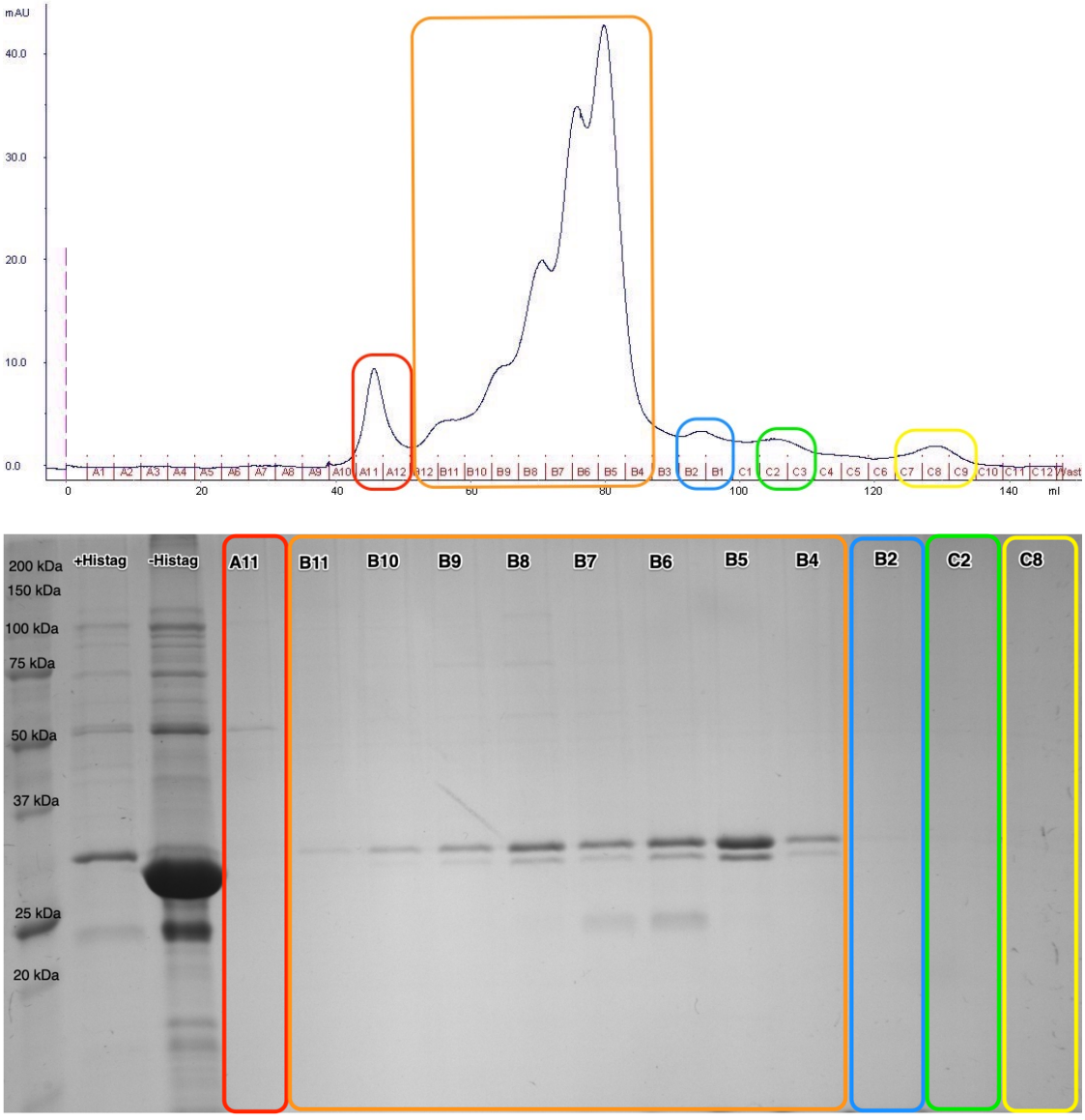


Figure 3.2.13: Chromatogram and SDS-PAGE gel of gel filtration column output for His-cleaved $FliD_{cj}D1D4$. Left to Right SDS-gel: 1: Protein gradient ladder, 2: Pre-thrombin incubation sample with His-tag, 3: Post-thrombin incubation sample with reduced molecular weight due to His-tag cleavage, 4-15: Elution fractions. Peaks colours correspond to the fractions of the gel.

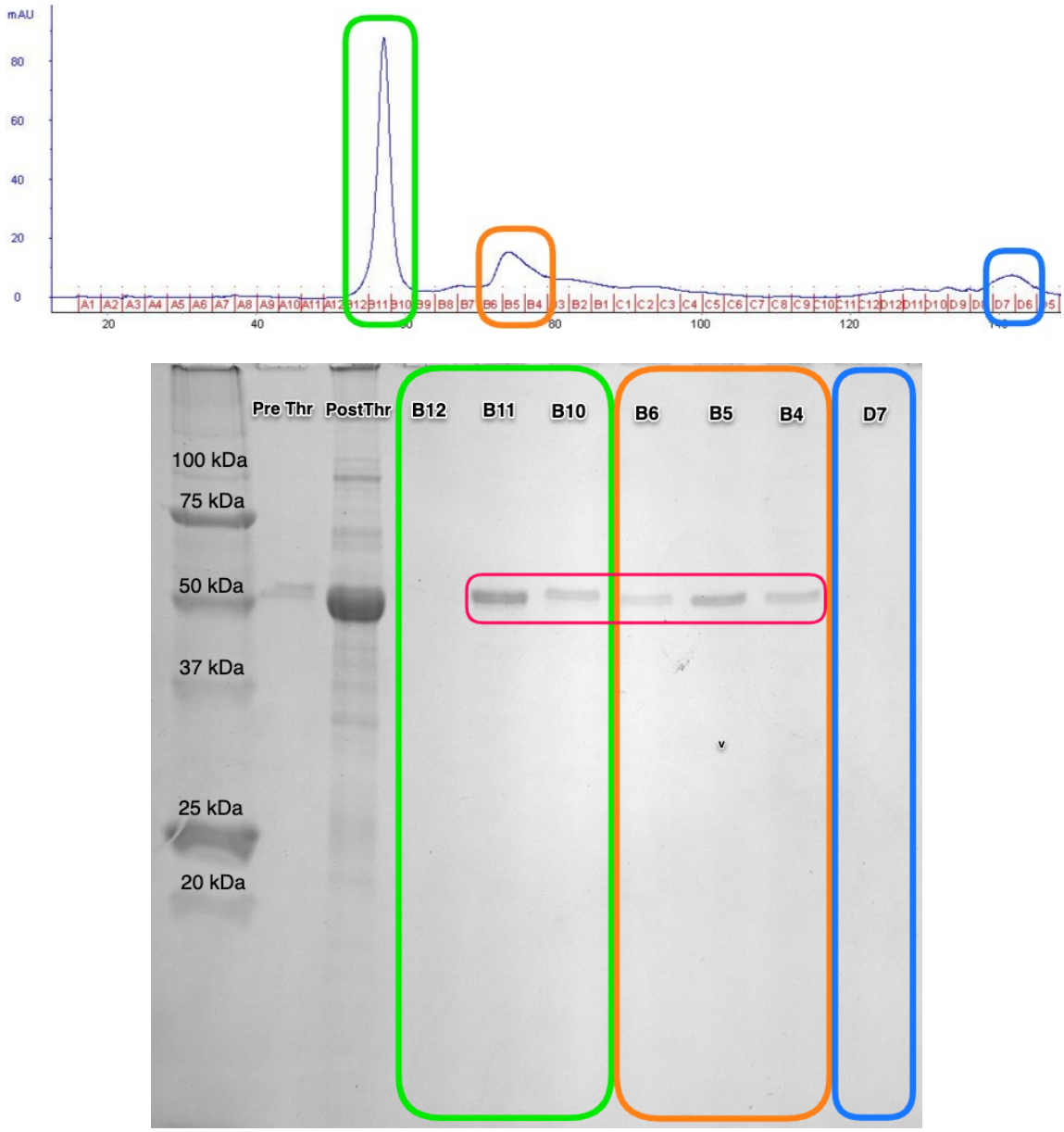


Figure 3.2.14: Chromatogram and SDS-PAGE gel of gel filtration column output for His-cleaved $FliD_{cj}$noD4. Left to Right SDS-gel: 1: Protein gradient ladder, 2: Pre-thrombin incubation sample with His-tag, 3: Post-thrombin incubation sample with reduced molecular weight due to His-tag cleavage, 4-10: Elution fractions. Peaks colours correspond to the fractions of the gel.

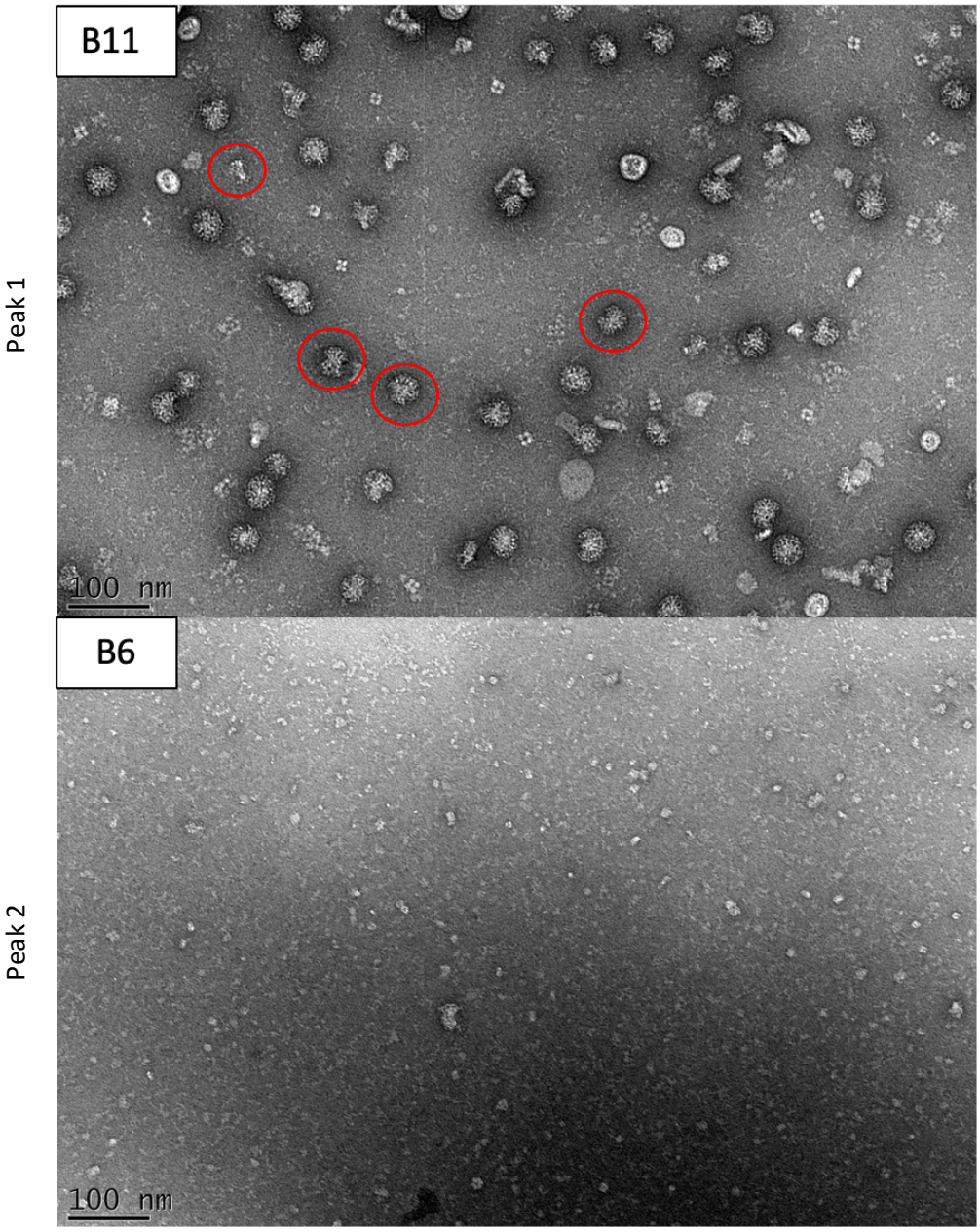


Figure 3.2.15: Negative stain images of $FliD_{cj}noD4$ from Peak 1 (B11) and Peak 2 (B6). Potential decamer clumps are shown in red.

3.2.5 Discussion

The structure in this section, at 4.7 Å, represents the highest resolution of FliD full-length complex to date. Nonetheless the crystal structure of the head domain, corresponding to domains D2–D3, has been reported for a range of species, including *S. enterica*, *P. aeruginosa*, *E. coli*, *S. marcescens*, and *H. pylori* (Cho et al., 2017, 2019a;

Maki-Yonekura et al., 2003; Postel et al., 2016; Song et al., 2017). In all orthologues, the structure is very similar, with RMSD values ranging from 1.5 Å to 2.5 Å to that of $FliD_{cj}$ as illustrated in Figure 3.2.16. In the *E. coli* orthologue, domain D1 was also present in the structure. It consists of a 4-helix bundle, and this structure is very similar to that of $FliD_{cj}$, with a RMSD of 1.5 Å between the two structures. When comparing the dimensions of these structures, the diameters of all complexes are similar, around ~140 Å. However, the dimension of the lumen differs significantly between structures, with $FliD_{cj}$ and $FliD_{se}$ having a central lumen of ~20 Å, while $FliD_{pa}$ and $FliD_{ec}$ have a lumen of ~50 Å and ~40 Å, respectively, and $FliD_{sm}$ a ~15 Å lumen, as shown in Figure 3.2.17.

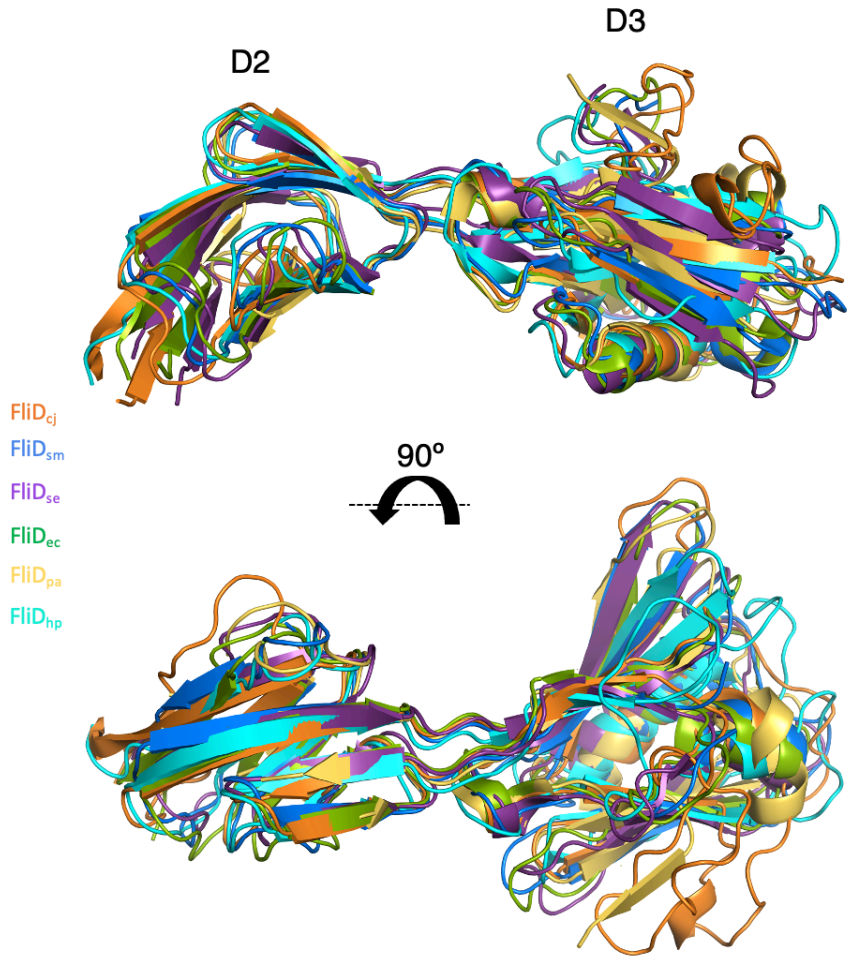


Figure 3.2.16: Overlay of the FliD D2–D3 domains structures from *C. jejuni* ($FliD_{cj}$, 6SIH—this study, orange), *S. marcescens* ($FliD_{sm}$, 5XLJ, blue), *S. enterica* ($FliD_{se}$, 5H5T, purple), *E. coli* ($FliD_{ec}$, 5H5V, green), *P. aeruginosa* ($FliD_{pa}$, 5FHY, yellow), and *H. pylori* ($FliD_{hp}$, 6IWY, cyan).

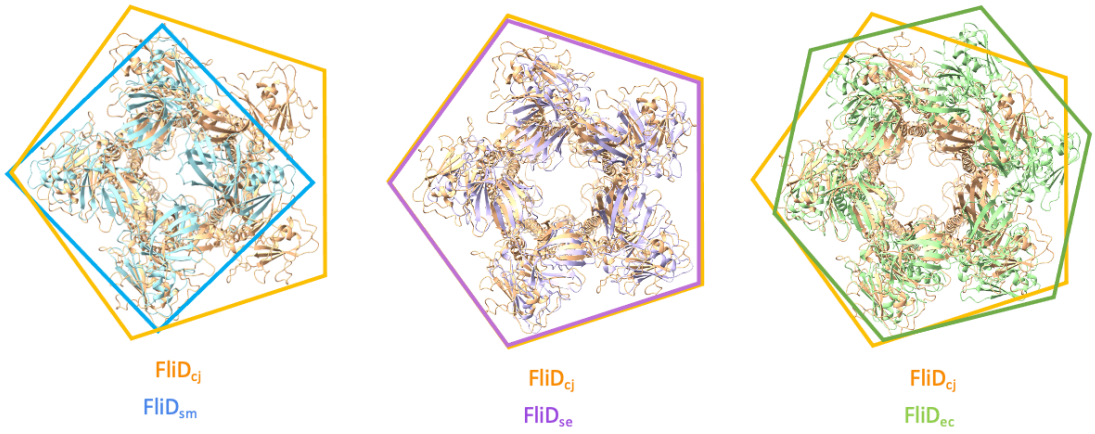


Figure 3.2.17: Alignments of X-ray crystallography derived oligomeric structures of the head domains from $FliD_{sm}$, $FliD_{se}$, and $FliD_{ec}$, to that of the $FliD_{cj}$ pentamer. The diameter of the lumen as well as the outer diameter of the capping protein is similar in the pentamer formed by $FliD_{se}$, but significantly smaller in the tetramer formed by FliDsm, and larger in the hexamer formed by $FliD_{ec}$.

Nonetheless, the position of D1 relative to that of D2–D3 is different in $FliD_{ec}$ compared to $FliD_{cj}$ as illustrated in Figure 3.2.18. This could occur because the hinge between D1 and D2 is flexible. Alternatively, it could be due to the fact that in $FliD_{cj}$ the D4 domain protrudes from the D1 helical bundle and thus a more planar conformation, as observed in $FliD_{se}$, may be sterically clashing with D2–D3 in the case of $FliD_{cj}$, leading to a conformation distinct to that of other species. In the cryo-EM map, FliD forms a pentameric architecture, consistent with the low-resolution cryo-EM structure of $FliD_{se}$, with a similar overall architecture consisting of two pentamers in a head-to-tail arrangement. Even in the case of $FliD_{se}$, which crystallized as a pentamer, while the overall dimensions are similar to that of the head domains of the $FliD_{cj}$ pentamer, in the *S. enterica* orthologue the pentamer is flattened compared to that of $FliD_{cj}$. This led to a conclusion that there is a difference between the cryo-EM structure in this work and crystallography determined FliD structures available to date, not only in the number of subunits but also in the angle of their interaction. It remains to be verified if this corresponds to a general artifact of the crystal structures of D2–D3 domains in isolation or corresponds to a difference in structure between orthologues.

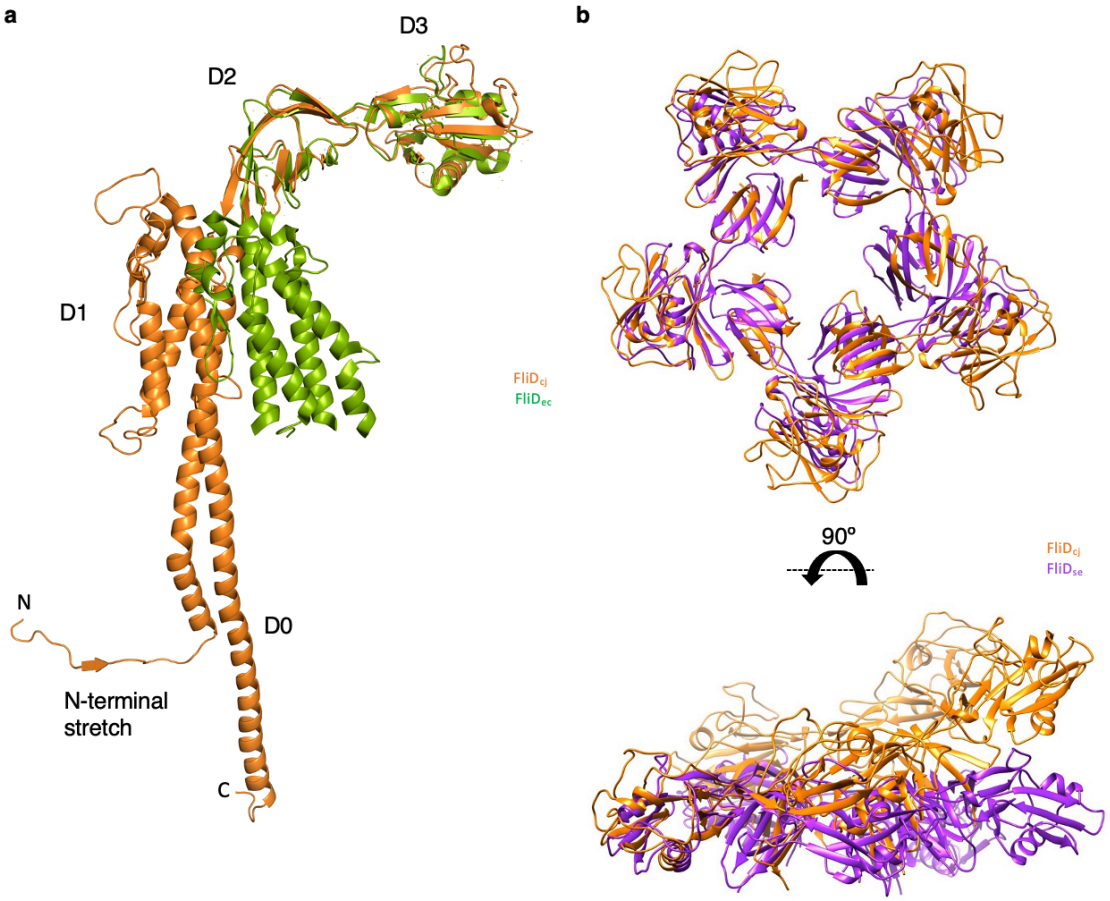


Figure 3.2.18: (a) Overlay of the FliD structures from *C. jejuni* ($FliD_{cj}$, this study, orange) and *E. coli* ($FliD_{ec}$, 5H5V, green). (b) Alignments of X-ray crystallography derived oligomeric structure of the D2-D3 domains from $FliD_{se}$ (purple) to that of the $FliD_{cj}$ pentamer (orange). It is aligned using Chimera Matchmaker and $FliD_{cj}$ was aligned to Chain A of $FliD_{se}$. The angle at which the D2 and D3 domains interact are different with the $FliD_{se}$ being more planar than $FliD_{cj}$.

In some of the previously reported FliD D2–D3 crystal structures, there are some subtle differences between subunits of the oligomer. A recently published paper on structure of the hook-basal body with the hook capped with FlgD also showed a mismatch in symmetry which allowed for conformational change of the protein while building the flagellar hook (Johnson et al., 2021). As the function and pentameric oligomerization of FliD is similar to FlgD, as illustrated in Figure 3.2.19, with the exception of building the flagellar filament instead of the hook, the same mismatch may be applied to $FliD_{cj}$ while attached to the growing filament end. An extensive structural study of native flagellar tips needs to be undertaken to confirm if this

symmetry mismatch occurs in native $FliD_{cj}$ as well, as it was not observed in our reconstruction.

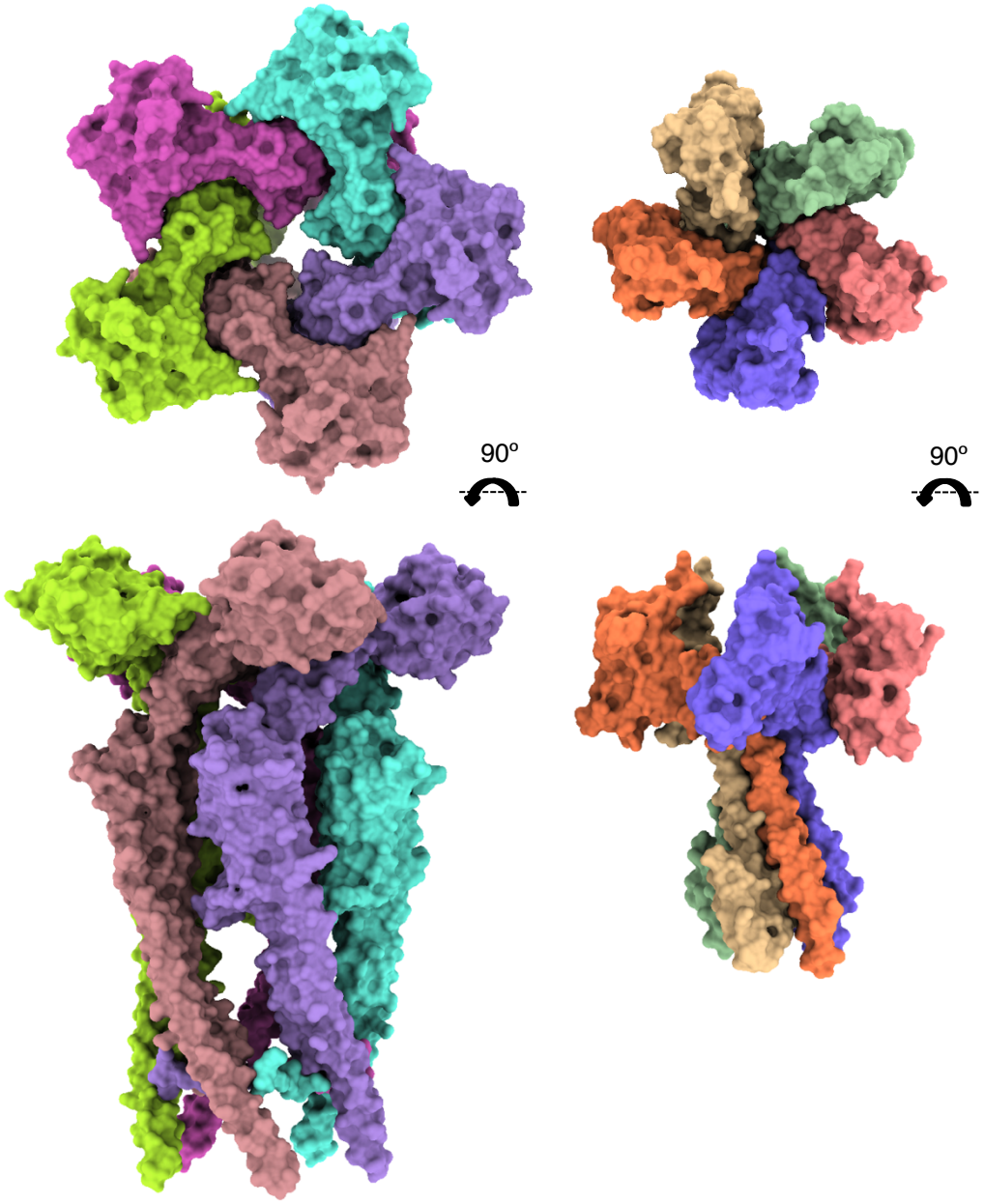


Figure 3.2.19: Structures of $FliD_{cj}$ (PDB: 6SIH) and FlgD (PDB: 7BHQ) compared. $FliD_{cj}$ pentamer is on the left and FlgD is on the right. Individual monomers are represented by different colours. Top view on top and side view on the bottom.

Through cloning, it was determined that $FliD_{cj}$ can still form decamer-like structures in solution upon deletion of the D4 domain, however clumped together, potentially due to the exposure of the hydrophobic surface of the D1 domain. It is

intriguing that $FliD_{cj}$ oligomerises without D4, although it is not certain into what oligomer or if the sample is homogenous. As the EM structure of $FliD_{cj}$ has a different angle of interaction between D1 and the head domains, will the head domains adapt a more planar conformation now that the D4 domain is removed? Or is the angle a species-specific factor of $FliD_{cj}$? These questions should be investigated through a purification and structure determination of the D4-less construct. D4 domain localization in the D1 helical bundle is *C. jejuni* specific, as in *H. pylori* the D4 domain is connected to the D3 head domain instead (Cho et al., 2019a). While both FliD have the same function of flagellar filament elongation, the extra domain may be important in some sort of *C. jejuni* specific function. Generating a D4-less construct could upon further experimentation be useful to test this hypothesis.

3.3 Structure/Function Analysis of $FliD_{CJ}$

3.3.1 Introduction

Protein structure is often directly linked to its function, so analysing one could often provide insights into the other. The structure of $FliD_{cj}$, solved in the previous section, has an overall domain architecture similar to that of the flagellin and hook, and is in agreement with the previously reported structures of the FliD head domains (Imada, 2018). It also reveals previously uncharacterized D0 domain, which consists of N and C terminal coiled coils and interacts with the growing end of the filament (Maki et al., 1998). However, this $FliD_{cj}$ structure is incomplete, as a ~140 amino acid *C. jejuni* specific domain D4 was observed in our map as a protrusion out of the D1 helical bundle, but the resolution was too low to build it de-novo.

The main function of FliD capping protein is the assembly of a functional flagellum, which in turn is responsible for motility. If FliD is not present, the cells lose their motility completely (Diószeghy et al., 2004; Imada et al., 1998). However, it can be complemented by adding purified recombinant FliD which was observed to be interchangeable in between certain species (Inaba et al., 2013). To investigate the FliD phenotype in motility, a genetically stable *C. jejuni* 81116 strain was used, from which the recombinant *fliD* sequence was isolated and used to solve $FliD_{cj}$ structure. A knockout plasmid was generated to remove the gene from *C. jejuni* genome and a complement plasmid to introduce the gene back. This complement plasmid was also used to introduce point mutations into the gene and observe the effect of different mutants on motility and flagellar morphology.

3.3.2 Construction of Mutant Strains in C. jejuni

$FliD_{cj}$ knockout mutant (*ΔfliD*) strain was designed and cloned as per section 2.4.1. The vector was designed such that spontaneous double crossover into the pseudogene region *cj0046*, by homologous regions present in the pCmetK plasmid with the *C. jejuni* 81116 genome would result in the replacement of the majority of the open reading frame of *fliD* with the kanamycin resistance cassette, allowing a means of selection. Amplified fragments for the plasmid were tested in 1% agarose gels and

extracted as per Figure 3.3.1. pGEM32ZFΔ$fliD$Kan^R plasmid was ligated and transformed into *E. coli*. The colonies were tested using the agarose gel for containing the plasmid of the correct length as shown in Figure 3.3.2. The positive control (isolated knockout plasmid) should be the same size as the tested plasmid while the negative control must be the size of the pGEM3ZF plasmid without any inserts (150 bp), which is the case in our test. All colonies contained the full-length plasmid but the direction of the inserted FliD$_{cj}$ top and bottom inserts still needed to be tested.

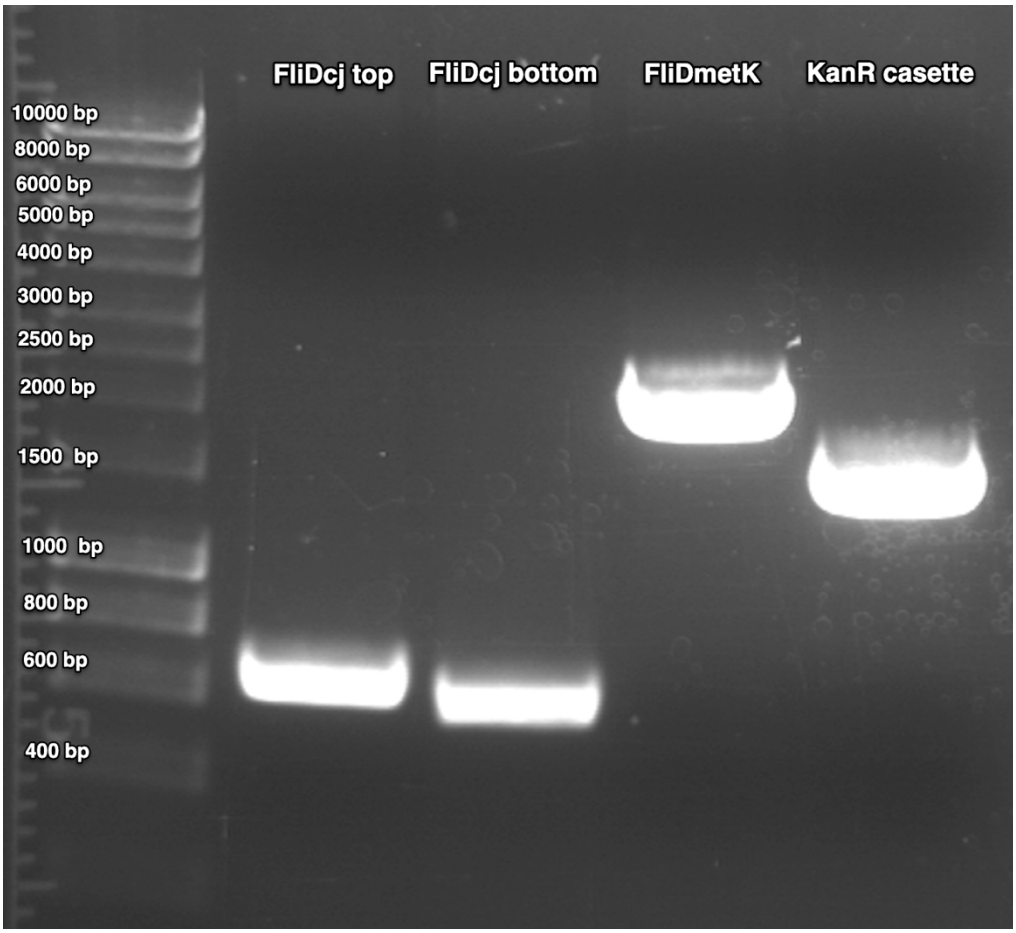


Figure 3.3.1: Agarose gel containing amplified fragments for the pGEM32ZFΔ$fliD$Kan^R plasmid. From left to right: DNA ladder, top FliD$_{cj}$ fragment containing adaptor region and upstream of *fliD* gene (600 bp), bottom FliD$_{cj}$ fragment containing adaptor region and downstream of *fliD* gene (550 bp, FliDmetK – a full *fliD* sequence isolated from the genome to be used for complement plasmid (2000 bp) and KanR kanamycin resistance cassette (1500).

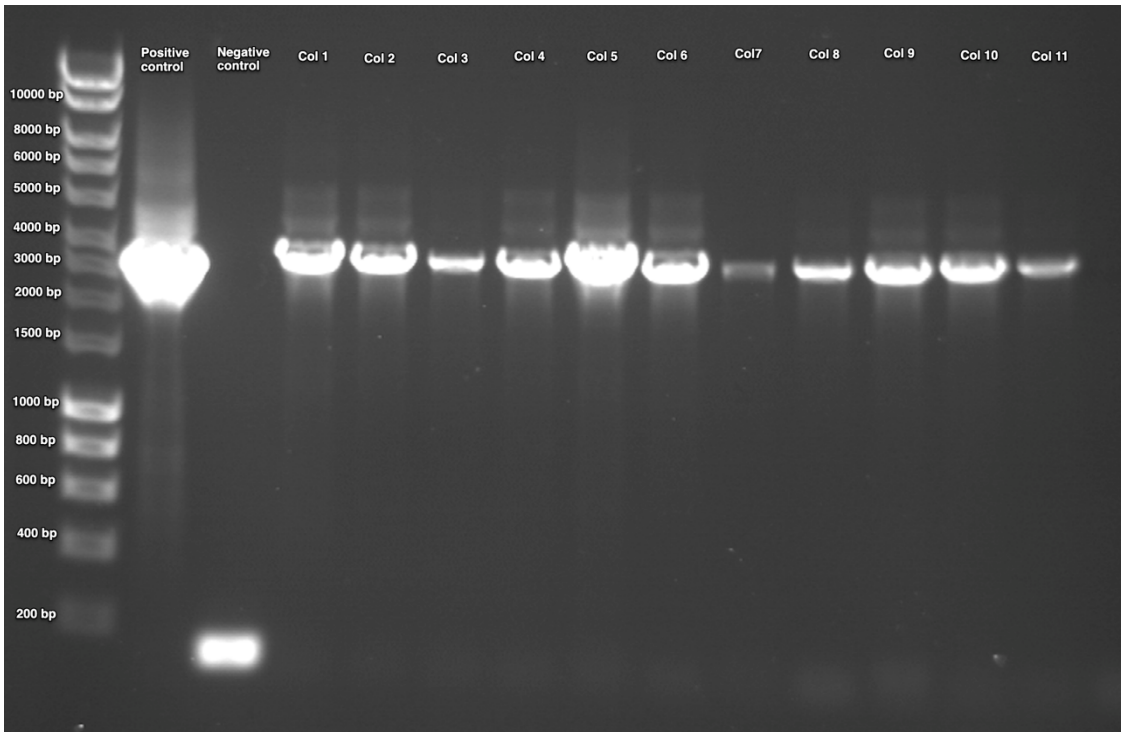


Figure 3.3.2: Agarose gel containing PCR amplified pGEM32ZFΔ$fliD$Kan^R plasmids transformed into *E. coli*. Left to right: DNA ladder, positive control (pGEM32ZFΔ$fliD$Kan^R plasmid), negative control (pGEM32ZF), Tested colonies 1-11.

The quality of the plasmid was tested and shown in Figure 3.3.3. Negative control was set up of pGEM32ZF plasmid with the standard 32ZF forward and reverse primers. This was to test if the core plasmid was intact. Another negative control consisted of pGEM32ZF plasmid with FliD$_{cj}$ top and bottom primers that should not produce a product as the vector should be empty. The tests were conducted as per the setup in Table 2.4.3 with pGEM32ZFΔ$fliD$Kan^R knockout plasmid with a combination of standard 32ZF primers and FliD$_{cj}$ primers to observe correct flanks insertion. Unlike the controls, all the tested primers should result in a 2500 bp product if FliD$_{cj}$ top and bottom flanks inserted correctly, which is the case on the gel.

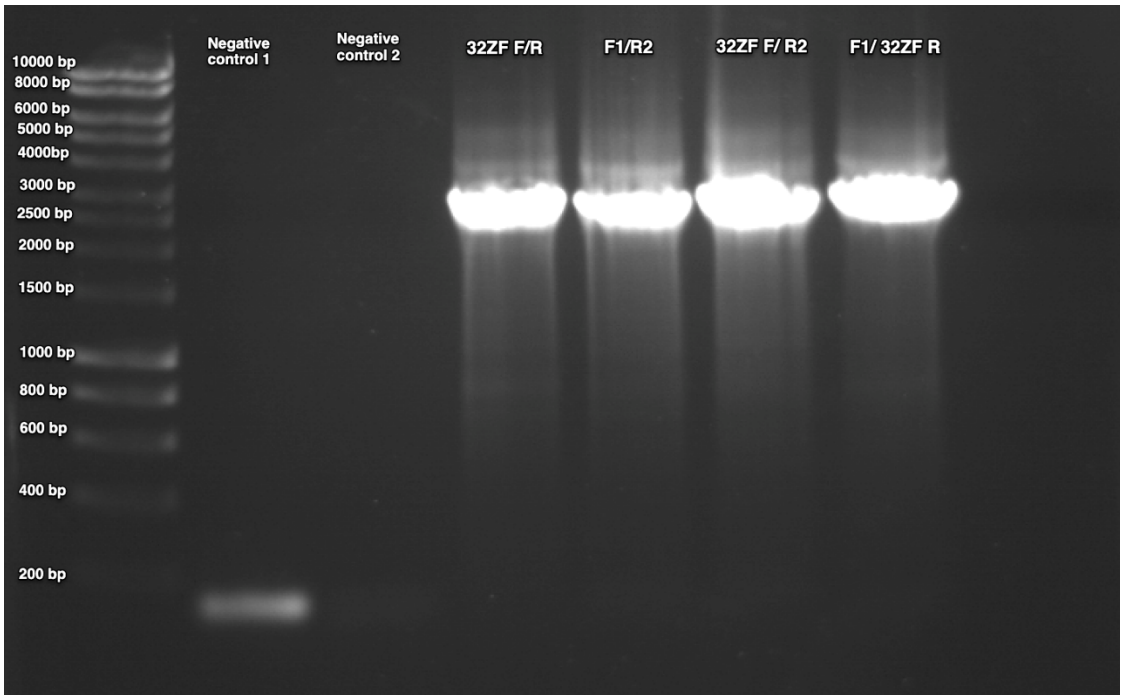


Figure 3.3.3: Agarose gel containing PCR amplified contents of the pGEM32ZFΔ$fliD$Kan^R plasmid. Left to right: DNA ladder, negative control 1 (pGEM32ZF with 32ZF forward and reverse primers), negative control 2 (pGEM32ZF with FliD$_{cj}$ top forward and FliD$_{cj}$ bottom reverse primers), test 1 for all inserts present (knockout plasmid with 32ZF forward and reverse primers), test 2 for both FliD$_{cj}$ flank inserts present (knockout plasmid with FliD$_{cj}$ top forward and FliD$_{cj}$ bottom reverse primers), test 3 for correct FliD$_{cj}$ bottom fragment insert (knockout plasmid with 32ZF forward and FliD$_{cj}$ bottom reverse primers), test 4 for correct FliD$_{cj}$ top fragment insert (knockout plasmid with FliD$_{cj}$ top forward and 32ZF reverse primers).

Now having tested the plasmid and confirmed the correct orientation of all inserts as illustrated in Figure 3.3.4, the plasmid was transformed into *C. jejuni* and grown on kanamycin selective media.

FliD$_{cj}$ complement (Δ*fliDφ*metK)) strain was designed and cloned as per section 2.4.2. The vector was designed such that spontaneous double crossover with the *C. jejuni* 81116 genome would result in reincorporation of the *fliD* gene, expression of which was driven by *metK* promoter and a chloramphenicol resistance cassette for selection. Amplified fragments for the plasmid were tested in 1% agarose gels and extracted as per Figure 3.3.5. pCmetK0509Chl^R plasmid was ligated and transformed into *E. coli*. Six colonies were tested using the agarose gel for containing the plasmid

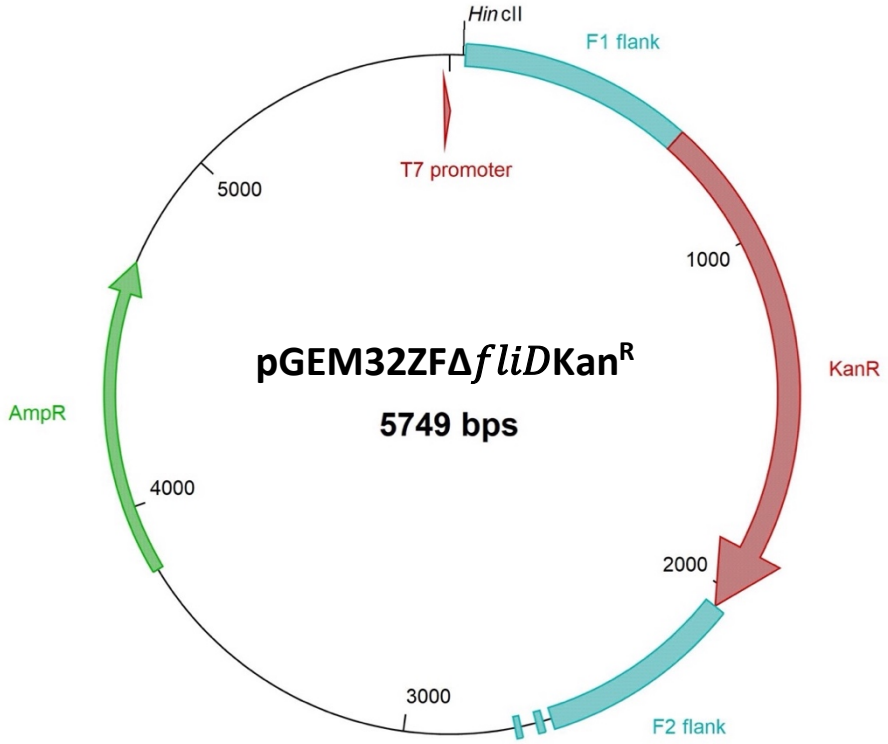


Figure 3.3.4: Schematic of the constructed pGEM32ZFΔ$fliD$Kan^R plasmid. Made via Clone Manager 9.

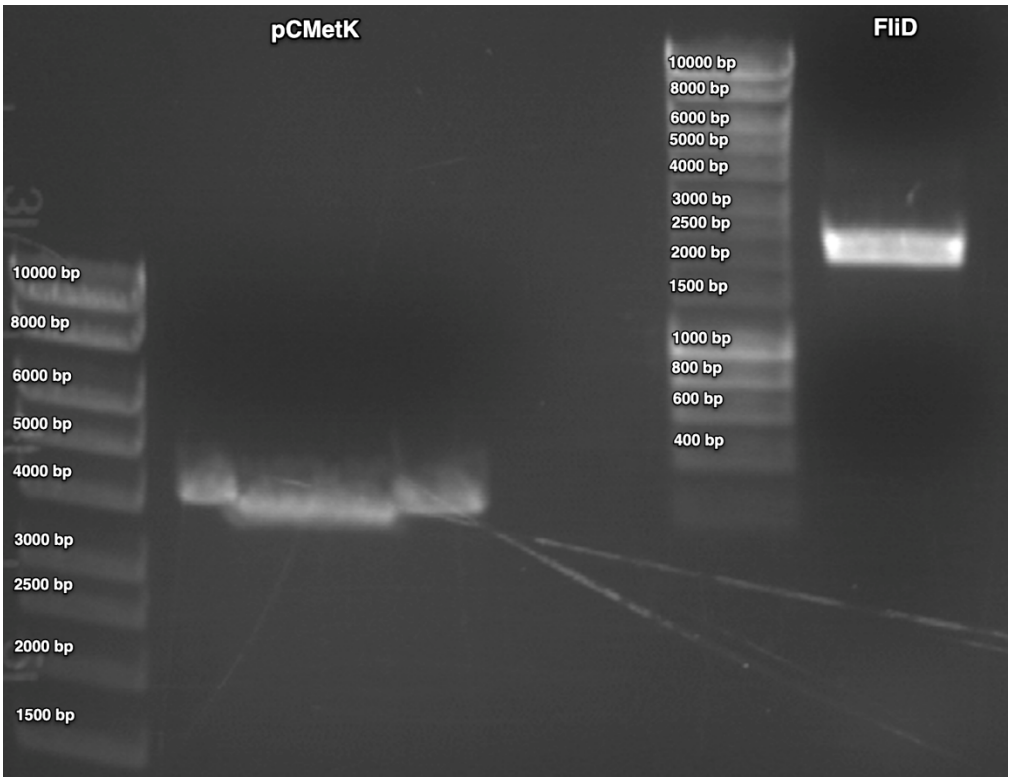


Figure 3.3.5: Agarose gel containing amplified fragments for the pCmetK0509Chl^R plasmid. From left to right: DNA ladder, pCmetK vector plasmid digested by *BsmBI* (4000 bp), FliD$_{cj}$ fragment amplified from genomic DNA of *C. jejuni* 81116 (2200 bp).

of the correct length as shown in Figure 3.3.6. Colonies 1-4 had the correct plasmid, while colonies 5-6 did not. The 4 colonies were checked for correct plasmid quality and shown in Figure 3.3.7. Negative control (pCmetK vector) and the 4 lysates were tested with primer set in Table 2.4.4 with the first 2 rows checking for fragment insert, 3rd row checking for correct orientation and last row for incorrect one. Negative control had no visible bands as expected. The first 3 colonies had the correct insert orientation but the 4th colony, having a band in the 4th row, did not. The correct plasmid, as illustrated in Figure 3.3.8, was the transformed into *C. jejuni* and grown on kanamycin and chloramphenicol selective media.

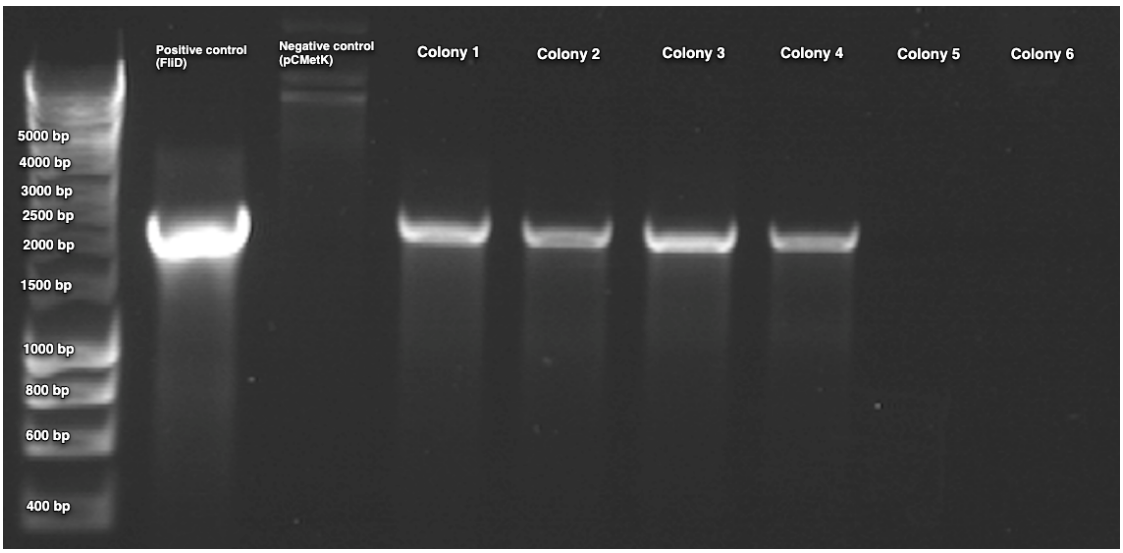


Figure 3.3.6: Agarose gel containing PCR amplified pCmetK0509Chl^{R} plasmids transformed into *E. coli*. Left to right: DNA ladder, positive control (FliD$_{cj}$ fragment), negative control (pCmetK), Tested colonies 1-6.

The three strains generated for *C. jejuni*: wild type 81116 grown on standard media, knockout (Δ*fliD*) on kanamycin and complement (Δ*fliDφ*metK) on kanamycin and chloramphenicol, were tested using growth curves from liquid media to access the viability of the cells and via RNA qRT-PCR to determine the level of expression of FliD$_{cj}$ and other genes within the operon. FliD expression is coupled with FliS expression (FliD chaperone) as well as an unlabelled protein, so interfering with it (knockout or complementation) may affect expression of those genes.

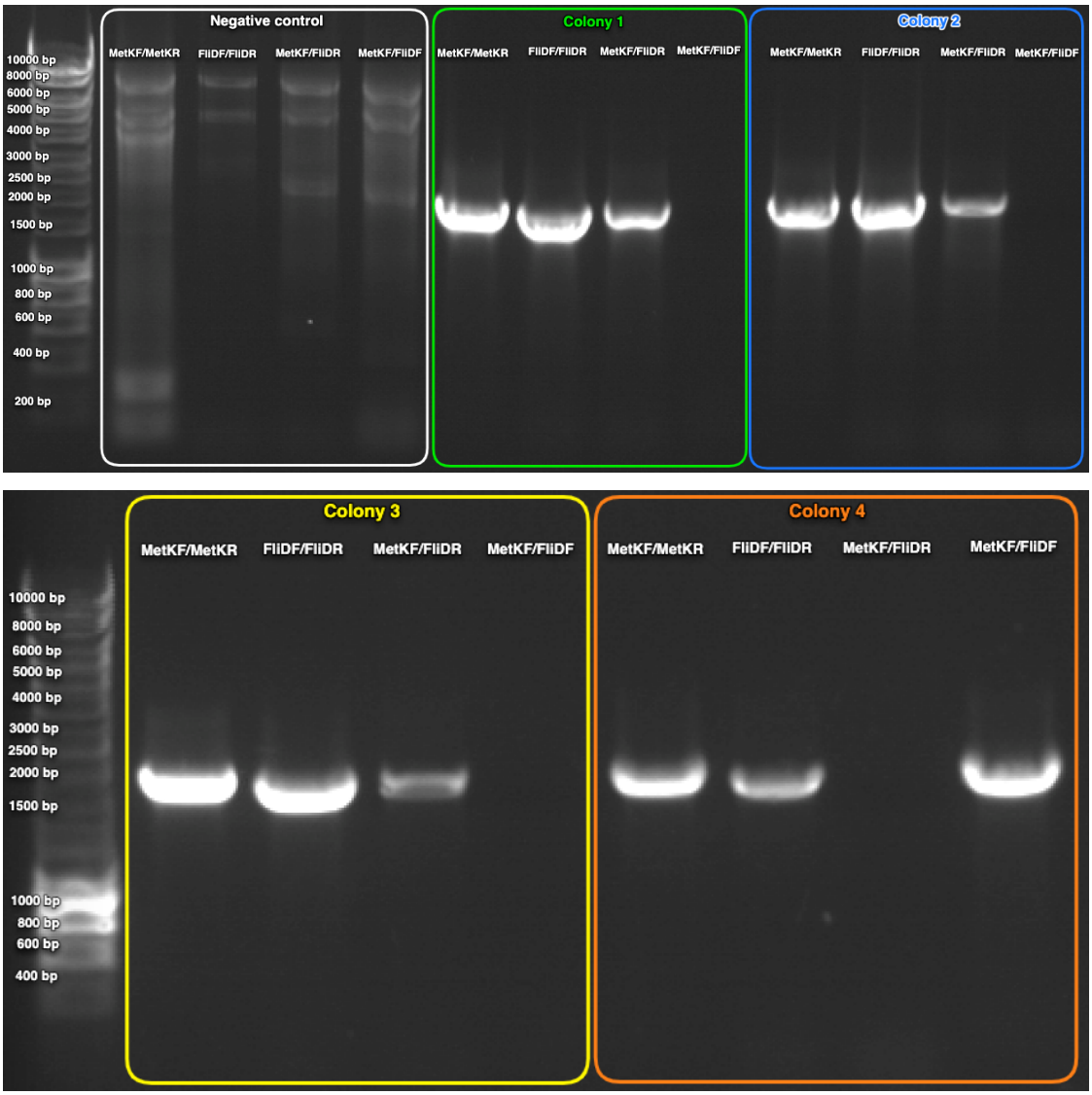


Figure 3.3.7: Agarose gel containing PCR amplified contents of the transformed lysate. Left to right: DNA ladder, negative control digests (1. metK forward and reverse fragments to check for fragment insert into vector, 2. $FliD_{cj}$ top forward and $FliD_{cj}$ bottom reverse primers to check for $FliD_{cj}$ insert, 3. metK forward and $FliD_{cj}$ bottom reverse primers to check if $FliD_{cj}$ insert is in the correct orientation, 4. metK forward and $FliD_{cj}$ top reverse primers to check if the insert is in the incorrect (backwards) orientation), colony digests (same primers as above).

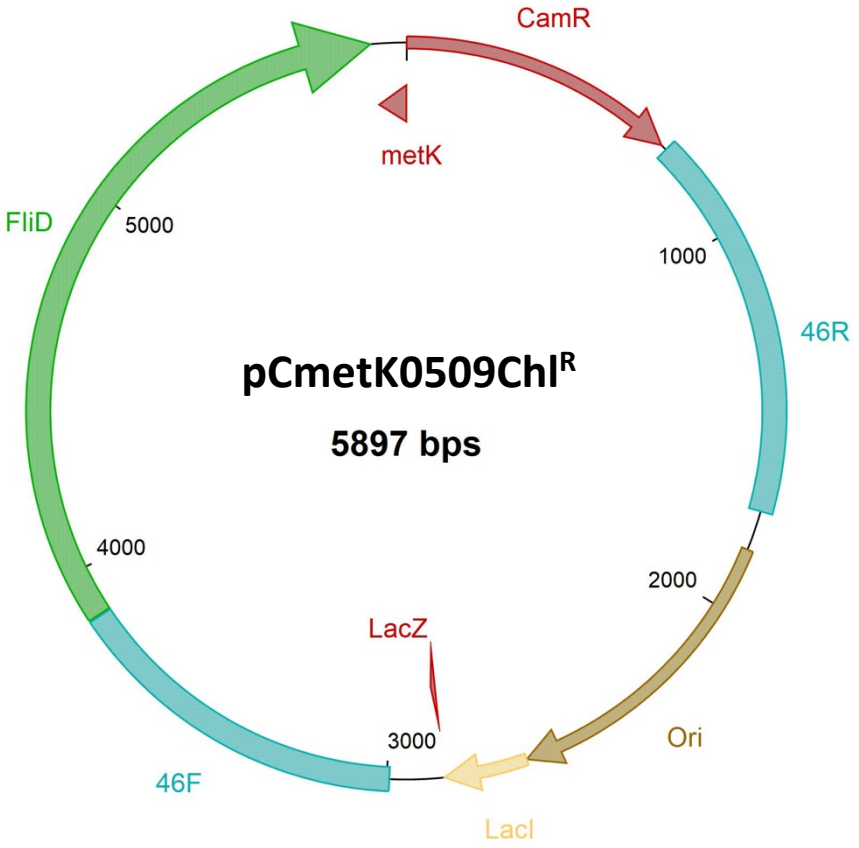


Figure 3.3.8: Schematic of the constructed pCmetK0509Chl^{R} plasmid. Made via Clone Manager 9.

Growth curves were measured through recording OD_{600} values every 2 hours for 12 hours, which are listed in Appendix 11. The curves themselves for all strains are almost identical, indicating that they are all viable and healthy as illustrated in Figure 3.3.9. RNA qRT-PCR was conducted by collaborator Dr Aidan Taylor and results illustrated in Figure 3.3.10. The qRT-PCR shows that the expression of *fliD* gene is as expected. 100% in WT, 0% in knockout and 77% in complement. The surrounding genes, *flaG* and unknown *0511* expression was not significantly reduced. However, the chaperone gene *fliS* expression was reduced in both knockout and complement. This appears to be due to the slight disruption of the operon. However, as there is no significant difference between the complement and knockout, then for the purpose of point mutations removing the motility trait these strains are suitable. As we used the

complement strain as the baseline and knockout strain as the 100% phenotype, this slight drop in expression was tolerated.

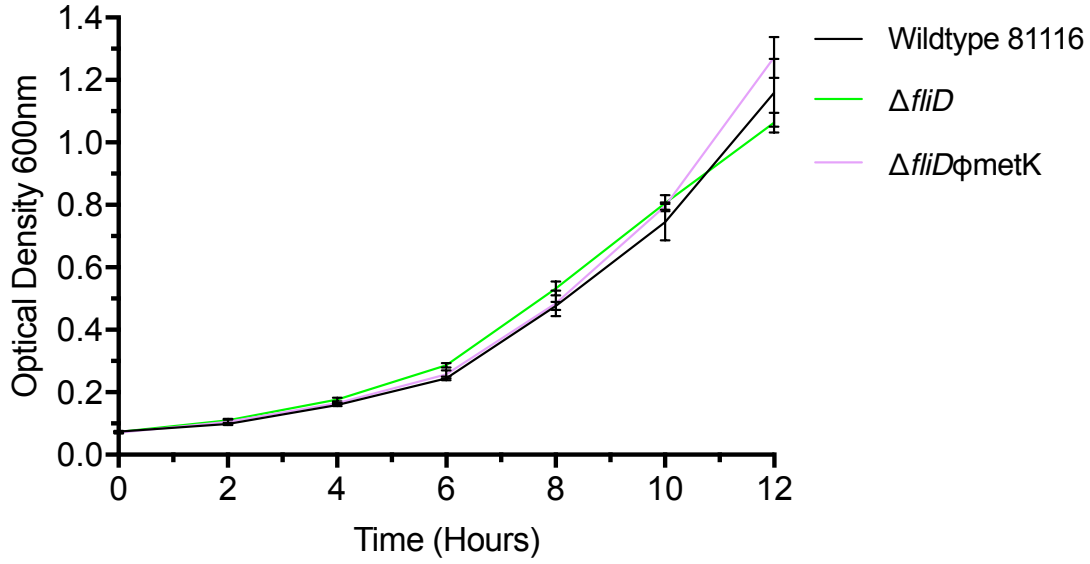


Figure 3.3.9: Growth curves of the wild type and generated strains. Error bars show standard deviation with N = 3.

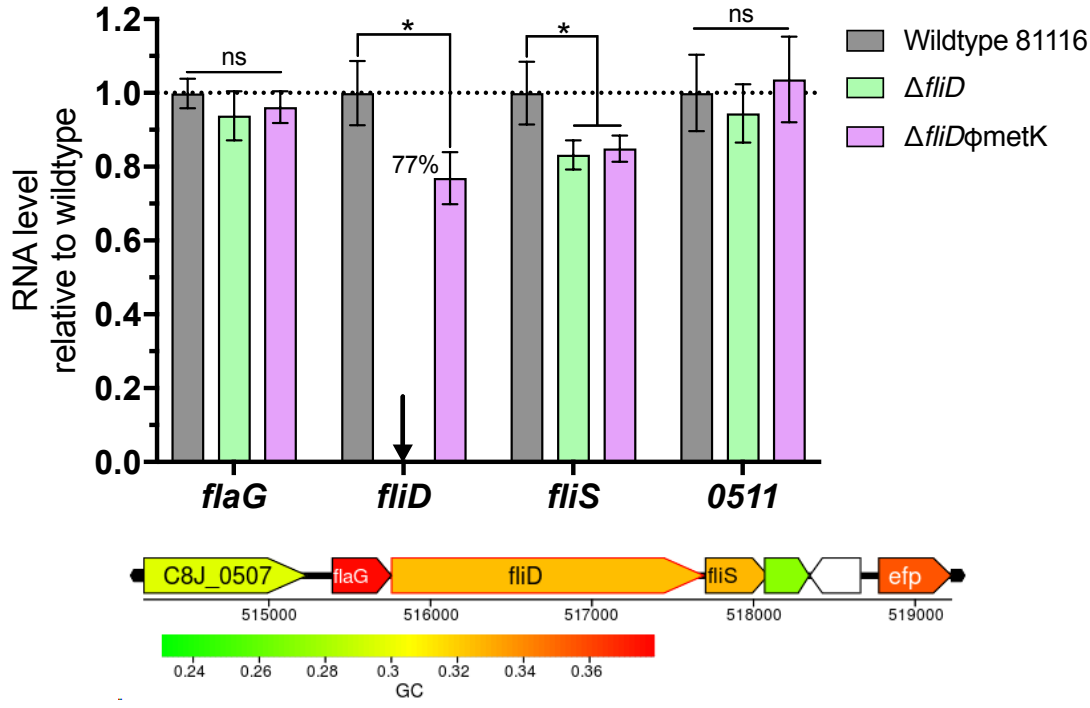


Figure 3.3.10: RNA qRT-PCR results of the expression of the *fliD* operon genes in the wild type and generated strains. Operon with the relevant genes is shown below the graph. Error bars show standard deviation with N = 3. (*) shows significant difference between samples. (ns) stands for non-significant difference between samples. Black arrow shows the lack of the bar as there was no expression.

Point mutations in terminal regions were chosen according to both their conservation across bacterial species and their location at the N and C terminal interfaces as shown in Figure 3.3.11. Chosen residues were substituted with conservative leucine or alanine and polar serine residues. N-terminal interface residues (F3, L9 and F11) and their interaction with the C-terminal interface of the adjacent monomer (W614 and Y617) were tested to observe if the close-knit interactions present in our sample are important for $FliD_{cj}$ intra-pentamer stability. C-terminal interface residues (I620, L624, L628 and I635) were chosen to test the interface observed in the decamer between the top and bottom pentamers, or inter-pentamer interactions, which according to the co-evolution data in Table 3.1.1 may be responsible for interacting with the flagellin monomers. Other conserved residues

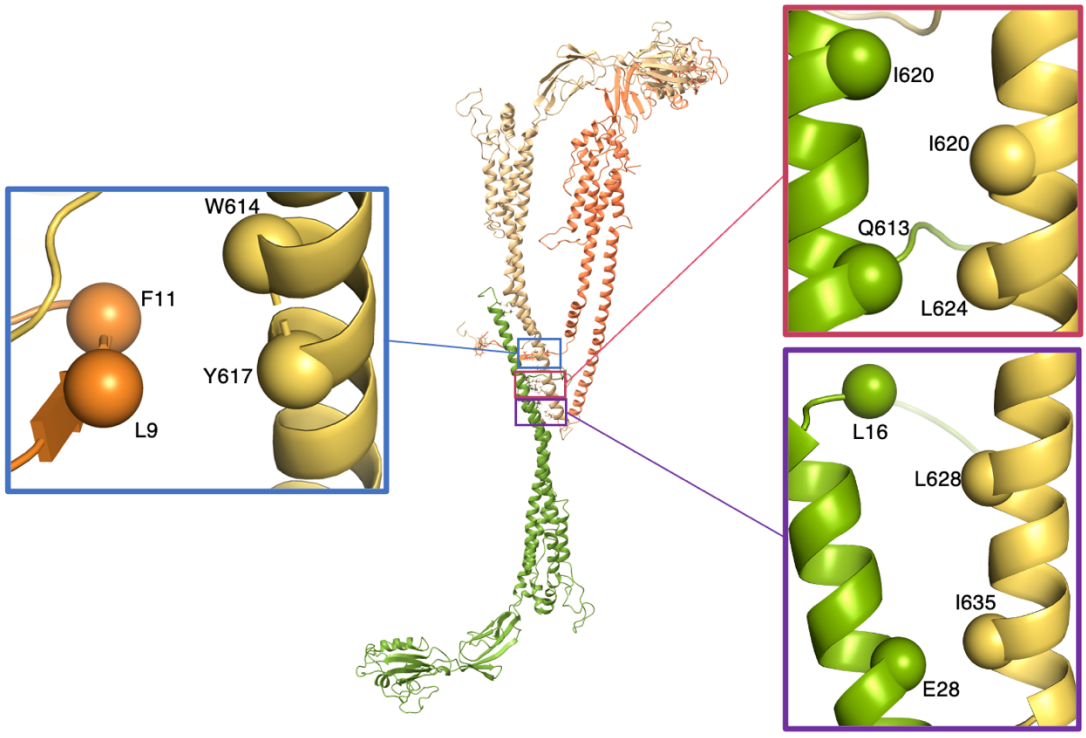


Figure 3.3.11: Location of the mutated residues on the $FliD_{cj}$ structure, chosen due to their conserved nature and approximate location in the interacting interface. Adjacent subunits (Yellow, Orange) were chosen to represent the interaction of the N-terminus residues with that of the next subunit C-terminus and which might be important in the formation of the pentamer. The residues in the lower C-terminus were chosen to represent the interaction interface with flagellin upon flagellar elongation, as the conserved residues in that region which bind to the bottom subunit C-terminus might mimic the interactions with the flagellin monomer. Adapted from (Al-Otaibi et al., 2020).

were also tested, primers for which are listed in Appendix 4, to observe if mutations in other domains also have a significant effect on motility. The mutants were made using the complement plasmid (pCmetK0509Chl^R) and transformed into the knockout *C. jejuni* strain (Δ*fliD*) as per section 2.4.3. Thus, the selection media contained both the kanamycin resistance from the inserted cassette in Δ*fliD* strain and chloramphenicol resistance from the pCmetK0509Chl^R plasmid.

3.3.3 MOTILITY ASSAYS

After generating all the strains, the impact of mutations in $FliD_{cj}$ was tested through motility assays set up as per section 2.4.4. The use of triphenyltetrazolium chloride (TTC) in the soft agar plates allowed for clearer observation of the viability of the strains, as it reacts to the oxidation of the media showing red colour upon cellular respiration (Bochner and Savageau, 1977). Thus, while the knockout strain is immotile, its viability was observed through the red colour at the stab site. The diameter of the motility swarm from the stab was measured in mm after 18-hour incubation and as expected the phenotype was successfully restored to the knockout mutant through complementation as illustrated in Figure 3.3.12. The diameter of the complement was similar to than that of the wild type strain and with previously described growth curves and RNA qRT-PCR we conclude that the 2 generated strains are not impaired in growth.

The effect of point mutations on *C. jejuni* motility was also observed through motility assays. Samples were tested in triplicate for each mutation and final values calculated as percentage of the complement strain control as detailed in Appendix 13. Mutations in residues located in the terminal interfaces detailed in Figure 3.3.11 were picked from the full range of mutants tested and summarized in Figure 3.3.13. Mutation of residues on the N-terminal interface and those interacting with them L9, F11, W614 and W617 resulted in reduction of motility, particularly for mutations to polar residues (F11S, W614S, L9S, or Y617S). This led to a proposal that the hydrophobic nature of this interface may be one of the properties affected by the point mutations that influences motility, suggesting that the interaction formed by the N-terminal stretch contributes to $FliD_{cj}$ function of building the filament and thus cell

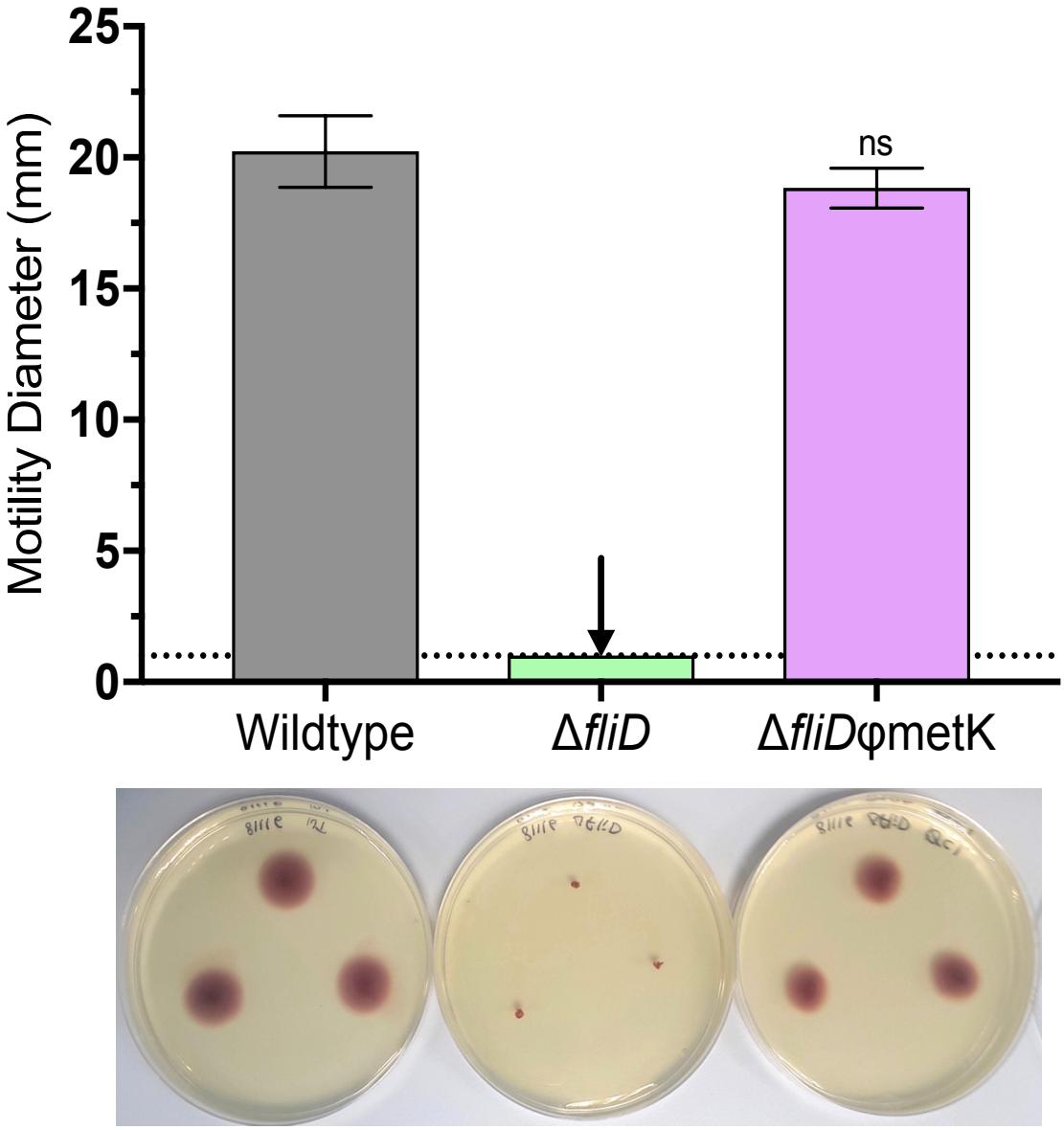


Figure 3.3.12: Motility assay testing the wild type and 2 core generated strains (knockout Δ*fliD* and complement Δ*fliD*ϕmetK. The bar chart above shows swarm diameter that is visually represented by the plate sample below. Each strain was grown on respective antibiotic selective agar and in triplicate. Error bars show standard deviation. Raw data for the chart is in Appendix 12.

motility. The C-terminal interface residues L628, I635, L624, and I620 when mutated impacted motility in a similar manner to that of the residues of the pentameric interface, confirming that the hydrophobicity of this interface also contributes to $FliD_{cj}$ function in filament assembly. While some individual mutations may not reduce motility significantly, their contribution to the overall hydrophobicity of the interfaces is the essential property in motility.

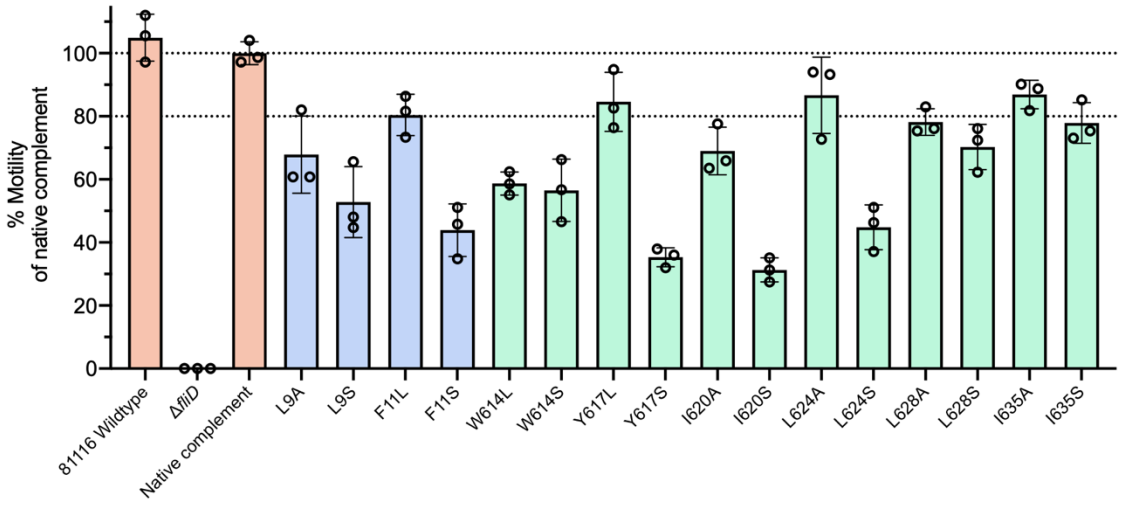


Figure 3.3.13: Motility assay results for point mutants represented as the mean percentage of the native complement strain, based on swarm diameter on soft agar. Controls (WT, deletion mutant and complement) are in orange. C-terminal mutants in light green and N-terminal mutants in purple. Error bars show standard deviation with N = 3. The corresponding data points are represented as dot plots. Raw data for the chart is in Appendix 13. Adapted from (Al-Otaibi et al., 2020).

3.3.4 Mutant Strain Morphology via TEM

To visualize the effect of knocking out *fliD,* the generated strains were tested via TEM as illustrated in Figure 3.3.14. Wild type cell, used as a reference, was helical in shape and had two polar flagella as expected. At higher magnification the filament subunits could be observed as well as an indent at the tip. The knockout strain had no visible flagellar filaments but still contained formations at the polar ends of the cell which could be the hook and hook-filament junction. As the filament formation was resumed in the complement strain and the filaments looked nearly identical to that of wild type *C. jejuni*, it appears that FliD is involved in building the filament and in its absence even if flagellin is expressed it cannot be assembled into a functional flagellum.

To verify if motility was affected because the point mutations prevented filament assembly, the corresponding bacteria were visualized by TEM. All mutations

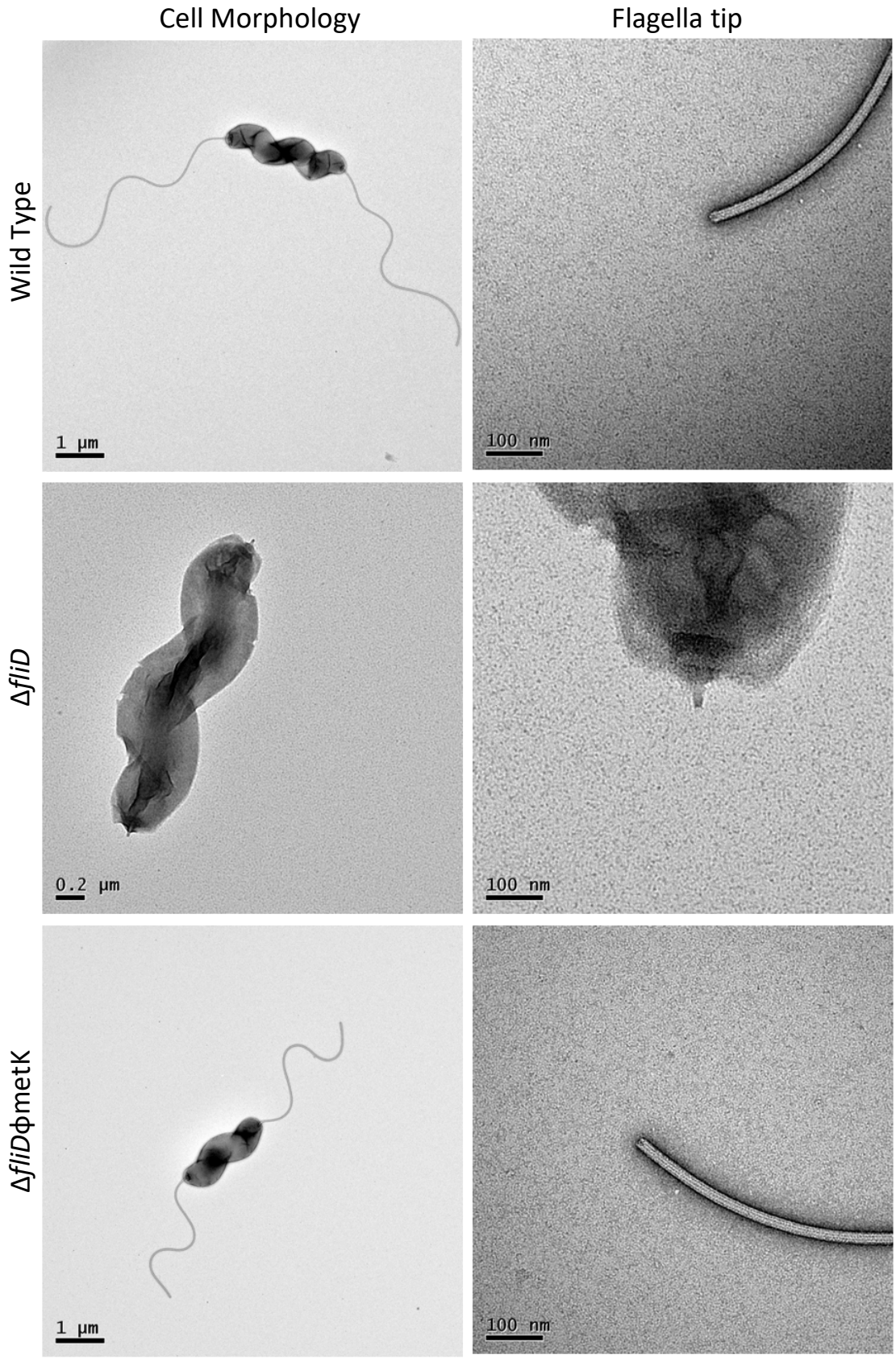


Figure 3.3.14: Morphology of *C. jejuni* wild type and generated knockout ΔfliD and complement ΔfliDϕmetK strains visualized through negative stain TEM.

still led to bacteria with assembled filaments, of length and morphology similar to that of WT bacteria as illustrated in Appendix 14. However, there were some cells with shorter flagella, with less curvature and many loose flagella in the solution. Some point mutant strains contained flagella broken off at the hook or at a random distance along the filament as illustrated in Figure 3.3.15.

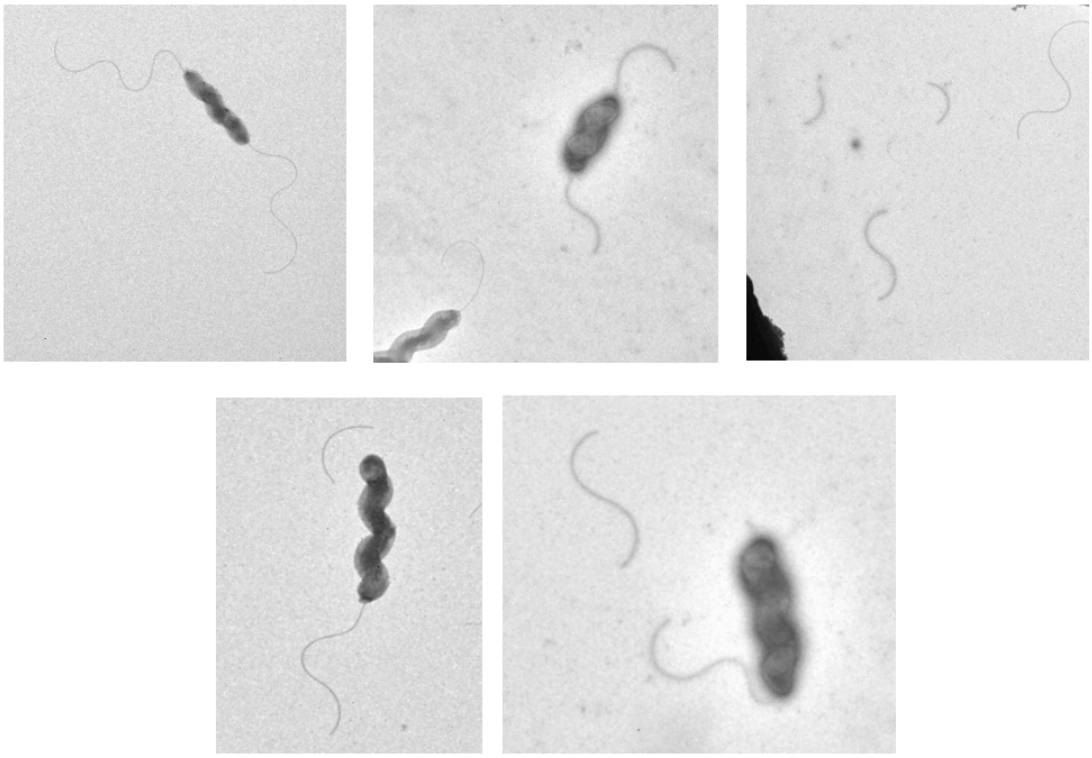


Figure 3.3.15: Illustrations of brittle filaments observed in our samples. From top Left to right: fully attached flagella, short stubby flagella with less than 2 inflection points, cut off flagella and fragments. From bottom left to right: flagellum cut off at the base, snapped off flagellum not at the base but further along the filament.

Flagella length for each point mutant sample was measured with ImageJ and compared to the wild type and complement control samples as shown in Figure 3.3.16. There appeared to be no significant difference between the controls and samples, indicating that the point mutations do not affect flagellar length. Thus, the reduction in motility may be due to unstable flagellin packing in the filament causing brittleness and shearing, so flagella attachment was quantified next. It was observed that the filaments are less stable in the mutants, with between 60 and 80% of filaments found broken off at varying lengths from the cell body (apart from W614 mutations), versus ~20% in control samples as illustrated in Figure 3.3.17. This led to a hypothesis that

while the overall flagellar formation is not affected by these point mutants, the integrity and packing of the filaments is, thus making them more brittle and likely to shear off.

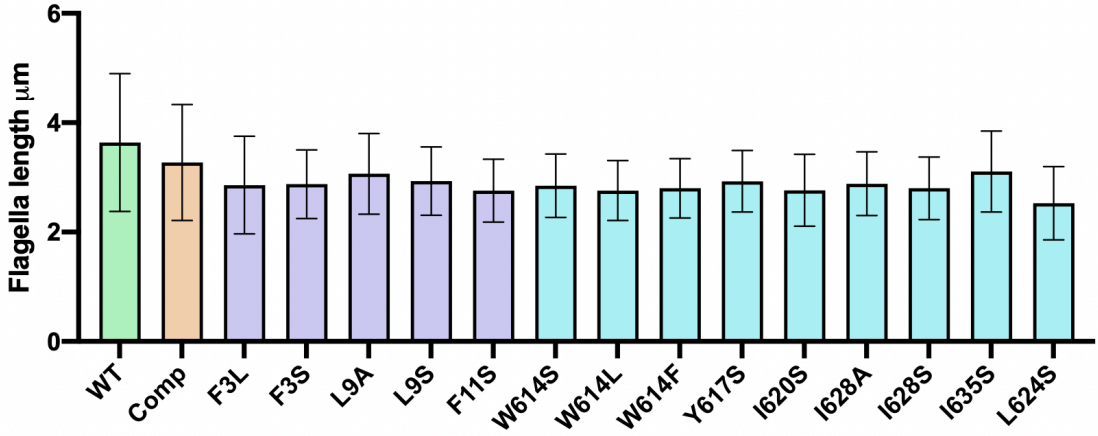


Figure 3.3.16: Bar chart of average flagellar filament length observed in each point mutant C. jejuni strain. Error bars represent standard deviation. Raw data for the chart is in Appendix 15.

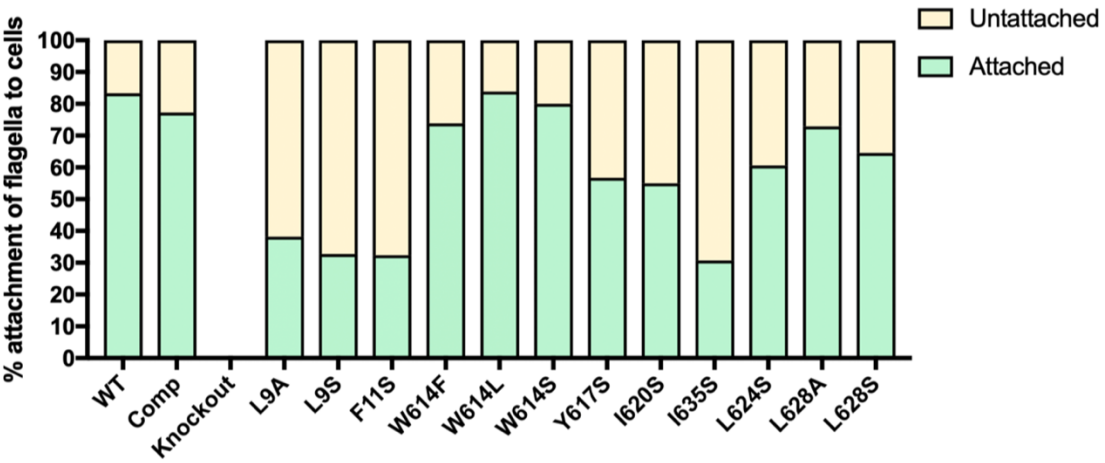


Figure 3.3.17: Bar chart of flagella attachment to cells calculated from micrographs for each point mutant. Values are calculated as percentage of all flagella observed per sample. Attached flagella are coloured green, unattached coloured yellow. Raw data for the chart is in Appendix 16.

3.3.5 Discussion

As mentioned above, the structural characterization of the cap complex indicates an unusual architecture of the N-terminus, which forms a stretch that wraps

around and forms contacts with two adjacent subunits, with mainly hydrophobic residues at the interface. This stretch of residues interacts with the C-terminal α-helix of the adjacent molecule. This is of particular interest since it was shown that the C-terminus contributes to the oligomerization of FliD (as well as interaction with its chaperone)(Kim et al., 2017). To confirm the role of these interface interactions in FliD function, a *C. jejuni* fliD knockout strain was engineered (Δ*fliD*), leading to a loss of chemotaxis observed in a soft agar swarm assay. Accordingly, no filament was observed in this strain, and thus this phenotype was attributed to lack of motility. Genetic complementation by expressing the *fliD* gene at a distal site on the chromosome fully rescued motility, and this was exploited to engineer point mutations in select highly conserved residues within the interacting interfaces to assess their impact on motility. Mutations to polar residues lead to significant reduction in motility and lower flagellar attachment. This led to a proposal that the hydrophobic nature of this interface may be one of the properties affected by the mutations that influence motility, suggesting that the interaction formed by the N-terminal stretch contributes to FliD function of building the filament and thus cell motility. To verify if motility was affected because the mutations prevented filament assembly, the corresponding bacteria was visualized by TEM. All the mutations still led to bacteria with assembled filaments, of length like that of WT bacteria, demonstrating that the corresponding FliD proteins are still able to promote filament elongation. The N-terminal ~20 residue stretch corresponds to the secretion signal in flagellar filaments of *S. enterica*, so potentially a similar signal exists for FliD to be secreted through the flagellum T3SS (Kovács et al., 2018). The observation that in the mutants described above, the filament is still formed, is a confirmation that these mutations did not interfere with FliD secretion, but rather with its function to promote filament assembly.

The second interacting interface observed in the D0 domain, is formed between the C-terminus of FliD in the pentamer-to-pentamer interface. The role of this interface was investigated in FliD function. To that end, a series of mutations was engineered in some of the conserved residues forming this complex. While not done within the scope of this work, to investigate this aspect the point mutants should be cloned in *E. coli* and purified for cryo-EM analysis to observe if the mutations disrupt

the inter-pentamer interface. As the protein complex was recombinantly expressed and assembled *in vitro*, the decameric interface is not physiological and instead the C-terminal region of FliD interacts with the flagellar filament lumen, as supported by previous low-resolution data of intact flagellum tips (Furukawa et al., 2002; Maki et al., 1998; Vonderviszt et al., 1998; Yonekura et al., 2000).

While these mutations were performed in the terminal domains of $FliD_{cj}$ and attributed to the phenotype of reduced motility, it is unclear whether it is the FliD itself that is disrupted, the assembly of the filament or $FliD_{cj}$ interactions with major and minor flagellins in *C. jejuni.* The disruptions in the structure of $FliD_{cj}$ itself would have to be subtle, as due to the presence of flagella in all point mutants, the protein should have localized at the tip and grown the filament (Appendix 14). While the FliD knockout strain generated in this section appeared to not form any filaments it did have a wild type flagellin gene expression and secreted it into the supernatant (see next section). While, this has not been done in this work, the supernatant of the FliD point mutants needs to be tested for flagellin to observe if these mutations cause deformations in the cap which in turn cause flagellin leaks from the cell as well.

Flagella packing is another reason why the point mutations in FliD result in reduction of motility. Flagellins pack into protofilaments with a highly conserved D0-D1 domain core with conserved helical parameters, indicating that the residues localized in these regions are essential for flagellum filament integrity (Montemayor et al., 2021). A single mutation in D1 domain of *S. enterica* FliC and *C. jejuni* FlaA causes the filaments to lose their curvature and become straight enough for data collection, so it is possible that interference or incorrect packing of the terminal region residues would have a significant effect on motility. As the leg domains of FliD have high homology with flagellin domains, multiple elongation mechanisms have been proposed based on the flagellin D0-D1 domains substituting for flagellin D0-D1 domains in a growing filament, moving out of the way for the actual flagellin to bind and elongate the filament. So, the interactions between FliD and flagellin in the filament mimic flagellin-flagellin interaction. All of these similarities lead to a possibility that point mutation in FliD leg domains are incapable of binding or interacting with the flagellar filament properly, so that when a new flagellin approaches the assembly point

the disturbance prevents it assembling in a stable way causing flagella “brittleness” and easier shearing.

It has been observed in *S. enterica* that sheared flagellar filaments can still regrow to wildtype length. As both flagellin and $FliD_{se}$ are late substrates in the flagellar T3SS export system, but FliD is expressed in class 2 genes prior to class 3 flagellar gene activation, it is assumed that in a sheared filament FliD expresses once more through a class 3 promoter to relocate on top of a sheared flagellar filament and regrow it to the wild type length, controlled by injection-diffusion mechanism (Paradis et al., 2017; Renault et al., 2017). In *C. jejuni* the flagellar filament assembly is more complex as there are two flagellins, with the minor flagellin FlaB being a part of class 2 genes and forming a short segment at the base of the filament, and major flagellin FlaA forming the rest of the flagellar filament. The length control is also more complex, as the standard anti-sigma factor FlgM only inhibits FliA, a class 3 promoter with the major flagellin responsible for the length, at 42 °C (Wösten et al., 2010). FlaG is an alternative anti-sigma factor that appears to inhibit FliA at all temperatures and when exported similar to FlgM, releases the sigma factor and allows for FlaA expression and elongation of flagellar filament. In a FlaG knockout mutant the filament length increased from that of the wild type *C. jejuni* of 3.5 µm to that of almost two-fold, indicating that the filament length is capped at 3.5 µm by FlaG (Inoue et al., 2018). Assuming that the majority of the length of the filament is FlaA, when sheared, FlaG would dissociate from FliA inducing more FlaA expression and inhibit it once again when the filament has recovered its length. The FliD cap would express from the class 3 promoter and attach to the growing FlaA filament and facilitate re-growth.

A more *C. jejuni* specific possibility of mutations in FliD causing reduction in motility, is that FliD interacts with FlaA and FlaB differently. FlaB only mutants have been observed to form a highly curved filament, while FlaA only mutants formed a long straight filament. FlaB localizes to the 0.5 µm section at the base of the filament and transfers the curvature to FlaA, which forms a straight filament of approximately 3 µm (Inoue et al., 2018). With such differences in function and assembly it is surprising that the only difference between the two flagellins sequence-wise is a 5 amino acid insert in a D2 area in FlaA, not adjacent to the highly conserved sequences. From

observations of the morphology of the point mutants, shown in Figure 3.3.18, some mutants such as L9S and F11S appeared to have a straighter filament located closer to the HBB than the rest. As $FliD_{cj}$ binds and facilitates both FlaA and FlaB integration, the point mutants may affect the formation of one flagellin section differently from the other. As FlaB section is highly curved, $FliD_{cj}$ point mutants might not be capable of creating a correct folding environment for the base part, while FlaA is not affected.

Another reason for "brittleness" and how $FliD_{cj}$ interacts with FlaA and FlaB differently might be represented by the I635S mutant, where the flagellum morphology is similar to the wild type but the flagella attachment value is low, as shown in Figure 3.3.18. From studies of localization of FlaB and FlaA within the filament, while visual appearance of proximal and distal parts of the filament differed, illustrating the consistency of one type of flagellin, the boundary was not obvious, suggesting FlaB may cause conformational change of FlaA in that region or the mixing region is very narrow (Inoue et al., 2018). Thus, mutations in $FliD_{cj}$ would affect its interaction with FlaA, destabilizing the mixing area and causing increased shearing while not affecting the morphology of the curvature significantly. The fact that these two types of point mutants are localized in different interfaces might hint onto their flagellin-specific function. While these observations are intriguing, they still require extensive research to determine the true effect of $FliD_{cj}$ on *C. jejuni* flagellin packing.

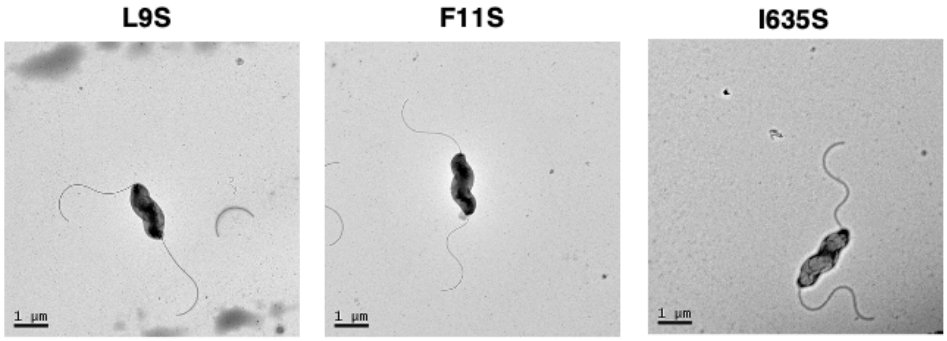


Figure 3.3.18: Morphology of F9S, F11S and I635S mutant strains of *C. jejuni* observed using negative stain.

3.4 *C. JEJUNI* FILAMENT STRUCTURE DETERMINATION VIA CRYO-EM

3.4.1 INTRODUCTION

The flagellar filament consists of multiple repeats of one type of flagellin and can be >20 μm in length in *S. enterica*. In *C. jejuni* the filament is composed of two types of flagellin: initial short segment made up of FlaB and a long filament of FlaA restricted by FlaG to 3.5 μm total flagellar length in *C. jejuni*. It has been studied extensively by cryo-EM, and high-resolution structures have been reported in a range of bacteria, including *B. subtilis*, *P. aeruginosa* and *S. enterica*. In all of these, the filament was shown to consist of 11 protofilaments (Maki-Yonekura et al., 2010; Wang et al., 2017). However, a low-resolution cryo-EM study of the *C. jejuni* flagellar filament suggested the presence of 7 protofilaments (Galkin et al., 2008). Taken together with the range of oligomeric states observed in the FliD crystal structures, these observations have led to a model where in different bacterial species, the cap complex has different oligomeric states (N), and in the corresponding filament, the number of protofilaments is 2N + 1(Cho et al., 2017).

After solving $FliD_{cj}$ structure (see 3.2) and demonstrating that $FliD_{pa}$ and $FliD_{sm}$ can form 5-fold oligomers in solution (see 3.1), the aims were to solve the structure of *C. jejuni* flagellar filament using single particle cryo-EM to either confirm or disprove the 7-protofilament model. After multiple attempts to purify the filaments through shearing and flagellin precipitation failed, wild type cells were used to determine the structure of a native flagellar filament to have 11 protofilaments. This lead to a proposal of a model of filament elongation involving a 5-to-11 ratio of capping protein to protofilament based on previously suggested mechanisms (Maki-Yonekura et al., 2003; Maki et al., 1998; Song et al., 2017; Yonekura et al., 2000).

3.4.2 Attempts at Filament Purification

As recombinant expression of flagellin did not yield soluble results (see 3.1), wild type *C. jejuni* strains 11168 and 81-176 were used instead and shear the flagella using a modified Avanti Polar Lipids apparatus as detailed in section 2.2.4. 11168 is a common laboratory strain isolated from human host used for a variety of microbiology experiments, while 81-176 is yet another human derived strain which excels in motility and biofilm formation. The resulting sample was concentrated and tested for presence of flagellin via SDS-PAGE gel as illustrated in Figure 3.4.1. As the SDS-PAGE gel was inconclusive, the samples were observed via negative stain EM as illustrated in Figure 3.4.2. 11168 strain had visible flagella, but the buffer contained multitude of contaminants making it difficult to use this sample for cryo-EM. The flagella also appeared to have bundled together, making it difficult to distinguish between

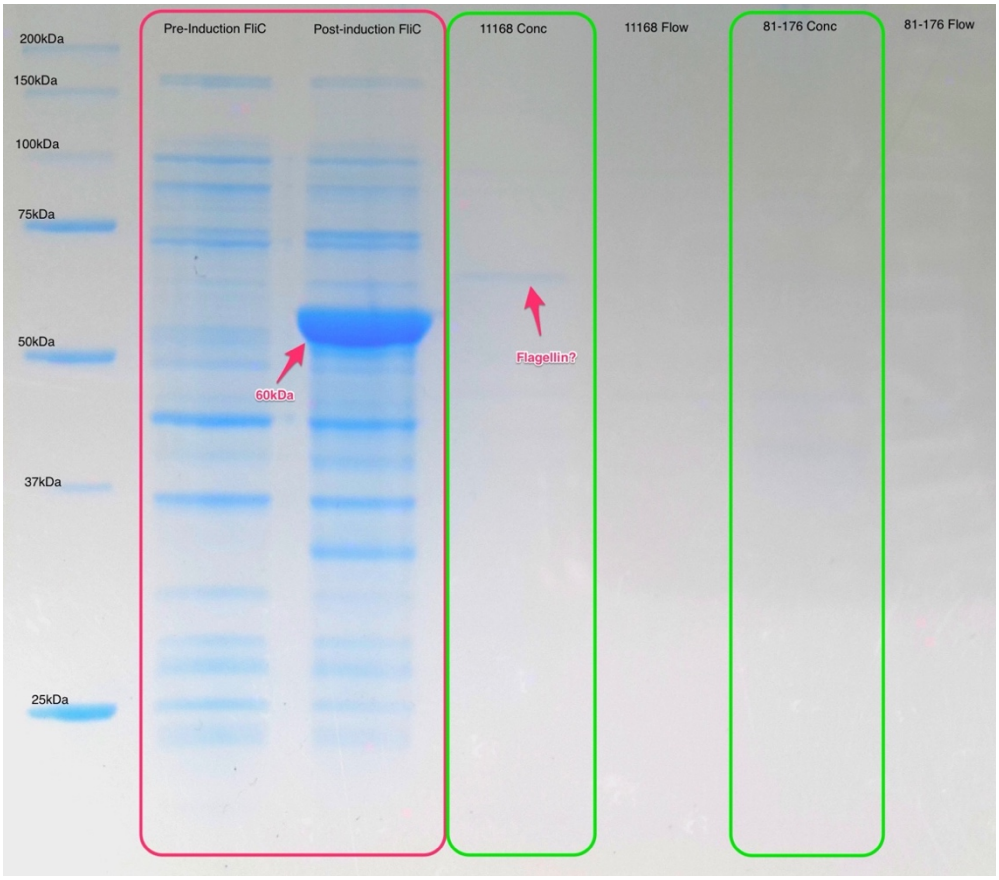


Figure 3.4.1: SDS-PAGE gel of flagella shearing experiments. **Red:** Control samples from recombinant $FliC_{cj}$ expression (see 3.1.3) with 60 kDa band representing the flagellin. **Green:** Concentrated supernatant samples that should contain sheared flagella. Potential flagellin band is shown with magenta arrow.

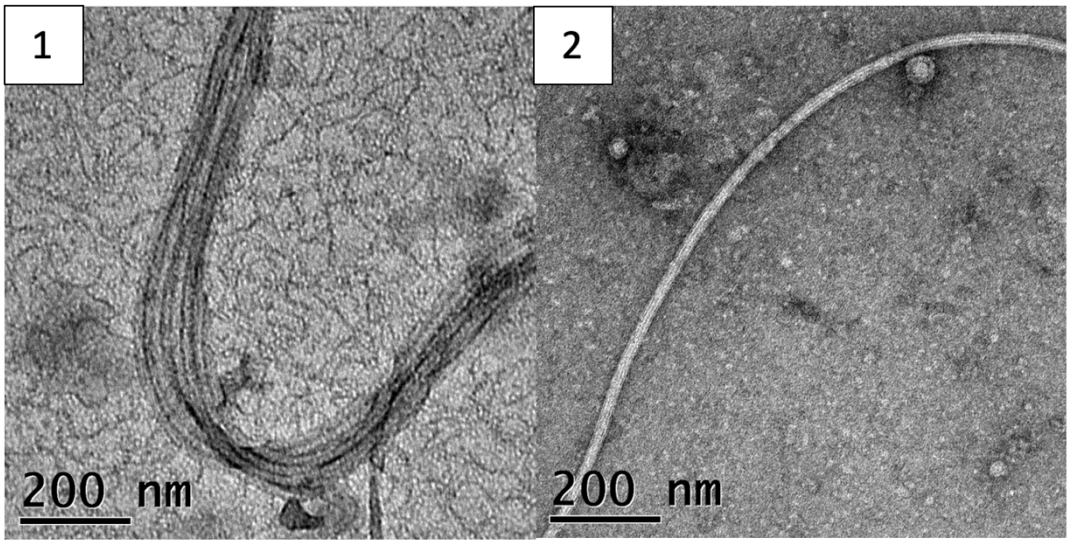


Figure 3.4.2: Negative stain images of sheared flagella from 1) *C. jejuni* 11168 strain and 2) 81-176 *C. jejuni* strain.

individual filaments. 81-176 strain, while having clear individual filaments visible, still contained a lot of contaminants in the background.

To purify the 81-176 strain sample further, it was ultracentrifuged to pellet down the filaments and get rid of smaller contaminants. While this successfully concentrated the filaments it also concentrated the contaminants as illustrated in Appendix 17. We attempted to further purify this through an anion exchange column, which uses charged beads to capture protein of opposite charge in the absence of salt and eluting it in high salt buffer. However, this did not work either as detailed in Appendix 17.

C. jejuni exports its flagellar proteins, one of them being flagellin, into the media as it grows. In *S. enterica* it is possible to polymerise the filaments from the monomeric flagellin in the media using supersaturation and adding very high concentrations of sodium sulphate salt. These non-native polymerised filaments can be collected through ultracentrifugation and depolymerised through addition of Mg^{+2} and Ca^{+2} ions to isolate monomeric flagellins (Wakabayashi et al., 1969). $81116\Delta fliD\text{Kan}^{R}$ knockout strain cloned in section 2.4.1 while not forming flagellar filaments still had wild type level of flagellin expression and thus the hypothesis was that it would all be ejected into the media and serve as a more concentrated source of flagellin.

Liquid media from $81116\Delta fliD$Kan^R knockout strain was isolated and filtered prior to the initial ultracentrifugation and polymerization. All subsequent steps were observed for flagellin via SDS-PAGE gel as illustrated in Figure 3.4.3. A 60 kDa band was clearly observed across all the steps indicating the presence of flagellin throughout the experiment. The final centrifugation step samples were observed via negative stain as illustrated in Figure 3.4.4. It appears that instead of forming ordered filaments flagellin polymerizes into aggregates large enough to be pelleted down. In addition, the presence of bacterial cells in the sample caused issue with any further purification using columns. The supernatant fraction had visible small proteins which could potentially be free monomeric flagellin, which could be purified further. A control experiment, that was not performed in the scope of this work, to determine if that is

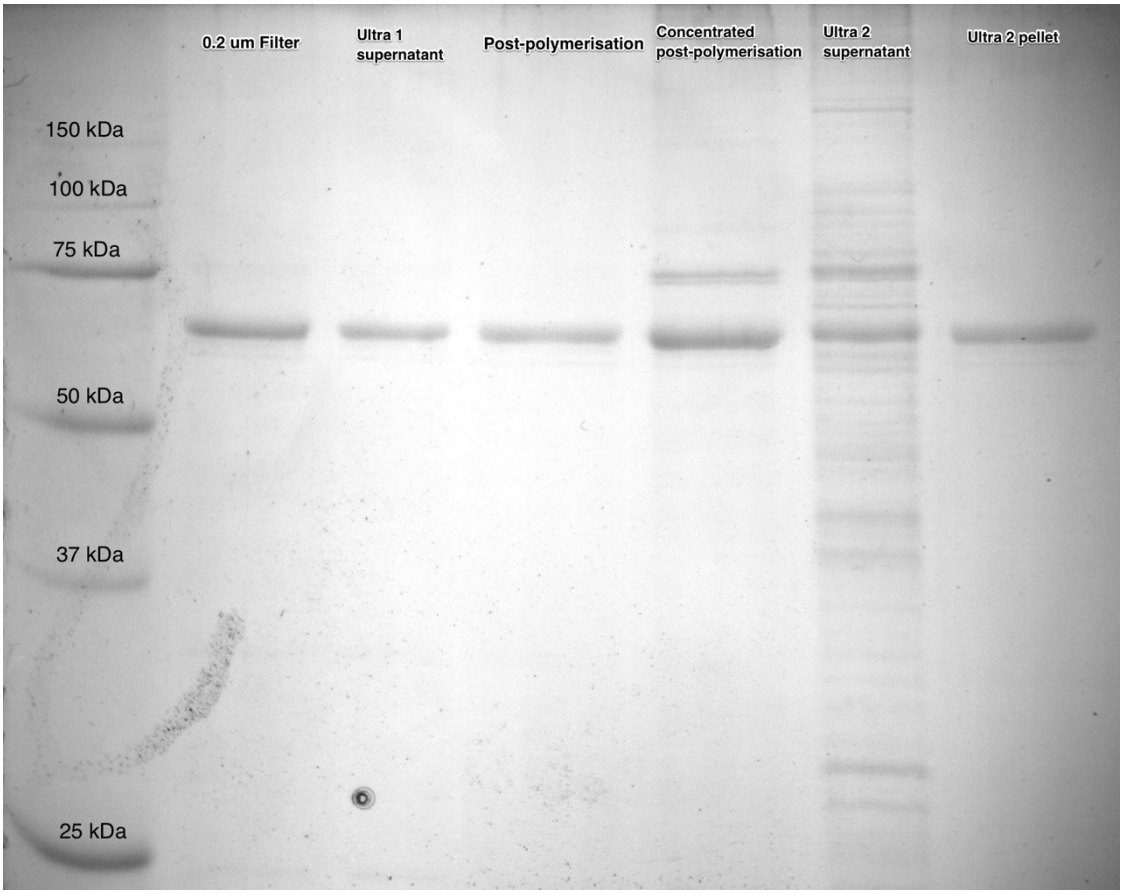


Figure 3.4.3: SDS-PAGE gel of 81116Δ$fliD$Kan^R knockout strain media and subsequent polymerization experiment steps. Lane 1: Protein ladder, Lane 2: Media filtered through 0.2 µm filter, Lane 3: 1st ultracentrifugation supernatant, Lane 4: post 2 days of polymerization, Lane 5: lane 4 sample concentrated, Lane 6: 2nd ultracentrifugation supernatant, Lane 7: 2nd ultracentrifugation pellet.

truly monomeric flagellin in the sample would be running a Western Blot incubated with anti-FlaA and anti-FlaB primary antibodies. Not only will this confirm the presence of flagellin, but also how much of each *C. jejuni* flagellin is exported into the media.

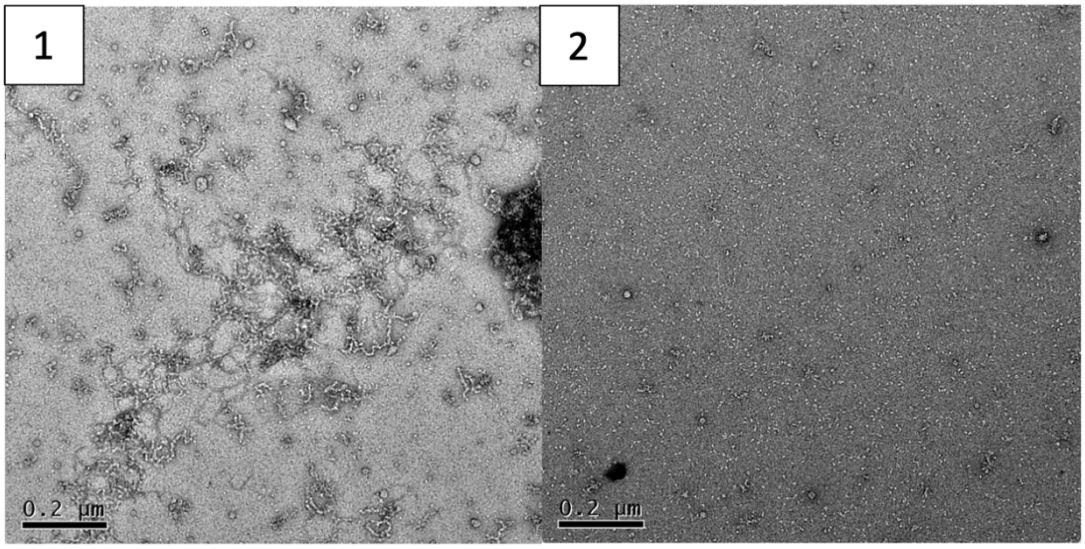
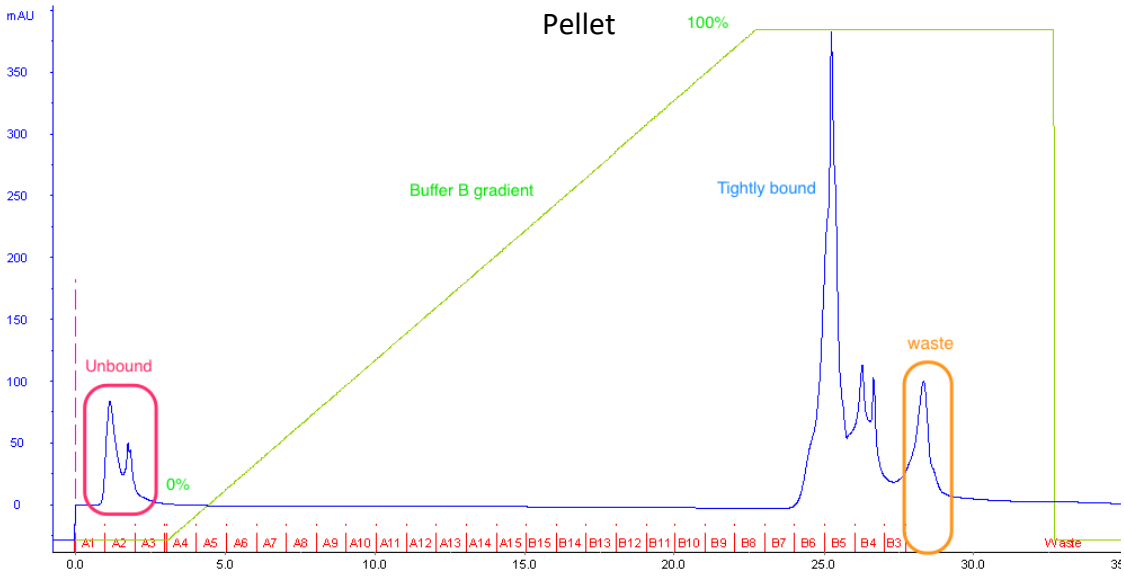
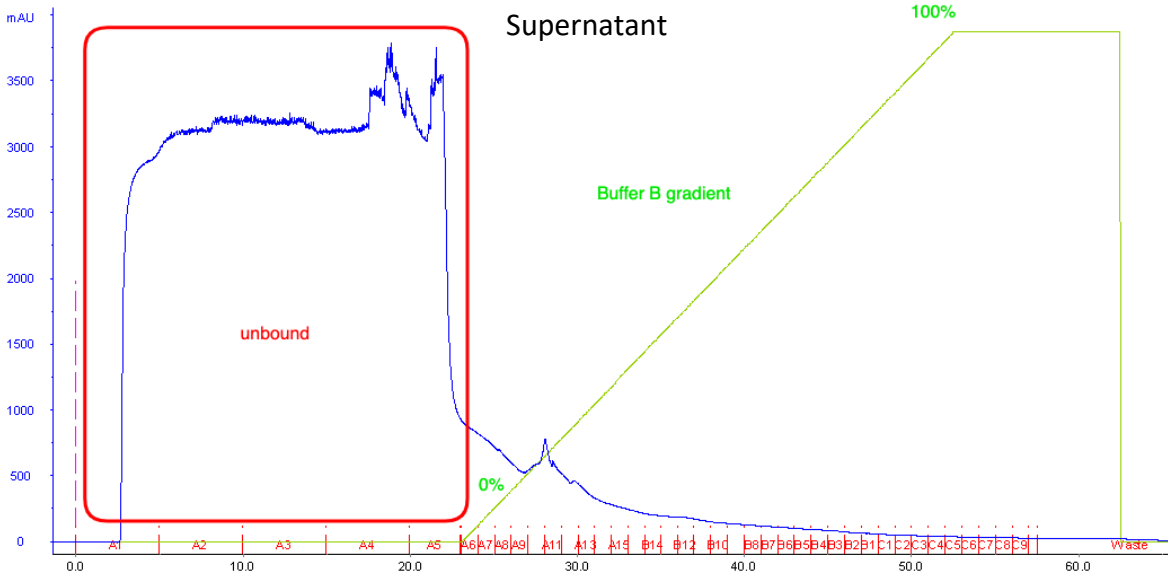


Figure 3.4.4: Negative stain images of flagella polymerization experiment. 1) Final ultracentrifugation pellet and 2) supernatant.

Both resuspended pellet and supernatant were purified using anion exchange columns as detailed in 2.2.4 and illustrated in Figure 3.4.5. The pellet eluted in 4 different peaks: the unbound peak at 0% Buffer B2, the very tightly bound peaks B6-B4 at 100% B2 buffer and the uncollected waste peak. The supernatant elutes mostly in an unbound state (majority of contaminants) and a single peak at A10-A11 at 20% B2 buffer. The pellet SDS-PAGE gel shows that none of the fractions corresponding to the peaks contained the 60 kDa flagellin band, indicating that it is not flagellin eluting in those fractions. The supernatant gel shows the majority of flagellin still present but elutes in the initial fractions. After observing the fractions via negative stain, A1-A4 were pooled together for further gel filtration as shown in Appendix 18. Gel filtration was conducted as described previously in section 2.2.3 and results tested via SDS-PAGE gel as illustrated in Figure 3.4.6. As flagellin does not have many aromatic residues it did not show as a high peak on the UV trace, while having a visible protein band on the gel. Fractions B7 and B6 were observed via negative stain and each fraction had relatively uniform small protein formations which we thought to be monomeric


Pellet
mAU
100%
Buffer B gradient
Tightly bound
Unbound
waste
0%
Waste


Supernatant
mAU
100%
Buffer B gradient
unbound
0%
Waste

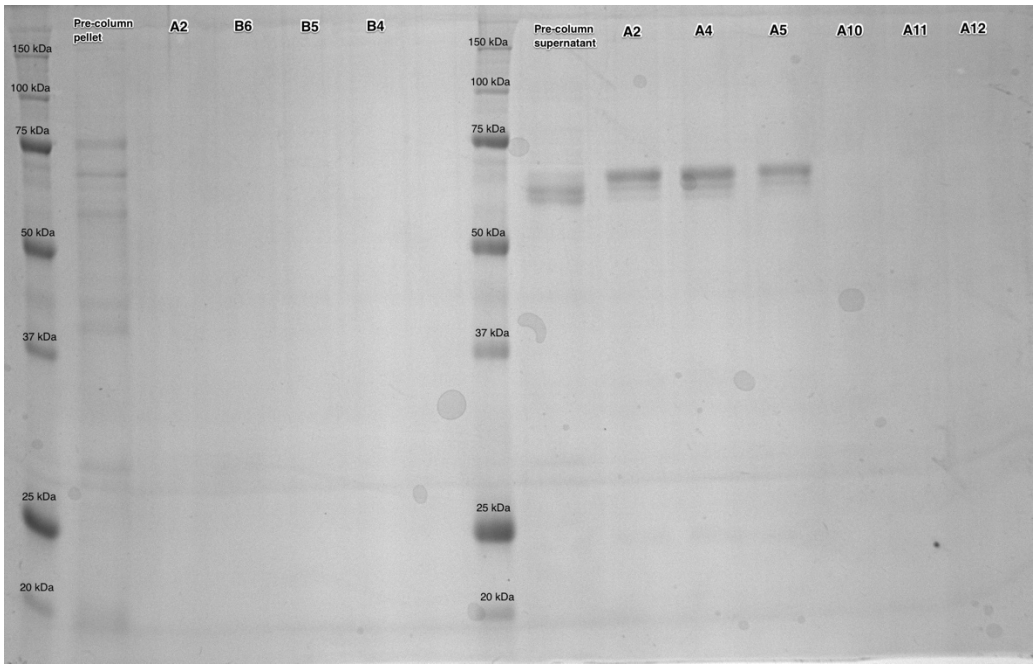

Pre-column pellet
A2
B6
B5
B4
Pre-column supernatant
A2
A4
A5
A10
A11
A12
150 kDa
100 kDa
75 kDa
50 kDa
37 kDa
25 kDa
20 kDa

Figure 3.4.5: Chromatograms and SDS-PAGE gel of anion exchange column output for pellet and supernatant from flagellin polymerization. Top chromatogram: ultracentrifuge pellet sample, bottom chromatogram: ultracentrifuge supernatant sample. **Red:** Unbound fraction, **Green**: Buffer B2 gradient. Left to Right SDS-gel: 1: Protein gradient ladder, 2: Pre-column pellet, 3-6 pellet elution fractions, 7: Protein gradient ladder, 8: Pre-column supernatant, 9-14: supernatant elution fractions.

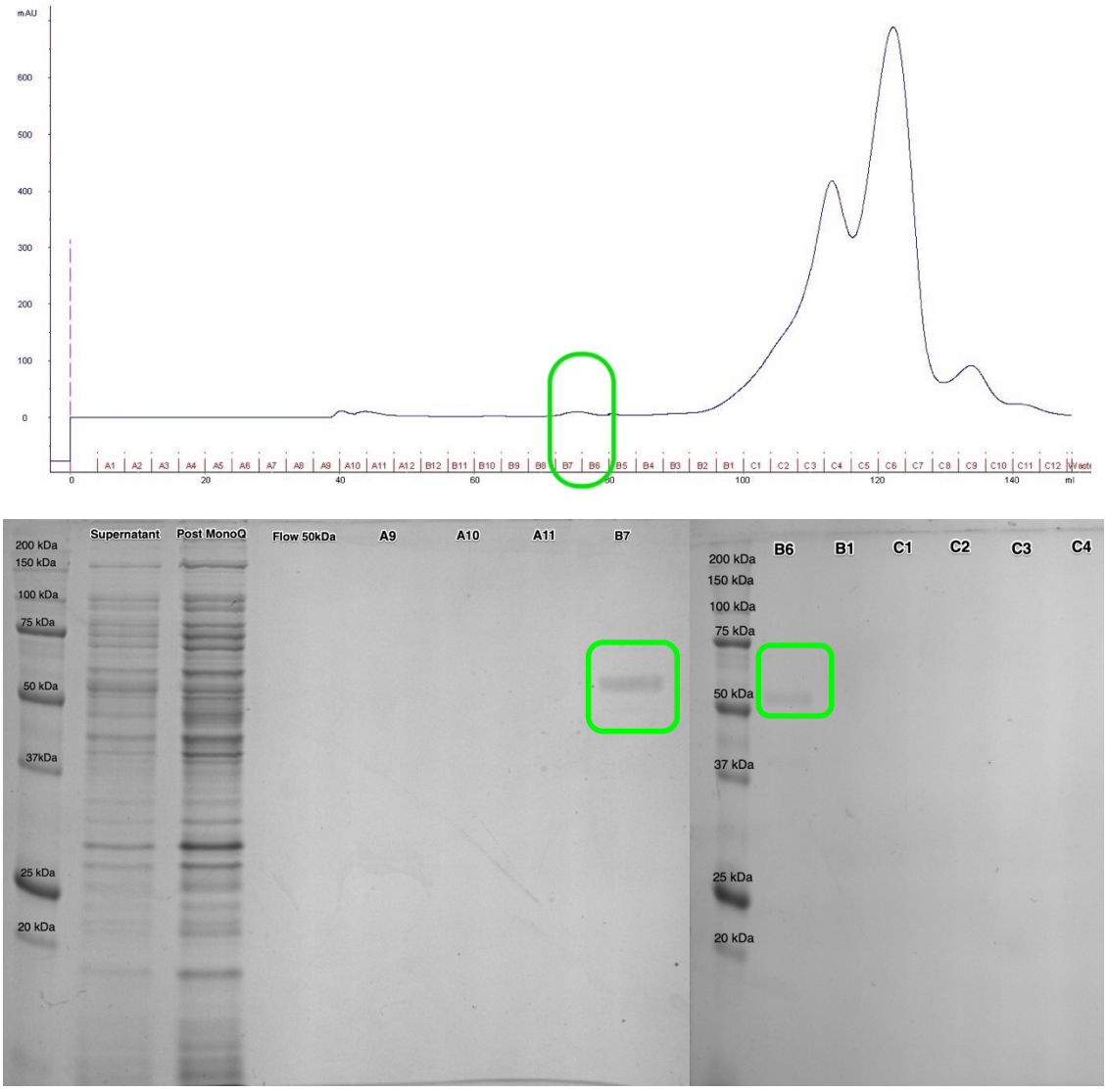


Figure 3.4.6: Chromatograms and SDS-PAGE gel of gel filtration column output for A1-A4 pooled fractions from flagellin anion column. Left to Right SDS-gel: 1: Protein gradient ladder, 2: Polymerization experiment supernatant sample, 3: Post-anion exchange column sample, 4: concentrator flow, 5-8: elution fractions, 9 Protein gradient ladder, 10-15: elution fractions continued.

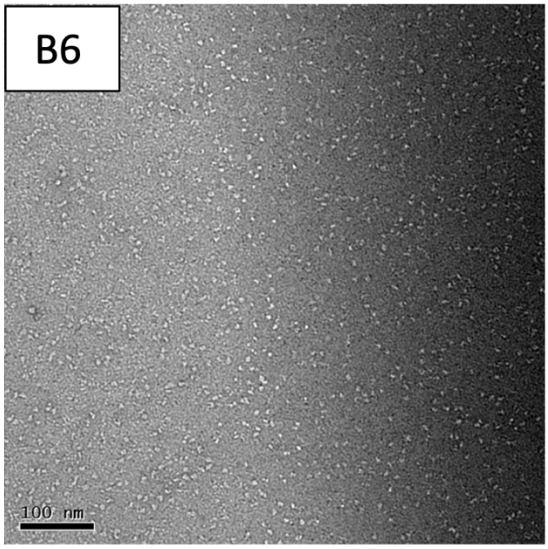


Figure 3.4.7: Negative stain images of gel filtration fraction showing monomeric flagellin protein of 60 kDa.

flagellin as illustrated in Figure 3.4.7, yet to be confirmed by a Western Blot. Now that there was a potentially pure flagellin sample, it could be used in combination with $FliD_{cj}$ to observe their interactions via ITC or other biophysical methods. However, polymerization experiments only resulted in aggregates and not ordered filaments. This could potentially occur due to the differences with *S. enterica* flagellar filaments in glycosylation and post-translational modifications. The control polymerization experiment resulting in polymerisation of *S. enterica* flagellar filament was not performed in this work, so the capability of isolated flagellins to form filaments with native helical conformation could not be confirmed. And if they did form, the packing of the filament will not be native to the flagellum, as essential proteins required for flagellar elongation are absent (FliD, FliS, FlgK platform).

3.4.3 Single Particle Cryo-EM Data Collection and Procession

The structure of the native *C. jejuni* filament was studied directly from wild-type cells. The manual dataset of 100 micrographs was collected inhouse and processed as detailed in sections 2.3.2-2.3.4 and the processing pipeline shown in Figure 3.4.8. The flagella segments were picked manually without differentiating between the FlaB and FlaA sections of the bacterial flagellum and classified as

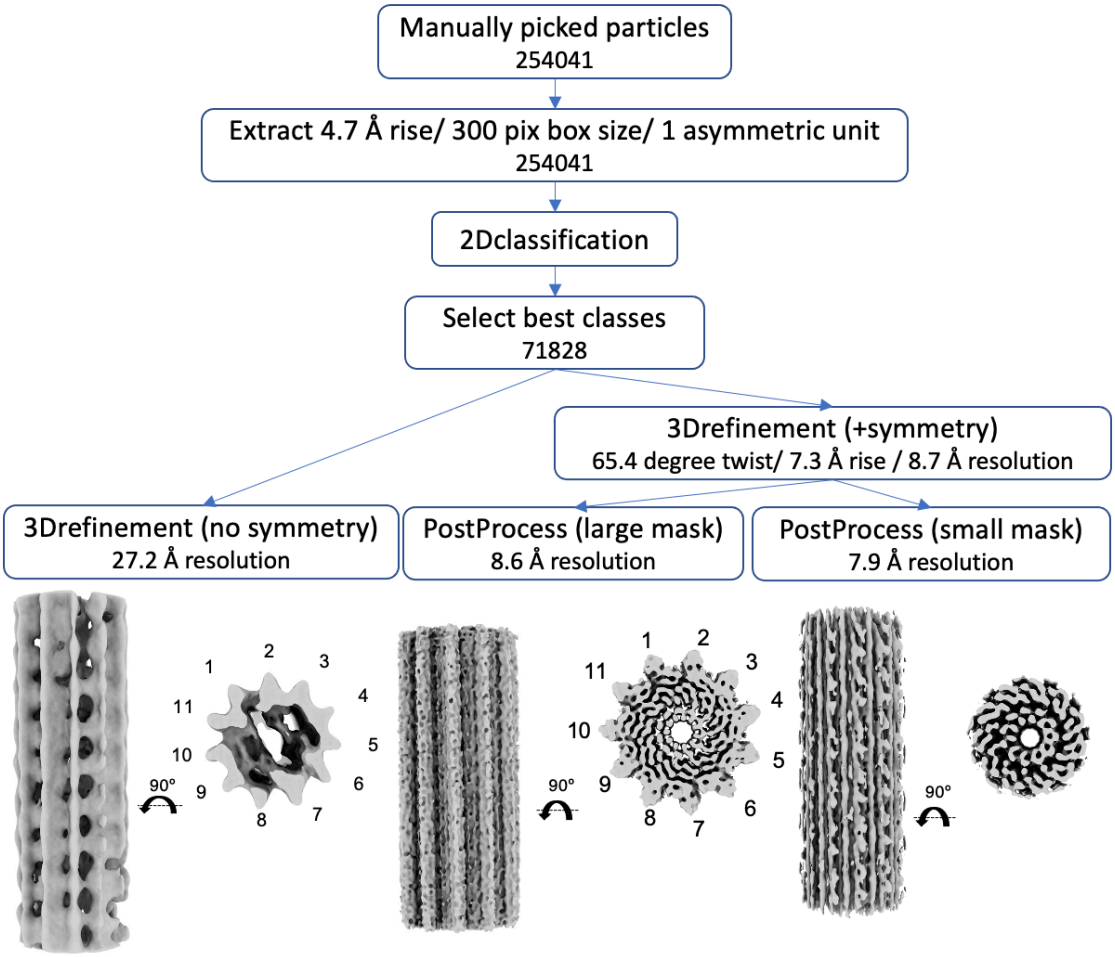


Figure 3.4.8: Workflow of manual filament data processing in Relion.

illustrated in Figure 3.4.9. To avoid biases due to symmetry, a 3D reconstruction was initially performed without any helical symmetry using a low pass filtered map of *P. aeruginosa* flagellar filament (EMDB:8855) as a reference. The obtained *C. jejuni* filament map clearly possessed 11-fold symmetry despite the low resolution (~ 27 Å), as illustrated in Figure 3.4.10. This demonstrates that the *C. jejuni* flagellar filament consists of 11 protofilaments with a lumen of ~25–30 Å and outer diameter of ~200 Å, alike that of other bacterial species. The map was therefore refined further by applying helical symmetry, initially with 65.5° twist and 4.7 Å rise as per the values used for the reconstruction of the *P. aeruginosa* flagellum filament (Wang et al., 2017). However, this reconstruction did not converge to a map with defined features and a smooth FSC curve. Multiple 3D refinements were performed with a search range for both

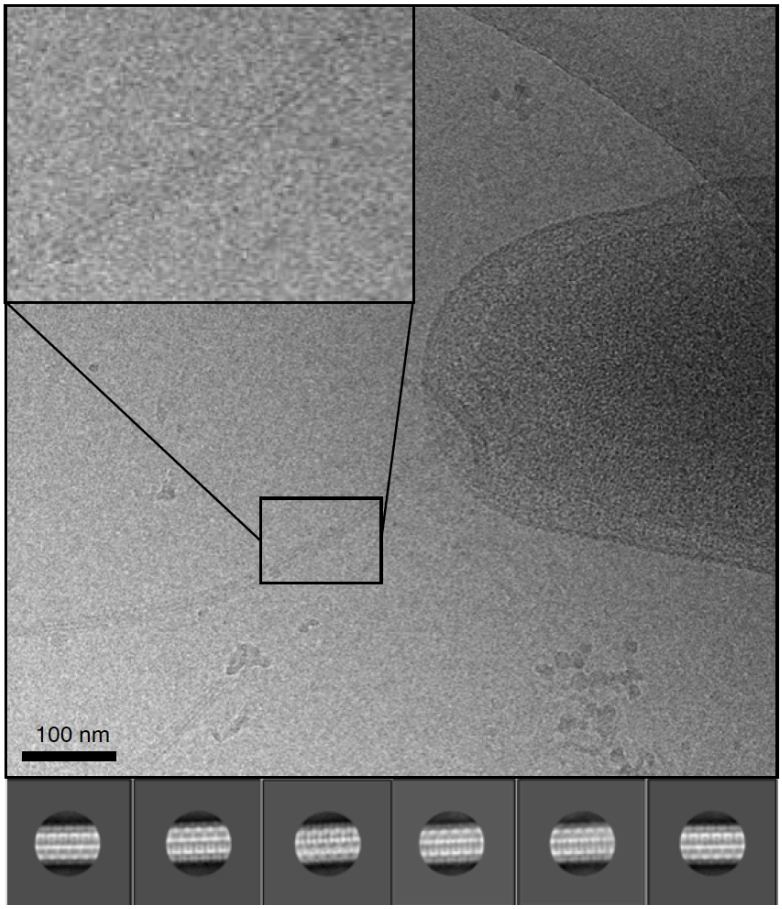


Figure 3.4.9: Cryo-electron micrograph of the native *C. jejuni* flagellum, used for the 3D reconstruction. 2D classes, generated from 71828 particles, are shown below. An enlarged image of a bacterial flagellum is portrayed in a panel at the top left corner of the micrograph. Adapted from (Al-Otaibi et al., 2020).

the twist and the rise, which converged on a 65.4° twist and 7.25 Å rise, which reached ~8.6 Å resolution. In this map the central D0–D1 domains are well resolved, with density for helices clearly visible, while domains D2 and D3 are visible, but less well resolved. A mask was used to exclude the D2 and D3 domain density and the resulting map of D0 and D1 domains was at a higher resolution with a less structured FSC curve.

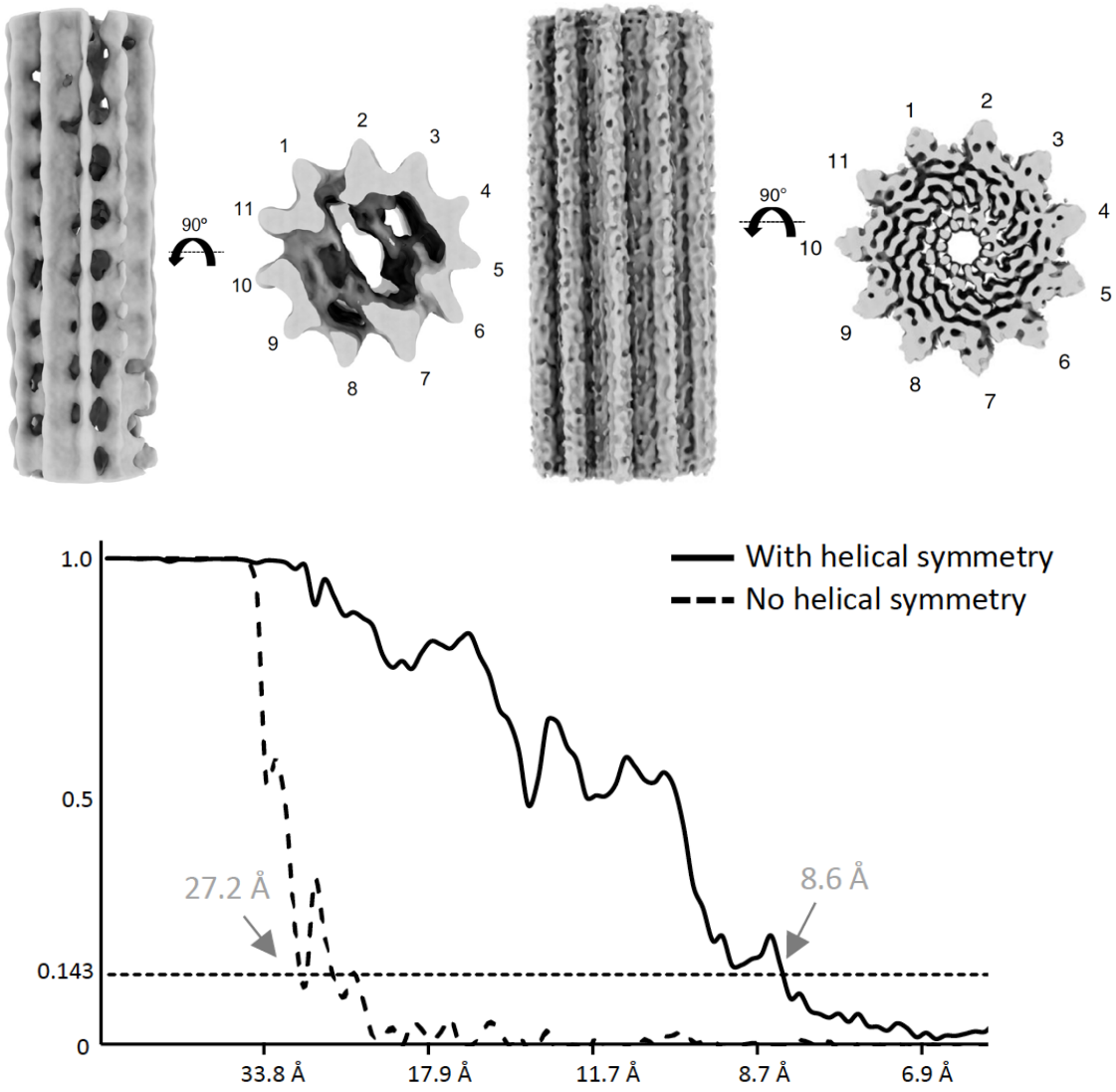


Figure 3.4.10: EM maps of the native flagellum, without (left) and with (right) helical symmetry applied, to 27.2 Å and 8.6 Å respectively as shown in the FSC below. Adapted from (Al-Otaibi et al., 2020).

An additional automatic dataset was collected from eBiC and processed as per sections 2.4.2-2.3.4 and the processing pipeline shown in Figure 3.4.11. While motion correction was done via Relion 3.0, the rest of the steps were done using CryoSPARC helical refinement function. Manually picked particles were classified into 2D classes with much better distinction between classes and detail than when processed in Relion 3.0, as illustrated in Figure 3.4.12. The best classes were used to generate a range of 3D refinement jobs each with a different search range for twist and rise and the best

four had the calculated values imposed on their symmetry. The ones with best features are shown in Appendix 19.

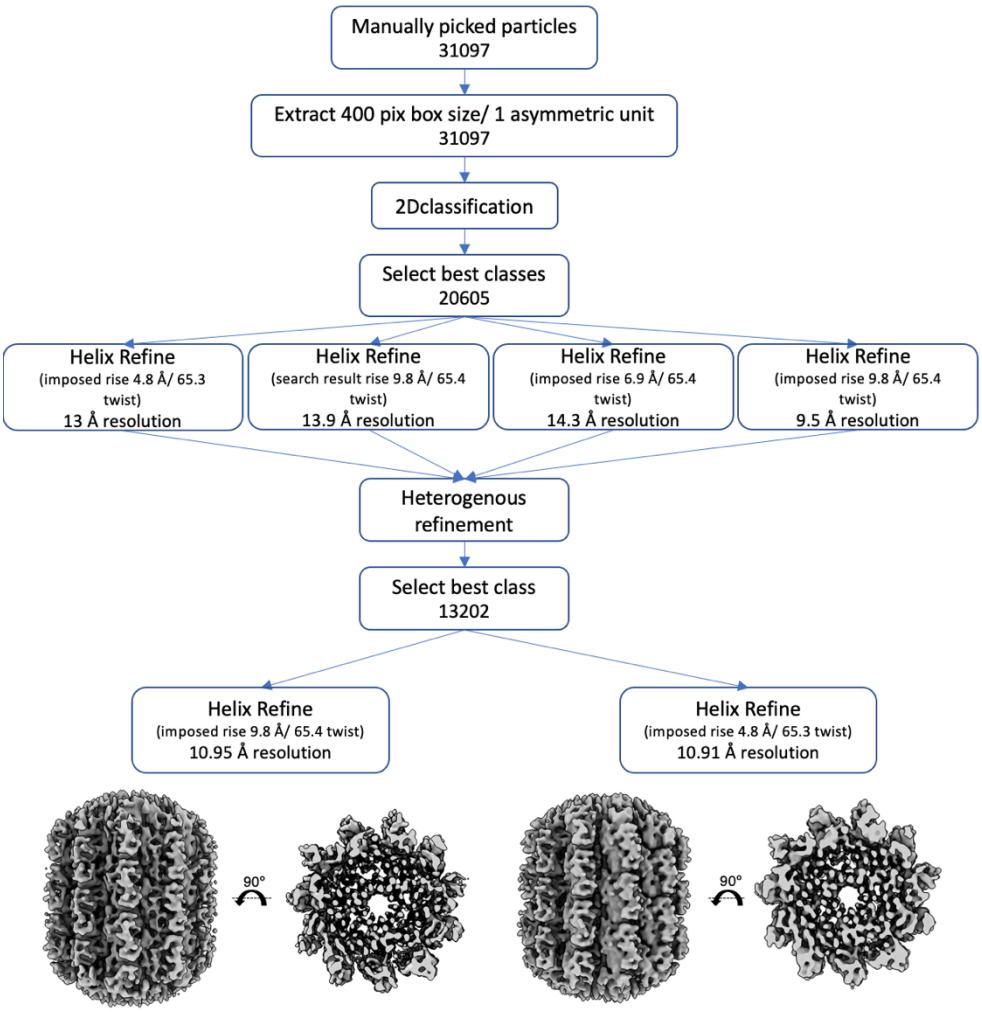


Figure 3.4.11: Workflow of Diamond Light Source filament data processed in CryoSPARC.

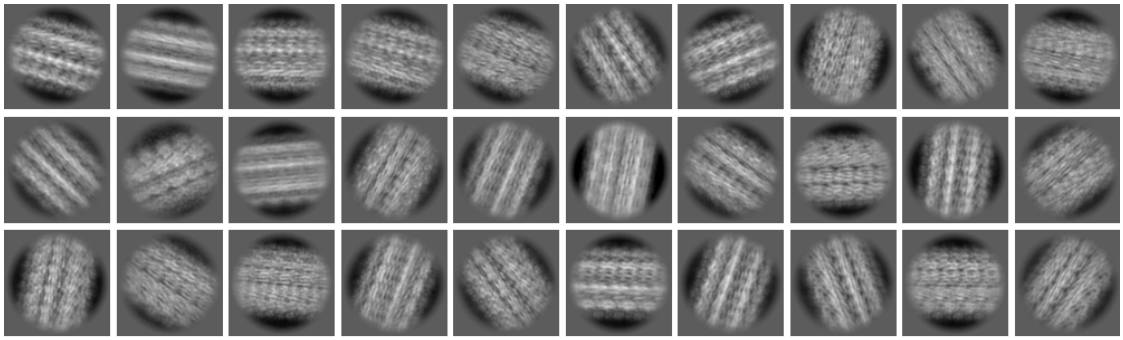


Figure 3.4.12: Selected best 2D classes of automated flagella data collection classified in CryoSPARC.

The different maps were input into heterogenous refinement function, as CryoSPARC does not have helical 3D classification unlike Relion 3.0. Best class was chosen with 13202 particles and two separate maps of highest resolution of 11 Å were achieved as shown in Figures 3.4.13 and 3.4.14. By the time of data collection, a *C. jejuni* flagellar filament structure has been published with a mutation in G508A causing straightening (Kreutzberger et al., 2020). As this work's dataset is from a native wild type filament, our aim was to compare the parameters provided by the paper and observe if there was any heterogeneity in 3D classes and filament conformations.

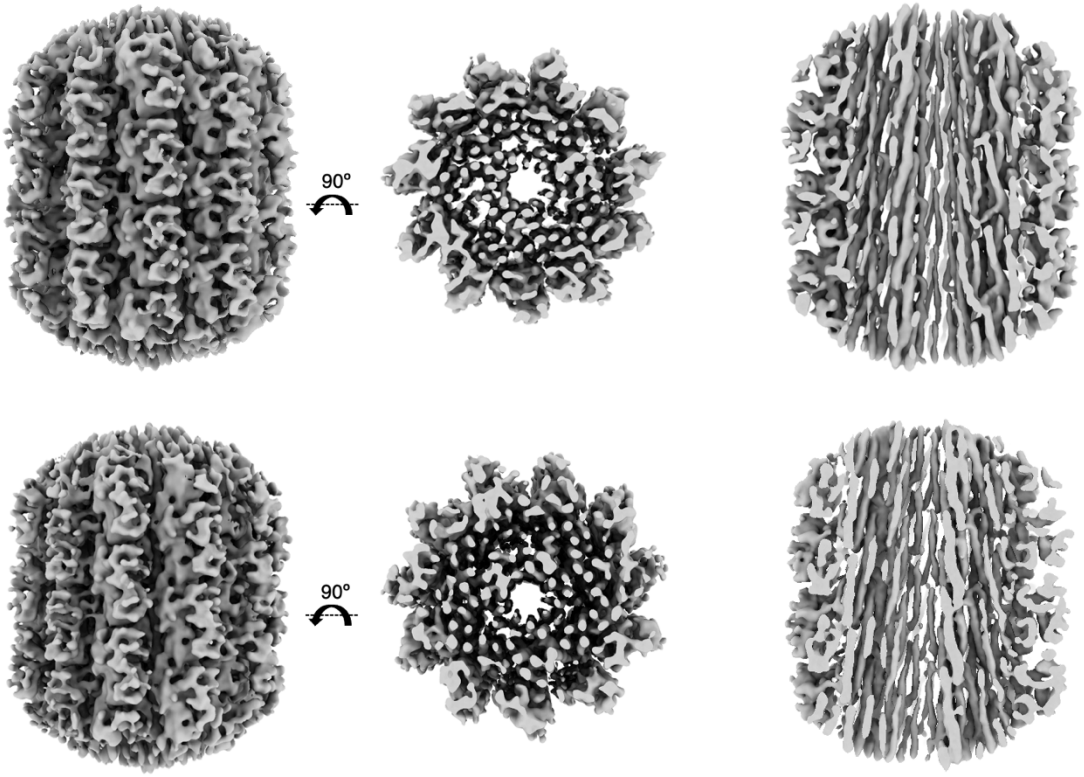


Figure 3.4.13: EM maps of the native flagellar filament processed in CryoSPARC. **Top:** 10.95 Å resolution Class 1 with 9.8 Å rise and 65.4° twist. **Bottom:** 10.91 Å resolution Class 2 with 4.8 Å rise and 65.3° twist. Side views are shown on the left and top views in the middle. A cross section view through the filament is shown on the left for each class.

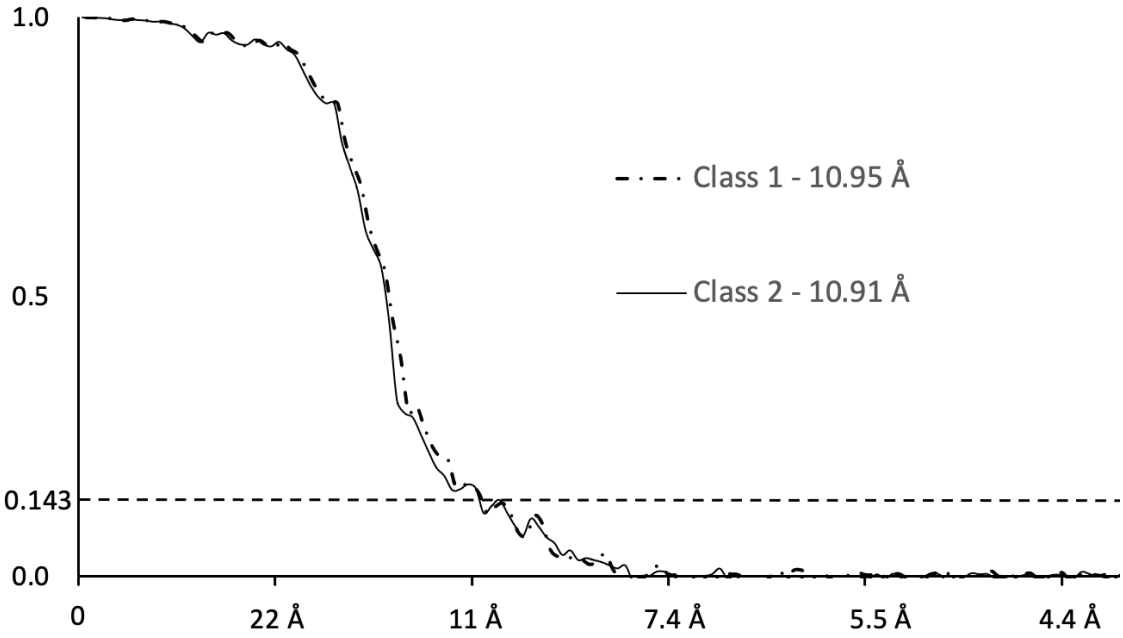


Figure 3.4.14: FSC of the EM maps in figure 3.4.13.

1st map thus had the referenced 4.8 Å rise and 65.3° twist applied as per the existing structure. The 2nd map converged on 9.8 Å rise and 65.4° twist, which had twice the rise but looked almost identical to the 1st map. After docking the existing *C. jejuni* flagellin structure (PDB:6X80) to both maps, as illustrated in Figure 3.4.15, the model fit both maps. This confirmed that the actual rise in both of our maps is 4.8 Å as in the published structure, but due to the complex helical symmetry of flagellar filaments, the software identifies multiple options for adjacent subunits to calculate the rise. However, as the resolution of the maps is too low to distinguish any further features, the minute differences between them could not be identified.

3.4.4 Model of Filament Elongation

After solving the structure of $FliD_{cj}$ (see section 3.2) a tomography map of FliD bound to the hook from *B. burgdorferi* was used to position the structure of the *P. aeruginosa* filament, as *C. jejuni* filament structure was not published at that time, and the cap complex structure of $FliD_{cj}$ reported here (Wang et al., 2017; Zhang et al., 2019). The reason why this structure was chosen was due to *B. burgdorferi* core flagellins being FlaA and FlaB, which are homologous to each other with FlaB forming

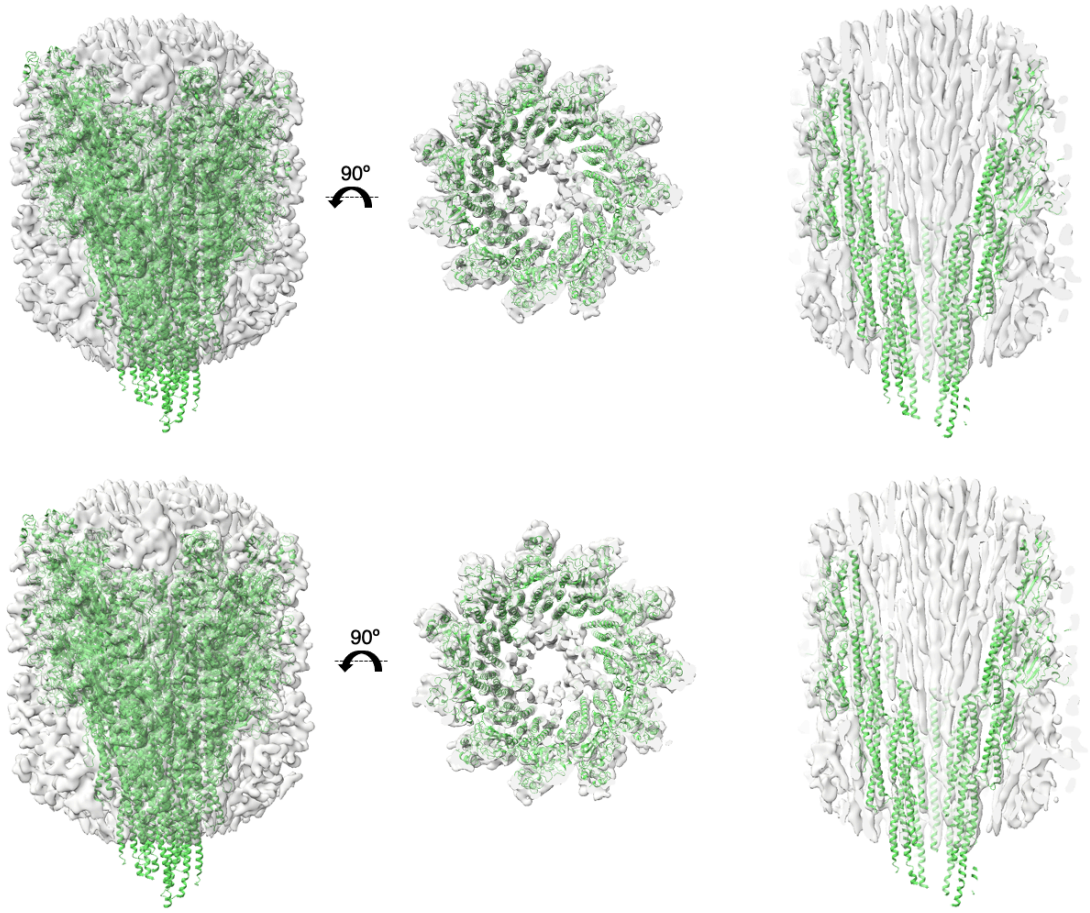


Figure 3.4.15: EM maps in Figure 3.4.13 fit to *C. jejuni* flagellar filament (PDB:6X80). **Top:** 10.95 Å resolution Class 1 with 9.8 Å rise and 65.4° twist. **Bottom:** 10.91 Å resolution Class 2 with 4.8 Å rise and 65.3° twist. Side views are shown on the left and top views in the middle. A cross section view through the filament is shown on the left for each class.

the distal filament and FlaA localized around the base, reminiscent of the case in *C. jejuni* (Sal et al., 2008). Three different possible positions of the FliD pentamer in relation to the flagellar filament could be fitted within the density, with one position showing no major steric clashes and with contacts between FliD and the filament consistent with the decameric interface of FliD as illustrated in Figure 3.4.16.

This allowed for suggestion of a model for FliD-flagellin interaction, where the C-terminus of FliD forms broadly non-specific, hydrophobic contacts with exposed regions of the filament, similar to flagellin-flagellin interactions. A gap between adjacent FliD molecules, on the side of the leg domain, is positioned in a suitable location for the insertion of a flagellin molecule and is the likely site of exit for nascent molecules. This however remains to be verified experimentally via cryo-EM data

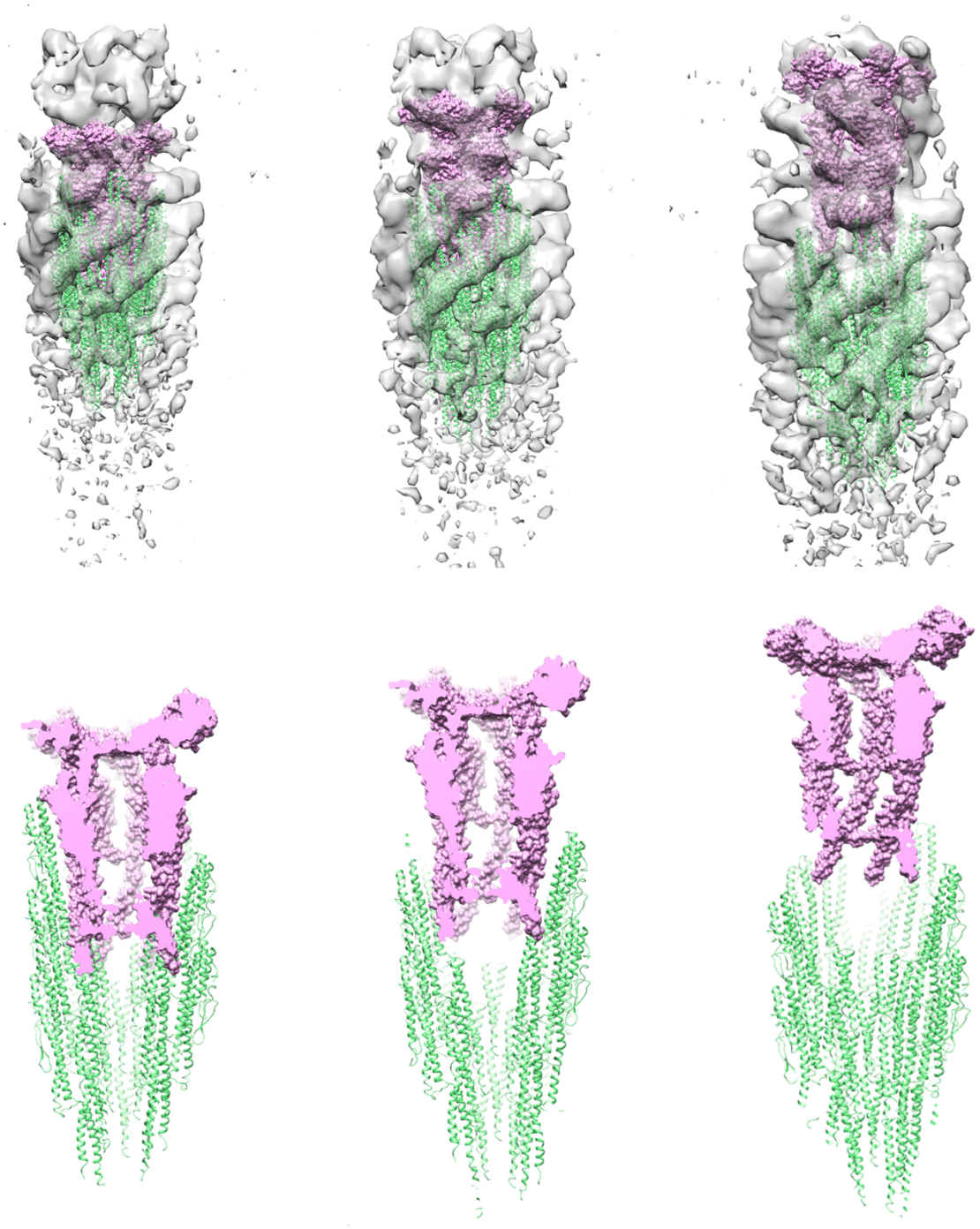


Figure 3.4.16: The 3 different fits of $FliD_{cj}$ (PDB ID: 6SIH) and the P. aeruginosa filament (PDB ID:5WK6) to the tomography map of B. burgdorferi (EMDB ID: 0525). Under each fit there is a cross section of the models illustrating the overlap of each fit.

collection focused on the flagellar tips to pinpoint the different conformations of FliD in filament elongation. While single particle analysis would work well for studying such

a small section of the filament, there might be an issue with preferred orientation in the ice. This could be solved by RCT and OTR techniques outlined in section 1.8.

3.4.5 DISCUSSION

The flagella purification attempts, including expression and purification of recombinant $FliC_{cj}$, while not resulting in a method to purify a concentrated flagellar filament sample allowed for understanding of the expression and properties of *C. jejuni* flagellin and obtaining a sample of unpolymerized monomers of suitable purity for subsequent biochemical and structural analysis. The ability of the $81116\Delta fliD\text{Kan}^R$ knockout strain generated in section 2.4.1 to act as an "open tap" for flagellin monomer export into the media, which in presence of FliD would have been controlled and incorporated into the filament instead, is essential to obtain large amounts of protein without resorting to recombinant expression. While the purification of media is a multistep process complicated by the flagellin not having many aromatic residues to show a clearly visible UV trace during HPLC columns and not binding the anion exchange column, it did result in a pure monomeric sample in the end which could be used for biochemical and biophysical studies.

While the purification attempts did not reconstitute the filament *in vitro*, we bypassed that fact by using native *C. jejuni* wild type cells and collecting a single particle dataset. The distribution of filaments across the grid as well as their curvature posed a challenge during processing and generation of the map. Due to sample heterogeneity, limited data and post-translational modifications prevalent in bacterial flagella, the resolution of the resulting map was not high enough during refinement without symmetry to identify the twist and rise from it. While it did confirm at the time that the *C. jejuni* filament has 11 and not 7 protofilaments as was published previously, trial and error was used to identify suitable parameters, which were not the same as the ones shown by a later paper from the same lab (Galkin et al., 2008; Kreutzberger et al., 2020). This could be most likely due to the sample used in this work having a range of different properties to the straightened mutant in the paper. The main difference is that the filament is in a native curved state and thus have a range of twists and rises,

causing difficulty in 3D reconstruction. While the twist and rise of the structure was not solved, a large dataset was collected and used to attempt to classify the potential different conformations of the filament in its native state. The result was two nearly identical maps at ~ 11 Å resolution with the same 65 ° twist but two different rises of 4.8 Å (as reported in (Kreutzberger et al., 2020)) and 9.8 Å. The fact that the latter is almost twice the size of the prior and the maps are nearly identical lead to a hypothesis that this occurs due to the complex helical symmetry of flagellin monomers in the filament and the processing software identifies adjacent individual subunits differently to calculate its rise. This would explain the difficulty of estimating the helical parameters using Relion 3.0 in the previous dataset. This also would explain why the lower resolution maps have more features than the 8.6 Å map shown in Figure 3.4.13, as they have the correct rise and twist.

While there is not yet a cryo-EM supported model for filament elongation, the interaction of FliD and flagellar filament was modelled using existing structures. A tomography map for FliD interacting with the hook in *B. burgdorferi* (EMDB:0525) was used to dock $FliD_{cj}$ structure solved in this work and the only high-resolution filament structure existing at the time, *P. aeruginosa* flagellar filament (PDB: 5WK6). This allowed for a proposal of a model of filament elongation as follows. Prior to assembling the filament, FliD is exported through the lumen of the flagellar hook and assembles at the end of the hook-filament junction (Minamino et al., 2008). After the successful folding of FliD at the tip of the hook, the filament elongation mechanism is initiated as illustrated in Figure 3.4.17. New unfolded flagellin molecules are secreted through the filament, and the model suggests that they would emerge in a chamber inside the cap complex (1) (Xing et al., 2018). According to the model, there is a gap between the D0–D1 domains of FliD that is not obstructed by the filament on one side. Unfolded flagellin molecules might exit the complex through this cavity (2) (Maki et al., 1998). Once outside of FliD, the exposed hydrophobic residues act as a chaperone, and promote flagellin folding in its insertion site (Motojima, 2015; Xing et al., 2018). In order to accommodate the next flagellin subunit, conformational changes need to occur to open an adjacent binding pocket. The folding of the new flagellin protomer leads to dislodging of the cap complex, that rotates by ~65° (4), thus positioning an adjacent

cavity of the cap complex close to the next flagellin insertion site (Maki et al., 1998; Song et al., 2017; Yonekura et al., 2000). This hypothesis agrees with previously proposed mechanisms of flagellar elongation (Maki-Yonekura et al., 2003).

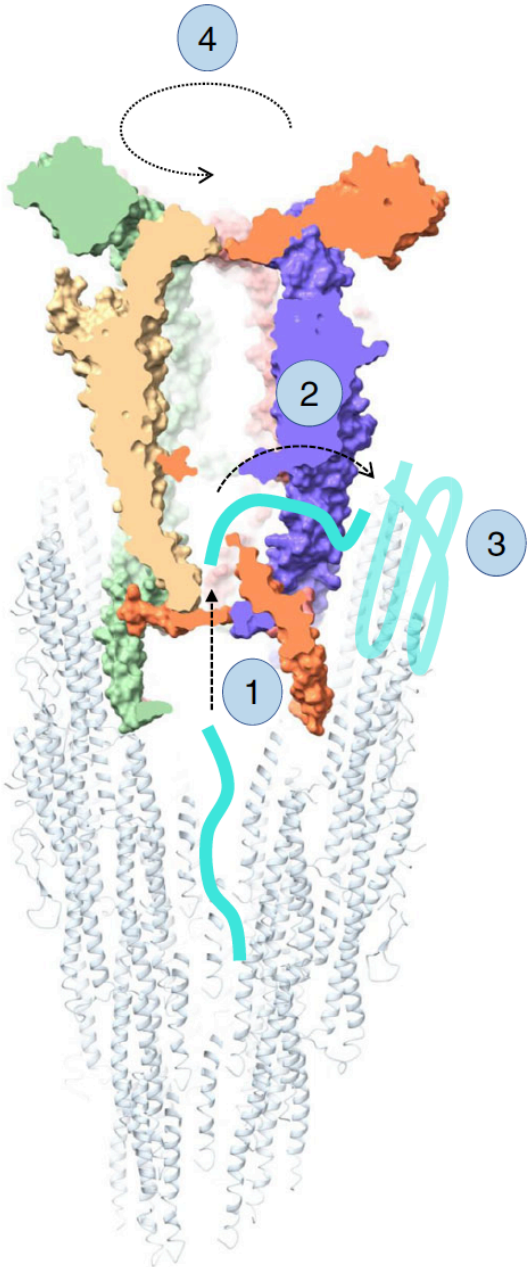


Figure 3.4.17: Model of cap-mediated filament elongation. Nascent flagellin molecules are secreted through the filament and enter the cap complex chamber (1). They then exit the cap complex through a side cavity (2), which positions them near the site of insertion. There, hydrophobic patches, composed of both exposed flagellin molecules, as well as FliD D0 domain, act as chaperones, and promote flagellin folding (3). Following the insertion of a new flagellin subunit, the cap complex rotates (4), positioning it to have the open cavity towards the next site of insertion.

Previous studies, based on low-resolution tomography and single particle data, have suggested that the D0 domain of FliD might be dynamic, with the leg domains opening and closing to promote filament elongation (Postel et al., 2016; Vonderviszt et al., 1998; Yonekura et al., 2000). In contrast, in the structure of the cap complex in this work, the D0 domain is rigid, locked in position by the N-terminal stretch. This structure therefore could provide evidence for a different mechanism, which had been proposed previously, whereby the cap complex acts as a rigid cog that rotates during flagellum elongation (Song et al., 2017). It has been observed that FlgD hook capping protein, which builds the flagellar hook, has a mismatch in symmetry when bound to the rod (prior to building the hook) allowing for conformational change in the D0-D1 domains of the protein during hook formation (Johnson et al., 2021). As this flexible model was observed using focused refinement on a large single-particle dataset to a high resolution it could be that the static model in this work is an artifact of FliD forming decamers in solution and in its native state the terminal domains move as well. Nonetheless, further experiments to characterize the flagellum-cap complex at high resolution will be required to confirm this model, and to observe if the rigidity of the D0 domain, as observed in the structure of the cap complex in this work, is also present when FliD is bound to the filament.

3.5 Exploiting Structure of $FliD_{Cj}$ for New Therapeutic Development

3.5.1 Introduction

FliD as a capping protein is not only essential for filament formation and growth but has been shown to bind glycosaminoglycans (GAGs) specifically heparin in a dose dependent fashion (Freitag et al., 2017). Heparin and many other GAGs are present in abundance on intestinal epithelial cells, thus allowing *C. jejuni* cells to adhere to the cell lining and increase the infection rate. The binding site was narrowed down to a potential interface between D2, D3 and D1 domains, but without a $FliD_{cj}$ structure it was not possible to confirm this.

ITC is a biophysical technique that determines thermodynamic parameters of the interaction of two experimental samples in solution (Pierce et al., 1999). As one of the samples (ligand) is introduced to the other (sample) via titration, the heat is released or absorbed as the result of the interaction. The instrument detects this heat by measuring the changes in power needed to maintain isothermal conditions between the reference and sample cell. As more ligand is introduced, the sample becomes saturated and peaks in heat change decrease. Eventually the peak size remains constant upon full saturation. Instrument software integrates the peaks and presents them in a Wiseman plot. After an appropriate binding model is chosen, the enthalpy change (ΔH), the dissociation constant (K_d), association constant (K_a) and stoichiometry (n) is calculated. From these Gibbs free energy (ΔG) and entropy (ΔS) changes can be calculated. Entropy is a measure of randomness and disorder of the system. Gibbs free energy is a thermodynamic potential that can be used to calculate the maximum reversible work performed by a closed system at a constant temperature and pressure. These three values along with the absolute temperature (T) measurement and gas constant (R) form an equation below:

$$\Delta G = -RT \ln K_a = \Delta H - T\Delta S$$

For an accurate measurement of binding affinity, the thermogram must be sigmoidal. The curve profile is determined by c-value which is calculated as follows, with M being the concentration of protein in the cell:

$$c = nK_a M$$

While the location of the binding site between two samples is not possible to determine via ITC, the stoichiometry of the interaction and the K_a as a measure of the strength of binding can be measured. This is what was aimed to be determined for $FliD_{cj}$ purified sample and heparin.

FliD itself is also an antigen, responsible for immunoglobulin response in 97% of infected patient sera in *H. pylori* and inducing a strong IgG response (Ghasemi et al., 2018). SIgA is the main antibody in mucous secretions and is highly involved in the infection of enteric pathogens (Perruzza et al., 2020). As we have already solved the structure of $FliD_{cj}$, observing binding of these antibodies to it could allow us to determine $FliD_{cj}$ quality as a drug target.

3.5.2 $FLID_{CJ}$ BINDING TO HEPARIN

To quantify the binding affinity between $FliD_{cj}$ and heparin, the ITC experiments were conducted as detailed in section 2.3.6. To set up the experiment, a set concentration of $FliD_{cj}$ and heparin in mM was placed into the sample cell and syringe respectively. Raw heat rate (µJ/s) was measured over time (s), plotted and analysed using NanoAnalyze software (TA). While multiple experiments with varying parameters were conducted as per table 2.3.2 in the methods section, results of which are shown in Appendix 20, the best representative has been shown in Figure 3.5.1. In the analyses below $FliD_{cj}$ will be referred to as “sample” and heparin as “ligand”.

All raw data showed negative peaks in the sample cell upon ligand addition, indicating that the reaction is exothermic. This is caused by decrease in power to maintain the constant temperature in the cell, as the exothermic reaction releases thermal energy into the surrounding solution. The software integrated the peaks and converted them to Wiseman plot, fitting the model that passes through most of the raw data points, from which the following values were measured: ΔH as the difference between the highest and lowest raw data points, molar ratio from the midpoint of the sigmoidal curve and K_d was measured from the slope of the graph.

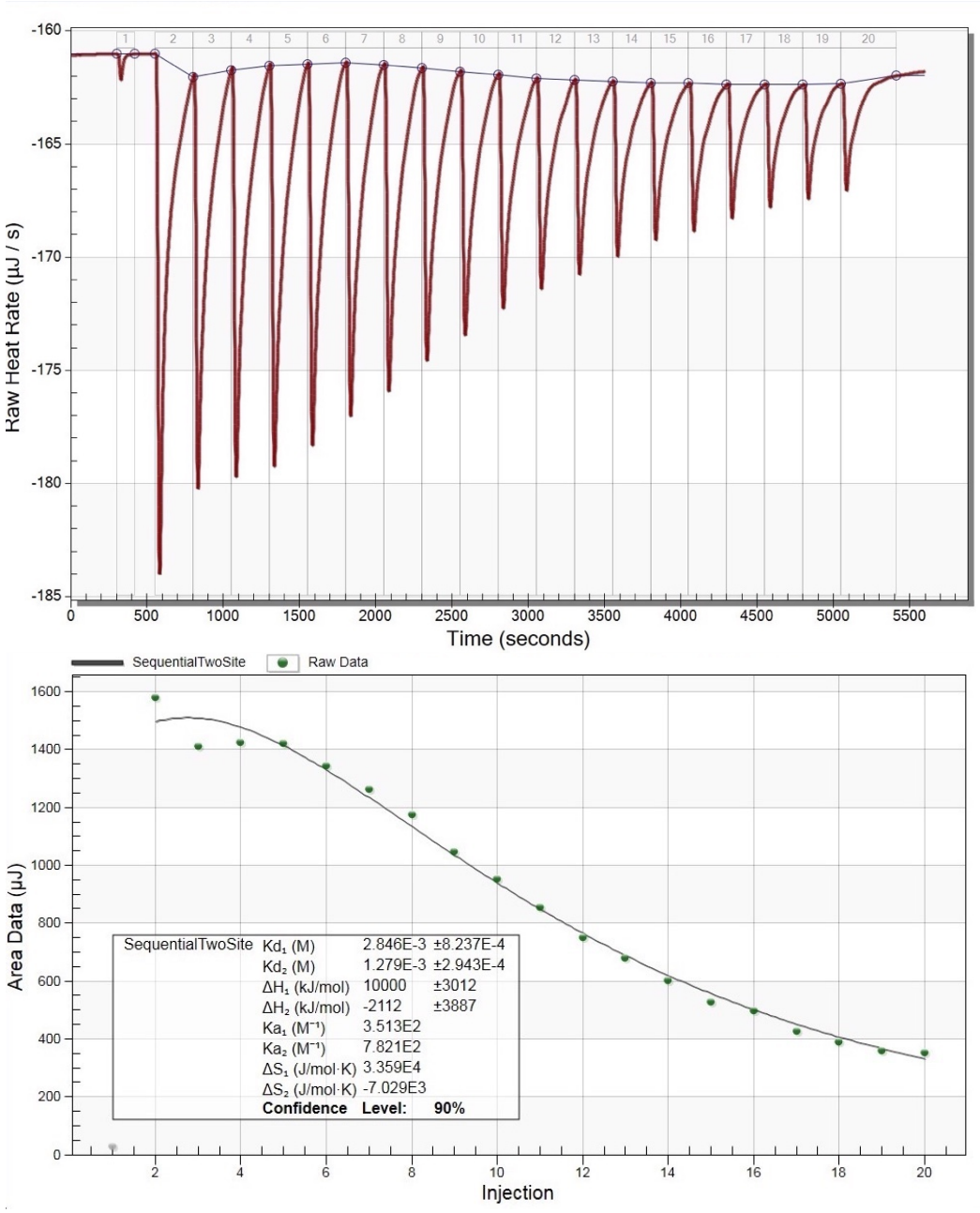


Figure 3.5.1: Raw (above) and Wiseman plot (below) of ITC data for 0.01 $FliD_{cj}$ in cell, 0.01 mM Heparin in the syringe, 250 second injection length, 2.5 µl injection volume and 20 injections.

To correctly choose a model fit for the Wiseman plot, there were a few points needed to be considered. As previous data for $FliD_{cj}$ and heparin suggests that the wide region encompassing some of the D2-D3 head domains and the majority of the D1 helical bundle has the potential to bind heparin (Freitag et al., 2017), in the native pentamer state of FliD there could be a possibility of multiple binding sites on the flagellar tip. As it often occurs in other biological systems, there is a possibility that

binding of heparin to one site will influence whether the second heparin would bind to the other, considering the large size of the polymer. In addition to this, the recombinant sample of $FliD_{cj}$ used was a decamer and thus had two sets of pentamers facing in opposite directions, which would affect the stoichiometry of the reaction, meaning that per molecule of $FliD_{cj}$ in solution we would have at least 2 heparin chains bound at each end. This made it difficult to identify the stoichiometry of the reaction accurately and using a generalized model, such as sequential two site model, was preferable. The sequential two site model corresponds to more than two ligand binding sites that might be identical or non-identical, independent or cooperative. This is a general model for any possible scenario with n ligand binding sites. While multiple other binding models exist, they were attempted to be fitted by the software but did not pass enough raw data points and had a low fit confidence level and thus were not included. The only model that fit the plots was the sequential two site model. As we did not know the stoichiometry of the interaction, we had to assume the 1:2 binding model and vary the protein concentration, injection volume/number and length as indicated by methods section table 2.3.2.

The fitting of the model resulted in parameters listed in Table 3.5.1. K_d dissociation constant measures the tightness of binding interactions. The smaller the value the tighter the binding. Across all experiments, even if the values varied the K_d was in the mM range, indicating weak binding across all sites. While the ΔH_1 was in the range of 10000 kJ/mol, some experiments included a negative ΔH_2 value. In exothermic reactions ΔH is always negative because the total energy of the products is less than the total energy of the reactants. Thus the last 2 experiments and their fitted model have a theoretically correct ΔH_2 even with the large error. The positive value of the other ΔH values across the experiments, while indicating endothermic reaction and intake of heat from the system, are not supported by the thermograms. While this could be an experimental error, this could also mean that the initial interaction with one site is exothermal and releases energy, while interaction with the other site(s) requires an additional input.

Table 3.5.1: Calculated parameters for the fitted models. Run numbers correspond to the ones in table 2.3.2. Run 1 is the only one with a 0.02 mM concentration of $FliD_{cj}$ the rest have 0.01 mM. Outliers are coloured red.

Run number	**Run 1***	**Run 2**	**Run 3**	**Run 4**	**Run 5**
K_{d1} (M)	6.32±1.27 x10^{-3}	0.72±2.95 x10^{-3}	3.76±0.54 x10^{-3}	3.23±0.82 x10^{-3}	2.85±0.82 x10^{-3}
K_{d2} (M)	8.18±2.2 x10^{-3}	2.12±3.70 x10^{-3}	2.20±0.61 x10^{-3}	1.36±0.32 x10^{-3}	1.28±0.29 x10^{-3}
ΔH_1 (kJ/mol)	8928±1720	1492±6709	10000±699	10000±2671	10000±3012
ΔH_2 (kJ/mol)	4621±2761	5000±8903	5000±939	-239±3559	-2112±3877
K_{a1} (M^{-1})	1.58 x10^2	1.40 x10^3	2.66 x10^2	3.10 x10^2	3.51 x10^2
K_{a2} (M^{-1})	1.22 x10^2	4.74 x10^2	4.55 x10^2	7.34 x10^2	7.82 x10^2
ΔS_1 (J/mol.K)	2.99 x10^4	5.06 x10^3	3.36 x10^4	3.36 x10^4	3.36 x10^4
ΔS_2 (J/mol.K)	1.55 x10^4	1.68 x10^4	1.68 x10^4	-7.48 x10^2	-7.03 x10^3
Model fit confidence level	**90%**	**95%**	**95%**	**95%**	**90%**

Overall, while the ITC data was not usable to calculate the accurate binding affinity of $FliD_{cj}$ and heparin, it did indicate that the reaction is exothermic and there is weak binding between the two. To confirm the biophysical data, there was an attempt to generate a cryo-EM structure of recombinant purified $FliD_{cj}$ and heparin as detailed in section 2.3.2. The same processing parameters were used as that of the solved $FliD_{cj}$ apo structure, but from the initial 2D classes and 3D refinement maps there was no visible additional density that could correspond to heparin and the structures simply looked like unbound $FliD_{cj}$ decamers but of much lower resolution than the already solved structure as illustrated in Figure 3.5.2. This probably occurred due to the low affinity of binding as observed in the biophysical data above.

There are multiple other techniques, both qualitative and quantitative, that could be used to test this binding event. One of such quantitative methods is affinity fractionation of FliD on Sepharose column with covalently linked heparin. Protein bound to the heparin will be eluted with different salt concentrations proportional to

the K_d of the interaction. The higher the affinity, the higher the salt concentration required to displace the bound ligand. Low affinity ligands would elute early in the column with NaCl concentrations below 150 mM. This technique could also be used in reverse, with FliD immobilized to the column and passing a range of GAG's, thus identifying heparin affinity in addition to other potentially better binders among the glycosaminoglycans. This might be a good approach as most GAG interacting proteins interact with heparin sulphate or heparin out of all the other GAGs (Varki et al., 1999).

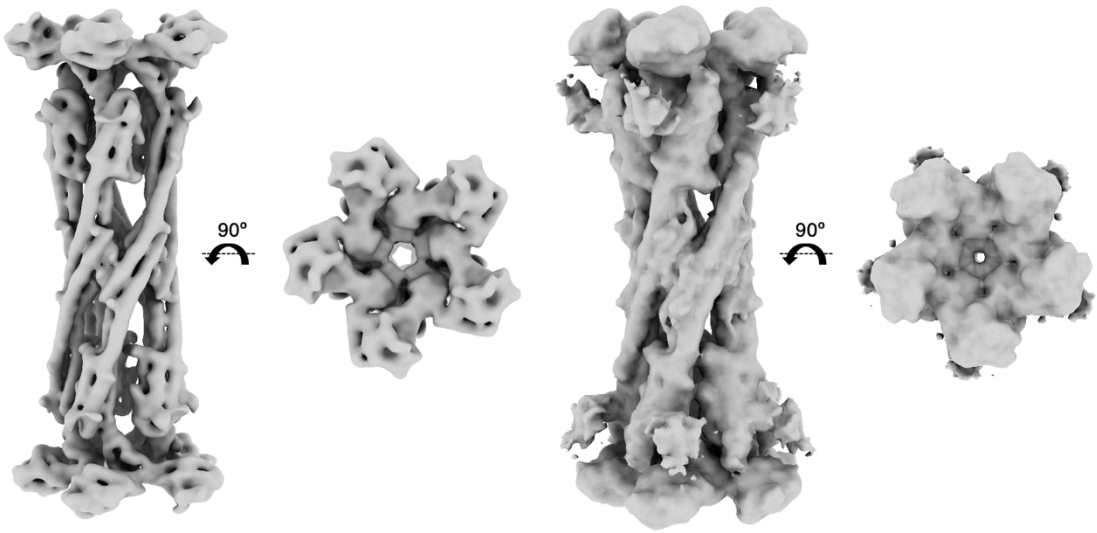


Figure 3.5.2: 3D map of $FliD_{cj}$ and Heparin dataset at regular (left) and low (right) contour value. Side views are to the left and top views to the right of each structure.

A both qualitative and quantitative analysis of FliD/heparin binding is affinity co-electrophoresis, where radioactively tagged GAGs are electrophoresed through acrylamide gels containing FliD. Association slows down the mobility and this can be visualized by imaging. Protein concentration variation can help measure the affinity but would require multiple gels (Varki et al., 1999). Another technique capable of measuring binding affinity, if access to it is available, is surface plasmon resonance (SPR). It is based on an optical biosensor technique measuring binding events on a flat metal surface by detection of change in local refractive index upon binding. A heparin or heparan sulphate sample with known length can be biotinylated and immobilised on to the streptavidin chip. FliD diluted in buffer is then injected and flowed over the chip. This method has the added advantage of mimicking the biological system, as

heparin sulphate is immobilized on the surface of the epithelial cells and FliD would be bound to the flagellar filament and flowing over it in solution (Zhang et al., 2015). As the structure of *C. jejuni* FliD is now available, modelling of interactions with GAG-dock method is possible without the need for a crystal structure of heparin bound to the protein to start (Griffith et al., 2017).

3.5.3 FLID$_{CJ}$ AS AN ANTIBODY TARGET

With antibodies provided by Matteo Pizzuto and Davide Corti from HumAbs BioMed, the purified FliD$_{cj}$ was used to make negative stain grids and observe initial binding. FliD-reactive monoclonal antibodies CAA1 and CCG4 were isolated from IgA^{+} and IgG^{+} memory B cells from tonsillar donors and recombinant Fab only constructs generated (Perruzza et al., 2020). Samples were prepared as in section 2.3.6 and negative stain data collected and processed as per section 2.3.1. Constructs tested were generated from human monoclonal antibodies CCG4 and CAA1, isolated for their reactivity against FliD (Perruzza et al., 2020). Each had two variations: full IgG and rFab fragment. Two SlgA antibodies CAA SlgA1 and CCG4 SlgA2 were also tested. While representative negative stain micrographs can be observed in Appendix 21, the 2D classes generated from cisTEM are shown in Figure 3.5.3.

For the CCG4 IgG sample through various classifications the two distinct classes are visible: A non-bound FliD$_{cj}$ and FliD$_{cj}$ with a density around the leg domain area (shown in red on the figure). This density could potentially be the bound antibody. However, unlike the rest of the samples this is the only one with a distinguishable class that has an additional density. No other classes provided clear additional density, precluding us to conclusively determine if the antibodies are bound, and if so, where.

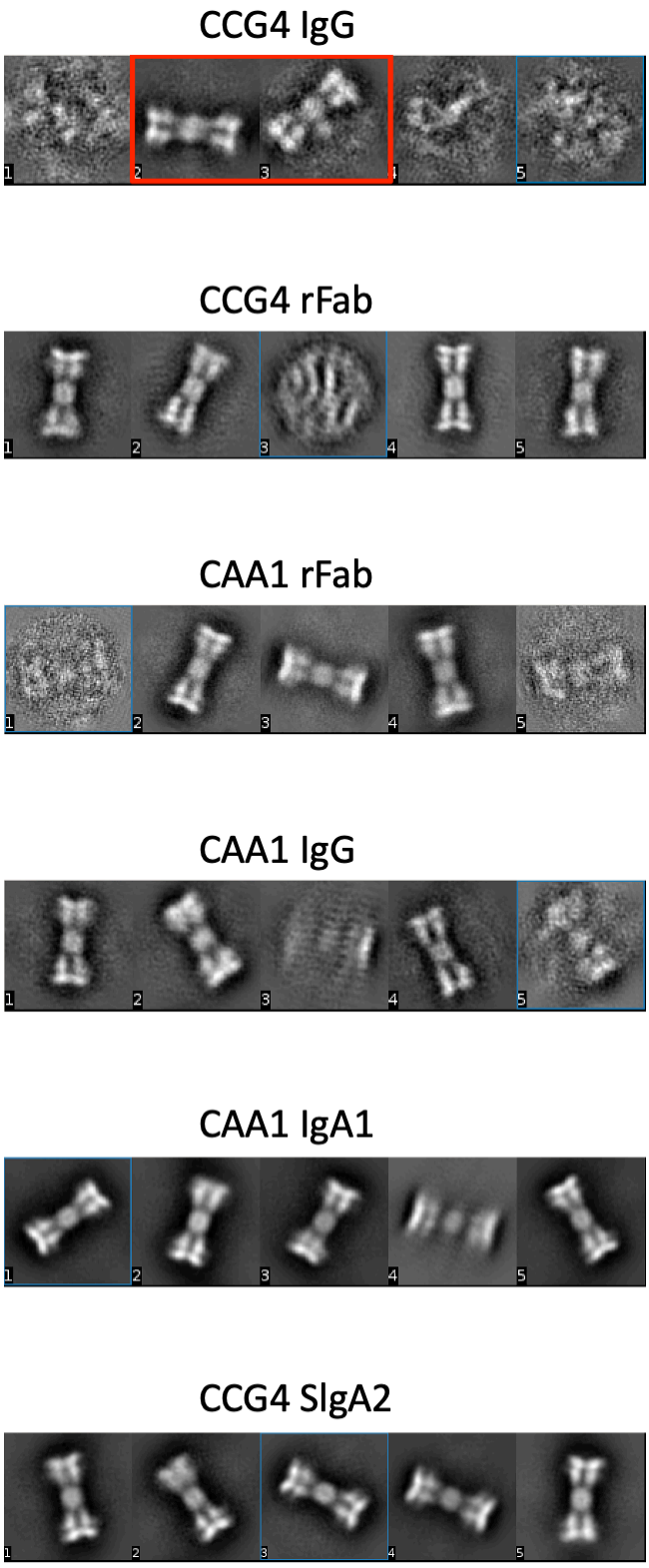


Figure 3.5.3: 2D classes generated by cisTEM for each antibody sample negative stain data collection. **Red:** potential antibody binding instance.

This could occur due to multiple reasons, however the most likely one is that the antibodies are capable of binding, but the experimental conditions make it inefficient, so that the percentage of bound particles to unbound is too low. Thus, during 2D classification the bound class is overshadowed by the unbound. It could also be possible that the binding site for these antibodies is obscured in the decamer and only available in the functional FliD pentamer. To confirm whether the density is the antibody or not, the next step would be co-purifying the antibodies with $FliD_{cj}$ and observing the complex via negative stain once more.

To improve this binding and optimise the experiments, the ratio of protein to antibody could be the first factor to be increased. However, prior to data collection, higher concentrations of antibodies were used with the same concentration of $FliD_{cj}$ and the negative stain grid was overloaded with antibodies making it difficult for FliD particles to be observed. But as the efficiency in a diluted sample was low, these overloaded samples could be revisited.

Other factors that can be optimised to improve antibody binding to protein are salt, pH and temperature. Antibody binding is more efficient at low temperatures and a range of 6.5 to 8.4 pH values. Low affinity bodies are expected to be significantly enhanced by low salt concentration in the buffer (Reverberi and Reverberi, 2007). Apart from varying these factors and observing their effect, the length of incubation could be changed prior to negative staining.

3.5.4 DISCUSSION

Heparan sulphate (HS) is a highly sulphated glycosaminoglycan present on the walls of epithelial cells. Heparin is a structurally similar polysaccharide found within mast cells and shown to differ in composition with some mammalian forms of HS. D-glucosamine residues are N-acetylated in HS and N-sulphated in heparin. All the above make heparin more sulphated and more charged than HS in addition to being smaller with the maximum molecular weight of 20 kDa. While the above would make it seem like HS and heparin are different, polysaccharides isolated from some organisms appear to be hybrid with heparin being a tissue-specific form of HS. Heparin is thus

usually employed as an experimental proxy for HS, as its preparation process leaves it a more homogenous mixture (Lima et al., 2015).

Negatively charged GAG bind basic lysine, arginine and histidine residues in stretches containing uncharged non-polar asparagine and glutamine (Sarkar and Desai, 2015). From the study of Freitag et al, 2017, $FliD_{cj}$ binds both heparin GAG *in vitro* and HS on the surface of epithelial cells in a dose dependent fashion. They also identified the region of $FliD_{cj}$ in which the binding site is located as located between amino acids 209 and 418 (Freitag et al., 2017). This region includes multiple potential sites in the D3, the N-terminal part of D2 domain and the majority of the D1 helical bundle. This is a wide possible area for both heparin and HS binding, but when considering the specific sequence of basic and non-polar residues, the head domain loops located on top of the pentamer, the loop coming off the hairpin in the D1 domain and a few small helices on the out-facing part of the D1 domain are potential binding sites.

While the cryo-EM attempt at solving the structure of $FliD_{cj}$ to heparin did not work as planned, the ITC data still suggests that there is weak and potentially non-specific binding happening between them. This could be the reason that the EM grid did not have many particles bound to heparin and thus was averaged out during classification and refinement. Also, the non-specific nature of heparin polymer binding to $FliD_{cj}$ could be too heterogenous to be able to solve with a single data collection and would require vastly more data to be able to separate out the different bound conformations. The GAG binding ability of $FliD_{cj}$ might be species specific, as the potential binding sites are located within a *C. jejuni* specific loop in D3 and D1 helical bundle near the D4 domain that was observed in the cryo-EM structure in this work. As the D4 domain was shown to be capable of adapting multiple conformations due to the flexible linker connecting it to the D1 bundle, the HS/Heparin polymer might interact with it as well when binding the basic residue rich stretch in the adjacent domain. $FliD_{cj}$noD4 construct generated in section 3.2.4 could be tested using ITC under the same parameters and any change in binding affinity observed. If D4 is interacting with the heparin binding site, then co-crystallization of D1-D4 isolated domains with heparin could yield structural information. As the binding sites could be spread across a wide range of protein sequence, the constructs generated in this work,

namely the D2-D3 head domains and D1-D4, could be used to narrow the HS/Heparin binding site further.

The quality of $FliD_{cj}$ as an antibody target, while tested using a variety of methods and shown to be effective, was not observed to a high degree using our negative stain data (Ghasemi et al., 2018; Perruzza et al., 2020). This could be due to low affinity binding of the antibody to the purified recombinant protein, causing loss of density in the averaged classes. Another reason could be that the antibodies bind to $FliD_{cj}$ specific D4 domain, which is flexible and often averaged out during 2D and 3D classifications. There could also potentially be preferred orientation in the negative stain and the antibody density is not visible in the views observed. Overall, extensive optimisation of antibody-protein complex formation on a negative stain grid need to be attempted before resorting to a co-purification experiment. Additional biophysical techniques such as ITC or SPR could be useful to ascertain the binding affinity as well.

4. DISCUSSION

This work presents the first near-atomic resolution structure of the flagellum cap complex showing the previously unresolved D0 terminal domain, which has been proposed to interact with the flagellin monomers in the growing filament. In addition to *C. jejuni* FliD, flagellar capping protein from a variety of Gram-negative bacterial species were purified and $FliD_{cj}$, $FliD_{pa}$ and $FliD_{sm}$ were observed to form 5-fold complexes. The role of $FliD_{cj}$ in building the flagellar filament was characterized by generating mutants based on the new terminal domain interfaces observed in the structure and determining their effect on motility. While many of the mutants significantly reduced motility, the formation and length of the observed filaments was not significantly reduced. Thus, the decrease in motility could occur due to difference in structural integrity of the filament. The hypothesis is that instead of preventing filament formation outright, the mutations cause an unstable interaction between FliD and the filament causing the flagellins to be packed less tightly or at a wrong angle. This would affect the integrity of the flagellum and interfere with the thrust exerted by the motor causing a defect in motility.

While previous work indicated that $FliD_{cj}$ binds heparin, the ITC data in this work confirms it but shows that the binding is weak and potentially non-specific. This observation might be because the epithelial cell GAG heparan sulphate forms much larger chains than the experimentally accepted alternative heparin, thus conforming around the D1-D2-D3 potential binding sites more efficiently. ITC is a technique where both the ligand and the substrate are suspended in solution, while GAGs are usually bound to a scaffold on the surface of the epithelial cells. This mechanistic difference might also affect the efficiency of binding between $FliD_{cj}$ and HS. A way to test this would be to utilize SPR, where HS/heparin could be immobilized on a streptavidin chip and protein flowed over it, or an affinity column with HS/heparin covalently linked to the beads and $FliD_{cj}$ eluted with different salt concentrations. Attempts at determining the structure of FliD bound to heparin did not reveal the presence of the ligand, potentially due to the weak non-specific binding creating too much heterogeneity in

the cryo-EM sample, causing difficulty with isolating a state while bound to heparin. Another reason could be the fact that heparin concentration had to be kept limited as the polymer was too dense in the cryo-EM grid, thus creating issues with obtaining optimal vitreous ice conditions. $FliD_{cj}$ specific antibody binding, observed by negative stain EM, showed only one out of six samples potentially binding, which could be due to preferred orientation and issues with sample preparation.

To understand the interaction of $FliD_{cj}$ and the filament further, the structure of wild type flagellar filament of *C. jejuni* was studied and an 11-protofilament conformation was confirmed, contrary to previously reported 7-protofilament structure (Galkin et al., 2008). While the data did not yield a high-resolution map, the structure obtained was similar to other filament structures and was later confirmed by an independent study of a straight mutant (Kreutzberger et al., 2020). After solving the structure of $FliD_{cj}$, showing that other FliD complexes from bacterial species can form pentamers and having evidence that the *C. jejuni* filament is made up of 11 protofilament, a model of filament elongation was proposed based on these results.

4.1 Novel $FliD_{cj}$ Structure

While this work focuses equally on the structure and function of $FliD_{cj}$, the structure itself is where the novelty lies. There is plenty of information available on the highly structurally conserved D2-D3 head domains due to extensive X-ray crystallography studies, but the $FliD_{cj}$ structure in this work is the first to show the structure and conformation of the leg domains. While the D1 helical bundle and D2-D3 domains are similar in fold to that of the other crystal structures of FliD, their packing in relation to each other is different, potentially to accommodate the *C. jejuni* specific D4 domain. The newly characterized D0 terminal domain forms a coiled coil as predicted in the C-terminus, but a short N-terminal residue stretch including a β-strand bends away from the C-terminus and interacts with the adjacent subunits in the pentamer, which has not been observed previously. This allows the C-terminus to interact with another C-terminal coiled coil from the bottom pentamer, which is a non-

physiological interaction that is present in the stabilized decamer. The newly characterized conformation of the N-terminus observed in our decamer structure is an interesting observation of the interconnectivity of the terminal domains within the pentamer. It is also to be noted that flagellar axial proteins such as FlgG, FlgE, FlgL, FlgK and flagellin also have a similar coiled coil terminal domain conserved across bacterial species, structures of which are not readily available due to the terminal domains hindering crystallization process and having to be truncated. However, the cryo-EM structures of these proteins do have the N and C-terminal coiled coils observed in the $FliD_{cj}$ structure. As the axial proteins interact with each other to form helical assemblies this common domain architecture may play an important role in this function and building of the flagellum. In T3SS injectisome the SctF needle forming protein as well as the hydrophilic translocator cap SctA also have similar coiled coil terminal domain, indicating that the mechanism of elongation of ordered helical components may not only be conserved across bacterial species but also across systems (Blocker et al., 2008; Hu et al., 2019).

The $FliD_{cj}$ terminal domain was also analysed in this work for its function in bacterial motility. The newly determined structure of the D0 terminal domain contained two interfaces of interest: one between N-terminal stretch and the C-terminal helix of the adjacent subunits in the pentamer and the one between the C-terminal helices in the top and bottom pentamers. The first interface, studied through point mutations in highly conserved residues from hydrophobic to polar, appeared to affect motility upon the loss of hydrophobicity. This led to a hypothesis that the interface is important in maintaining $FliD_{cj}$ pentamer stability rather than filament formation. A potential *C. jejuni* specific effect of these mutations is the differential interaction of $FliD_{cj}$ with the two flagellins: FlaB minor flagellin forming a 0.5 µm section of the filament close to the HBB and FlaA major flagellin forming a 3.0 µm section of the distal filament. The expression of the latter is controlled by a different mechanism than in enteric γ-proteobacteria. FlaG anti-sigma factor represses FlaA expression through inhibiting FliA promoter and limits the flagella length. Upon mutation in the N-terminal D0 section of $FliD_{cj}$ from hydrophobic to polar interface, the interactions between FliD and FlaB might be perturbed resulting in a normally curved section at the

base of the filament to alter the packing of flagellins and appear straighter. The FlaA appears to be unaffected, resulting in a distal filament of the wild type length.

The second interface, studied under the same conditions, was hypothesized to mimic the physiological interaction of the FliD C-terminus with the flagellar filament lumen. While the actual positioning of the two interacting C-termini would not be exactly the same in the growing filament, the general properties of the interface showed that mutations from polar to hydrophobic also reduced motility and caused brittleness in flagellar filaments. This event also has a possible *C. jejuni* specific explanation, where mutation of the C-terminal interface affects FlaA interaction with $FliD_{cj}$ and causes "brittleness" in the boundary between the two flagellins, resulting in easier shearing of the filament. This leaves a short stub of FlaB in sheared filaments, but an overall native looking filament in the filaments still attached.

4.2 Oligomeric state of Filament and Cap

This work focused on the FliD from *C. jejuni,* but it also demonstrated that FliD from other species can form pentamers of similar dimensions contrary to crystallography-derived structures. Even when crystallography yielded high resolution structures for FliD from other bacterial species, they were all lacking the D0 terminal domains observed in the structure solved in this work. Thus, their oligomerization within the crystal could be an artefact of crystal packing and not the native complex. While there is evidence that $FliD_{pa}$ can form hexamers in solution, this work shows that it is also capable of forming a pentamer of similar dimensions to that of previously observed $FliD_{se}$ (Postel et al., 2016). However, the only other cryo-EM structure of FliD in *S. enterica* showed a decamer oligomerization in solution similar to that of $FliD_{cj}$ reported in this work, indicating that the pentamer may likely be the physiological state of FliD. This preliminary evidence shows alternative oligomerisation states from crystallographic data, suggesting different stoichiometries for the cap complex depending on the species and contributes to proposing a mechanism for flagellar elongation viable across different bacterial species. Solving the cryo-EM structure of

FliD from other bacterial species will help test this hypothesis. The intriguing part is that in both cross-linking data as well as this work $FliD_{pa}$ is capable of forming a stable 6-fold oligomer in solution. It is yet unknown if this is a functional state or it serves an alternative purpose.

Due to the different oligomerization states of FliD structures across bacterial species and the instance of *C. jejuni* flagellar filament having 7 protofilaments instead of 11, a model was proposed where depending on the bacterial species, the oligomerization of FliD and the number of protofilaments would change while maintaining a certain ratio (Cho et al., 2017). This work, in addition to pre-existing models of filament elongation by FliD in *S. enterica*, strongly argues for a universal 11 protofilament to 5 FliD subunits stoichiometry model unaffected by the bacterial species. In addition, a new structure of *C. jejuni* filament was published showing that it also had 11 protofilaments and supporting the pentamer $FliD_{cj}$ physiological state (Kreutzberger et al., 2020).

The initial localization of FliD to the top of the growing flagellum is facilitated by its chaperone FliT binding to the flagellar ATPase and promoting secretion of FliD monomers through the 2 µm lumen of the axial components that were already assembled (Minamino et al., 2021). It is known that FliD interacts with FlgL, but how FlgK and FlgL hook-filament junction proteins assemble on top of the flagellar hook has not been visualized using structural biology and so far has only been modelled (Imada, 2018; Yonekura et al., 2000). So, in turn how FliD assembles on top of FlgL into a pentamer cap is not yet clear. However, the hook filament at 55 nm long most likely is not planar and flat for FlgK to assemble into a fully symmetrical 11-subunit ring, requiring some asymmetry. Same lack of following the helical pitch at starting and growing flagellar filament ends is likely due to the helical symmetry observed in EM structures. This is indicative that the initial flagellins would have to bind a base with a similar asymmetry to grow from FlgL. How do the hook-filament junction proteins facilitate the visually seamless transition from the hook to the filament in flagella is a curious assembly question, yet to be determined. One hypothesis would be that starting from the final FlgE protein at the end of the hook, the rest of the axial structures sit with one subunit not following the helical pitch to not create gaps during

transitions, causing FliD bound to FlgL to also adapt a kinked conformation akin to that observed in FlgD bound at the rod tip (Johnson et al., 2021).

The flagellar filament itself is highly diverse across bacterial species. While the high-resolution EM data suggests an 11-protofilament conformation, many of the filaments were obtained through mutations leading to straightening of the filament and consisted of a single flagellin type. In nature, bacterial flagella are often composed of a mixture of different types of flagellin depending on the bacterial species and physiological requirements for survival. For example, in *S. enterica* filaments can form from FliC or FljB, while *C. crescentus* has FljK as one of six flagellins. However, only one flagellin can be expressed at a time, with the switch between them being called phase variation (Montemayor et al., 2021). In *Clostridium haemolyticum* flagella are made up of main structural flagellin and FliA(H) flagellinolysin which acts as a zinc-metallopeptidase (Eckhard et al., 2017). In ε-proteobacteria, like *C. jejuni* and *H. pylori*, FlaA and FlaB are not used together in phase variation but are responsible for building different sections of the flagellar filament. FlaB is a highly curved minor flagellin at the base and FlaA is a straight major flagellin responsible for the majority of the distal filament. In *Spirochetes,* the FlaB flagellin appears to be responsible for general structural integrity of the filament and FlaA localizes to the base and affects curvature. There are additional genus specific FcpA and FcpB flagellins responsible for supercoiling, virulence and *Spirochete* specific functions (Gibson et al., 2020).

Apart from having different flagellins incorporated into the filament, the protein sequence itself is highly variable within the same species. While the terminal regions are conserved, the D2-D3 domains are hypervariable with added insertions like that observed in the FlaA and FlaB flagellins of *C. jejuni* (see figure 3.1.2). The hypervariable regions contain internal duplication which results in stepwise phylogenic size increase in α, γ, δ and ε-proteobacteria. Giant flagellins, such as the 120 kDa *Bermanella marisrubri* flagellin, are 300 Å in diameter, 25% thicker than that of *S. enterica* (230 Å) due to the large non-homologous insertion in the D2-D3 domain (Thomson et al., 2018). To increase variability, many bacteria including *C. jejuni*, have flagella that are glycosylated. Glycosylation of flagella involves O-link attachment of pseudaminic or legionaminic acids to Serine and Threonine residues on the surface of

the flagellum (see section 1.2). It prevents exposure of a sensitive peptide region which triggers an immune response in majority of bacterial species, except for *C. jejuni* where it has no effect as the evasion of TLR5 is an innate feature of the recognition site (Kreutzberger et al., 2020; Rossez et al., 2015; Young et al., 2007). Some bacteria express multiple flagella like *Vibrio parahaemolyticus, P. aeruginosa* and *Bradyrhizobium japonicum,* with thick polar flagellum for liquid swimming and thin smaller filaments for swarming on surfaces (Atsumi et al., 1992; Jarrell and McBride, 2008; Kanbe et al., 2007). Others have a unique sheathed flagellar filament as an extension of the outer membrane, like *Vibrio alginolyticus* (Zhu et al., 2017).

FliD and other flagellar proteins across bacterial species were also observed to contain such domain duplication as in the larger flagellins in *C. jejuni* and *H. pylori.* It would be intriguing to observe the flagellar filament formation in species with significant size differences in flagellins and determine if the mechanism of elongation is the same.

4.3 Putative Mechanism of Filament Assembly

From the core findings in this work, a model of filament elongation was proposed based on the previous models in *S. enterica.* Due to symmetry mismatch between pentamer FliD and the 11-fold flagellar filament, a gap between the D0-D1 domains of FliD is not obstructed on one side. The unfolded flagellin occupies and folds within that gap and to accommodate the next subunit, the whole FliD assembly rotates by ~65° and displaces to a higher position with the cavity at the next flagellin insertion site. While multiple previous models suggested the leg domain of FliD being dynamic due to the flexible linker between it and the head domains, resulting in an elongation model requiring displacement of the legs to the sides while building the filament, the structure of FliD in this work shows rigid closely interacting leg domains connected through the N-terminal stretch (Maki-Yonekura et al., 2003; Yonekura et al., 2000). The rigid-legs model, based to the symmetry mismatch of the FliD pentamer and flagellar 11 protofilament, would have enough space to accommodate the flagellin molecule

upon growth and displace the cap like a cog freeing up the next position, which has been proposed before (Song et al., 2017).

The existing flexible leg model of filament elongation is based on the first low resolution cryo-EM structure of the FliD cap in solution and FliD localized at the tip of the filament in *S. enterica* (Yonekura et al., 2000). While the resolution was too low to distinguish the leg domains and any of their conformations, the pentamer head at the tip of the filament was visible. Due to the symmetry mismatch, this model suggests that the legs of each monomer of FliD must have different conformations from one another to fit within the 11-fold filament, which is possible because of the flexible linker between the D1 and D2 domain as well as the flexible terminal regions. The main source of this is the low-resolution single particle map of cap-filament interaction, which contained "gaps" visible at different leg domain locations. The gaps could be the evidence of leg domain displacement but could also be an artefact of the high contour value and low resolution cryo-EM map (Yonekura et al., 2000). This relies on the concept that the head domains are planar and symmetrical with the leg domains moving horizontally but not up and down. Due to the flagellar filament having an empty slot for the next flagellin to slot in below the level of the cap legs, that position is not interacted with as tightly by FliD. Flagellin assembly proceeds in a CCW direction ever 65.5° and the cap rotates 6.5° CW (Yonekura et al., 2000)..

Another model was based on a crystal structure of $FliD_{se}$ distorted pentamer of D2-D3 domains with unequal length between the monomers causing one side to be wider than the rest (Song et al., 2017). Like the previously discussed model, this mechanism was based on the symmetry mismatch between the filament and the cap, but instead of flexible legs the distorted pentamer FliD would create a large enough gap to fit the flagellin monomers through. However, no other crystal structures of truncated head domains had similar distortion. The EM structure of $FliD_{cj}$ head domains was based on a map reconstructed with C5 symmetry and thus yielded a symmetric pentamer. However, further classification with C1 symmetry, using a bigger dataset might be necessary to potentially isolate any asymmetric states present.

A recent study of the flagellar basal body showed FlgD hook tip protein in different conformations while bound to the FlgG rod of *S. enterica*. As pentameric FlgD, with similar D0 coiled coil domains to FliD, builds an 11-protofilament hook, the mechanism could be similar to that of FliD building the flagellar filament. This mechanism relies on the helical regions blocking potential filament sites and moving up and down in a piston-like movement to free up the space and allow the filament to grow (Johnson et al., 2021). This is possible due to the flexible linker between the legs and the head domains also observed in FliD structures. Thus, this is yet another potential mechanism of filament elongation.

The similarities between the T3SS injectisome and flagellum may also elucidate the mechanism of filament elongation, considering that many axial proteins have a terminal coiled coil conformation similar to that of flagellin and FliD. However, the needle complex protein SctF is much smaller than flagellin and amenable to more conformational changes. The T3SS injectisome needle tip complex also has a significantly different conformation while binding to the filament from that observed in FliD as illustrated in Figure 4.1.1 (Lunelli et al., 2011). While the SctA needle tip does have similar coiled coil terminal regions to FliD, the rest of the domains have a different conformation. Due to the structural evidence of SctA needle tip binding SctF needle forming protein, a model was proposed where 5 monomers of SctA would bind to the sides of the needle as illustrated in Figure 4.1.1 (c) (Lunelli et al., 2011). $FliD_{cj}$ does not bind the filament as a monomer but as an oligomer. It also attaches at the tip, slotting in the leg domains in between the flagellin monomers and not at the side of the flagellar filament.

C. jejuni flagellar T3SS exports toxins as well as the flagellar structural components. The positioning of $FliD_{cj}$ on top of the filament during elongation needs to account for secretion as well. This is where the Yonekura model fits well, as the FliD leg domains being asymmetrical in binding to the growing filament allows for one leg to be shifted sideways to open a space for flagellin to polymerise. As T3SS is highly specific, when the toxin approaches the FlhAB gate the flagellin secretion is probably paused and the toxin is threaded through the lumen, most likely followed by another flagellin monomer. Within the lumen, injection-diffusion mechanism transports

partially helical flagellins and the toxin through to FliD, where after polymerising the last flagellin and shifting one leg to a side to allow for a space to open up, the toxin might pass through the gap between flagellins and thus be exported. For this, the model proposed in this work might not seem ideal, as with the legs locked in a stable but rigid conformation the only gap in the interface between flagellin and FliD is the one required for the formation of next flagellin monomer. However, *C. jejuni* FliD also possesses D4 domain attached to the D1 helical bundle through a flexible loop that would cause $FliD_{cj}$ to sit higher in the growing filament than other homologous flagellar caps and might move D4 around to free a space for the toxin to be exported and move back to prevent flagellins to escape. $FliD_{cj}$ is unlikely to export effectors through the lumen in between the head domains D2-D3, as that distance is roughly 12 Å. The width of the lumen at the bottom of FliD pentamer in-between the N-terminal stretches of sequence is 20 Å, same as the rod, hook and flagellar filament lumens, which allows for a secretion of an unfolded or partially helically folded protein into the middle of the $FliD_{cj}$ pentamer with limited space to go. This is when the narrow space left by the 5-11 asymmetry allows the toxin to escape the filament and be secreted or flagellin to fold, shifting the cap upwards. But what is to stop flagellin from escaping as well? This could be due to the N- and C-terminal hydrophobic interfaces interacting with flagellin and folding it, while allowing the effector to pass through. However, all of the above is speculative until a clearer elongation mechanism is determined.

Further work is required to determine the exact interaction *in vivo*. One way to do so would be to undertake an extensive cryo-EM study of wild type flagellar filament tips to observe FliD at different stages of filament building. As it is yet unknown whether FliD remains in a rigid state or has dynamic leg movements, a large volume of particles at different orientations is required to be able to resolve these conformations. Using *C. jejuni* for this study would be particularly challenging due to it only having two flagella at cell poles and thus a limit of two potential tips per cell. A peritrichous strain such as *S. enterica* would facilitate this process. As the hypothesis is that the mechanism is conserved across bacterial species, potentially other peritrichous bacteria could be observed as well.

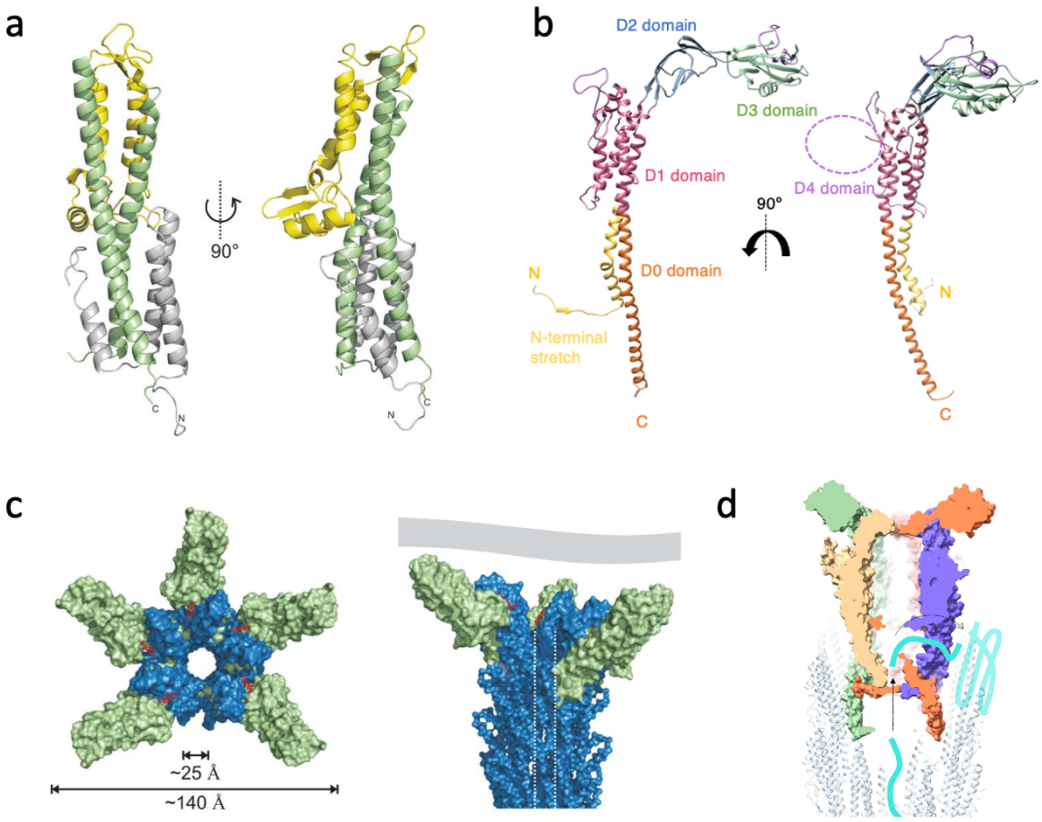


Figure 4.1.1: Comparison between the T3SS injectisome needle tip complex SipD from *S. enterica* (standard nomenclature SctA) and FliD from *C. jejuni*. (a) Crystal structure of SipD monomer. (b) Cryo-EM structure of $FliD_{cj}$ monomer. (c) Model of the open state of the needle tip with SipD in green and needle in blue (PrgI/SctF). Left top view, right side view. (d) Model of the FliD binding flagellar filament in a side cross section view. FliD is multicoloured, flagellin grey. Adapted from (Lunelli et al., 2011).

4.4 Concluding Remarks

To test the hypothesis that FliD physiological oligomeric state is that of a pentamer regardless of the bacterial species, the cryo-EM structures of FliD from other flagellated bacteria need to be determined with the D0 terminal domains intact. As *C. jejuni* and *H. pylori* appear to have extra domains localized in D1 in $FliD_{cj}$ and the head domains in $FliD_{hp}$, it would be interesting to observe how these additional structures interact upon filament elongation and if they influence the mechanism itself apart from serving a species-specific function. So far, the similarity between 5-11 FliD-

protofilament *S. enterica* and *C. jejuni* models of elongation lead to a hypothesis that the elongation mechanism is conserved across bacterial species. But is it actually identical, or are there species-specific adaptations resulting in structural differences, as observed in the motors of *Vibrio*, *Campylobacter* and *Borrelia*? The additional domains of $FliD_{cj}$ and $FliD_{hp}$ must have some sort of role in elongation and secretion, and it would be interesting to investigate how these domains affect the agreed-upon elongation mechanism. While additional domains are common in flagellar axial proteins due to internal duplication of the head domains while the terminal domains are conserved, unlike *H. pylori* D4, *C. jejuni* D4 is in an odd position between the head and terminal domains. What significance does this position have? Is its purpose *C. jejuni* specific? Does $FliD_{cj}$ interact with major FlaA and minor FlaB flagellins differently, as they form conformationally distinct sections of the flagellar filament? Is this the case in all bacteria utilizing multiple flagellins to build the filament such as *H. pylori* and *B. burgdorferi*? There are many questions left about the mechanism of filament elongation by the cap and with the recent developments in cryo-EM techniques some of them might be answered. It would be beneficial to analyse the way FliD fits on top of a growing filament, which would require a large dataset of native filaments, with the intention to classify the heterogeneity to isolate the different states of FliD in relation to the filament. This could be done by SPA, but the potential problem of preferred orientation might become an issue.

Another question still largely unanswered about flagellar filament elongation: how the first few flagellins form? It has been widely accepted that it is impossible to reconstitute flagellar filaments with native parameters from recombinantly purified flagellin, so the first few monomers must require a scaffold. This brings up another interesting point in flagellar assembly: while the majority of the structural components in the flagellum have been characterized in their oligomeric state, FlgK and FlgL hook-filament junction, the C-ring and the flagellar ATPase oligomer structures have not been determined. As FlgK and FlgL form on top of the hook and connect it to the filament, how does FlgK and FlgL interact together, with the hook (the end of which has an FlgE monomer not on the same plane as the rest), with FliD and the first few flagellins? Determining FlgK and FlgL complex structures on their own and together in

vitro would be the first point to start with, followed by observation of their formation in a native growing filament. Once these proteins structures have been solved, this would complete the structural understanding of flagellar axial region.

In *C. jejuni,* FliD was found to interact with heparin using the broad interface of D2-D3 and D1 domains, which remains to be shown via cryo-EM (Freitag et al., 2017). An alternative hypothesis involves *C. jejuni* specific D4 domain playing a role in such binding, and using the constructs generated in this work some further biochemical and structural studies could be done to investigate this binding further. Heparin is a laboratory alternative to HS, which might require an interaction analysis using biophysical methods relying on scaffolding the HS prior to introducing FliD, such as SPR and affinity chromatography. The different domain constructs might assist in narrowing down the HS binding site so their purification and potential role of D4 domain binding to heparin needs to be investigated.

To delve further into $FliD_{cj}$ specific function, the antibody binding assay could be optimised via adjusting the stoichiometry of the components and potential co-purification to find the specific binding site. After HS binding of $FliD_{cj}$ is characterised, a synthetic peptide could be designed, which could mimic this interaction, binding with a higher affinity than HS, and acting as an inhibitor of *C. jejuni* attachment to epithelial cells. This in combination with anti-$FliD_{cj}$ antibodies could be used to design a robust method to combat virulence of *C. jejuni* flagellum. As *C. jejuni* relies heavily on its flagella both for motility and pathogenesis, the newly solved structure of $FliD_{cj}$ antigen could potentially lead to a better understanding of how *C. jejuni* infects the intestinal epithelial layer and assist in developing a targeted therapeutic treatment against adherence and infection. The interaction of FliD and the filament discussed in this work could also potentially lead to an improved insight into the bacterial flagellum and its role in infection across different species.

APPENDICES

APPENDIX 1

Protein	Codon optimised DNA sequence from Bio Basic
FliDcj	ATGGCCTTCGGTAGCCTGAGCAGCCTGGGCTTCGGTAGCGGCGTTCTGACCCAGG ATACCATCGATAAACTGAAAGAAGCGGAACAGAAAGCTCGCATCGACCCGTATAC CAAGAAAATCGAAGAAAACACCACCAAACAGAAAGATCTGACCGAAATTAAAACC AAACTGCTGTCTTTCCAGACCGCGGTTTCCTCCCTGGCGGACGCGACCGTGTTCGC GAAACGTAAAGTGGTTGGCAGCATCAGCGATAACCCGCCGGCAAGCCTGACCGTT AACTCTGGTGTGGCGCTGCAGTCCATGAACATTAACGTGACCCAGCTGGCTCAGA AAGATGTGTATCAGTCCAAAGGTCTGGCGAACGACAGCGGTTTTGTTAACGCTAA CCTGACCGGCACCACCGATCTGACCTTCTTCTCTAACGGCAAAGAATACACCGTGA CCGTGGACAAAAACACCACCTACCGTGATCTGGCGGATAAAATCAACGAAGCTAG CGGCGGTGAAATCGTTGCGAAAATCGTAAACACGGGCGAAAAAGGCACCCCGTA CCGTCTGACCCTGACCTCCAAAGAAACCGGCGAAGATAGCGCGATCTCCTTCTACG CGGGTAAAAAAGACGCGCAGGGTCAGTACCAGAGCGACCCGGAGGCGGAAAAC ATCTTCTCCAACCTGGGCTGGGAACTGGACAAAACCACCCAGACCATCGACCCGG CAAAAGATAAGAAAGGCTACGGCATTAAAGATGCGTCTCTGCACATCCAGACCGC GCAGAACGCGGAATTTACCCTGGACGGTATCAAAATGTTCCGTTCCTCCAACACCG TGACCGACCTGGGTGTTGGTATGACCCTGACCCTGAACAAAACCGGCGAAATCAA CTTCGATGTTCAGCAGGACTTCGAAGGTGTTACCAAAGCCATGCAGGACCTGGTA GATGCTTATAACGATCTGGTTACCAACCTGAACGCGGCGACCGATTACAACAGCG AAACCGGCACCAAAGGCACCCTGCAGGGCATTTCTGAAGTTAACAGCATCCGTAG CTCTATTCTGGCGGATTTGTTCGACTCTCAGGTGGTTGACGGTACCACCGAAGATG CTAACGGTAACAAAGTTAACACCAAAGTTATGCTGTCCATGCAGGACTTCGGCTTA TCTCTGAACGATGCGGGCACCCTGAGCTTCGACTCTTCCAAATTCGAACAGAAAGT TAAAGAAGATCCGGATTCTACCGAATCCTTCTTTTCTAACATCACTAAATACGAAG ATATTAACCACACCGGTGAAGTGATCAAACAGGGCAGCCTGAACCAGTACCTGGA TAGCAGCGGTACCGGCAACAAAGGTCTGGATTTCAAACCAGGCGACTTCACCATC GTTTTCAACAACCAGACCTATGACCTGTCCAAAAATAGCGACGGCACCAACTTCAA GCTGACCGGTAAAACCGAAGAAGAATTGCTGCAAAACCTGGCAAACCACATCAAC TCAAAAGGTATCGAAGGTCTGAAAGTTAAAGTGGAATCGTACGATCAGAACGGC GTGAAAGGTTTCAAGCTGAACTTCTCTGGTGATGGTAGCTCTGATTTCTCTATTAA AGGCAACGCGACCATCTTGCAGGAACTGGGCCTGTCTGATGTTAAtATCACCTCCA AACCGATCGAAGGCAAAGGCATCTTCTCCAAACTGAAAGCTACCCTGCAGGAAAT GACCGGTAAAGACGGTAGCATCACCAAATATGATGAAAGCCTGACCAATGACATT AAGTCACTGAACACTAGTAAAGATAGCACGCAGGCAATGATTGACACCCGCTATG ATACCATGGCTAACCAATGGCTGCAGTACGAATCGATCCTGAATAAACTGAACCA GCAGCTGAATACCGTGACTAACATGATTAACGCAGCTAATAATAGTAATAACTAA
FliCcj	ATGGGCTTCCGTATTAACACCAACGTTGCGGCACTGAACGCTAAAGCAAACGCTG ACCTGAACTCTAAATCTCTGGATGCTAGCCTGAGCCGTCTGTCTAGCGGCCTGCGC ATCAACTCCGCAGCTGATGATGCATCCGGTATGGCTATTGCAGATAGCCTGCGTTC ACAGGCTAATACCCTGGGTCAAGCTATTTCTAACGGCAACGACGCTCTGGGTATCC TGCAGACCGCTGATAAAGCGATGGATGAACAACTGAAAATTCTGGATACTATTAA AACCAAAGCAACCCAGGCCGCGCAGGATGGCCAATCCCTGAAAACTCGTACCATG CTGCAGGCGGATATTAACCGTCTGATGGAAGAACTGGACAACATCGCAAACACCA CCTCCTTTAACGGGAAACAACTGCTGAGCGGTAATTTCATTAACCAAGAATTTCAG ATTGGTGCCTCTTCTAACCAGACCGTGAAAGCAACCATCGGTGCAACCCAGTCCAG CAAAATCGGCCTGACCCGCTTCGAGACCGGCGGCCGTATTTCTACCAGCGGTGAG

	GTTCAGTTCACCCTGAAAAACTATAACGGTATCGATGATTTCCAGTTCCAGAAAGT TGTTATTAGCACCTCCGTTGGTACCGGCCTGGGTGCACTGGCGGATGAAATCAAC AAAAACGCAGATAAAACCGGCGTGCGTGCGACCTTCACCGTTGAAACCCGTGGCA TCGCTGCGGTTCGTGCGGGCGCGACCTCTGACACCTTCGCGATCAACGGTGTTAA AATCGGTAAAGTTGATTACAAAGATGGCGACGCTAACGGTGCGCTGGTTGCAGCA ATCAACTCCGTGAAAGATACCACCGGCGTTGAAGCCTCCATCGATGCGAACGGTC AGCTGCTGCTGACCTCTCGTGAAGGCCGCGGCATAAAAATCGATGGTAACATCGG CGGCGGTGCATTCATCAACGCGGATATGAAAGAAAACTACGGCCGTCTGTCTCTG GTGAAAAACGACGGTAAAGACATTCTGATCAGCGGTTCCAACCTGAGCAGCGCG GGTTTCGGTGCTACTCAGTTCATCTCTCAGGCGAGCGTTTCCCTCCGTGAATCTAA AGGCCAGATCGACGCTAACATCGCGGACGCAATGGGTTTCGGCTCTGCGAACAAA GGTGTTGTTCTGGGTGGCTACTCCAGCGTGTCTGCGTATATGAGCTCCGCAGGTTC TGGTTTCTCTTCGGGTTCTGGTTACTCTGTGGGCAGCGGCAAAAACTACAGCACCG GCTTCGCTAACGCGATCGCAATCTCCGCGGCGAGCCAGCTGTCGACCGTTTATAAC GTTAGCGCTGGTAGCGGCTTCAGCTCTGGCAGCACCCTGAGCCAGTTCGCGACCA TGAAAACCACCGCATTCGGCGTTAAAGATGAAACCGCGGGTGTGACCACCCTGAA AGGCGCGATGGCGGTTATGGATATAGCGGAAACCGCGATCACCAACCTGGATCA GATCCGCGCAGACATCGGTTCTGTTCAGAACCAGGTTACCAGCACCATCAACAAC ATCACCGTGACCCAGGTTAACGTTAAAGCGGCGGAATCTCAGATCCGTGATGTTG ACTTCGCGGCTGAAAGCGCTAACTATTCTAAAGCTAACATCCTGGCTCAGAGCGG CAGCTATGCGATGGCCCAGGCGAACAGCGTTCAGCAGAACGTCCTGCGTCTGCTG CAGTAA
FliDpa	ATGGCGGGTATCTCTATCGGCGTTGGTTCTACCGACTACACCGACCTGGTTAACAA AATGGTTAACCTGGAAGGTGCGGCTAAAACCAACCAGCTGGCTACCCTGGAAAAA ACCACTACCACCCGTCTGACCGCTCTGGGTCAGTTCAAATCTGCGATTTCTGCTTTC CAGACCGCTCTGACCGCGCTGAACTCTAACGCGGTGTTCATGGCTCGTACCGCTAA ATCTTCCAACGAAGATATCCTGAAAGCGTCTGCTACCCAGTCTGCAGTTGCTGGTA CTTACCAGATCCAGGTTAACTCTCTGGCGACCTCTTCTAAAATCGCGCTGCAGGCG ATCGCTGATCCGGCGAACGCTAAATTCAACTCCGGTACTCTGAACATCTCCGTTGG CGACACCAAACTGCCGGCGATTACTGTTGATTCTTCTAACAACACTCTGGCGGGTA TGCGTGATGCAATCAACCAGGCGGGTAAAGAAGCTGGCGTTTCTGCTACCATCAT CACTGACAACTCCGGCTCTCGTCTGGTTCTGTCTTCTACCAAAACTGGTGATGGTA AAGACATTAAAGTTGAAGTCTCTGATGACGGTTCTGGTGGTAACACCTCTCTGTCC CAGCTGGCTTTCGACCCGGCAACCGCGCCGAAACTGTCCGATGGTGCAGCTGCAG GCTATGTTACTAAAGCTGCTAACGGTGAAATCACTGTTGATGGTCTGAAACGTTCT ATCGCTTCCAACAGCGTATCTGACGTTATTGACGGTGTTTCCTTCGATGTGAAAGC AGTTACCGAAGCAGGTAAACCAATCACTCTGACCGTTTCTCGTGACGACGCAGGT GTTAAAGATAACGTTAAAAAATTCGTTGAAGCTTATAACACCCTGACCAAATTCAT CAACGAACAGACTGTAGTTACTAAAGTTGGTGAAGATAAAAACCCGGTTACTGGT GCACTGCTGGGTGACGCATCTGTGCGTGCTCTGGTTAACACCATGCGTAGCGAAC TGATCGCGTCTAACGAAAACGGTTCTGTTCGTAACCTGGCGGCTCTGGGTATCACT ACCACTAAAGATGGTACCCTGGAAATCGATGAGAAAAAACTGGACAAAGCTATTT CTGCTGATTTCGAAGGTGTTGCTTCCTACTTCACCGGTGACACCGGTCTGGCGAAA CGCCTGGGTGACAAAATGAAACCGTACACCGACGCTCAGGGCATTCTGGATCAGC GTACCACCACTCTGCAGAAAACCCTGTCTAACGTTGATACCCAGAAAGCGGACCT GGCTAAACGTCTGGCGGCACTGCAGGAAAAACTGACCACCCAGTTCAACCTGCTG TCTGCGATGCAGGACGAAATGACCAAACGCCAGAAATCTATCACCGATAACCTGG CGTCTCTGCCGTACGGCTCTGGTAAGAAAACCTAA
FliDsm	ATGGCGACCATCTCTTCTCTGGGCCTGGGTTCTGGCCTGGACCTGAACGGTCTGCT GGATAAACTGACCAAAGCTGAACAGCAGCGTCTGACCCCGTACACCACTAAACAG TCTTCTTATAACGCTCAGCTGACCGGTTACGGTACCCTGAAAGGTGCGCTGGAAAA

	ATTCGACAACCTGTCTAAAGAAATGGCAAAAGAAGATTTCTTCAAAGCTACTACCG CTACCGAACACGACGCGTTCAAAATCACCACCAACGCAAAAGCTGTTCCGGGTAA CTACGTTGTGGAAGTTAAAAACCTGGCACAGGCTCAGACCCTGACCACCCAGGCG AAAGTTAGCGACCAGGGTGCTAAACTGGGTGCGGAAGGCGTGACCGATCGTTCT CTGACCATCACCGCTGGTAACCCGCCGAAAGAAACCAAAATCCCGCTGTCTGACG ATCAGACCAGCCTGGTTGAACTGCGTGATGCAATTAACGGTGCGAAAGCTGGCGT TACCGCGTCTATTATGCGTGTGGGTGATAACGACTACCAGCTGGCGGTTTCTTCTT CTACCACCGGTGAAAACAACAAAATTAGCCTGCAGGTTGATAACGATGATAAACT GGGTGACATCCTGAACTACAACGCGACTCGTGGTACCTCCACTGCTATGAAACAG ACTGTTGCTCCGCAGGACGCGGAACTGATGGTTAACGGCACCGCGATCAAACGTT CTACCAACTCTATCTCTGATGCGCTGCAGGGTGTTACTATCGATCTGAAAACCAAA ACCAAAACCGATGAACCGCAGCACCTGGTGATTTCCACCAACACCGCGGGTACTA CCGACAAAATCAAAGAATGGGTTGACTCTTACAACTCTCTGCTGGACACCTTCAAC GCACTGTCTAAATTCACCCCGGTTAAAACCGGTGAAGCGCCGAACCCGTCTAACG GTCCGCTGCTGGGTGACAACACCCTGCGTGGCGTTCAGTCTTCCATCAAATCTGCA CTGTCTGCTGCGCAGGACAACCCGGAACTGAAAGGTCTGGGCAACCTGGGTATCT CTACCAACACTAAAACCGGCAAACTGGAAATCGATTCTGCTAAACTGAAAAAAGC TATGGACGAAAAACCGGACCAGGTTTCTAACTTCTTCGTTGGTAACGGTAAAGAC ACCGGTATGGCGACCGAAATCCACAACGAAATCCAGTCCTACATCAAAAGCGGCG GTATCATCGAAAACTCTACTAAATCTATCAACACCAACCTGGACCGCCTGAACTCT CAGATCACTACCGTTACTGCGTCTATCCAGAACACCATCGACCGTTACAAACAGCA GTTCGTTCAGCTGGATACCATGATGTCTAAAATGAACGGTACCTCTAACTACCTGG CACAGCAGTTCAAATAA
FliDhp	ATGGCGATCGGTTCCCTGTCTAGCCTGGGTCTGGGTTCTAAAGTTCTGAACTACGA CGTTATCGACAAACTGAAAGATGCTGATGAAAAAGCACTGATCGCGCCGCTGGAT AAGAAAATGGAACAGAACGTTGAAAAACAGAAAGCGCTGGTTGAAATCAAAACC CTGCTGTCCAGCCTGAAAGGCCCGGTGAAAACCCTGTCTGACTATTCTACTTACAT CTCTCGTAAATCTAACGTTACCGGTGACGCTCTGTCTGCGTCCGTGGGTGCTGGTG TTCCGATTCAGGACATCAAAGTTGATGTTCAGAACCTGGCGCAGGGTGACATCAA CGAACTGGGTGCAAAATTCTCTAGCCGTGACGACATCTTCAGCCAGGTGGATACT ACCCTGAAATTCTACACTCAGAACAAAGACTACGCTGTGAACATCAAAGCGGGCA TGACCCTGGGTGACGTTGCACAGTCTATCACCGACGCTACCAACGGCGAAGTTAT GGGCATCGTTATGAAAACCGGTGGTAACGACCCGTACCAGCTGATGGTTAACACC AAAAACACCGGCGAAGATAACCGTATCTACTTCGGTTCTCACCTGCAGTCTACCCT GACCAACAAAAACGCGCTGTCTCTGGGTGTGGATGGCTCCGGCAAATCCGAAGTT TCTCTGAACCTGAAAGGTGCGGACGGCAGCATGCACGAAGTGCCGATCATGCTGG AACTGCCGGAATCTACCAGCATCAAACAGAAAAACACCGCAATCCAGAAAGCTAT CGAACAGGCGCTGGAAAACGATCCGAACTTTAAAGACCTGATCGCAAACGGTGAC ATCTCTATCGACACCCTGCACGGTGGCGAATCTCTGATCATCAACGACCGTCGCGG TGGTAACATCGAAATCAAAGGTTCTAAAGCGAAAGAACTGGGCTTCCTGCAGACC GCGACCCAGGAATCCGATCTGCTGAAATCTTCCCGTACCATCAAAGAAGGTAAAC TGGAAGGTGTTATCTCTCTGAACGGTCAGAAACTGGACCTGAAAGCGCTGACCAA AGAATCTAACACCTCTGAAGAAAACACCGACGCGATCATCCAGGCGATCAACGCT AAAGAAGGTCTGAACGCGTTCAAAAACGCGGAAGGCAAACTGGTTATCAACTCTA AAACCGGTATGCTGACCATCAAAGGTGAAGATGCGCTGGGTAAAGCTTCCCTGAA AGACCTGGGTCTGTCCGCGGGTATGGTTCAGTCTTACGAAGCGTCCCAGGGTACC CTGTTCATGTCCAAAAACCTGCAGAAAGCGTCTGATTCTGCGTTCACCTACAACGG TGTTTCTATCACCCGTCCGACCAACGAAGTTAACGACGTTATCTCCGGCGTTAACA TCACCCTGGAACAGACCACCGAACCGAACAAACCGGCGATCATCTCCGTTATCCGT GACAACCAGGCGATCATCGACTCTCTGACCGAGTTCGTTAAAGCGTACAACGAAC TGATCCCGAAACTGGACGAAGATACCCGTTACGACGCGGACACCAAAATCGCTGG

	TATCTTCAACGGTGTTGGCGACATCCGTACCATCCGTTCTTCCCTGAACAACGTTTT CTCTTACAGCGTTCACACCGATAACGGTGTTGAATCCCTGATGAAATACGGCCTGT CTCTGGACGATAAAGGTGTTATGAGCCTGGACGAAGCGAAACTGTCCTCTGCTCT GAACTCTAACCCGAAAGCGACCCAGGACTTCTTCTACGGTTCTGACTCTAAAGACA TGGGTGGTCGTGAAATCCACCAGGAAGGTATCTTCTCTAAATTCAACCAGGTTATC GCTAACCTGATCGACGGTGGCAACGCGAAACTGAAAATCTACGAGGACTCTCTGG ACCGTGACGCGAAATCTCTGACCAAAGACAAAGAAAACGCGCAGGAACTGCTGA AAACCCGCTACAACATCATGGCGGAACGTTTCGCGGCGTACGACTCTCAGATCTCT AAAGCTAACCAGAAATTCAACTCTGTTCAGATGATGATCGACCAGGCGGCGGCGA AGAAAAACTAA
FliDec	ATGGCTTCTATTTCTTCCCTGGGTGTTGGTTCTGGTCTGGATCTGTCCTCTATCCTG GATTCCCTGACCGCTGCTCAAAAAGCTACCCTAACCCCGATCAGCAACCAGCAGTC TAGCTTCACCGCGAAACTGTCCGCATACGGTACCCTGAAATCTGCGCTGACCACCT TCCAGACCGCTAACACCGCGCTGTCTAAAGCTGATCTCTTCAGCGCTACCTCTACC ACCTCCTCCACCACTGCGTTCTCTGCCACCACCGCAGGTAACGCGATCGCCGGTAA ATACACTATCTCCGTTACCCACCTGGCCCAGGCGCAGACCCTGACCACTCGCACCA CCCGTGATGATACCAAAACCGCTATTGCGACCTCTGACAGCAAACTGACCATTCAG CAGGGTGGTGACAAAGACCCGATCACCATCGATATCTCCGCTGCTAACAGCTCCCT GTCTGGTATCCGTGATGCTATCAACAACGCAAAAGCTGGTGTGTCCGCGTCCATCA TCAACGTGGGTAACGGTGAATACCGTCTGAGCGTAACTTCCAACGACACCGGTCT GGATAACGCGATGACTCTGAGCGTTAGCGGCGATGACGCGCTGCAGTCTTTCATG GGTTACGACGCAAGCGCTAGCTCTAACGGCATGGAAGTGTCTGTTGCCGCTCAGA ACGCGCAGCTGACCGTTAACAACGTTGCGATCGAAAACTCTTCTAACACCATCTCT GATGCGCTGGAAAACATAACCCTGAACCTGAACGACGTTACCACCGGTAACCAGA CCCTGACCATCACGCAGGACACCAGCAAAGCTCAGACCGCGATCAAAGACTGGGT TAACGCGTACAACTCTCTTATCGACACCTTCTCTAGCCTGACCAAATACACCGCGGT GGACGCCGGTGCAGACTCCCAGTCCAGCAGCAACGGCGCGCTGCTGGGTGACTCT ACCCTGCGTACCATCCAGACCCAGCTGAAATCCATGCTGTCCAACACCGTTTCCAG CAGCTCCTATAAAACCCTGGCTCAGATCGGTATCACCACCGACCCATCTGATGGCA AACTGGAACTGGACGCTGATAAACTGACCGCAGCGCTGAAGAAAGATGCATCTG GCGTAGGCGCTCTGATCGTTGGTGATGGCAAGAAAACCGGCATCACCACCACCAT CGGTTCTAACCTGACTTCCTGGCTGAGCACCACTGGCATTATCAAAGCGGCTACCG ACGGTGTTTCTAAAACTCTGAACAAACTGACCAAAGATTACAACGCAGCGTCCGAT CGTATCGATGCTCAGGTTGCTCGTTACAAAGAACAGTTCACTCAGCTGGATGTGCT GATGACCTCTCTGAACTCCACCTCTTCTTACCTGACCCAGCAGTTCGAAAACAACTC TAACTCTAAATAA
FliDcjnoD4	ATGGCGTTCGGTTCTCTGTCTAGCCTGGGTTTCGGTTCTGGTGTTCTGACCCAGGA CACCATCGATAAACTGAAAGAAGCTGAACAGAAAGCGCGTATCGATCCGTACACC AAGAAAATCGAAGAAAACACCACCAAACAGAAAGATCTGACCGAAATCAAAACCA AACTGCTGTCCTTCCAGACCGCGGTTTCCTCTCTGGCTGACGCCACCGTTTTCGCTA AACGTAAAGTTGTGGGTTCTATCAGCGACAACCCGCCGGCTTCCCTGACTGTTAAC TCTGGCGTGGCTCTGCAGAGCATGAACATCAACGTGACCCAGCTGGCTCAGAAAG ACGTTTACCAGTCTAAAGGTCTGGCTAACGACTCTGGCTTCGTTAACGCTAACCTG ACCGGTACCACCGATCTGACCTTCTTCTCTAACGGTAAAGAATACACCGTTACCGT TGATAAAAACACCACCTACCGTGATCTGGCTGATAAAATCAACGAAGCGTCTGGC GGCGAAATCGTTGCTAAAATCGTTAACACCGGTGAAAAAGGTACCCCGTACCGCC TGACCCTGACCTCTAAAGAAACCGGTGAAGATAGCGCGATCTCCTTCTACGCGGG TAAGAAAGACGCTCAGGGTCAGTACCAGTCTGATCCGGAAGCGGAAAACATCTTC TCCAACCTGGGTTGGGAACTGGACAAAACCACCCAGACCATCGACCCGGCTAAAG ATAAGAAAGGCTATGGTATCAAAGACGCGTCCCTGCACATCCAGACCGCGCAGAA CGCGGAATTCACCCTGGACGGCATCAAAATGTTCCGTTCTTCTAACACCGTTACTG

	ATCTGGGTGTGGGCATGACCCTGACCCTGAACAAAACCGGCGAAATCAACTTCGA CGTGCAGCAGGATTTCGAAGGCGTGACCAAAGCGATGCAGGACCTGGTGGACGC TTACAACGACCTGGTTACCAACCTGAACGCTGCAACCGACTACAACTCTGAAACTG GTACTAAAGGCACCCTGCAGGGTATCTCTGAAGTTAACAGCATCCGTTCTTCTATC CTGGCTGACCTGTTCGACTCTCAGGTTGTTGACGGCACCACCGAGGACGCCAACG GTAACAAAGTTAACACCAAAGTAATGCTGAGCATGCAGGACTTCGGTCTGTCTCT GAACGACGCGGGTACTCTGTCTTTCGACTCTTCCAAATTCGAACAGAAAGTTAAAG AAGATCCGGATTCTACCGAATCTTTCTTCGGCATCTTCTCTAAACTGAAAGCGACCC TGCAGGAAATGACCGGTAAAGATGGTTCTATCACCAAATACGATGAATCCCTGAC CAACGACATCAAATCCCTGAACACTTCTAAAGATTCTACCCAGGCGATGATCGACA CCCGTTATGACACCATGGCTAACCAGTGGCTGCAGTACGAATCCATCCTGAACAAA CTGAACCAGCAGCTGAACACCGTAACCAACATGATCAACGCTGCGAACAACTCTA ACAACTAA

APPENDIX 2

Construct	Primer name	Sequence
D2-D3	75K-300DFwdtop	5'-CCGCGCGGCAGCCATAAACGTAAAGTGGTTG-3'
	75K-300DRvrstop	5'-CAACCACTTTACGTTTATGGCTGCCGCGCGG-3’
	75K-300DFwdbot	5'-CTTCGATGTTCAGCAGGACTAATAATTCGAGCTCCGTC-3'
	75K-300DRvrsbot	5'-GACGGAGCTCGAATTATTAGTCCTGCTGAACATCGAAG-3’
D1-D4	301F-585SFwdtop	5'-GCCGCGCGGCAGCCATTTCGAAGGTGTT-3'
	301F-585SRvrsbot	5'-AACACCTTCGAAATGGCTGCCGCGCGGC-3’
	301F-585SFwdbot	5'-CATCACCAAATATGATGAAAGCTAATAATTCGAGCTCCGTCGAC-3'
	301F-585SRvrsbot	5'-GTCGACGGAGCTCGAATTATTAGCTTTCATCATATTTGGTGATG-3’
D4	446G-579IFwdtop	5'-CGCGCGGCAGCCATGGTACCGGCAA-3’
	446G-579IRvrstop	5'-TTGCCGGTACCATGGCTGCCGCGCG-3’
	446G-579IFwdbot	5'-CGGTAAAGACGGTAGCATCTAATAATTCGAGCTCCGTC-3'
	446G-579IRvrsbot	5'-GACGGAGCTCGAATTATTAGATGCTACCGTCTTTACCG-3’

APPENDIX 3

Construct	Name	Sequence 5' – 3'
FliDcj top fragment	fliDmutantF1	GAGCTCGGTACCCGGGGATCCTCTAGAGTCgtcgatataagcttttaactagc
	fliDmutantR1	AAGCTGTCAAACATGAGAACCAAGGAGAATgtaatttagttttgatttctgttaa
FliDcj bottom fragment	fliDmutantF2	GAATTGTTTTAGTACCTAGCCAAGGTGTGCatacctctaaagactcaactcag
	fliDmutantR2	AGAATACTCAAGCTTGCATGCCTGCAGGTCactgtttcattgttatgcac
Standard Kanamycin cassette	KanF	ATTCTCCTTGGTTCTCATGTTTGACAGCTTAT
	KanR	GCACACCTTGGCTAGGTACTAAAACAATTCAT
Amplified *fliD* cassette from *C. jejuni* 81116	fliDcompF	AATATTCGTCTCACATGgcatttggtagtctatctagttta
	fliDcompR	AATATTCGTCTCACATGgcttgattttgagaataagc

Table of primers for construction of *fliD* deletion mutant and complemented strains. The uppercase sequences of the *fliD* mutant primers are the adaptor regions used in the Gibson assembly cloning, while the lowercase sequences are the regions annealing to a region upstream of *fliD* (F1) and just inside the *fliD* coding region (R1) or at the end of the *fliD* coding region (F2) and downstream of *fliD* (R2). The KanF and KanR primers are adaptors that also amplify the *kan* gene from pJMK30

APPENDIX 4

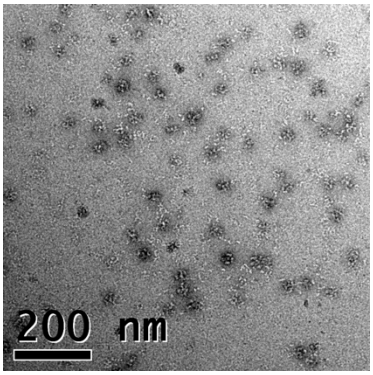
Name	Sequence 5' – 3'
F3L_F	TCATGGCATTaGGTAGTCTATC
F3S_F	TCATGGCATcTGGTAGTCTATC
F3_R	AAAAGTCCTTTCATTTAAAATGAAC
L9A_F	TCTATCTAGTgcAGGATTTGGTTC
L9S_F	TCTATCTAGTTcAGGATTTGGTTC
L9_R	CTACCAAATGCCATGAAAAAG
F11L_F	GTTTAGGATTaGGTTCTGGGG
F11S_F	GTTTAGGATcTGGTTCTGGGG
F11_R	TAGATAGACTACCAAATGC
Y315L_F	GGTGGATGCTctTAATGATTTAGTAAC
Y315S_F	GGTGGATGCTTcTAATGATTTAGTAAC
Y315_R	AAATCTTGCATGGCTTTTG
N316S_F	GATGCTTATAgTGATTTAGTAACCAATC
N316L_F	GATGCTTATctTGATTTAGTAACCAATC
N316_R	CACCAAATCTTGCATGGC
L318A_F	TTATAATGATgcAGTAACCAATCTTAATGC
L318S_F	TTATAATGATTcAGTAACCAATCTTAATGC
L318_R	GCATCCACCAAATCTTGC
L338A_F	AAAAGGAACTgcACAAGGCATC
L338S_F	AAAAGGAACTTcACAAGGCATC
L338_R	GTTCCAGTTTCACTATTATAG
D397N_F	TTTGAGTTTTaATTCTTCTAAATTTGAAC
D397L_F	TTTGAGTTTTctTTCTTCTAAATTTGAAC
D397_R	GTGCCTGCATCATTTAAAC
K400S_F	GATTCTTCTAgtTTTGAACAAAAAGTTAAAGAAGATC
K400L_F	GATTCTTCTttATTTGAACAAAAAGTTAAAGAAGATC
K400_R	AAAACTCAAAGTGCCTGC
L592A_F	TATTAAATCAgcAAATACCTCTAAA
L592_R	TCATTTGTCAAACTCTCATC
M602L_F	AACTCAGGCTcTGATTGATACAAG
M602_R	GAGTCTTTAGAGGTATTTAATG
W614L_F	GCGAATCAATtGTTGCAATATG
W614S_F	GCGAATCAATcGTTGCAATATG
W614F_F	GCGAATCAATtcTTGCAATATG
W614_R	CATTGTATCATATCTTGTATCAATC
Y617L_F	GGTTGCAATtaGAGAGTATTTTAAATAAAC
Y617S_F	GGTTGCAATcTGAGAGTATTTTAAATAAAC
Y617_R	ATTGATTCGCCATTGTATC
I620A_F	ATATGAGAGTgcTTTAAATAAACTCAATCAACAGC
I620S_F	ATATGAGAGTtcTTTAAATAAACTCAATCAACAGC
I620_R	TGCAACCATTGATTCGCC
L624A_F	TTTAAATAAAgcCAATCAACAGCTAAATACTGTAAC
L624S_F	TTTAAATAAAtcCAATCAACAGCTAAATACTGTAAC
L624_R	ATACTCTCATATTGCAACC
L628A_F	CAATCAACAGgcAAATACTGTAACTAATATG

L628S_F	CAATCAACAGtcAAATACTGTAACTAATATG
L628_R	AGTTTATTTAAAATACTCTCATATTG
I635A_F	AACTAATATGgcTAATGCGGCAAACAATTC
I635S_F	AACTAATATGtcTAATGCGGCAAACAATTC
I635_R	ACAGTATTTAGCTGTTGATTG

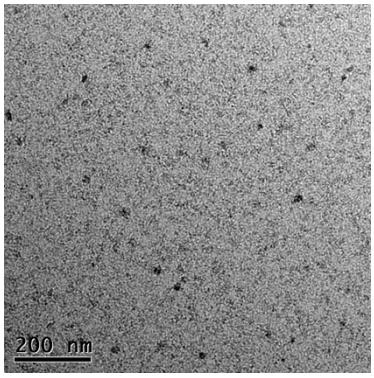
APPENDIX 5

Non-cleaved fraction samples

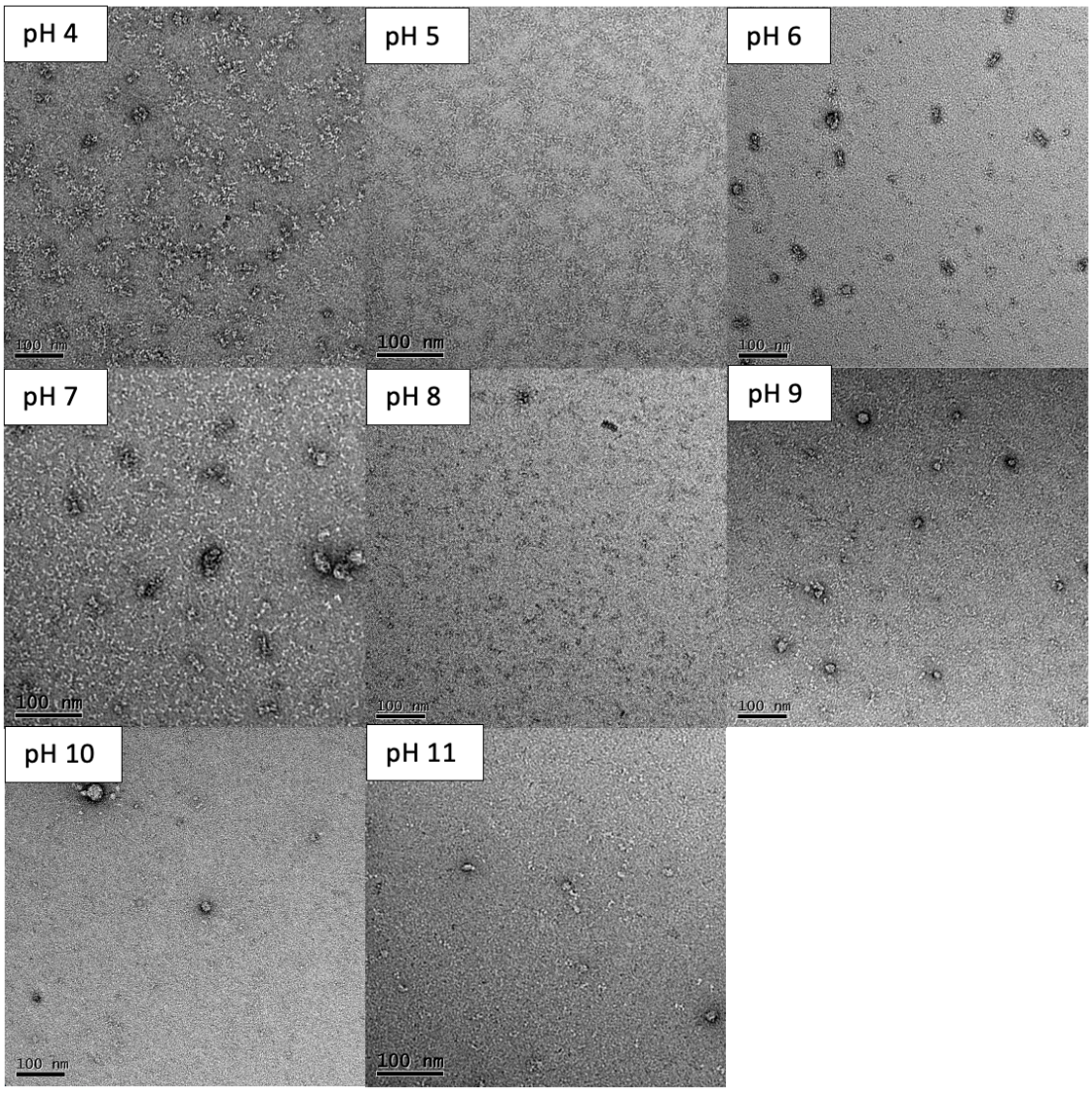
A10



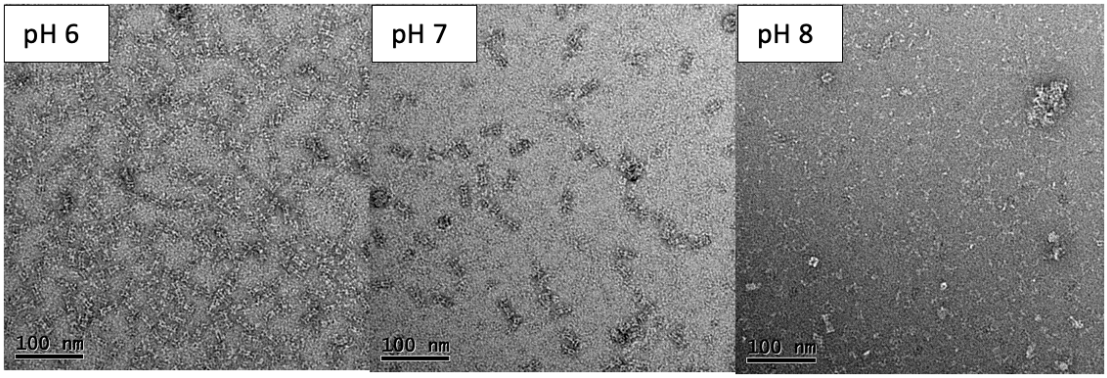
B12



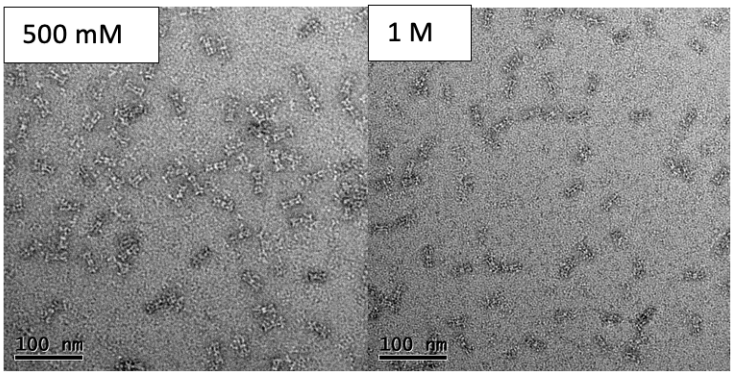
No Salt



150 mM Salt

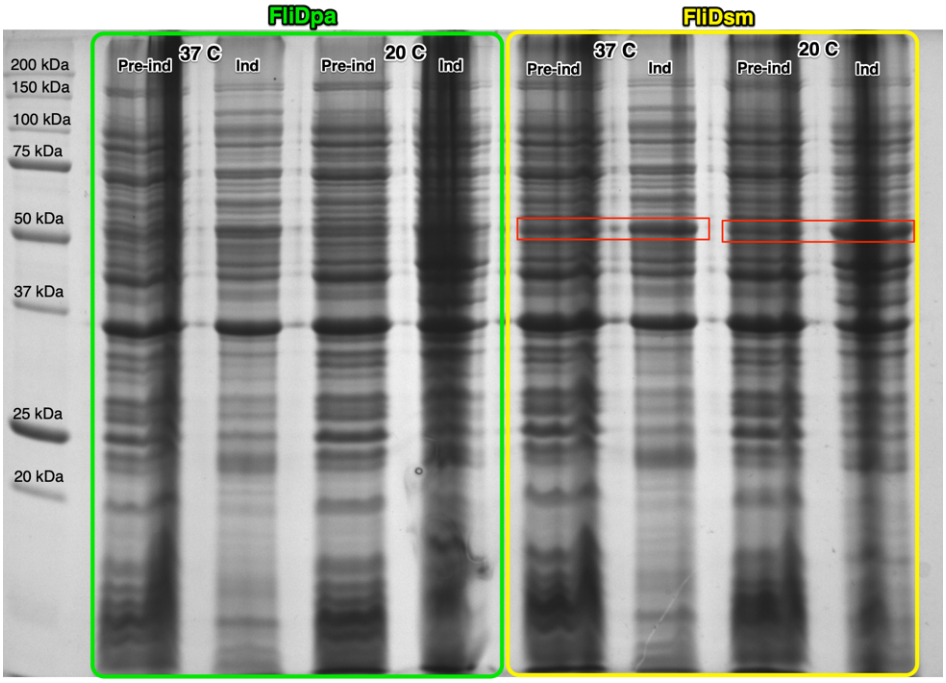


pH 7 high salt samples

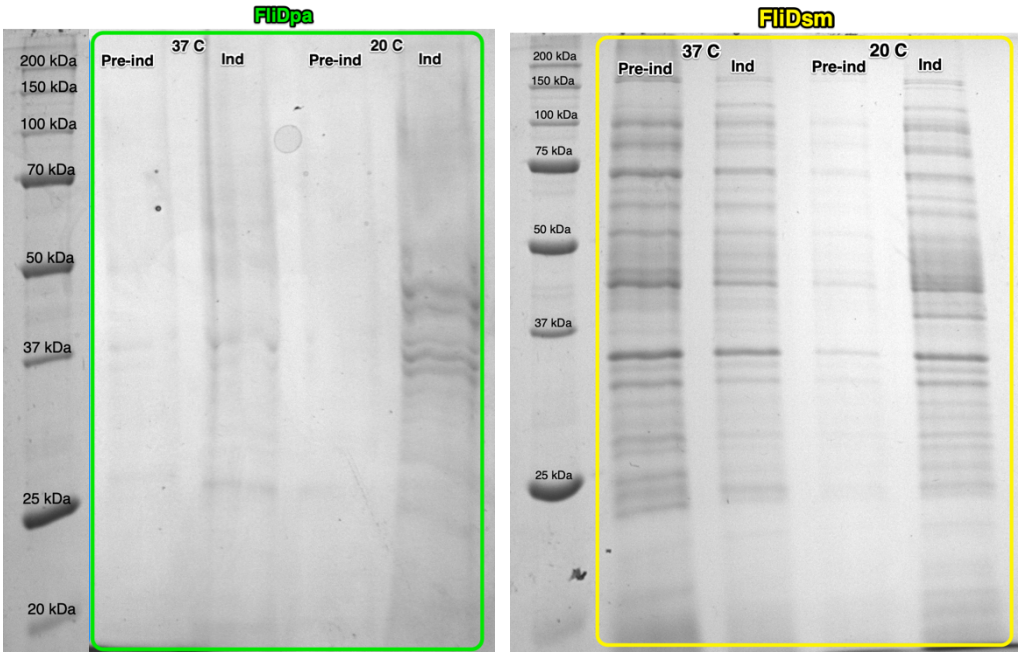


APPENDIX 6

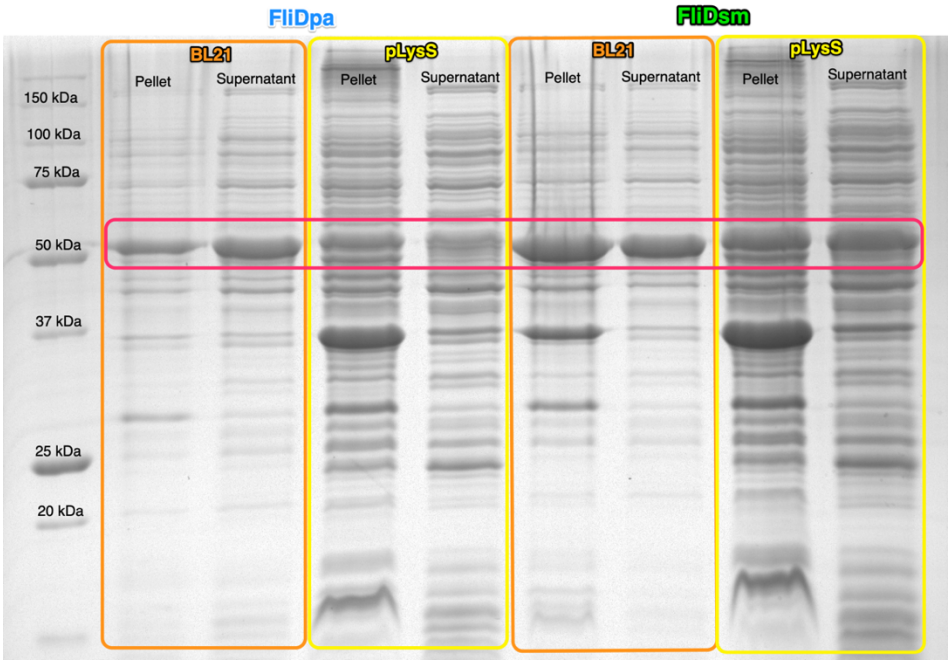
BL21 IPTG expression trial



pLysS IPTG expression trial

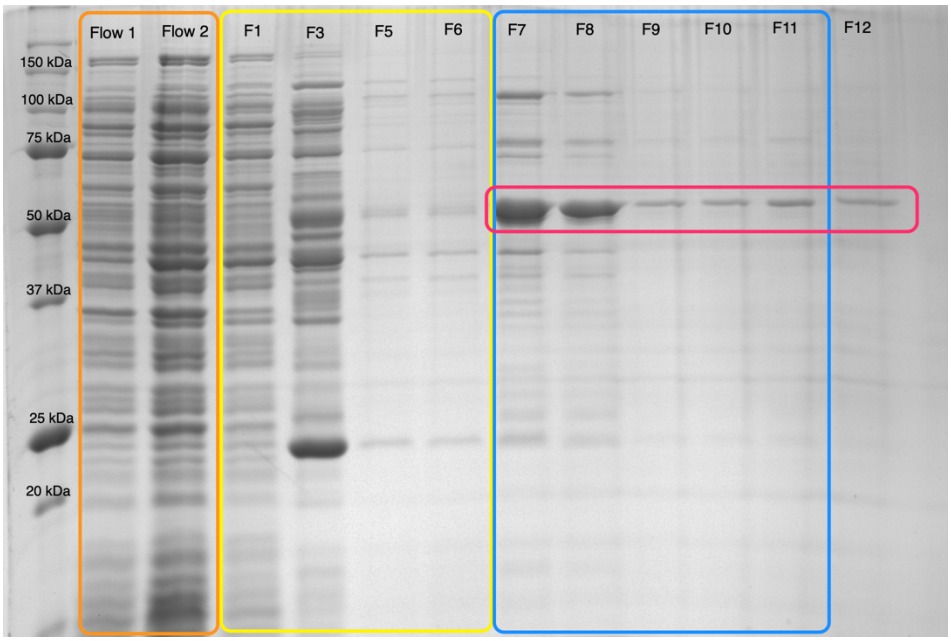


Expression trial 3: BL21 (Auto-induction), pLysS (IPTG)

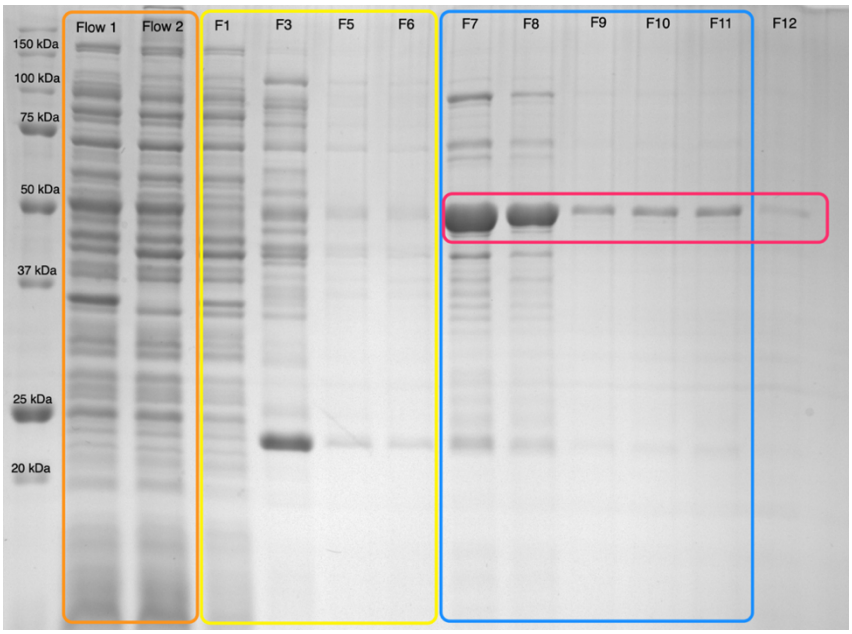


<u>Ni-NTA columns:</u>

FliDpa pLysS

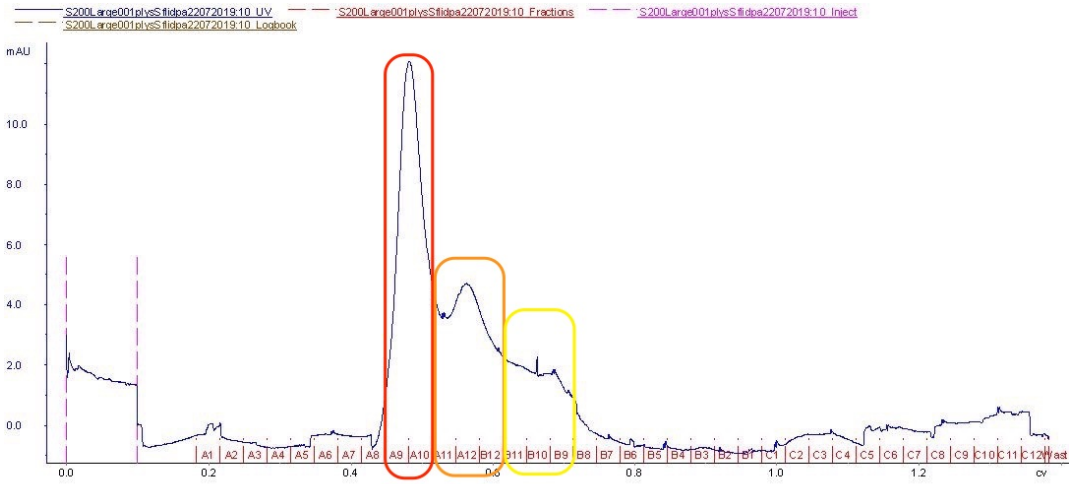


FliDsm pLysS

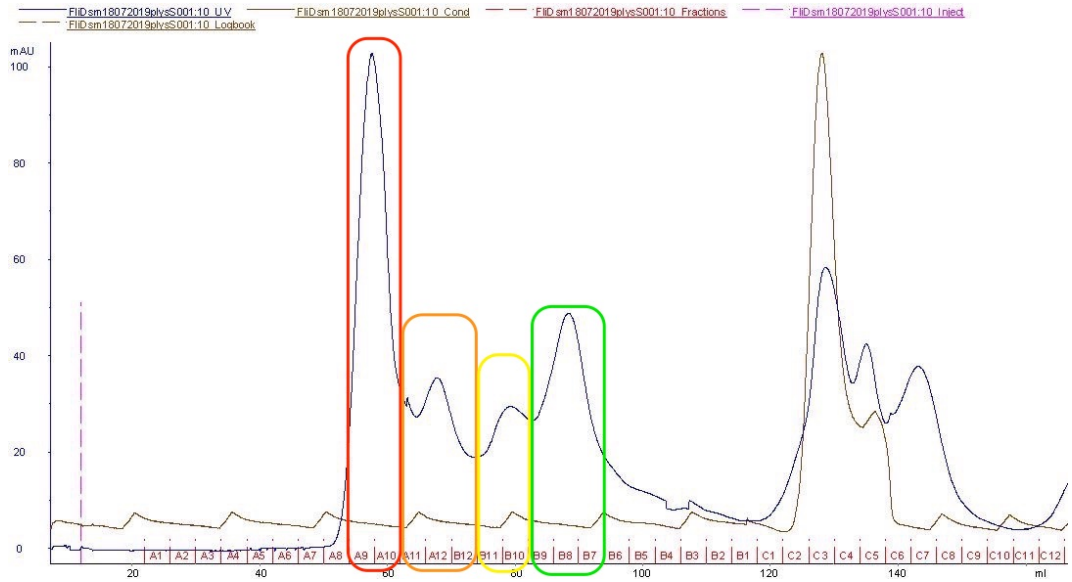


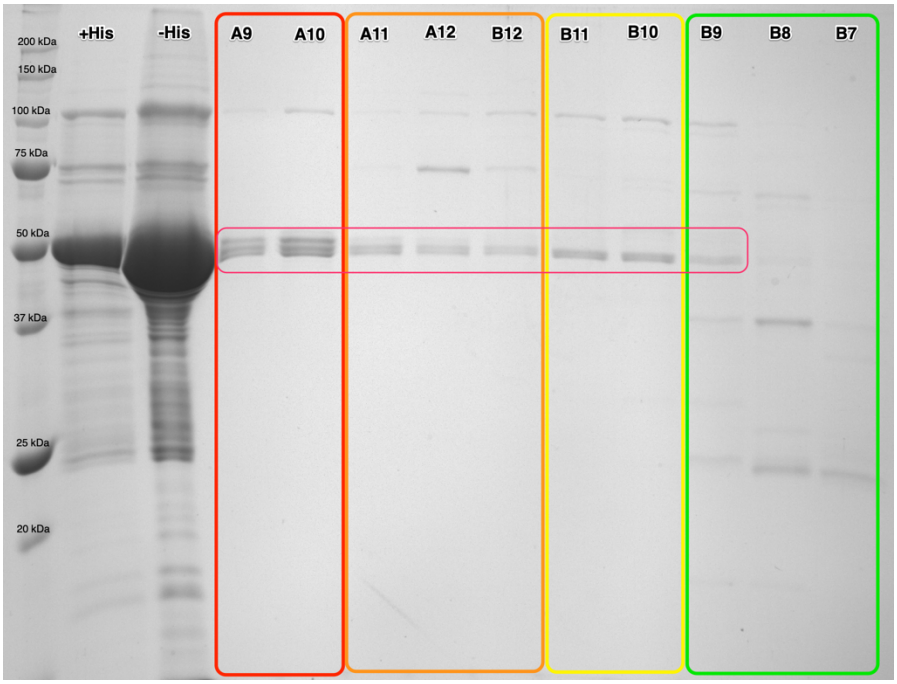
<u>Gel filtration columns:</u>

FliDpa pLysS



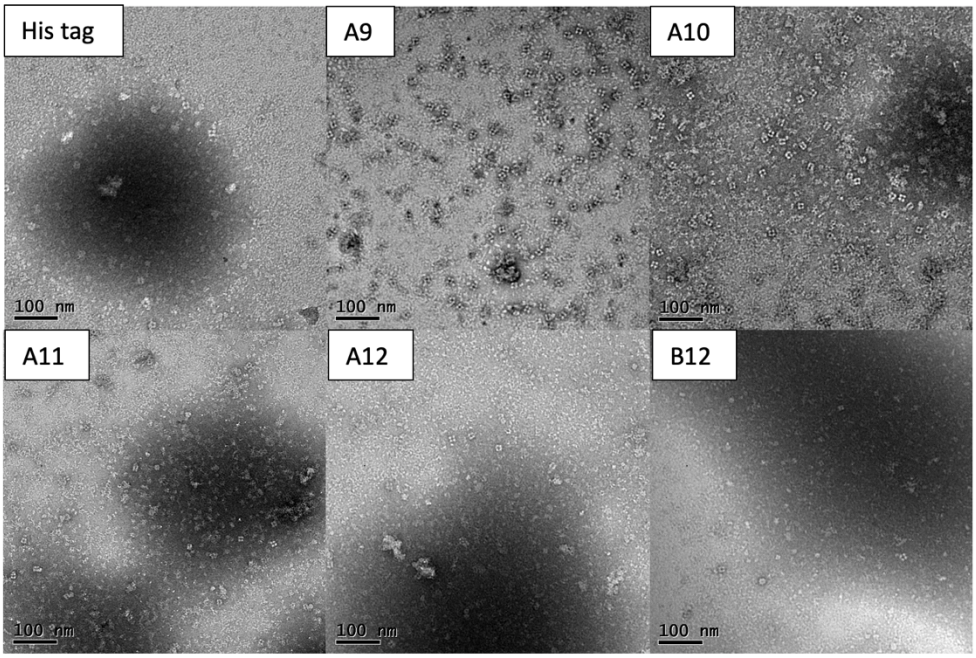
FliDsm pLysS



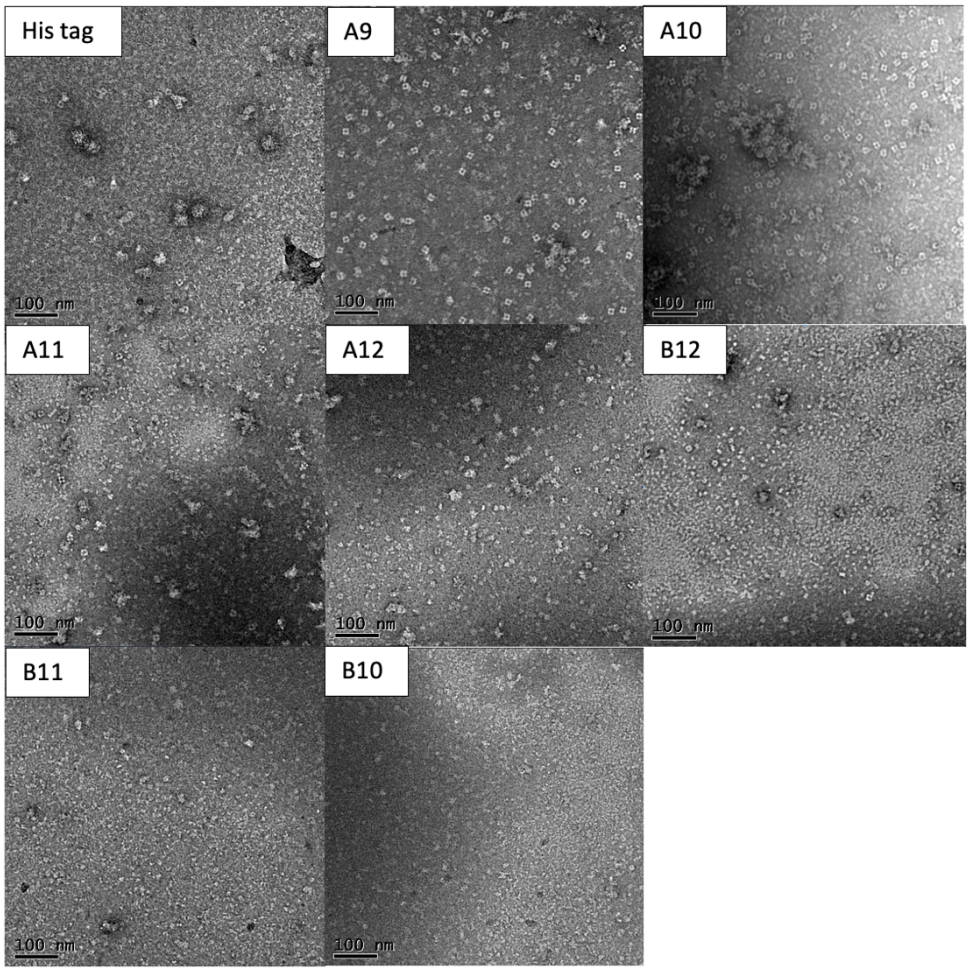


Negative stain:

FliDpa



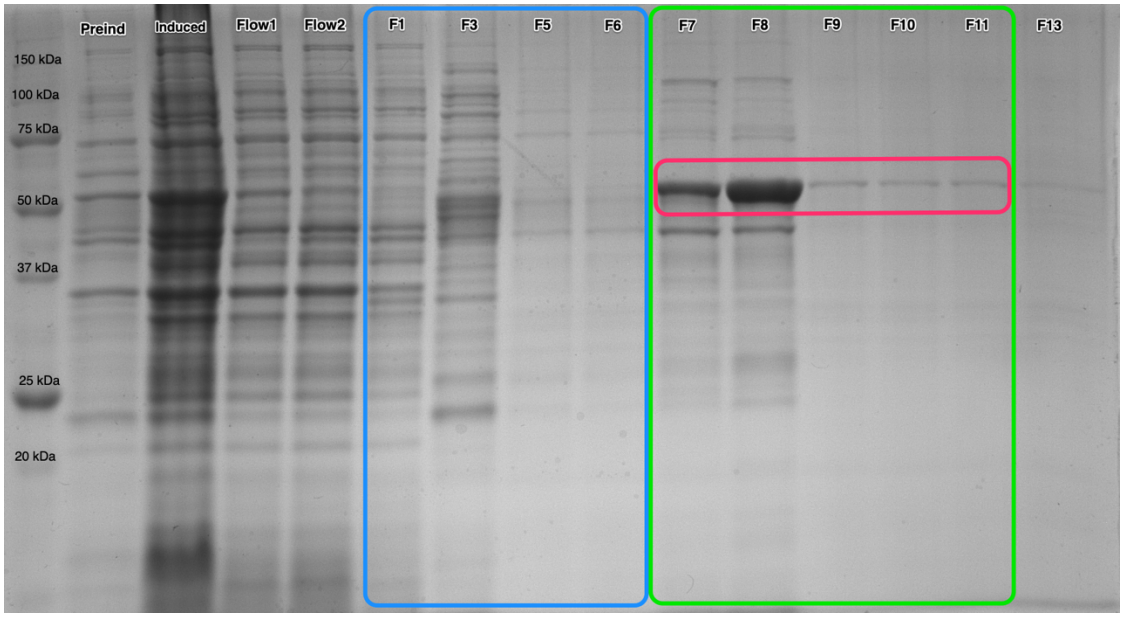
FliDsm



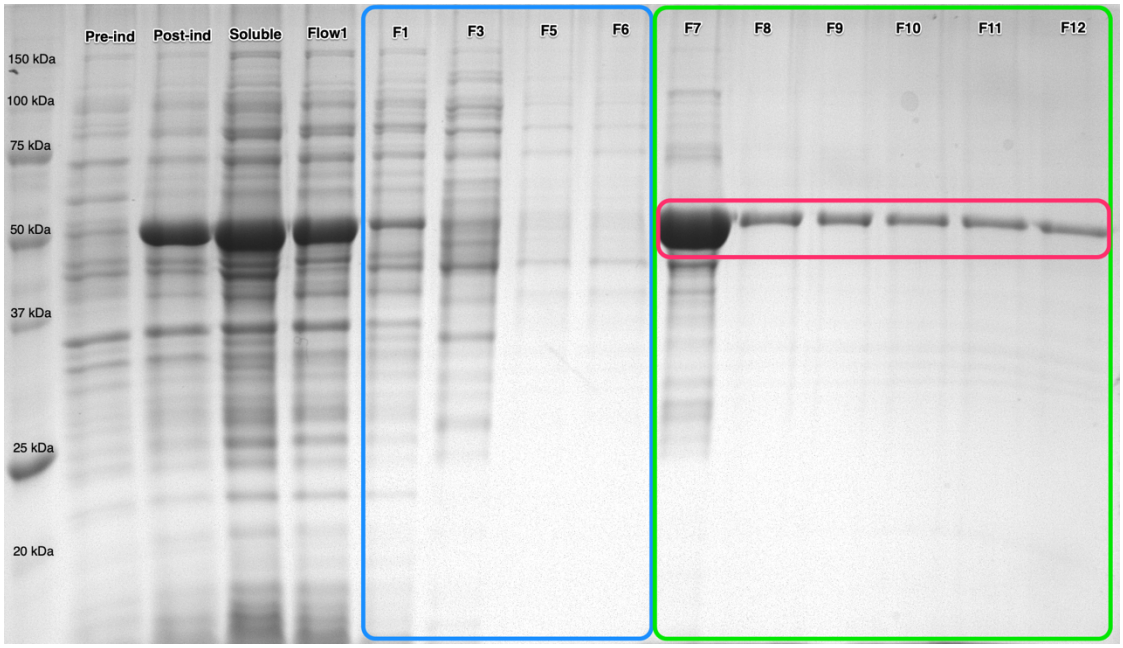
APPENDIX 7

Ni-Nickel column output

FliDpa

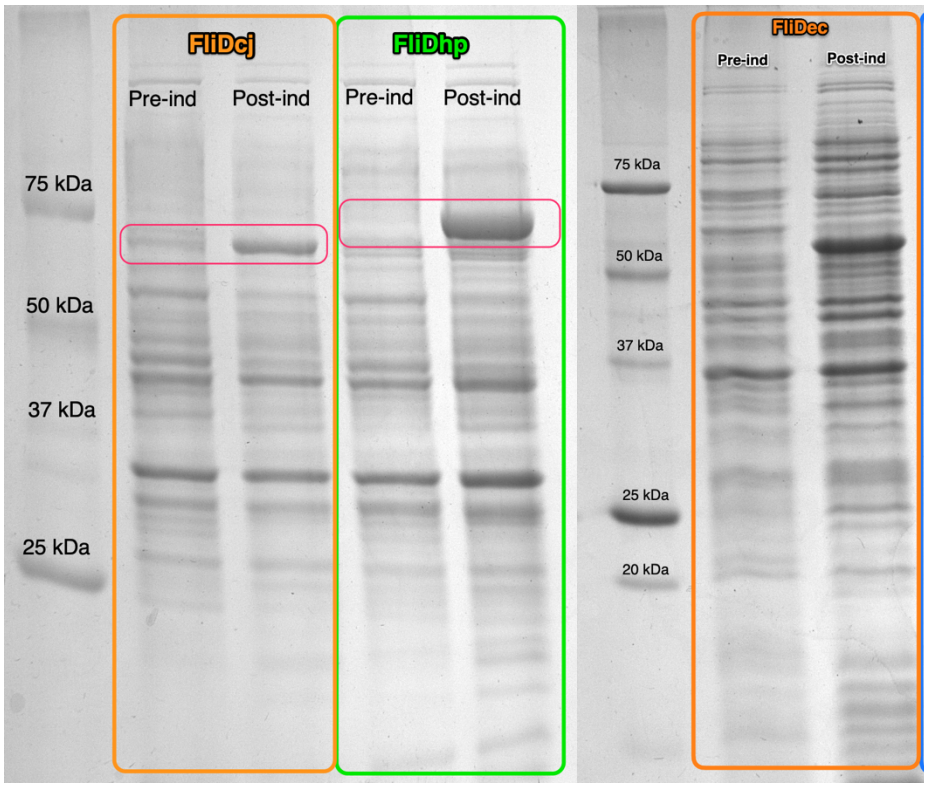


FliDsm

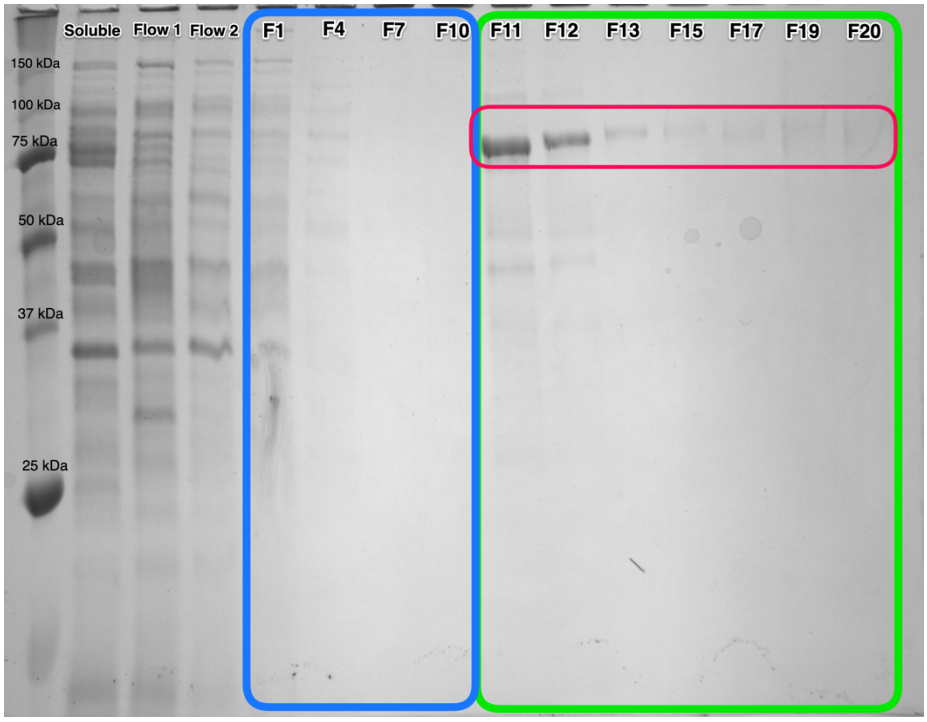


APPENDIX 8

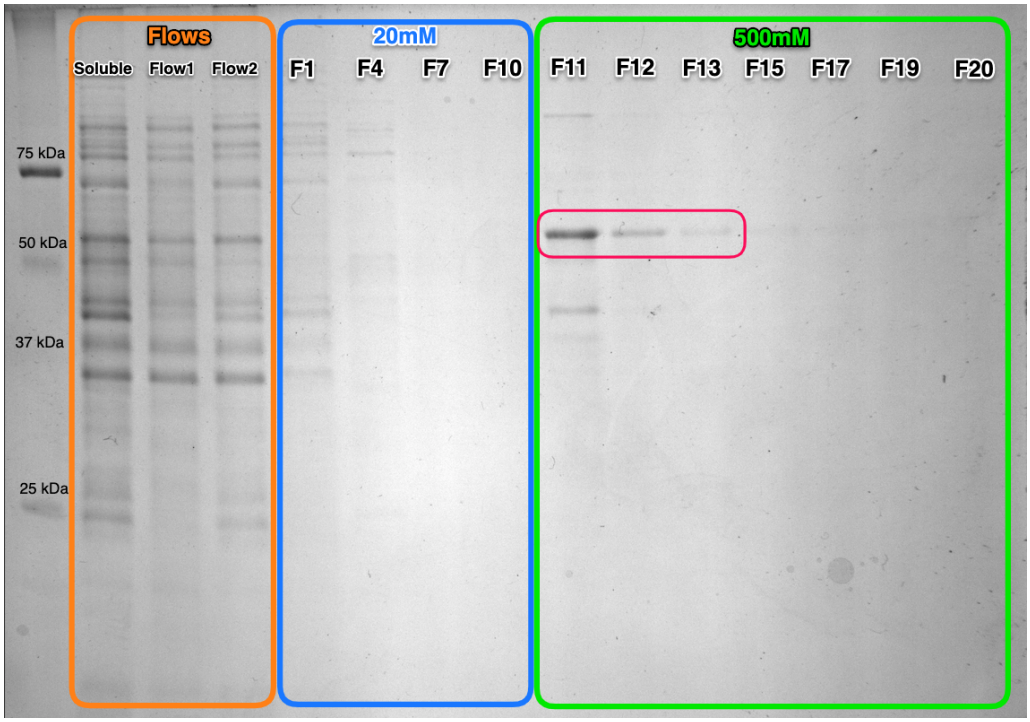
Expression gels



Ni-NTA column FliDhp

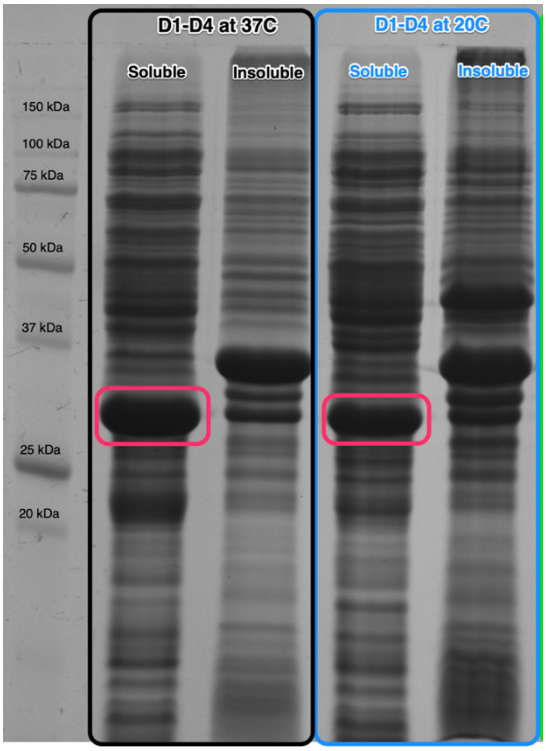


Ni-NTA column FliDec

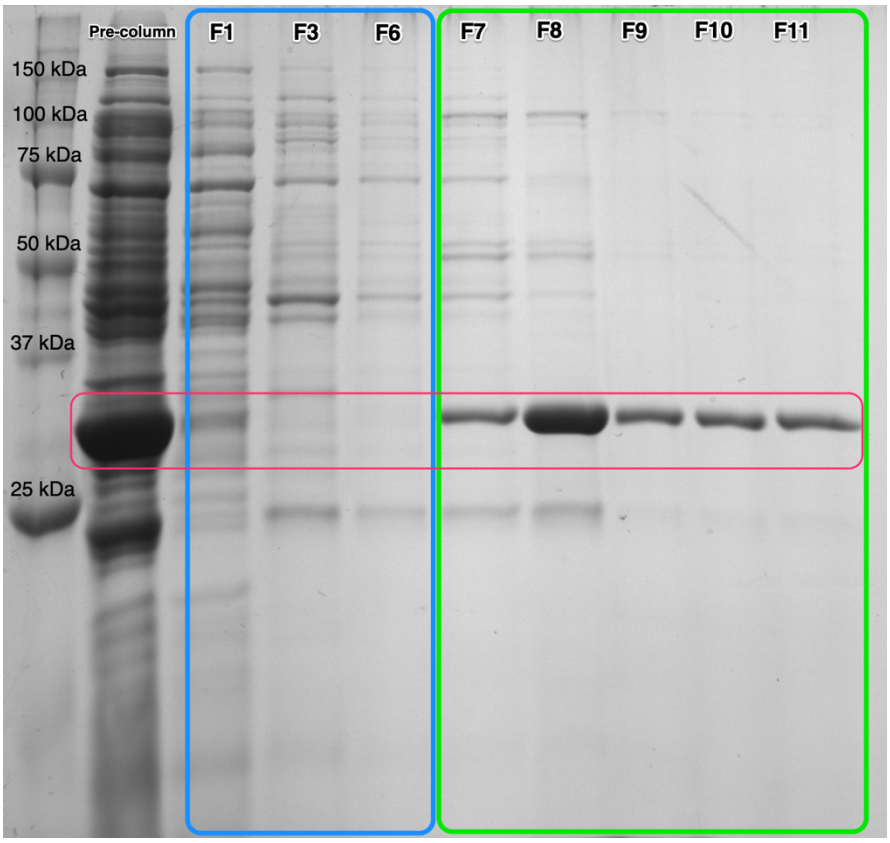


APPENDIX 9

Expression gel

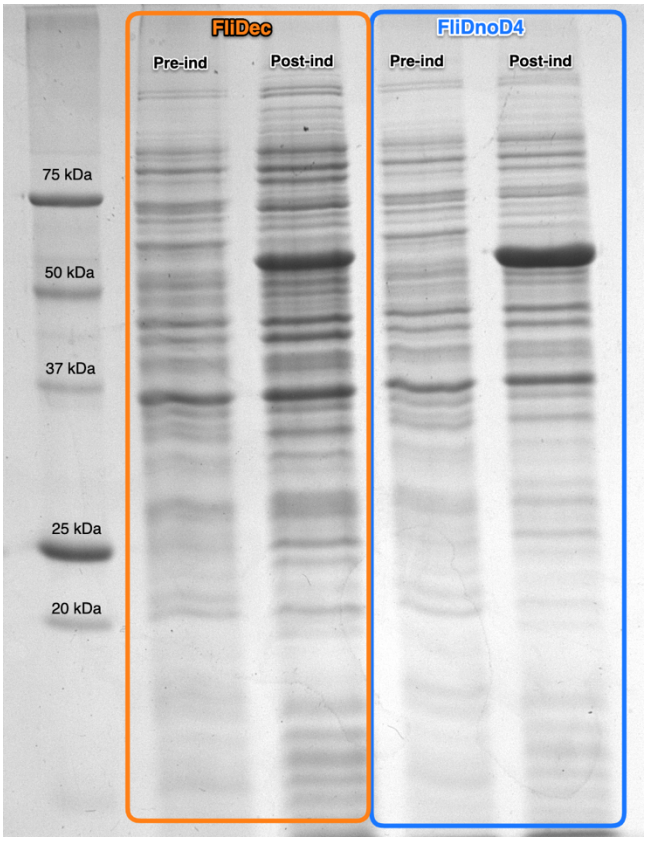


Ni-NTA column

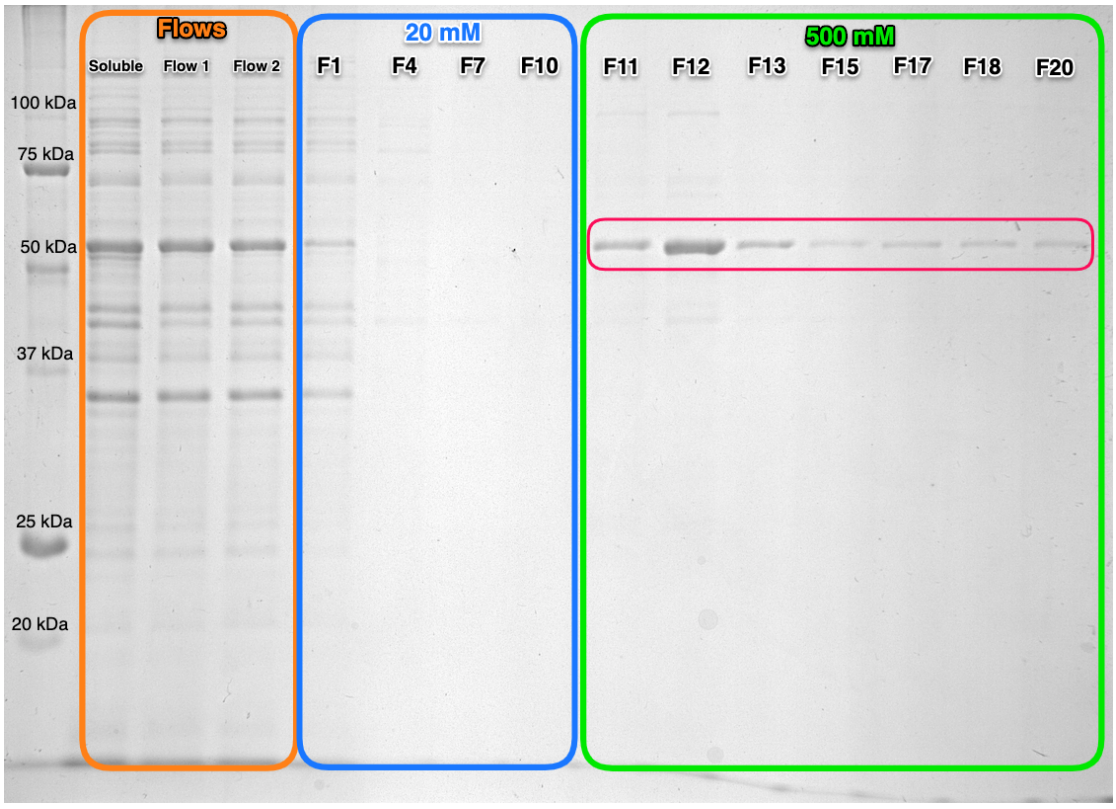


APPENDIX 10

Expression gel



Ni-NTA column



APPENDIX 11

Time	WT 81116			ΔfliD			ΔfliDϕmetK		
	1	2	3	1	2	3	1	2	3
0	0.076	0.072	0.073	0.074	0.077	0.069	0.068	0.071	0.074
2	0.104	0.096	0.097	0.107	0.116	0.108	0.114	0.105	0.099
4	0.163	0.158	0.156	0.178	0.182	0.170	0.164	0.168	0.162
6	0.250	0.238	0.245	0.291	0.289	0.278	0.245	0.257	0.270
8	0.492	0.467	0.471	0.551	0.539	0.507	0.441	0.490	0.522
10	0.810	0.697	0.726	0.828	0.813	0.777	0.793	0.785	0.809
12	1.284	1.090	1.104	1.055	1.099	1.037	1.348	1.234	1.236

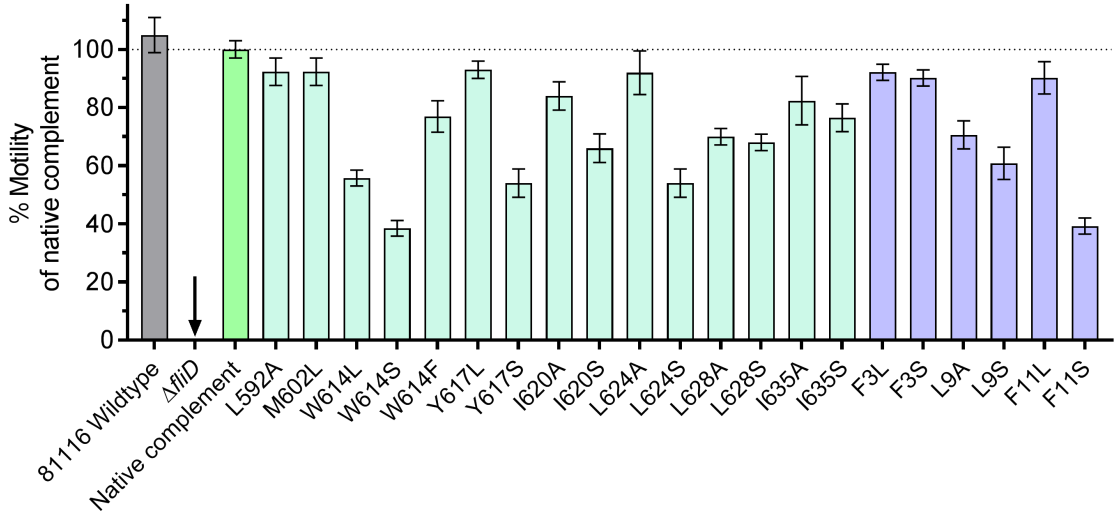
APPENDIX 12

Sample		Swarm diameter (mm)
Wild type	Stab 1	22.7
	Stab 2	19.7
	Stab 3	21.4
Δ*fliD*	Stab 1	0
	Stab 2	0
	Stab 3	0
Δ*fliD*ϕmetK	Stab 1	21.1
	Stab 2	19.7
	Stab 3	20

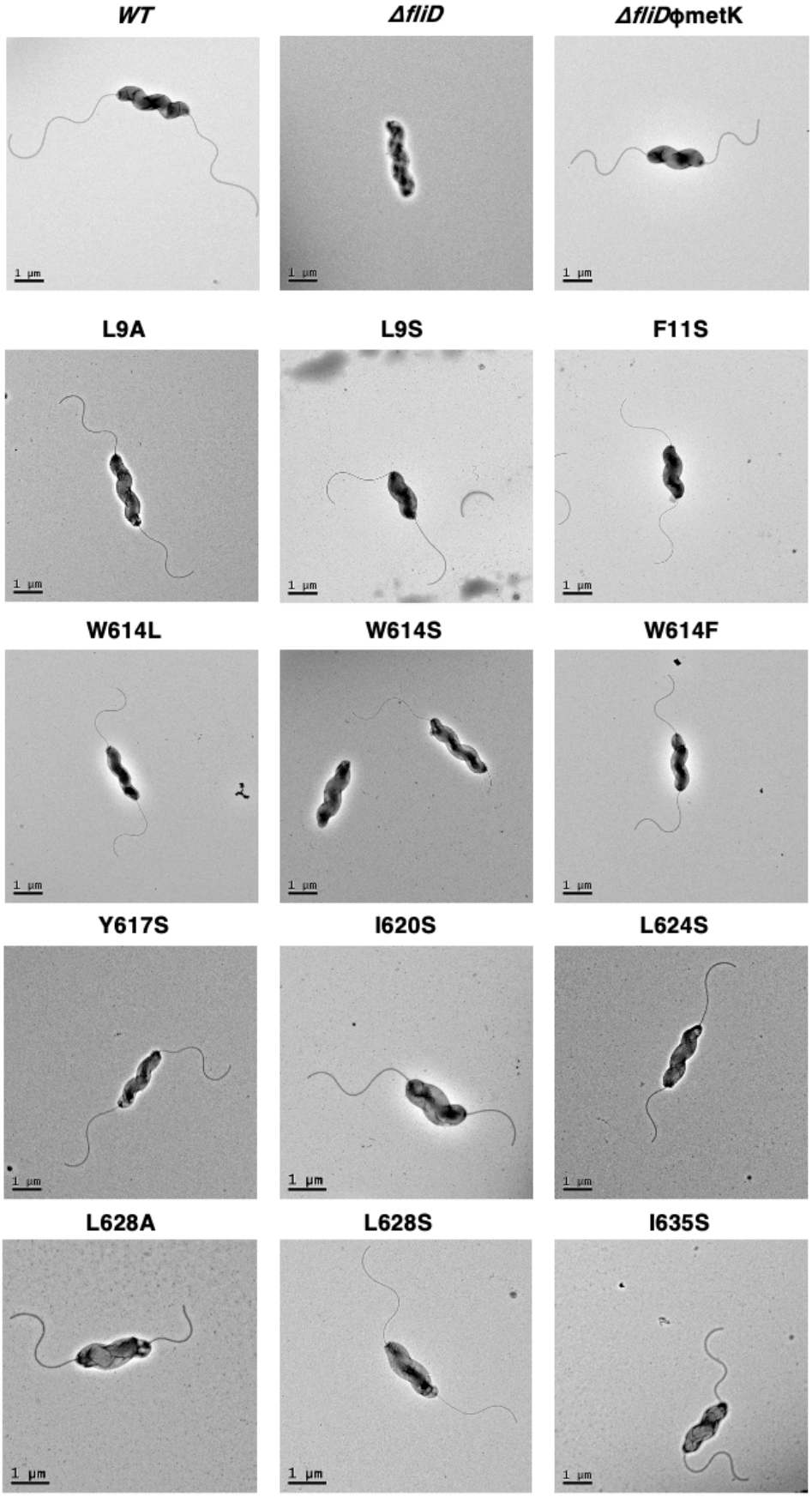
APPENDIX 13

	Diameter of swarm (mm)			Control average	% of Native complement				
Strain	**R1**	**R2**	**R3**		**R1**	**R2**	**R3**	**Mean**	**SD**
ΔfliDφ metK	***16***	***17***	***19***	***17.33***	***92.30769***	***98.07692***	***109.6154***	***100***	***7.195495***
L592A	15	16	17		86.53846	92.30769	98.07692	92.30769	4.710557
M602L	15	16	17		86.53846	92.30769	98.07692	92.30769	4.710557
W614L	9	10	10		51.92308	57.69231	57.69231	55.76923	2.719641
W614S	6	7	7		34.61538	40.38462	40.38462	38.46154	2.719641
W614F	12	14	14		69.23077	80.76923	80.76923	76.92308	5.439283
ΔfliDφ metK	***15***	***17***	***18***	***16.67***	***90***	***102***	***108***	***100***	***7.483315***
Y617L	15	16	21		90	96	126	93	3
Y617S	8	9	10		48	54	60	54	4.898979
I620A	13	14	15		78	84	90	84	4.898979
I620S	10	11	12		60	66	72	66	4.898979
L624A	14	15	17		84	90	102	92	7.483315
L624S	8	9	10		48	54	60	54	4.898979
L628A	11	12	12		66	72	72	70	2.828427
L628S	11	11	12		66	66	72	68	2.828427
I635A	12	15	15		70.58824	88.23529	88.23529	82.35294	8.318903
I635S	12	13	14		70.58824	76.47059	82.35294	76.47059	4.802921
F3L	15	16	16		88.23529	94.11765	94.11765	92.15686	2.772968
F3S	15	15	16		88.23529	88.23529	94.11765	90.19608	2.772968
L9A	11	12	13		64.70588	70.58824	76.47059	70.58824	4.802921
L9S	9	11	11		52.94118	64.70588	64.70588	60.78431	5.545936
F11L	14	16	16		82.35294	94.11765	94.11765	90.19608	5.545936
F11S	6	7	7		35.29412	41.17647	41.17647	39.21569	2.772968

Values on the bar graph were calculated as percentage of native complement control specified above the tested residues in italic. The point mutants were tested in 2 batches thus required 2 controls.



APPENDIX 14


WT
ΔfliD
ΔfliDϕmetK
L9A
L9S
F11S
W614L
W614S
W614F
Y617S
I620S
L624S
L628A
L628S
I635S
1 µm

APPENDIX 15

WT	Comp	F3L	F3S	L9A	L9S	W614S	F11S	W614L	W614F	Y617S	I620S	I628A	I628S	I635S	L624S
3.463	3.228	2.891	3.518	3.084	2.852	2.658	2.186	2.366	2.879	3.097	2.891	2.392	2.402	3.229	3.404
2.561	3.902	2.843	3.397	3.855	2.803	2.756	2.979	2.682	3.168	4.107	2.578	3.577	3.579	3.493	2.918
2.727	1.783	1.397	3.108	3.734	2.948	2.975	1.129	2.829	3.58	2.618	1.662	2.658	1.922	3.421	2.115
0.98	4.441	1.132	2.072	3.228	3.118	2.658	2.258	2.512	2.709	2.835	2.939	1.643	1.85	3.831	3.258
2.644	4.343	3.036	3.084	3.229	2.731	2.756	3.099	2.609	3.096	2.738	1.928	3.117	2.522	2.602	3.088
3.163	3.484	3.084	3.566	2.819	3.166	2.975	3.051	2.594	3.325	2.882	3.252	2.948	3.397	2.867	3.088
3.299	2.895	2.771	1.06	1.879	2.948	3.609	2.931	2.762	3.228	3.432	3.036	2.851	3.099	3.156	2.796
2.944	3.564	2.602	2.409	3.493	3.116	3.658	1.994	2.042	3.156	2.642	3.797	1.909	1.922	3.301	2.747
4.004	4.023	2.289	2.719	1.951	3.77	2.317	1.329	3.195	3.325	2.33	2.781	2.61	4.204	4.795	2.457
4.868	3.534	2.988	1.903	3.831	3.746	2.951	1.885	3.045	3.024	2.498	3.435	3.311	3.795	3.276	1.831
4.105	3.758	3.373	3.383	4.578	1.16	3.048	3.263	3.793	1.524	3.122	2.709	2.876	2.931	5.035	3.903
2.705	4.032	0.988	4.253	3.108	2.852	3.292	2.344	2.803	4.257	3.146	2.854	3.383	2.931	3.18	1.157
2.975	3.628	2.482	2.61	4.12	2.803	1.219	3.745	3.182	2.951	3.027	3.459	1.933	3.003	2.987	3.18
3.75	0.889	2.65	2.707	2.885	2.948	3.243	2.368	2.127	2.225	3.435	2.975	3.238	2.786	2.915	2.69
5.585	3.382	1.494	2.731	4.654	3.118	3.219	2.513	2.9	2.843	3.099	3.362	3.069	3.018	1.928	2.018
3.151	3.378	1.735	3.021	2.957	2.731	2.78	4.036	3.552	2.988	3.339	0.822	1.909	3.525	2.288	1.465
5.222	3.557	0.892	2.876	2.4	3.166	2.78	2.988	2.03	3.18	3.291	3.072	3.214	2.68	2.795	2.594
5.198	3.916	0.94	2.9	2.303	2.948	2.756	3.71	3.214	2.964	2.689	3	2.851	3.307	3.132	1.514
3.376	2.402	4.529	2.876	3.49	3.116	1.634	3.879	3.045	3.47	3.05	3.895	3.165	3.066	3.252	2.234
3.182	3.723	2.048	3.238	2.957	3.77	3.421	3.325	2.393	3.229	2.69	2.443	1.82	3.597	2.482	1.922
3.666	2.8	4.385	3.214	3.612	3.746	3.132	3.084	2.537	3.277	3.483	3.58	3.257	2.414	2.53	2.522
3.952	1.255	2.481	2.489	3.49	1.16		2.482	3.311	2.731	2.45	2.975	2.994	2.907	2.337	1.706
3.739	3.476	2.795	2.296	3.926	3.262		2.939	2.513	2.707	2.162	3.048	3.209	3.003	3.325	3.051
3.36	3.842	1.711	2.054	3.878	4.374		2.21	3.238	2.17	1.73	2.613	3.185	2.138	2.602	3.051
4.425	2.474	4.636	2.513	3.999	2.61		2.526	2.562	2.707	2.114	2.878	3.185	3.122	2.626	2.594
3.904	3.904	3.987	2.951	3.393	3.142		2.356	2.924	3.146	2.546	3.942	3.712	2.618	2.072	2.354
6.38	3.281	4.636	2.879	2.763	1.523		2.307	2.876	2.429	2.468	2.274	2.778	2.907	3.48	
6.281	5.532	3.627	2.927	3.418	3.722		2.72	2.344	3.376	2.976	1.742	2.97	2.282	2.779	
4.049	4.348	4.327	2.613	3.539	2.344		2.939	2.731	3.012	3.846	2.129	4	2.642	3.19	
1.153	2.57	2.979	3.193	3.199	3.237		3.376	2.659	2.065	3.12	2.685	2.203	3.051	3.576	
1.35	2.546	4.108	2.902	3.03	2.658		3.012	2.924	1.822	2.796	2.419	3.531	2.666	3.093	
4.367	1.405	2.042	3.121	3.612	2.223		3.137	3.199	1.749	3.193	3.774	2.81	2.834	2.827	
3.656	4.714	3.003	3.266	3.297	2.626		2.395	3.102	2.915	2.105	2.443	3.027	2.186	2.586	
3.827	3.818	2.21	3.145	3.297	2.626		3.161	2.521	2.72	2.637	3.814	3.051	2.306	4.035	
4.536	1.015	3.123	3.338	2.691	2.918		2.251	3.345	2.648	2.758	2.695	2.618	2.892	2.153	
4.658	4.543	3	3.217	3.155	2.723		4.455	2.885	3.398	2.869	2.89	3.603	2.533	2.274	
1.171	1.86	3.291	2.661	2.797	2.772		2.97	2.836	2.209	2.991	3.352	3.963	2.82	3	
3.853	3.141	1.609	4.233	3.131	1.629		3.233	3.224	2.429	2.212	2.452	3.363	2.677	3.072	
		3.843	3.097	2.366	3.744		2.898	2.763	2.501	2.626	2.062	1.586	2.51	3.12	

		2.978	3.024	3.203	2.747		3.281	1.41	2.526	2.383	2.502	3.003	2.294	3.314	
		3.099	2.588	3.654	1.537		3.185	3.088	3.011	4.953	3.206	2.859	2.366	2.25	
		2.186	2.153	3.299	3.355		2.778	3.063	1.613	2.981	2.89	3.387	2.701	1.863	
		3.531	3.168	2.94	3.501		2.299	2.334	2.915	2.116	3.132	2.258	2.247	3.943	
		4.42	3.12	2.82	3.647		2.682	4.571	2.384	3.006	1.493	3.699	2.964	3.145	
		3.195	3.263	2.581	3.671		2.251	2.426	2.83	3.005	2.216	2.762	2.748	3.605	
		4.276	2.592	2.385	2.747		1.988	2.546	3.145	2.356	2.165	3.145	3.37	3.096	
		2.763	2.4	2.605	4.109		3.195	2.594	2.096	2.867	2.602	2.879	2.797	1.766	
		1.994	2.472	4.278	2.728		2.954	2.882	2.819	4.24	2.626	2.153	2.81	3.531	
		4.516	2.855	2.557	3.184		2.267	2.498	2.819	2.65	3.349	2.25	3.099	3.145	
		3.267	2.904	1.984	2.588		2.45	1.946	1.976	2.987	3.47	3.096	3.075	4.547	
		2.354	3.072	2.581	2.443		2.402	2.907	2.987	3.036	2.12	2.806	2.186	2.878	
		3.05	3.12	2.462	2.903		1.922	0.961	3.397	3.349	2.698	1.234	3.627	6.58	
		1.225	2.784	2.127	2.878		3.147	2.81		3.734	2.939	3.629	2.883	2.661	
		3.195	2.88	3.657	2.661		3.075	2.834		3.132	1.036	2.879	2.474	3.338	
		2.883	2.448	2.844	2.951		3.373			3.276	2.819	2.975	2.931	3.459	
		1.874	1.824	2.844	3.157		1.903			2.771	2.771	3.048	3.458	3.096	
		2.714	2.855	2.412	1.723		1.542				2.867	2.637	2.882	2.796	
		3.747	2.904	2.701	2.672		2.554					2.782	2.498	2.699	
		2.374	3.072	2.604	3.734		2.385					2.516	2.474	2.358	
		1.465	3.12	3.25	2.915		2.963					3.483	3.147	2.942	
		3.027	2.784	2.496	3.373		3.036					2.927	3.291	3.331	
		2.354	2.88	2.808	3.084		2.747						3.045	2.796	
		3.411	2.448	2.496	2.795		2.241						1.933	3.161	
		2.975	1.824	2.088	2.771		2.626						0.556	3.015	
		3.072	2.855	2.904	3.277		3.205						2.489	3.361	
		2.998	2.784	2.4	3.466		2.747						1.305	3.926	
		3.316	2.89	6.144	2.852		2.771						3.19	3.656	
		1.805	2.866	3.239	2.508		2.53						3.165	4.097	
		3.511	3.206	3.719	2.95		3.445						3.117	3.116	
		2.317	2.647	2.712	3.467		3.712						3.601	3.239	
		3.634	3.668	2.688			2.898						3.456	4.368	
		3.829	2.405	2.544			2.994						2.996	2.625	
		2.414	3.473	2.4			3.496						3.238	3.091	
		2.804	2.963	3.335			2.539							2.576	
		2.683	3.18	1.08			2.915							2.061	
		3.536	1.931				3.108							3.165	
		2.511	0.723				1.663							3.165	
		2.95	3.277				2.854							3.435	
		3.414	1.686				2.781							2.55	
		2.899	3.132				3.054							1.987	
		3.951	3.108				2.06							3.386	

		2.902	2.963				2.981								
		2.366	1.879				2.569								
		2.707	1.952				2.885								
			2.144				3.204								
			3.421				3.277								
			4.765				2.385								
			2.226				2.096								
			3.483				2.554								
			2.927				2.53								
			3.435				3.084								
			4.217												
			2.588												
			3.314												
			2.685												
			2.347												
			3.286												
			2.513												
			2.296												
			2.996												
			2.972												
			3.207												
			3.262												
			2.731												
			3.673												
			5.099												
			2.827												
			2.44												
3.63	**3.27**	**2.86**	**2.88**	**3.07**	**2.93**	**2.85**	**2.76**	**2.76**	**2.80**	**2.93**	**2.76**	**2.88**	**2.80**	**3.11**	**2.53**
1.26	*1.06*	*0.89*	*0.63*	*0.74*	*0.62*	*0.58*	*0.576*	*0.55*	*0.54*	*0.56*	*0.66*	*0.58*	*0.57*	*0.74*	*0.67*
38	38	84	108	75	70	21	91	54	52	56	57	61	73	81	26

Bottom 3 rows in following order:

Mean

SD

N

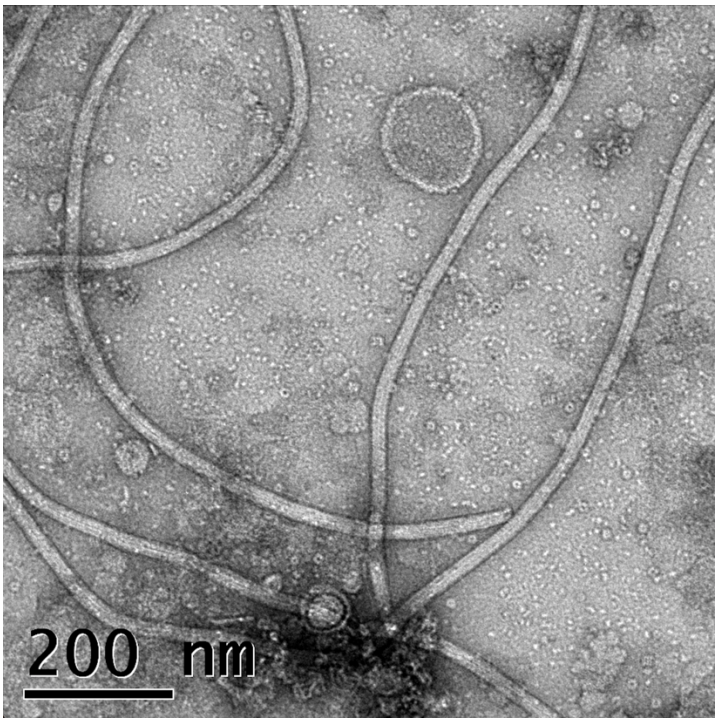
APPENDIX 16

Sample		Flagella	Percentage of total (%)
Comp	Total	48	100
	Attached	49	83.3
	Unattached	8	16.7
WT	Total	44	100
	Attached	34	77.3
	Unattached	10	22.7
Knockout	Total	0	0
	Attached	0	0
	Unattached	0	0
F11S	Total	139	100
	Attached	45	32.4
	Unattached	94	67.6
L9S	Total	125	100
	Attached	42	32.8
	Unattached	84	67.2
L624S	Total	33	100
	Attached	20	60.6
	Unattached	13	39.4
W614L	Total	62	100
	Attached	52	83.9
	Unattached	10	16.1
W614S	Total	20	100
	Attached	16	80
	Unattached	4	20
Y617S	Total	67	100
	Attached	38	56.7

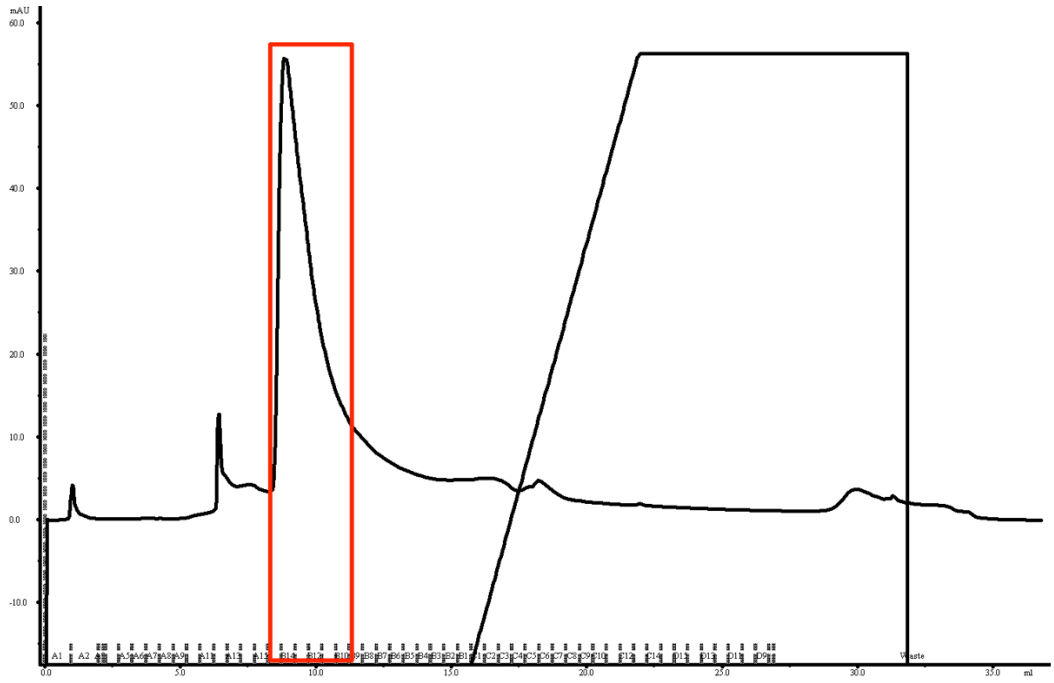
	Unattached	29	43.3
F3L	Total	109	100
	Attached	42	38.5
	Unattached	67	61.5
F3S	Total	166	100
	Attached	42	25.3
	Unattached	124	74.7
L9A	Total	110	100
	Attached	42	38.2
	Unattached	68	61.8
I635S	Total	140	100
	Attached	43	30.7
	Unattached	97	69.3
I620S	Total	80	100
	Attached	44	55
	Unattached	36	45
L628S	Total	79	100
	Attached	51	64.6
	Unattached	28	35.4
L628A	Total	85	100
	Attached	62	72.9
	Unattached	23	27.1
W614F	Total	65	100
	Attached	48	73.8
	Unattached	17	26.2

APPENDIX 17

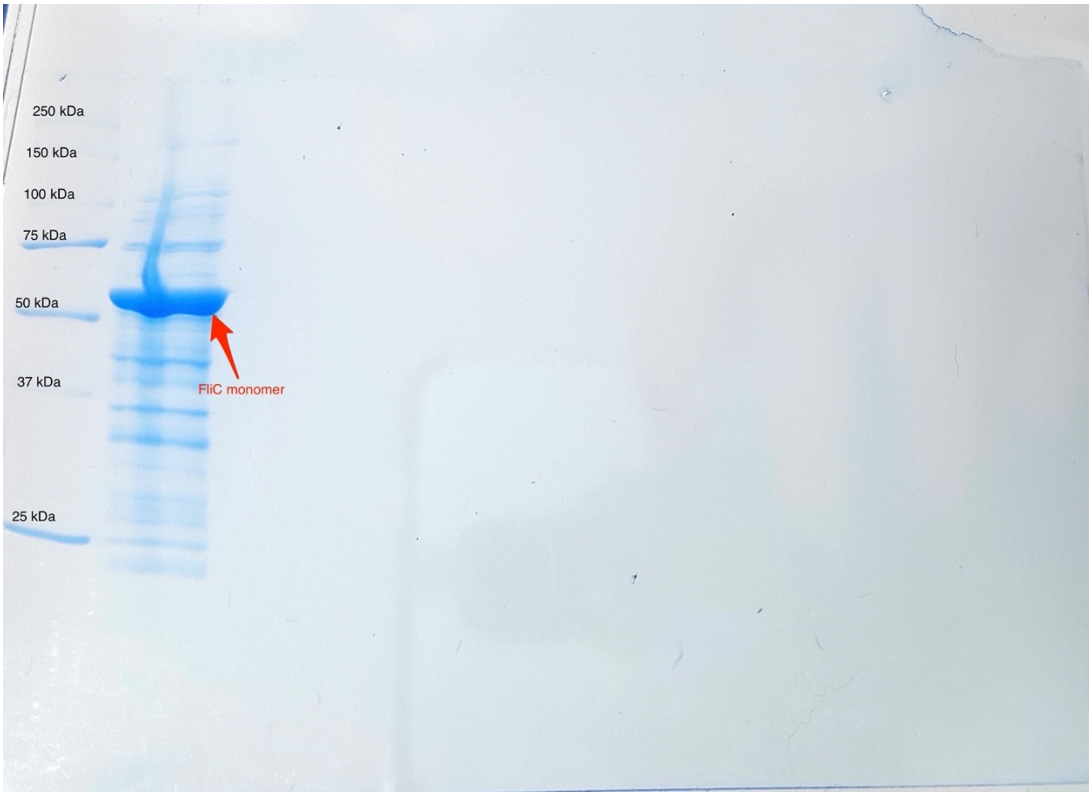
Ultracentrifuged 81-176 sample:



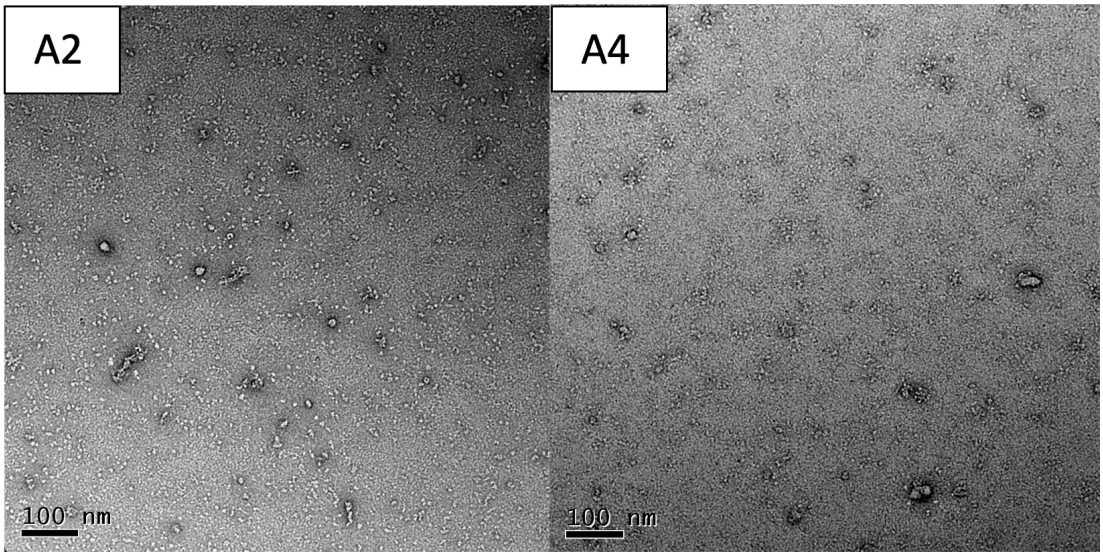
Anion exchange column



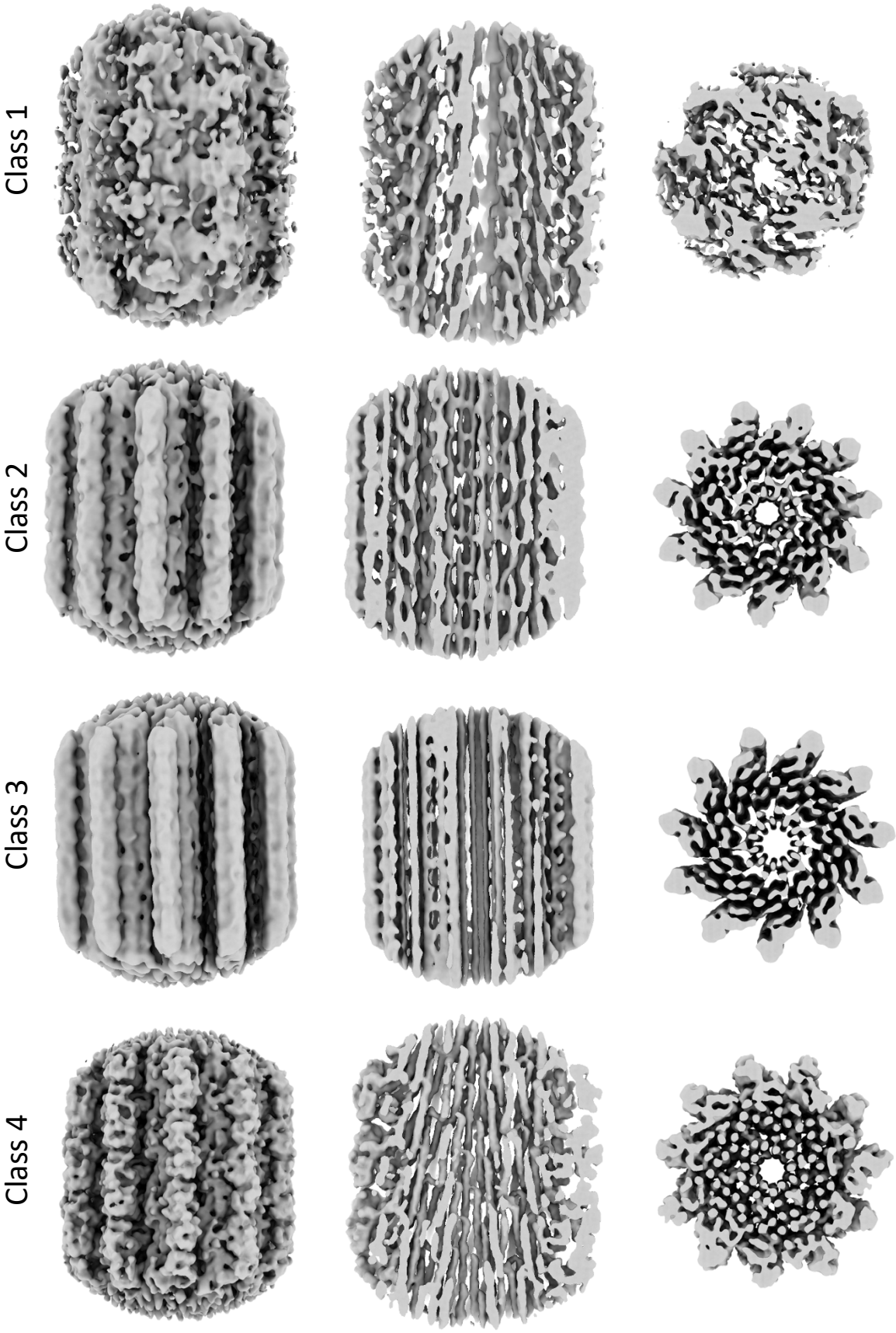
SDS-PAGE gel



APPENDIX 18



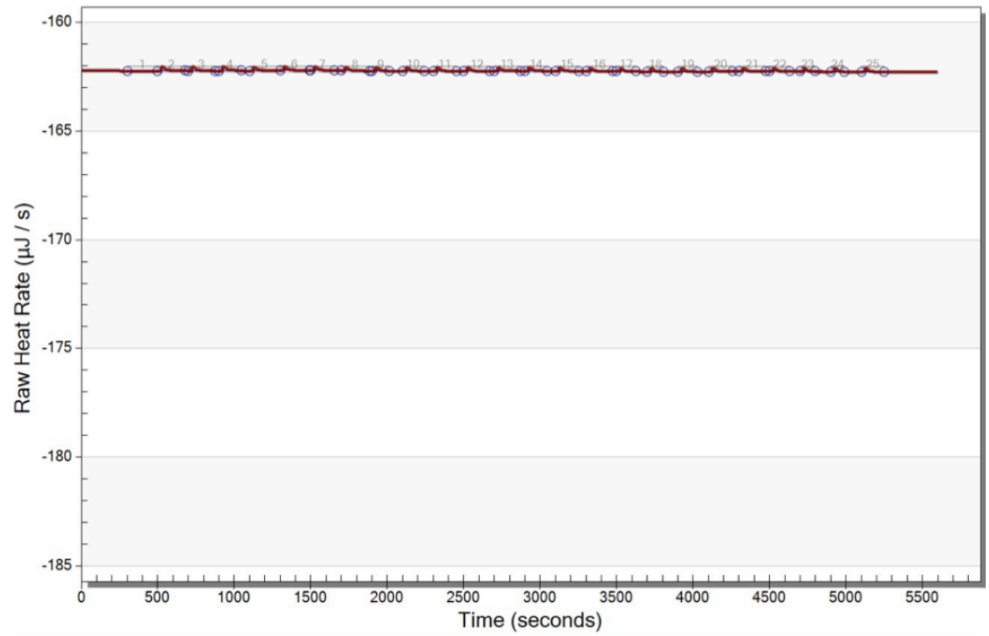
APPENDIX 19




1.0
0.5
0.143
0.0
0
22 Å
11 Å
7.4 Å
5.5 Å
4.4 Å
Class 1 - 13 Å
Class 2 - 13.9 Å
Class 3 - 14.3 Å
Class 4 - 9.5 Å

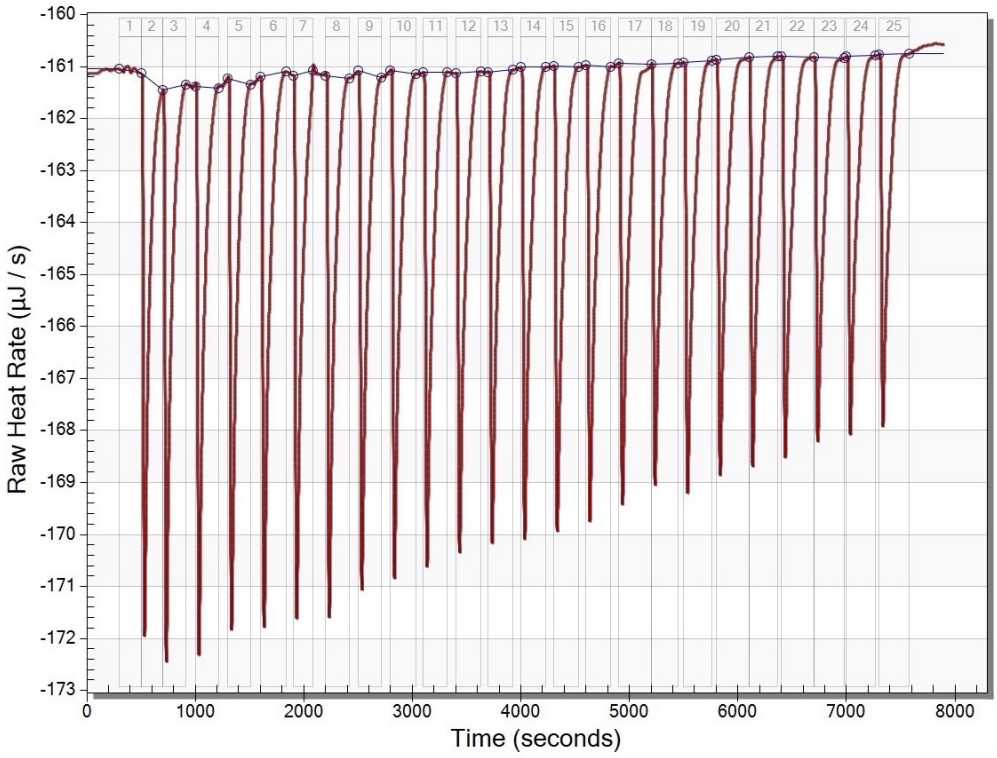
APPENDIX 20

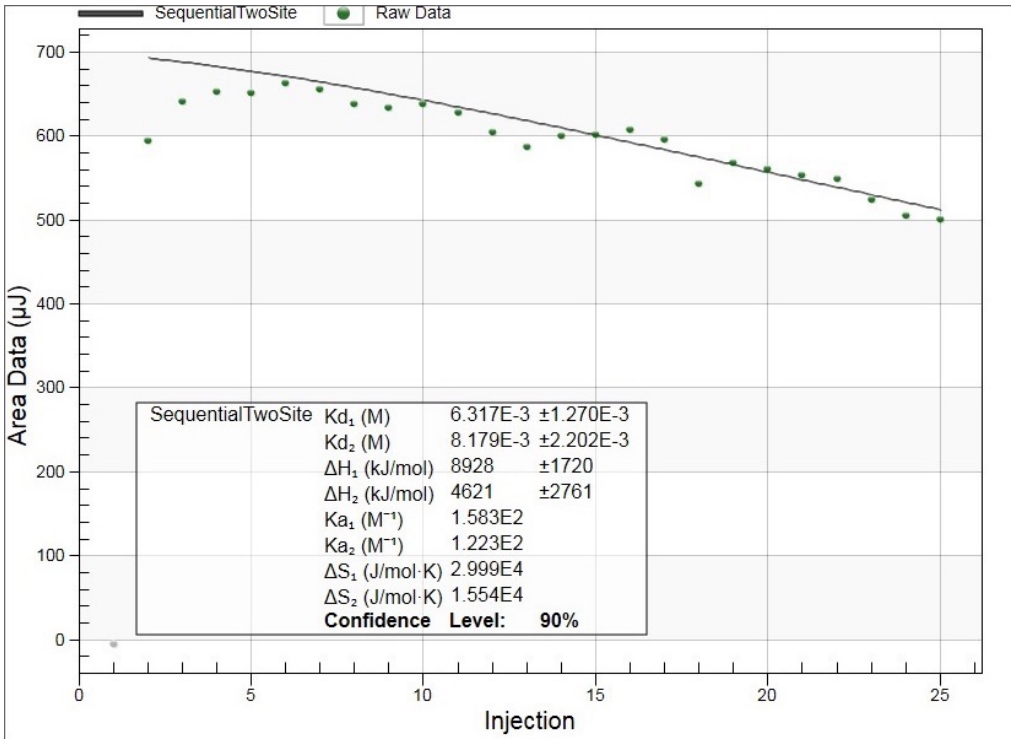
Raw data for a heat of dilution control experiment, with 0.02 mM $FliD_{cj}$ in the cell, 250 second injection length, 2 µl injection volume and 25 injections of buffer into the cell



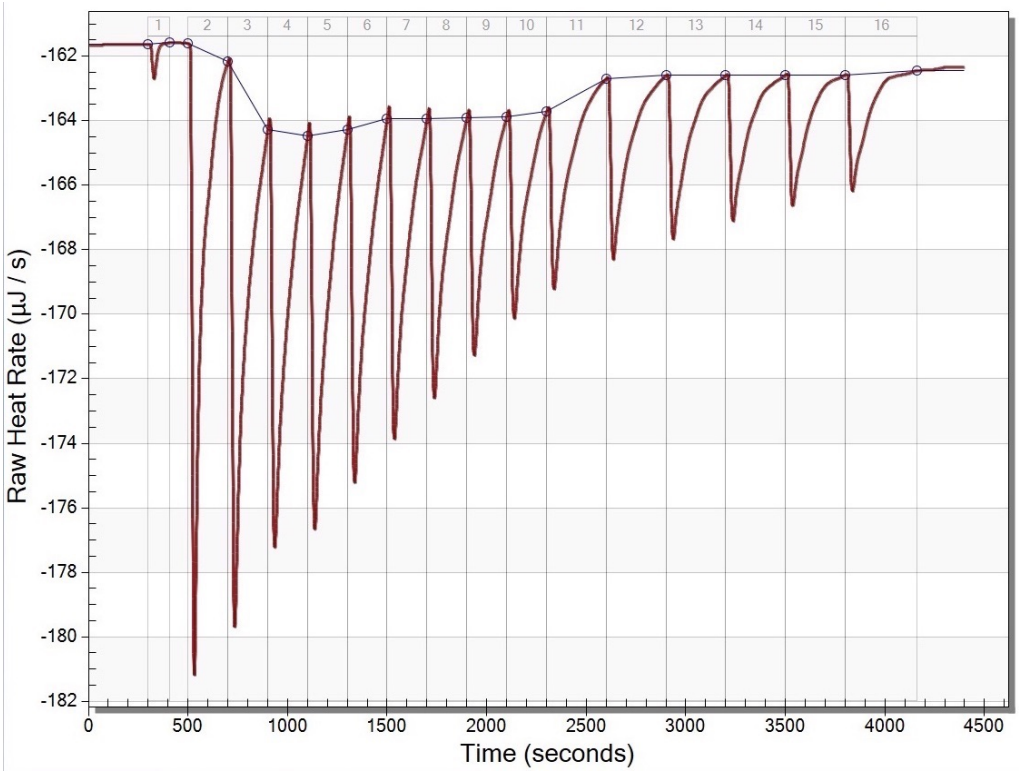
As per Table 2.3.2

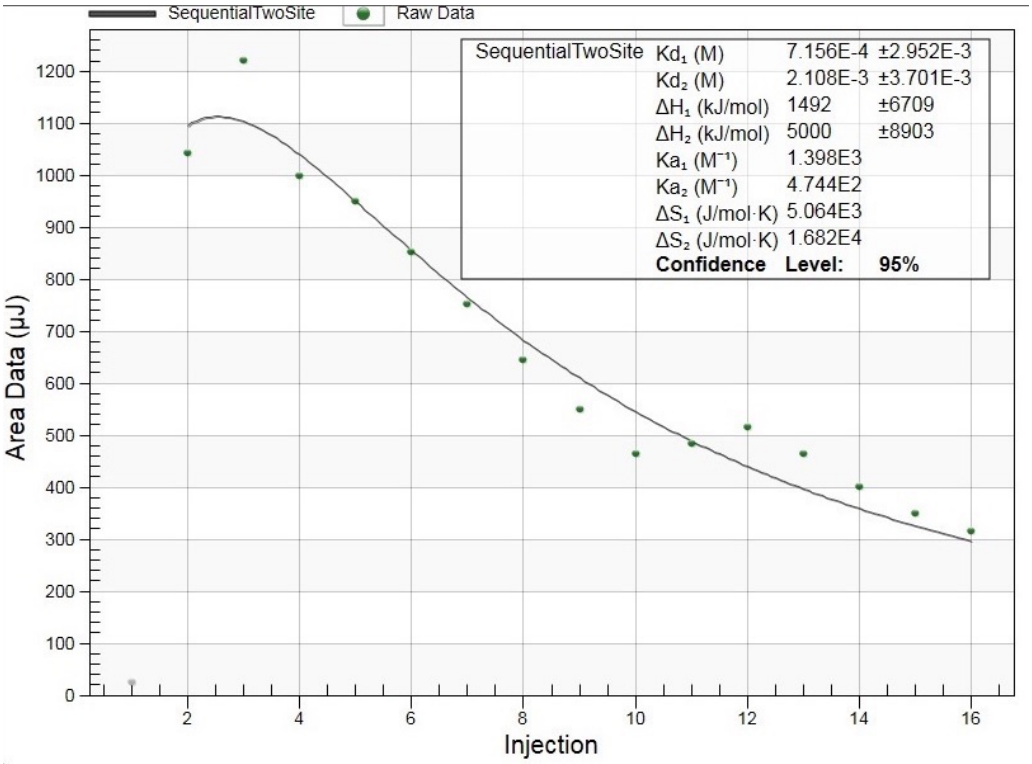
Run 1: Speed 300 s, $FliD_{cj}$ concentration in cell 0.02 mM, Heparin concentration in syringe 16.6 mM, Injection length 200 s, volume 1.5 µl, number of 25.



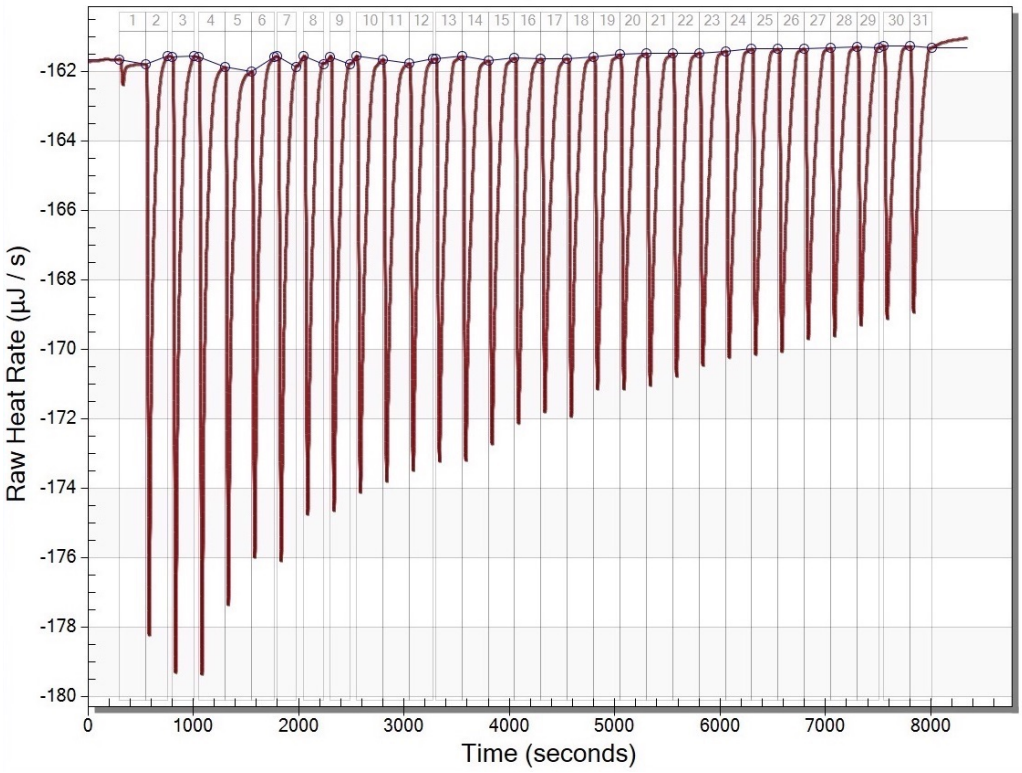


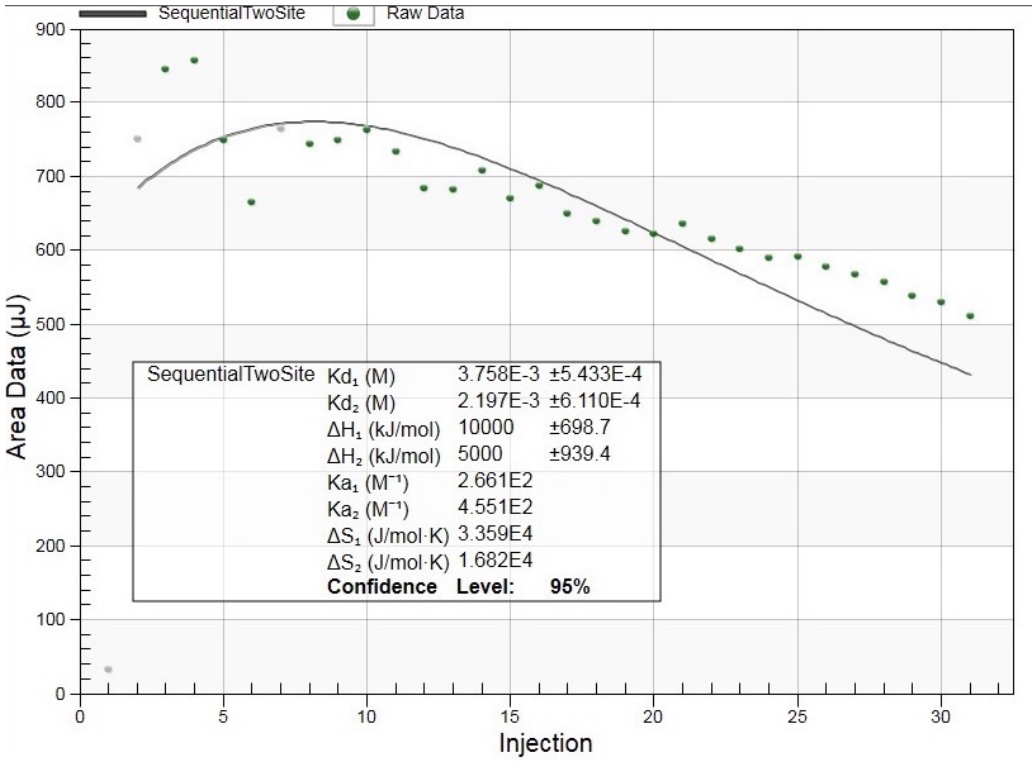
Run 2: Speed 300 s, $FliD_{cj}$ concentration in cell 0.01 mM, Heparin concentration in syringe 16.6 mM, Injection length 200 s, volume 3 µl, number of 16.



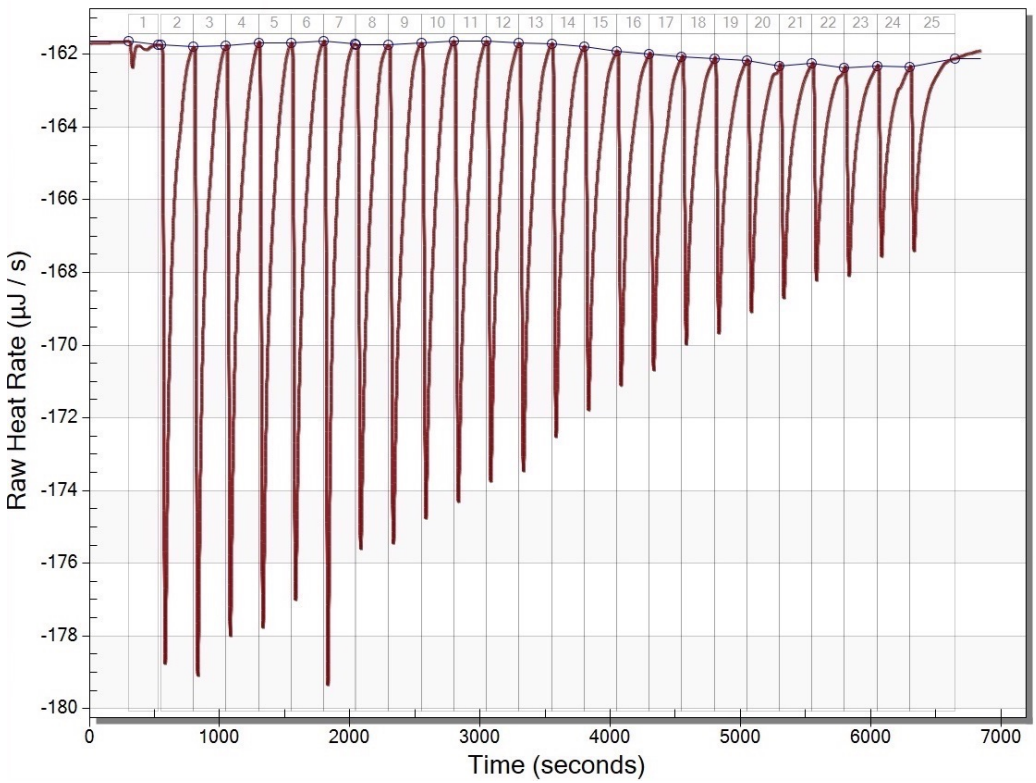


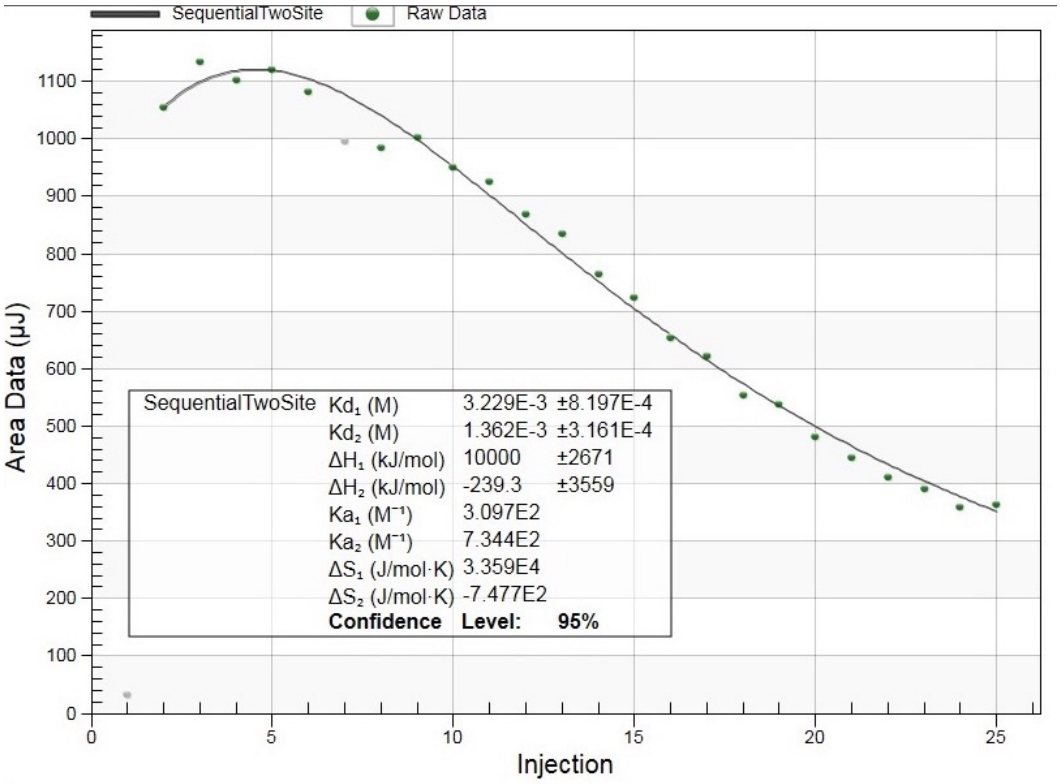
Run 3: Speed 300 s, $FliD_{cj}$ concentration in cell 0.01 mM, Heparin concentration in syringe 16.6 mM, Injection length 250 s, volume 1.5 µl, number of 31.





Run 4: Speed 300 s, $FliD_{cj}$ concentration in cell 0.01 mM, Heparin concentration in syringe 16.6 mM, Injection length 250 s, volume 2 µl, number of 25.

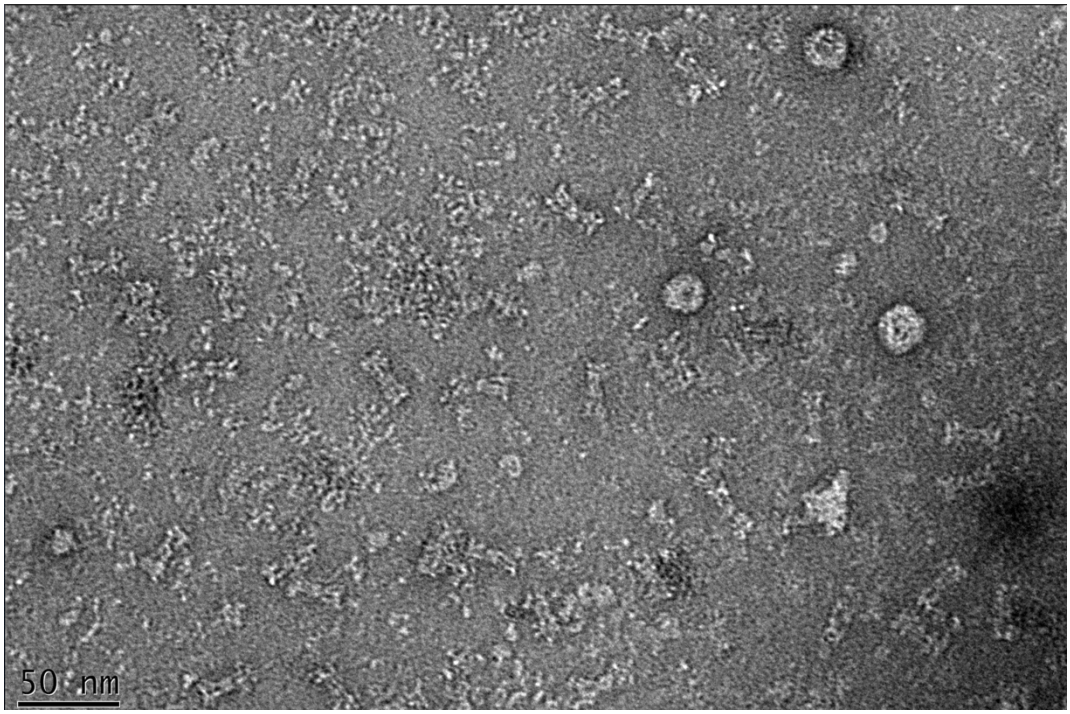



SequentialTwoSite
Raw Data
Area Data (µJ)
Injection
SequentialTwoSite
Kd1 (M) 3.229E-3 ±8.197E-4
Kd2 (M) 1.362E-3 ±3.161E-4
ΔH1 (kJ/mol) 10000 ±2671
ΔH2 (kJ/mol) -239.3 ±3559
Ka1 (M⁻¹) 3.097E2
Ka2 (M⁻¹) 7.344E2
ΔS1 (J/mol·K) 3.359E4
ΔS2 (J/mol·K) -7.477E2
Confidence Level: 95%

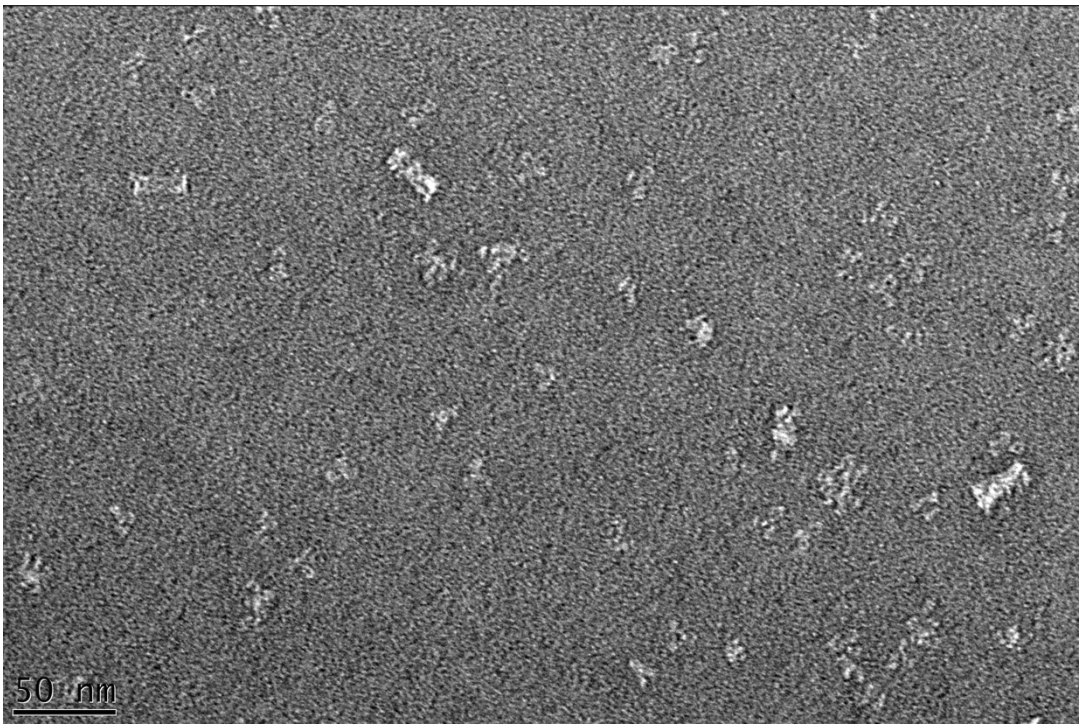
APPENDIX 21

IgG:

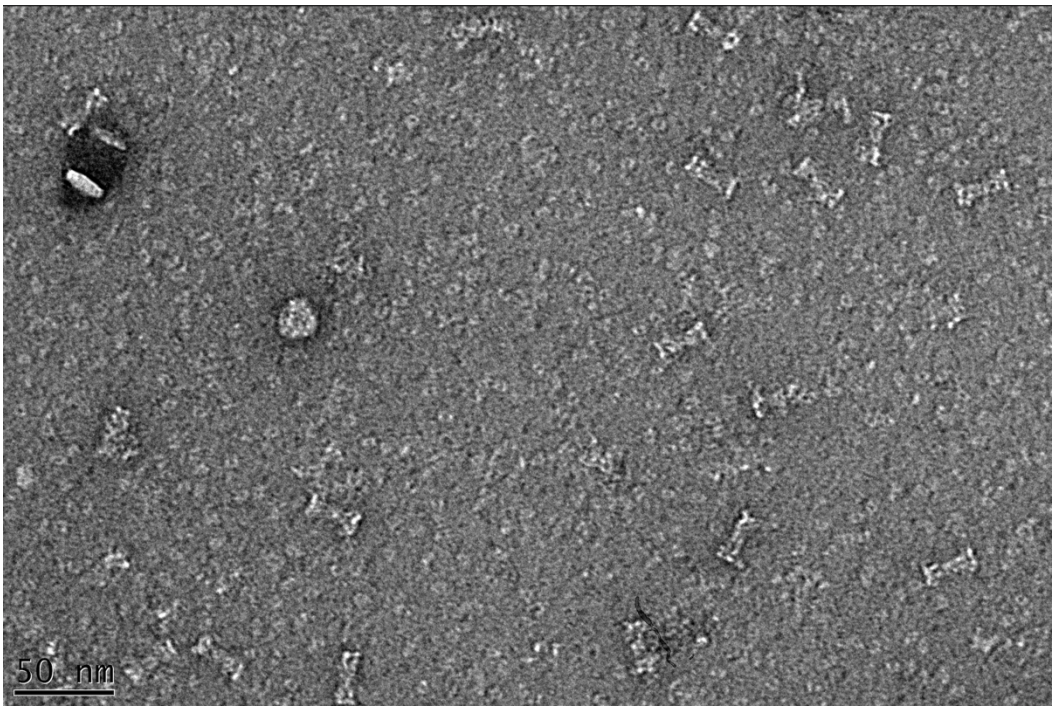
CAA1 Full IgG



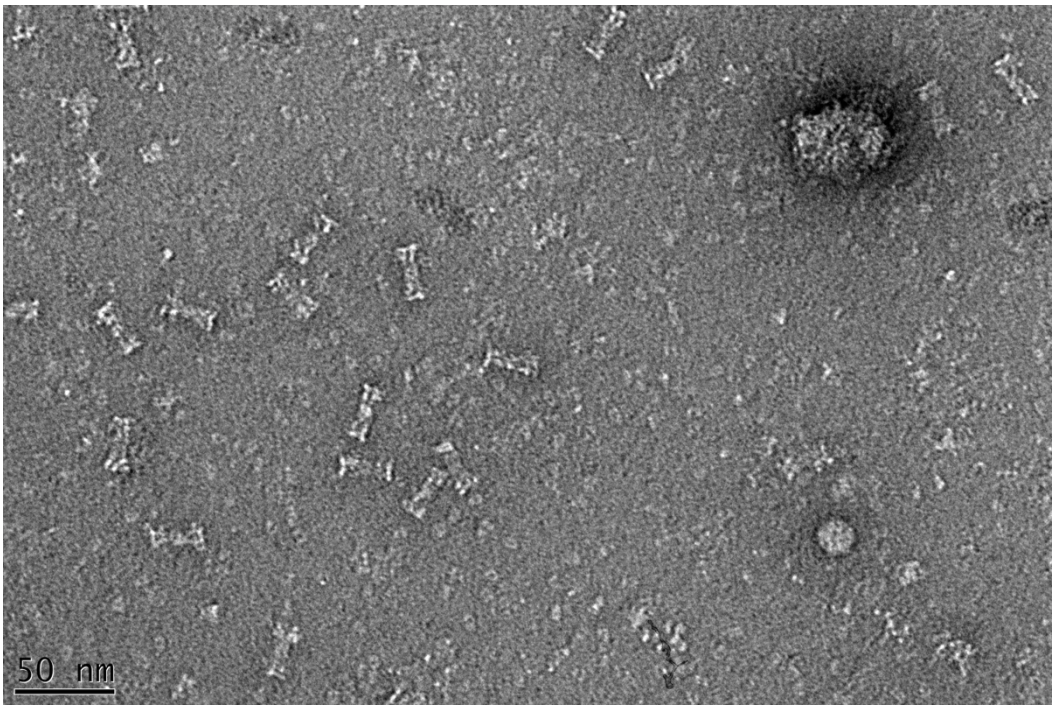
CCG4 Full IgG



CAA1 rFab

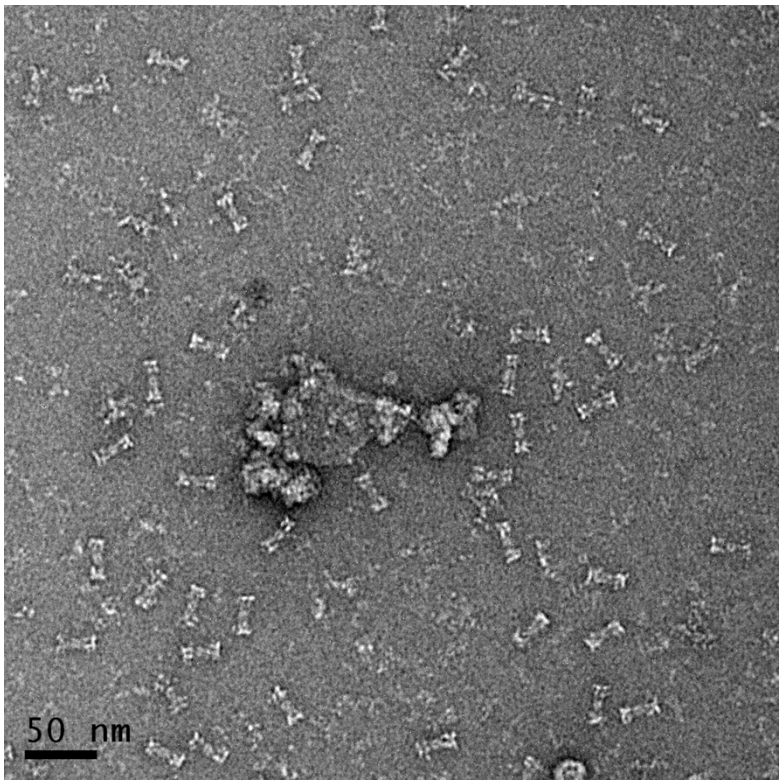


CCG4 rFab

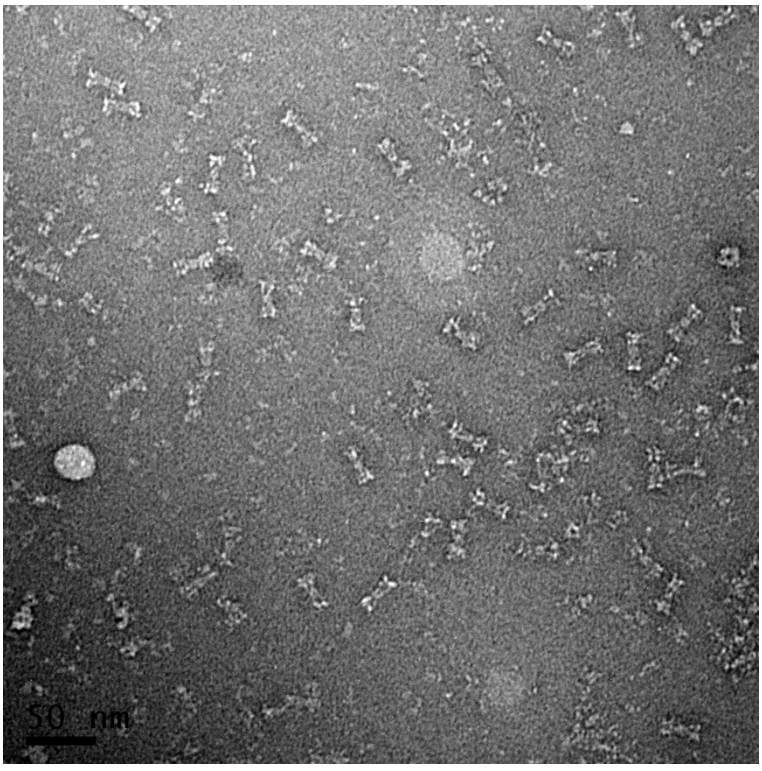


IgA:

CAA1 SlgA1



CCG4 SlgA2



REFERENCES

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

Adler, J. (1966). Chemotaxis in Bacteria Motile Escherichia coli migrate in bands that are. Science (80-.). *153*, 708–716.

Armitage, J.P., and Berry, R.M. (2020). Assembly and Dynamics of the Bacterial Flagellum. Annu. Rev. Microbiol. *74*, 181–200.

Atsumi, T., McCarter, L., and Imae, Y. (1992). Polar and lateral flagellar motors of marine Vibrio are driven by different ion-motive forces. Nature *355*, 1385–1387.

Backert, S., and Hofreuter, D. (2013). Molecular methods to investigate adhesion, transmigration, invasion and intracellular survival of the foodborne pathogen campylobacter jejuni. J. Microbiol. Methods *95*, 8–23.

Baker, L.A., and Rubinstein, J.L. (2010). Radiation damage in electron cryomicroscopy. Methods Enzymol. *481*, 371–388.

Bammes, B.E., Rochat, R.H., Jakana, J., and Chiu, W. (2011). Practical performance evaluation of a 10k×10k CCD for electron cryo-microscopy. J. Struct. Biol. *175*, 384–393.

Beeby, M. (2015). Motility in the epsilon-proteobacteria. Curr. Opin. Microbiol. *28*, 115–121.

Beeby, M., Ribardo, D.A., Brennan, C.A., Ruby, E.G., Jensen, G.J., and Hendrixson, D.R. (2016). Diverse high-torque bacterial flagellar motors assemble wider stator rings using a conserved protein scaffold. Proc. Natl. Acad. Sci. U. S. A. *113*, E2759.

Beeby, M., Ferreira, J.L., Tripp, P., Albers, S.V., and Mitchell, D.R. (2020). Propulsive nanomachines: The convergent evolution of archaella, flagella and cilia.

Berg, Howard C; Brown, D.A. (1972). Chemotaxis in E coli analysed by 3D tracking. Nature *239*, 500–504.

Bergeron, J.R.C., Fernández, L., Wasney, G.A., Vuckovic, M., Reffuveille, F., Hancock, R.E.W., and Strynadka, N.C.J. (2016). The Structure of a type 3 secretion system (T3SS) ruler protein suggests a molecular mechanism for needle length sensing. J. Biol. Chem. *291*, 1676–1691.

Blocker, A.J., Deane, J.E., Veenendaal, A.K.J., Roversi, P., Hodgkinson, J.L., Johnson, S., and Lea, S.M. (2008). What's the point of the type III secretion system needle? Proc. Natl. Acad. Sci. U. S. A. *105*, 6507–6513.

Blum, T.B., Filippidou, S., Fatton, M., Junier, P., and Abrahams, J.P. (2019). The wild-type flagellar filament of the Firmicute Kurthia at 2.8 Å resolution in vivo. Sci. Rep. *9*, 1–8.

Bochner, B.R., and Savageau, M.A. (1977). Generalized indicator plate for genetic, metabolic, and taxonomic studies with microorganisms. Appl. Environ. Microbiol. *33*, 434–444.

Brilot, A.F., Chen, J.Z., Cheng, A., Pan, J., Harrison, S.C., Potter, C.S., Carragher, B., Henderson, R., and Grigorieff, N. (2012). Beam-induced motion of vitrified specimen on holey carbon film. J. Struct. Biol. *177*, 630–637.

Bulieris, P. V., Shaikh, N.H., Freddolino, P.L., and Samatey, F.A. (2017). Structure of FlgK reveals the divergence of the bacterial Hook-Filament Junction of Campylobacter. Sci. Rep. *7*, 15743.

Carroni, M., and Saibil, H.R. (2016). Cryo electron microscopy to determine the structure of macromolecular complexes. Methods *95*, 78–85.

Chaban, B., Hughes, H.V., and Beeby, M. (2015). The flagellum in bacterial pathogens: For motility and a whole lot more. Semin. Cell Dev. Biol. *46*, 91–103.

Chang, Y., Zhang, K., Carroll, B.L., Zhao, X., Charon, N.W., Norris, S.J., Motaleb, M.A., Li, C., and Liu, J. (2020). Molecular mechanism for rotational switching of the bacterial flagellar motor. Nat. Struct. Mol. Biol. *27*, 1041–1047.

Chang, Y.W., Rettberg, L.A., Treuner-Lange, A., Iwasa, J., Søgaard-Andersen, L., and Jensen, G.J. (2016). Architecture of the type IVa pilus machine. Science (80-.). *351*.

Chen, S., Beeby, M., Murphy, G.E., Leadbetter, J.R., Hendrixson, D.R., Briegel, A., Li, Z., Shi, J., Tocheva, E.I., Müller, A., et al. (2011). Structural diversity of bacterial flagellar motors. EMBO J. *30*, 2972–2981.

Cheng, Y. (2015). Single-particle Cryo-EM at crystallographic resolution. Cell *161*, 450–457.

Cho, S.Y., Song, W.S., Hong, H.J., Lee, G.S., Kang, S.G., Ko, H.J., Kim, P.H., and Yoon, S. il (2017). Tetrameric structure of the flagellar cap protein FliD from Serratia marcescens. Biochem. Biophys. Res. Commun. *489*, 63–69.

Cho, S.Y., Song, W.S., Oh, H.B., Kim, H. ul, Jung, H.S., and Yoon, S. il (2019a). Structural analysis of the flagellar capping protein FliD from Helicobacter pylori. Biochem. Biophys. Res. Commun. *514*, 98–104.

Cho, S.Y., Song, W.S., and Yoon, S. il (2019b). Crystal structure of the flagellar cap protein FliD from Bdellovibrio bacteriovorus. Biochem. Biophys. Res. Commun. *519*, 652–658.

Christie, P.J. (2019). The Rich Tapestry of Bacterial Protein Translocation Systems. Protein J. *38*, 389–408.

Cohen-Bazire, G., and London, J. (1967). Basal organelles of bacterial flagella. J. Bacteriol. *94*, 458–465.

Cohen, E.J., Nakane, D., Kabata, Y., Hendrixson, D.R., Nishizaka, T., and Beeby, M. (2020). Campylobacter jejuni motility integrates specialized cell shape, flagellar filament, and motor, to coordinate action of its opposed flagella. PLoS Pathog. *16*, 1–24.

Cossart, P., and Sansonetti, P.J. (2004). Bacterial Invasion: The Paradigms of Enteroinvasive Pathogens. Science (80-.). *304*, 242–248.

Crine, S.L., and Acharya, K.R. (2021). Molecular basis of C-mannosylation – a structural

perspective. FEBS J. 1–18.

Dasti, J.I., Tareen, A.M., Lugert, R., Zautner, A.E., and Groß, U. (2010). Campylobacter jejuni: A brief overview on pathogenicity-associated factors and disease-mediating mechanisms. Int. J. Med. Microbiol. *300*, 205–211.

Deme, J.C., Johnson, S., Vickery, O., Muellbauer, A., Monkhouse, H., Griffiths, T., James, R.H., Berks, B.C., Coulton, J.W., Stansfeld, P.J., et al. (2020). Structures of the stator complex that drives rotation of the bacterial flagellum. Nat. Microbiol. *5*, 1553–1564.

Denise, R., Abby, S.S., and Rocha, E.P.C. (2019). Diversification of the type IV filament superfamily into machines for adhesion, protein secretion, DNA uptake, and motility.

Diószeghy, Z., Závodszky, P., Namba, K., and Vonderviszt, F. (2004). Stabilization of flagellar filaments by HAP2 capping. FEBS Lett. *568*, 105–109.

Dubochet, J., Adrian, M., Chang, J.-J., Homo, J., Lepault, J., McDowall, A.W., and Schultz, P. (1988). Cryo-electron microscopy of vitrified specimens. Q. Rev. Biophys. *21*, 129–228.

Eckhard, U., Bandukwala, H., Mansfield, M.J., Marino, G., Cheng, J., Wallace, I., Holyoak, T., Charles, T.C., Austin, J., Overall, C.M., et al. (2017). Discovery of a proteolytic flagellin family in diverse bacterial phyla that assembles enzymatically active flagella. Nat. Commun. *8*, 521.

Eisenbach, M. (2011). Bacterial Chemotaxis. In ELS, (John Wiley & Sons, Ltf: Chichester), pp. 1–17.

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

Epps, S.V.R., Harvey, R.B., Hume, M.E., Phillips, T.D., Anderson, R.C., and Nisbet, D.J. (2013). Foodborne Campylobacter: Infections, metabolism, pathogenesis and reservoirs. Int. J. Environ. Res. Public Health *10*, 6292–6304.

Erhardt, M., Singer, H.M., Wee, D.H., Keener, J.P., and Hughes, K.T. (2011). An infrequent molecular ruler controls flagellar hook length in Salmonella enterica. EMBO

J. *30*, 2948–2961.

Ewing, C.P., Andreishcheva, E., and Guerry, P. (2009). Functional characterization of flagellin glycosylation in Campylobacter jejuni 81-176. J. Bacteriol. *191*, 7086–7093.

Faïs, T., Delmas, J., Serres, A., Bonnet, R., and Dalmasso, G. (2016). Impact of CDT toxin on human diseases. Toxins (Basel). *8*, 18–20.

Ferreira, J.L., Coleman, I., Addison, M.L., Zachs, T., Quigley, B.L., Wuichet, K., and Beeby, M. (2021). The "Jack-of-all-Trades" Flagellum From Salmonella and E. coli Was Horizontally Acquired From an Ancestral β-Proteobacterium. Front. Microbiol. *12*, 1–8.

Frank, J. (2006). Three-Dimensional Electron Microscopy of Macromolecular Assemblies (New York: Oxford University Press).

Franken, L.E., Grünewald, K., Boekema, E.J., and Stuart, M.C.A. (2020). A Technical Introduction to Transmission Electron Microscopy for Soft-Matter: Imaging, Possibilities, Choices, and Technical Developments. Small *16*.

Freitag, C.M., Strijbis, K., and van Putten, J.P.M. (2017). Host cell binding of the flagellar tip protein of Campylobacter jejuni. Cell. Microbiol. *19*, 1–13.

Frenz, B., Walls, A.C., Egelman, E.H., Veesler, D., and Di Maio, F. (2017). RosettaES: A sampling strategy enabling automated interpretation of difficult cryo-EMmaps. Nat. Methods *14*, 797–800.

Fujii, T., Matsunami, H., Inoue, Y., and Namba, K. (2018). Evidence for the hook supercoiling mechanism of the bacterial flagellum. Biophys. Physicobiology *15*, 28–32.

Furukawa, Y., Imada, K., Vonderviszt, F., Matsunami, H., Sano, K.I., Kutsukake, K., and Namba, K. (2002). Interactions between bacterial flagellar axial proteins in their monomeric state in solution. J. Mol. Biol. *318*, 889–900.

Galkin, V.E., Yu, X., Bielnicki, J., Heuser, J., Ewing, C.P., Guerry, P., and Egelman, E.H. (2008). Divergence of Quaternary Structures Among Bacterial Flagellar Filaments. Science (80-.). *320*, 382–385.

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M.R., Appel, R.D., and Bairoch, A. (2005). Protein Identification and Analysis Tools on the ExPASy Server. Proteomics Protoc. Handb. 571–608.

Gavel, Y., and Heijne, G. Von (1990). Sequence differences between glycosylated and non-glycosylated asn-x-thr/ser acceptor sites: Implications for protein engineering. Protein Eng. Des. Sel. *3*, 433–442.

Ghasemi, A., Mohammad, N., Mautner, J., Karsabet, M.T., Ardjmand, A., and Moniri, R. (2018). Immunization with recombinant FliD confers protection against Helicobacter pylori infection in mice. Mol. Immunol. *94*, 176–182.

Gibson, K.H., Trajtenberg, F., Wunder, E.A., Brady, M.R., San Martin, F., Mechaly, A., Shang, Z., Liu, J., Picardeau, M., Ko, A., et al. (2020). An asymmetric sheath controls flagellar supercoiling and motility in the leptospira spirochete. Elife *9*, 1–24.

Gourlay, L.J., Thomas, R.J., Peri, C., Conchillo-Solé, O., Ferrer-Navarro, M., Nithichanon, A., Vila, J., Daura, X., Lertmemongkolchai, G., Titball, R., et al. (2015). From crystal structure to in silico epitope discovery in the Burkholderia pseudomallei flagellar hook-associated protein FlgK. FEBS J. *282*, 1319–1333.

Grant, T., Rohou, A., and Grigorieff, N. (2018). CisTEM, user-friendly software for single-particle image processing. Elife *7*, 1–24.

Griffith, A.R., Rogers, C.J., Miller, G.M., Abrol, R., Hsieh-Wilson, L.C., and Goddard, W.A. (2017). Predicting glycosaminoglycan surface protein interactions and implications for studying axonal growth. Proc. Natl. Acad. Sci. U. S. A. *114*, 13697–13702.

Gulbronson, C.J., Ribardo, D.A., Balaban, M., Knauer, C., Bange, G., and Hendrixson, D.R. (2016). FlhG employs diverse intrinsic domains and influences FlhF GTPase activity to numerically regulate polar flagellar biogenesis in Campylobacter jejuni. Mol. Microbiol. *99*, 291–306.

Hajra, D., Nair, A.V., and Chakravortty, D. (2021). An elegant nano-injection machinery for sabotaging the host: Role of Type III secretion system in virulence of different human and animal pathogenic bacteria. Phys. Life Rev. *38*, 25–54.

Halte, M., and Erhardt, M. (2021). Protein export via the type iii secretion system of the bacterial flagellum. Biomolecules *11*, 1–19.

Hong, H.J., Kim, T.H., Song, W.S., Ko, H.J., Lee, G.S., Kang, S.G., Kim, P.H., and Yoon, S. il (2018). Crystal structure of FlgL and its implications for flagellar assembly. Sci. Rep. *8*, 1–11.

Horvath, P., Kato, T., Miyata, T., and Namba, K. (2019). Structure of Salmonella Flagellar Hook Reveals Intermolecular Domain Interactions for the Universal Joint Function. MDPI Biomol. *9*, 1–13.

Hu, J., Worrall, L.J., Vuckovic, M., Hong, C., Deng, W., Atkinson, C.E., Brett Finlay, B., Yu, Z., and Strynadka, N.C.J. (2019). T3S injectisome needle complex structures in four distinct states reveal the basis of membrane coupling and assembly. Nat. Microbiol. *4*, 2010–2019.

Hughes, R. (2004). Campylobacter jejuni in Guillain-Barré syndrome. Neurology *3*, 644.

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

Imada, K., Vonderviszt, F., Furukawa, Y., Oosawa, K., and Namba, K. (1998). Assembly characteristics of flagellar cap protein HAP2 of Salmonella: Decamer and pentamer in the pH-sensitive equilibrium. J. Mol. Biol. *277*, 883–891.

Inaba, S., Hashimoto, M., Jyot, J., and Aizawa, S.I. (2013). Exchangeability of the flagellin (FliC) and the cap protein (FliD) among different species in flagellar assembly. Biopolymers *99*, 63–72.

Inoue, T., Barker, C.S., Matsunami, H., Aizawa, S.I., and Samatey, F.A. (2018). The FlaG regulator is involved in length control of the polar flagella of Campylobacter jejuni. Microbiol. (United Kingdom) *164*, 740–750.

Ito, K.I., Nakamura, S., and Toyabe, S. (2021). Cooperative stator assembly of bacterial flagellar motor mediated by rotation. Nat. Commun. *12*, 1–7.

Iwashkiw, J.A., Vozza, N.F., Kinsella, R.L., and Feldman, M.F. (2013). Pour some sugar

on it: The expanding world of bacterial protein O-linked glycosylation. Mol. Microbiol. *89*, 14–28.

Jacobsen, T., Bardiaux, B., Francetic, O., Izadi-Pruneyre, N., and Nilges, M. (2020). Structure and function of minor pilins of type IV pili. Med. Microbiol. Immunol. *209*, 301–308.

Jarrell, K.F., and McBride, M.J. (2008). The surprisingly diverse ways that prokaryotes move. Nat. Rev. Microbiol. *6*, 466–476.

Johnson, S., Fong, Y.H., Deme, J.C., Furlong, E.J., Kuhlen, L., and Lea, S.M. (2020). Symmetry mismatch in the MS-ring of the bacterial flagellar rotor explains the structural coordination of secretion and rotation. Nat. Microbiol. *5*, 966–975.

Johnson, S., Furlong, E.J., Deme, J.C., Nord, A.L., Caesar, J., Chevance, F.F.V., Berry, R.M., Hughes, K.T., and Lea, S.M. (2021). Molecular structure of the intact bacterial flagellar basal body. Nat. Microbiol. *6*, 712–721.

Kanbe, M., Yagasaki, J., Zehner, S., Göttfert, M., and Aizawa, S.I. (2007). Characterization of two sets of subpolar flagella in Bradyrhizobium japonicum. J. Bacteriol. *189*, 1083–1089.

Kato, T., Makino, F., Miyata, T., and Horváth, P. (2019). Structure of the native supercoiled flagellar hook as a universal joint. 1–17.

Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N., and Sternberg, M.J.E. (2015). The Phyre2 web portal for protein modeling, prediction and analysis. Nat. Protoc. *10*, 845–858.

Khanra, N., Rossi, P., Economou, A., and Kalodimos, C.G. (2016). Recognition and targeting mechanisms by chaperones in flagellum assembly and operation. Proc. Natl. Acad. Sci. *113*, 9798–9803.

Kim, H.J., Yoo, W., Jin, K.S., Ryu, S., and Lee, H.H. (2017). The role of the FliD C-terminal domain in pentamer formation and interaction with FliT. Sci. Rep. *7*, 1–11.

Knoll, M., and Ruska, E. (1932). Das Elektronenmikroskop. Zeitschrift Für Phys. *78*, 318–

339.

Kovács, N., Jankovics, H., and Vonderviszt, F. (2018). Deletion analysis of the flagellum-specific secretion signal in Salmonella flagellin. FEBS Lett. *592*, 3074–3081.

Kreutzberger, M.A.B., Ewing, C., Poly, F., Wang, F., and Egelman, E.H. (2020). Atomic structure of the Campylobacter jejuni flagellar filament reveals how e Proteobacteria escaped Toll-like receptor 5 surveillance. Proc. Natl. Acad. Sci. U. S. A. *117*, 16985–16991.

Kühlbrandt, W. (2014). Continuing the resolution revolution. Science (80-.). *343*.

van Leeuwenhoek, A. (1677). Observation, communicated to the publisher by Mr Anthony van Leewenhoeck, in a Dutch letter of the 9 Octob. 1676 here English'd: concerning little animals by him observed in rain-well-sea and snow water; as also in water wherein pepper had lain infused. Phil. Trans. R. Soc. *12*, 821–831.

Lertsethtakarn, P., Ottemann, K.M., and Hendrixson, D.R. (2011). Motility and chemotaxis in campylobacter and helicobacter. Annu. Rev. Microbiol. *65*, 389–410.

Li, H., and Sourjik, V. (2011). Assembly and stability of flagellar motor in Escherichia coli. Mol. Microbiol. *80*, 886–899.

Li, X., Ren, F., Cai, G., Huang, P., Chai, Q., Gundogdu, O., Jiao, X., and Huang, J. (2020). Investigating the Role of FlhF Identifies Novel Interactions With Genes Involved in Flagellar Synthesis in Campylobacter jejuni. Front. Microbiol. *11*, 1–11.

Lima, M.A., Cavalheiro, R.P., Meneghetti, M.C.Z., Nobre, L.T.D.B., Viana, G.M., Yates, E.A., and Nader, H.B. (2015). Heparan sulfate and heparin: Structure and function. Sulfated Polysaccharides 1–16.

Loconte, V., Kekez, I., Matković-Čalogović, D., and Zanotti, G. (2017). Structural characterization of FlgE2 protein from Helicobacter pylori hook. FEBS J. *284*, 4328–4342.

Lunelli, M., Hurwitz, R., Lambers, J., and Kolbe, M. (2011). Crystal structure of prgi-sipd: Insight into a secretion competent state of the type three secretion system needle tip

and its interaction with host ligands. PLoS Pathog. *7*, e1002163.

Lynch, M.J., Miller, M., James, M., Zhang, S., Zhang, K., Li, C., Charon, N.W., and Crane, B.R. (2019). Structure and chemistry of lysinoalanine crosslinking in the spirochaete flagella hook. Nat. Chem. Biol. *15*, 959–965.

De Maayer, P., and Cowan, D.A. (2016). Flashy flagella: Flagellin modification is relatively common and highly versatile among the Enterobacteriaceae. BMC Genomics *17*, 1–13.

Majander, K., Anton, L., Antikainen, J., Lång, H., Brummer, M., Korhonen, T.K., and Westerlund-Wikström, B. (2005). Extracellular secretion of polypeptides using a modified Escherichia coli flagellar secretion apparatus. Nat. Biotechnol. *23*, 475–481.

Maki-Yonekura, S., Yonekura, K., and Namba, K. (2003). Domain movements of HAP2 in the cap-filament complex formation and growth process of the bacterial flagellum. Proc. Natl. Acad. Sci. U. S. A. *100*, 15528–15533.

Maki-Yonekura, S., Yonekura, K., and Namba, K. (2010). Conformational change of flagellin for polymorphic supercoiling of the flagellar filament. Nat. Struct. Mol. Biol. *17*, 417–422.

Maki, S., Vonderviszt, F., Furukawa, Y., Imada, K., and Namba, K. (1998). Plugging interactions of HAP2 pentamer into the distal end of flagellar filament revealed by electuon microscopy. J. Mol. Biol. *277*, 771–777.

Matsunami, H., Barker, C.S., Yoon, Y.H., Wolf, M., and Samatey, F.A. (2016). Complete structure of the bacterial flagellar hook reveals extensive set of stabilizing interactions. Nat. Commun. *7*, 1–10.

McCallum, M., Burrows, L.L., and Howell, P.L. (2019). The Dynamic Structures of the Type IV Pilus. Microbiol. Spectr. *7*, 22–27.

Mendoza-Barberá, E., Merino, S., and Tomás, J. (2021). Surface Glucan Structures in Aeromonas spp. Mar. Drugs *19*, 649.

Milne-Davies, B., Wimmi, S., and Diepold, A. (2021). Adaptivity and dynamics in type III

secretion systems. Mol. Microbiol. *115*, 395–411.

Minamino, T., Imada, K., and Namba, K. (2008). Mechanisms of type III protein export for bacterial flagellar assembly. Mol. Biosyst. *4*, 1105–1115.

Minamino, T., Morimoto, Y. V., Kinoshita, M., and Namba, K. (2021). Multiple Roles of Flagellar Export Chaperones for Efficient and Robust Flagellar Filament Formation in Salmonella. Front. Microbiol. *12*.

Montemayor, E.J., Ploscariu, N.T., Sanchez, J.C., Parrell, D., Sillard, R.S., Shebelut, C.W., Ke, Z., Guerrero-ferreira, R.C., and Wright, R. (2021). Flagellar Structures from the Bacterium Caulobacter crescentus and Implications for Phage /CbK Predation of Multiflagellin Bacteria. J. Bacteriol. *203*, 1–20.

Motojima, F. (2015). How do chaperonins fold protein? Biophys. *11*, 93–102.

Naskar, S., Hohl, M., Tassinari, M., and Low, H.H. (2021). The structure and mechanism of the bacterial type II secretion system. Mol. Microbiol. *115*, 412–424.

Neal-McKinney, J.M., and Konkel, M.E. (2012). The Campylobacter jejuni CiaC virulence protein is secreted from the flagellum and delivered to the cytosol of host cells. Front. Cell. Infect. Microbiol. *2*, 31.

Nithichanon, A., Rinchai, D., Gori, A., Lassaux, P., Peri, C., Conchillio-Solé, O., Ferrer-Navarro, M., Gourlay, L.J., Nardini, M., Vila, J., et al. (2015). Sequence-and structure-based immunoreactive epitope discovery for Burkholderia pseudomallei flagellin. PLoS Negl. Trop. Dis. *9*, 1–20.

Nogales, E., and Scheres, S.H.W. (2015). Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity. Mol. Cell *58*, 677–689.

Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-coffee: A novel method for fast and accurate multiple sequence alignment. J. Mol. Biol. *302*, 205–217.

Paradis, G., Chevance, F.F.V., Liou, W., Renault, T.T., Hughes, K.T., Rainville, S., and Erhardt, M. (2017). Variability in bacterial flagella re-growth patterns after breakage. Sci. Rep. *7*, 1–10.

Perruzza, L., Jaconi, S., Lombardo, G., Pinna, D., Strati, F., Morone, D., Seehusen, F., Hu, Y., Bajoria, S., Xiong, J., et al. (2020). Prophylactic Activity of Orally Administered FliD-Reactive Monoclonal SIgA Against Campylobacter Infection. Front. Immunol. *11*, 1–16.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., and Ferrin, T.E. (2004). UCSF Chimera - A visualization system for exploratory research and analysis. J. Comput. Chem. *25*, 1605–1612.

Pierce, M.M., Raman, C.S., and Nall, B.T. (1999). Isothermal titration calorimetry of protein-protein interactions. Methods A Companion to Methods Enzymol. *19*, 213–221.

Poly, F., and Guerry, P. (2008). Pathogenesis of Campylobacter infection. Curr. Opin. Gastrointerology *24*, 27–31.

Postel, S., Deredge, D., Bonsor, D.A., Yu, X., Diederichs, K., Helmsing, S., Vromen, A., Friedler, A., Hust, M., Egelman, E.H., et al. (2016). Bacterial flagellar capping proteins adopt diverse oligomeric states. Elife *5*, 1–20.

Punjani, A., Rubinstein, J.L., Fleet, D.J., and Brubaker, M.A. (2017). CryoSPARC: Algorithms for rapid unsupervised cryo-EM structure determination. Nat. Methods *14*, 290–296.

van Putten, J.P.M., Alphen, L.B. Van, Wo, M.M.S.M., and Zoete, M.R. De (2009). Molecular Mechanisms of Campylobacter Infection. C. Sasakawa, ed. (Springer-Verlag Berlin Heidelberg), pp. 197–229.

Radermacher, M., Wagenknecht, T., Verschoor, A., and Frank, J. (1987). Three-dimensional reconstruction from a single-exposure, random conical tilt series applied to the 50S ribosomal subunit of Escherichia coli. J. Microsc. *146*, 113–136.

Ren, F., Lei, T., Song, Z., Yu, T., Li, Q., Huang, J., and Jiao, X. an (2018). Could FlhF be a key element that controls Campylobacter jejuni flagella biosynthesis in the initial assembly stage? Microbiol. Res. *207*, 240–248.

Renault, T.T., Abraham, A.O., Bergmiller, T., Paradis, G., Rainville, S., Charpentier, E., Guet, C.C., Tu, Y., Namba, K., Keener, J.P., et al. (2017). Bacterial flagella grow through

an injection-diffusion mechanism. Elife *6*, 1–22.

Reverberi, R., and Reverberi, L. (2007). Factors affecting the antigen-antibody reaction. Blood Transfus. *5*, 227–240.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. *42*, 320–324.

Rohou, A., and Grigorieff, N. (2015). CTFFIND4: Fast and accurate defocus estimation from electron micrographs. J. Struct. Biol. *192*, 216–221.

Rosenthal, P.B. (2015). From high symmetry to high resolution in biological electron microscopy: A commentary on Crowther (1971) 'Procedures for Three-Dimensional reconstruction of spherical viruses by Fourier synthesis from electron micrographs.' Philos. Trans. R. Soc. B Biol. Sci. *370*.

Rossez, Y., Wolfson, E.B., Holmes, A., Gally, D.L., and Holden, N.J. (2015). Bacterial flagella: twist and stick, or dodge across the kingdoms. PLoS Pathog. *11*, e1004483.

Saijo-Hamano, Y., Matsunami, H., Namba, K., and Imada, K. (2019). Architecture of the Bacterial Flagellar Distal Rod and Hook of Salmonella. Biomolecules *9*, 260.

Sal, M.S., Li, C., Motalab, M.A., Shibata, S., Aizawa, S.I., and Charon, N.W. (2008). Borrelia burgdorferi uniquely regulates its motility genes and has an intricate flagellar hook-basal body structure. J. Bacteriol. *190*, 1912–1921.

Salah Ud-Din, A.I.M., and Roujeinikova, A. (2018). Flagellin glycosylation with pseudaminic acid in Campylobacter and Helicobacter: prospects for development of novel therapeutics. Cell. Mol. Life Sci. *75*, 1163–1178.

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

Samatey, F.A., Matsunami, H., Imada, K., Nagashima, S., Shaikh, T.R., Thomas, D.R., Chen, J.Z., DeRosier, D.J., Kitao, A., and Namba, K. (2004). Structure of the bacterial flagellar hook and implication for the molecular universal joint mechanism. Nature

431, 1062–1068.

Samuelson, D.R., Eucker, T.P., Bell, J.A., Dybas, L., Mansfield, L.S., and Konkel, M.E. (2013). The Campylobacter jejuni CiaD effector protein activates MAP kinase signaling pathways and is required for the development of disease. Cell Commun. Signal. *11*, 1.

Santiveri, M., Roa-Eguiara, A., Kühne, C., Wadhwa, N., Hu, H., Berg, H.C., Erhardt, M., and Taylor, N.M.I. (2020). Structure and Function of Stator Units of the Bacterial Flagellar Motor. Cell *183*, 244-257.e16.

Sarkar, A., and Desai, U.R. (2015). A simple method for discovering druggable, specific glycosaminoglycan-protein systems. Elucidation of key principles from heparin/heparan sulfate-binding proteins. PLoS One *10*, 1–18.

Scheres, S.H. w. (2014). Beam-induced motion correction for sub-megadalton cryo-EM particles. Elife *3*, e03665.

Scheres, S.H.W. (2012). RELION: Implementation of a Bayesian approach to cryo-EM structure determination. J. Struct. Biol. *180*, 519–530.

Scheres, S.H.W., and Chen, S. (2012). Prevention of overfitting in cryo-EM structure determination. Nat. Methods *9*, 853–854.

Schwarz, F., and Aebi, M. (2011). Mechanisms and principles of N-linked protein glycosylation. Curr. Opin. Struct. Biol. *21*, 576–582.

Shibata, S., Matsunami, H., Aizawa, S.I., and Wolf, M. (2019). Torque transmission mechanism of the curved bacterial flagellar hook revealed by cryo-EM. Nat. Struct. Mol. Biol. *26*, 941–945.

Snijder, J., Borst, A.J., Dosey, A., Walls, A.C., Burrell, A., Reddy, V.S., Kollman, J.M., and Veesler, D. (2017). Vitrification after multiple rounds of sample application and blotting improves particle density on cryo-electron microscopy grids. J. Struct. Biol. *198*, 38–42.

Song, W.S., and Yoon, S. Il (2014). 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.

Song, W.S., Cho, S.Y., Hong, H.J., Park, S.C., and Yoon, S. (2017). Self-Oligomerizing Structure of the Flagellar Cap Protein FliD and Its Implication in Filament Assembly. J. Mol. Biol. *429*, 847–857.

Song, W.S., Hong, H.J., and Yoon, S. il (2020). Structural study of the flagellar junction protein FlgL from Legionella pneumophila. Biochem. Biophys. Res. Commun. *529*, 513–518.

Song, Y., Dimaio, F., Wang, R.Y.R., Kim, D., Miles, C., Brunette, T., Thompson, J., and Baker, D. (2013). High-resolution comparative modeling with RosettaCM. Structure *21*, 1735–1742.

Studier, F.W. (2005). Protein production by auto-induction in high density shaking cultures. Protein Expr. Purif. *41*, 207–234.

Tan, J., Zhang, X., Wang, X., Xu, C., Chang, S., Wu, H., Wang, T., Liang, H., Gao, H., Zhou, Y., et al. (2021). Structural basis of assembly and torque transmission of the bacterial flagellar motor. Cell *184*, 2665-2679.e19.

Taylor, K.A., and Glaeser, R.M. (1974). Electron Diffraction of Frozen , Hydrated Protein Crystals Sunday and Workday Variations in Photochemical Air Pollutants in New Jersey and New York. Science (80-.). *186*, 1036–1037.

Terwilliger, T.C., Sobolev, O. V., Afonine, P. V., and Adams, P.D. (2018). Automated map sharpening by maximization of detail and connectivity. Acta Crystallogr. Sect. D Struct. Biol. *74*, 545–559.

Thompson, R.F., Walker, M., Siebert, C.A., Muench, S.P., and Ranson, N.A. (2016). An introduction to sample preparation and imaging by cryo-electron microscopy for structural biology. Methods *100*, 3–15.

Thomson, N.M., Ferreira, J.L., Matthews-Palmer, T.R., Beeby, M., and Pallen, M.J. (2018). Giant flagellins form thick flagellar filaments in two species of marine γ-proteobacteria. PLoS One *13*, 1–12.

Varki, A., Cummings, R., and Esco, J. (1999). Glycosaminoglycan-binding Proteins. In Essentials of Glycobiology, (Cold spring harbor (NY): Cold Spring Harbor Laboratory

Press), p.

Vonderviszt, F., Imada, K., Furukawa, Y., Uedaira, H., Taniguchi, H., and Namba, K. (1998). Mechanism of self-association and filament capping by flagellar HAP2. J. Mol. Biol. *284*, 1399–1416.

Wakabayashi, K., Hotani, H., and Akaura, S. (1969). Polymerization of Salmonella Flagellin in the presence of High concentrations of salts. Biochim. Biophys. Acta *175*, 195–203.

Wang, F., Burrage, A.M., Postel, S., Clark, R.E., Orlova, A., Sundberg, E.J., Kearns, D.B., and Egelman, E.H. (2017). A structural model of flagellar filament switching across multiple bacterial species. Nat. Commun. *8*, 960.

Worrall, L.J., Lameignere, E., and Strynadka, N.C.J. (2011). Structural overview of the bacterial injectisome. Curr. Opin. Microbiol. *14*, 3–8.

Wösten, M.M.S.M., Van Dijk, L., Veenendaal, A.K.J., De Zoete, M.R., Bleumink-Pluijm, N.M.C., and Van Putten, J.P.M. (2010). Temperature-dependent FlgM/FliA complex formation regulates Campylobacter jejuni flagella length. Mol. Microbiol. *75*, 1577–1591.

Xing, Q., Shi, K., Portaliou, A., Rossi, P., Economou, A., and Kalodimos, C.G. (2018). Structures of chaperone-substrate complexes docked onto the export gate in a type III secretion system. Nat. Commun. 1–9.

Yamaguchi, T., Toma, S., Terahara, N., Miyata, T., Ashihara, M., Minamino, T., Namba, K., and Kato, T. (2020). Structural and functional comparison of salmonella flagellar filaments composed of FljB and FliC. Biomolecules *10*, 1–11.

Yonekura, K., Maki, S., Morgan, D.G., DeRosier, D.J., Vonderviszt, F., Imada, K., and Namba, K. (2000). The Bacterial Flagellar Cap as the Rotary Promoter of Flagellin Self-Assembly. Science (80-.). *290*, 2148–2152.

Yonekura, K., Maki-Yonekura, S., and Namba, K. (2003). Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. Nature *424*, 643–650.

Yoon, Y.H., Barker, C.S., Bulieris, P. V., Matsunami, H., and Samatey, F.A. (2016). Structural insights into bacterial flagellar hooks similarities and specificities. Sci. Rep. *6*, 1–11.

Young, K.T., Davis, L.M., and DiRita, V.J. (2007). Campylobacter jejuni: Molecular biology and pathogenesis. Nat. Rev. Microbiol. *5*, 665–679.

Zeng, H., Wang, S., Zhou, T., Zhao, F., Li, X., Wu, Q., and Xu, J. (2018). ComplexContact: A web server for inter-protein contact prediction using deep learning. Nucleic Acids Res. *46*, W432–W437.

Zhang, F., Lee, K.B., and Linhardt, R.J. (2015). SPR biosensor probing the interactions between TIMP-3 and heparin/GAGs. Biosensors *5*, 500–512.

Zhang, K., Qin, Z., Chang, Y., Liu, J., Malkowski, M.G., Shipa, S., Li, L., Qiu, W., Zhang, J.R., and Li, C. (2019). Analysis of a flagellar filament cap mutant reveals that HtrA serine protease degrades unfolded flagellin protein in the periplasm of Borrelia burgdorferi. Mol. Microbiol. *111*, 1652–1670.

Zheng, S.Q., Palovcak, E., Armache, J.P., Verba, K.A., Cheng, Y., and Agard, D.A. (2017). MotionCor2: Anisotropic correction of beam-induced motion for improved cryo-electron microscopy. Nat. Methods *14*, 331–332.

Zhu, S., Nishikino, T., Hu, B., Kojima, S., Homma, M., and Liu, J. (2017). Molecular architecture of the sheathed polar flagellum in Vibrio alginolyticus. Proc. Natl. Acad. Sci. U. S. A. *114*, 10966–10971.

De Zoete, M.R., Keestra, A.M., Wagenaar, J.A., and Van Putten, J.P.M. (2010). Reconstitution of a functional toll-like receptor 5 binding site in Campylobacter jejuni flagellin. J. Biol. Chem. *285*, 12149–12158.