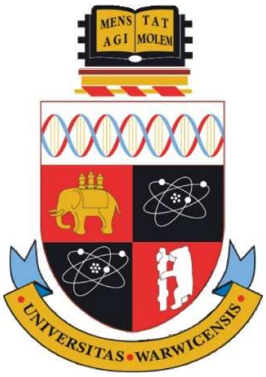




# Structural studies of the FliG protein of the bacterial flagellar rotor

**Dagnija Tupiņa**

Thesis submitted for the degree of Doctor of Philosophy in Molecular Biomedicine

University of Warwick, Warwick Medical School

September, 2022

# Table of contents

# List of figures

# List of tables

# Acknowledgements

Firstly, I would like to thank both of my supervisors Dr. Chrystala Constantinidou in University of Warwick and Dr. Peter John Bond in Bioinformatics Institute in A*STAR for their resilience in supervising my PhD both in person and through remote communication means throughout one of the most complicated times in recent world history while carrying burdens in each of our individual lives.

I thank Chrystala for her understanding, patience and kindness in her both attitudes to work and people. She helped me settle in a country previously unknown to me in my first study year and tackle some of the challenges a student faces in that time, as well for warmly receiving me back upon my return from Singapore for the final year, in the aftermath of perhaps the most challenging events in my life.

I am grateful to Pete for showing an excellent example of leadership that inspires respect without a need for a power distance and maintaining a high performing scientific group, while letting its members keep their passion for science unextinguished and integrating their interests and talents in the goals of the group, and for showing example that it is possible for a group leader to run a large group without losing touch with each of its projects and members.

My particular thanks to my day-to-day supervisor Dr. Alexander Krah who had the daunting task of getting me from a mainly laboratory background up to speed in a cutting-edge computational field and did a great job and set a great example of integrity and commitment towards scientific work and teaching. I would also like to thank members of Pete's group and their spouses who were with me in Singapore during the dark Covid-19 pandemic times – Alex, Jan, Lorena, Alister, Aishwary, Raghu, Pari, Julie and Eadelin and of course, Pete himself – for being the closest thing to a family I had for 1 year and 9 months spent in Singapore non-stop and for all the in person (particularly BBQs) and virtual gatherings.

My utmost gratitude to my mom and dad for tirelessly supporting me in all the mentally and emotionally difficult times and to my dad for providing me with a PC that proved crucial in finishing up the computational work during my final year. Thanks to my grandfather for the encouragement throughout the years as he has been very excited to have a scientist in the family from the very start of my studies.

*Dedicated to my grandmother Ausma who sadly did not see me return from my long travels in pursuit of science and education, but who kindled the spirit of curiousity in me by teaching me the passion for books and reading.*

# Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. The work carried out (data generated and analysed) is my own work and has not been submitted in any previous application for any degree in this or any other university. Parts of this work have been published in my article (Tupiņa et al., 2022). Work presented in this thesis has been performed by the author except in parts stated below:

The simulations of *A. aeolicus* FliG structure and corresponding RMSD analysis was done by Dr. Alexander Krah. Simulations of isolated $5TDY_{bent}$ and $5TDY_{ext}$ and corresponding RMSD analysis was done by Dr. Alexander Krah.

MDFF runs in NAMD of 3 copies of FliF-C/FliG/FliM-M and FliG/FliM-M complexes were performed by Dr. Jan K. Marzinek.

The script to generate 34-mer ring assembly from single input structure was written by Adam Moverley and edited by Lorena Zuzic.

CryoET data collection in Leicaster on Thermo Fisher Titan Krios G3 microscope was performed by Dr. Christos Savva. Microscope EOL 2100Plus during sample optimization tests was operated by Dr. Saskia Bakker. Flagellated cell counting on negative stain grids in EOL 2100Plus microscope was done by Dr. Saskia Bakker.

Some of the primers, plasmids and strains used in work with *C. jejuni* have been designed and created by Emily Stoakes and Charlotte Dixon.

MultiQC and analysis with VarScan2 of the sequencing data was performed by Dr. Chrystala Constantinidou.

Simulations were performed on both A*STAR Bioinformatics institute cluster resources by core funds and the A*STAR Graduate Academy (A*GA) and on resources of the National Supercomputing Centre, Singapore (https://www.nscc.sg).

## Abstract

Flagella are complex multiprotein structures that not only enable bacteria to move through the environment, but are also key virulence factors for pathogens. The flagellar rotor that drives the entire flagellum, with rotation that can happen in both directions, is composed of the C-ring (built of proteins FliG, FliM and FliN) and MS-ring (built of the FliF protein). Interaction between FliG and FliF proteins and their higher-order complex formation is central to the bacterial flagellum biosynthesis. This is because FliF and its co-folding partners are among the earliest structures to assemble during the flagella building process, and because they physically link all the flagellar-associated structures in the cytoplasm with those in the periplasmic space (in Gram-negative bacteria), outer-membrane and finally the filament that resides outside the cell. Meanwhile, FliG functions to transfer the torque from the membrane anchored stators, transducing the energy from ion flow to rotation of the rotor and thus the entire flagellar structure. To expand our mechanistic understanding of this molecular machine, a combination of methods was employed here. Molecular dynamics simulations were used to study the structure of the FliG, revealing that the linker between the FliG N-terminal and middle domains likely adopts an extended conformation in vivo, in contrast with crystallographic data. An integrative modelling approach was then taken, encompassing homology modelling and molecular dynamics flexible fitting approaches, to create full length viable FliG models in a C-ring assembly. Furthermore, no high-resolution structure of the FliG ring has been solved to date; thus, two cryo-electron tomography datasets were collected to visualize motor reconstructions from *C. jejuni* (a pathogen and a flagella research model organism), and preliminary reconstructions were obtained. Construction of a *ΔfliF* strain and a set of complementation strains was created to study a function of previously identified pseudorevertant mutations in FliF that restored motility to non-motile *C. jejuni*. Whole genome sequencing and negative stain electron microscopy highlighted the importance of the two genes coding FliG and FliF to be co-transcribed and co-translated for an efficient motor assembly. Sequence analysis also identified a novel mutation in FliF associated with increased motility. Collectively, these methods expanded the knowledge of the bacterial flagellar rotor, as well as providing models to generate new hypotheses that lay groundwork for future experiments.

## Abbreviations

Apr – apramycin
Cat – chloramphenicol
CCDA – Campylobacter blood-free selective agar
CG – coarse grained
cryoEM – cryo-electron microscopy
cryoET – cryo-electron tomography
CW – clockwise
CCW – counter-clockwise
F – forward (primer)
fT3SS – flagellar type III secretion system
gDNA – genomic deoxy ribo nucleic acid
INDEL – insertion and deletion
LB – Loui Bertani
Kan – Kanamycin
MD – molecular dynamics
MDFF – Molecular Dynamics Flexible Fitting
MH – Mueller Hinton
MS-ring – membrane/supramembrane ring
mRNA – messenger ribo nucleic acid
NMR – nuclear magnetic resonance
OD – optical density
PBS – phosphate-buffered saline
PCR – polymerase hain reaction
PDB-ID – protein database
*pmf* – proton motive force
R – reverse (primer)
RBM – Ring binding motif
RMSD – root mean square deviation
RT– room temperature
SNP – single nucleotide polymorphism
T3SS – type III secretion system
WT – wild-type

# Chapter 1 – Introduction

## 1.1. Bacterial flagella – movement and virulence

Bacteria move through the environment with help of flagella, which are complex multiprotein structures propelled by a molecular motor.

Flagella have multiple functions in the cell, the primary of which is movement, but others include involvement in virulence factor secretion, adhesion, biofilm formation, chemotaxis and host colonization (Chaban et al., 2015). Studies with flagellation interrupting mutants in multiple species have shown the impediment of adhesion and colonization, but the degree to which motility is needed varies from only early stages of invasion to stages throughout the infection cycle (Josenhans and Suerbaum, 2002). Chemotaxis is closely linked to virulence as it enables bacteria to actively seek out favourable environments in the host and surroundings, by sensing chemicals and switching the rotation direction of the flagellum (Chaban et al., 2015); for example *Helicobacter pylori* has been shown to directly migrate towards ulcerated stomach tissue, and disruption of motility severely affected stomach tissue colonization and improved ulcer healing (Aihara et al., 2014).

Not all bacteria have flagella, but major pathogens such as *Enterobacteriacea, Helicobacter pylori, Campylobacter spp., Salmonellae, Pseudomonas aeruginosa* (Scallan et al., 2011; Vouga and Greub, 2016) are flagellated. In *Clostridium difficile* epidemic strains with flagella aid adherence during colonization (Baban et al., 2013).

Some species that are not in general flagellated, may include select strains that although under most conditions do not express flagella, under specific conditions they are motile. In select strains, such as *Bordetella pertussis* flagella may be expressed in response to fetal bovine serum (Hoffman et al., 2019). Condition dependent flagellation has also been observed in several *Shigella* strains and isolates (Girón, 1995).

Bacterial species that normally express flagella under most conditions can be described by the number of flagella and their placement into the following groups: monotrichous, peritrichous, amphitrichous and lophotrichous (Figure 1.1) (Schuhmacher et al., 2015).

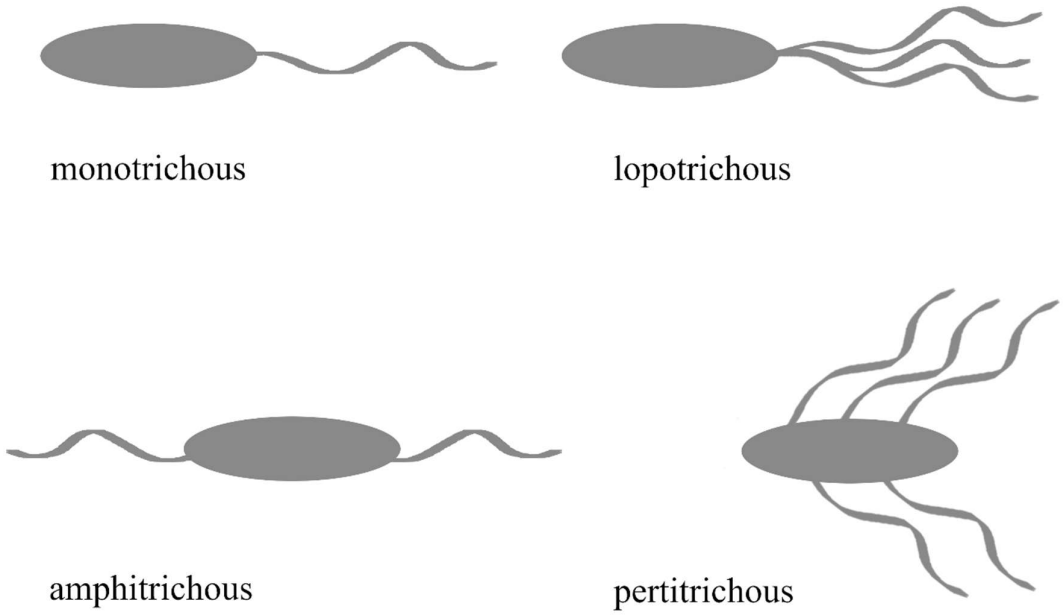



***Figure 1.1*. Types of flagellar placement in the bacterial cell.** Flagella placement on a bacterial cell can differ by location (poles or sides) or numbers (single or multiple).

Monotrichous bacteria have a single flagellum located on one of the cell poles, the most notable examples being *Caulobacter crescentus, Vibrio alginolyticus*, *Thermotoga maritima*. Similarly, lophotrichous bacteria have multiple flagella closely clustered on one cell pole, for example *Helicobacter* and *Aquifex aeolicus.* Amphitrichous bacteria have flagella on each pole, notably *Campylobacter* (Muller et al., 2014). Peritrichous bacteria have multiple flagella across the cell surface such as commonly researched organisms like *Escherichia coli*, *Salmonella enterica, Bacillus sp*. Exceptions of these main recognized categories exist as well, such as monotrichous *Schewanella* having an additional flagellum on the side of the cell or *Rhodobacter* having a single medial flagellum instead of placement at the cell pole, referred to as septal (Altegoer et al., 2014; Schuhmacher et al., 2015). An even more unusual flagella type is found in spirochetes, like *Borrelia burgdorferi*, that possess periplasmic flagella (Chang and Liu, 2019).

## 1.2. Structure and components of the bacterial flagellum

Major segments of the bacterial flagellum are the propeller, the molecular motor and in between them the joint, consisting of the hook and the rod (Terashima et al., 2008) (Figure 1.2). Flagellar components are the same in gram-negative and gram-positive bacteria where the only difference is that the latter do not have an outer membrane (Chevance and Hughes, 2008). Gram-negative bacteria have inner and outer membrane, and rod traverses the periplasmic space and outer membrane, while in gram-positive bacteria like *Bacillus subtilis* the rod is placed in the thick peptidoglycan layer above the single membrane the cell possesses (Sampriti and Kearns, 2016).

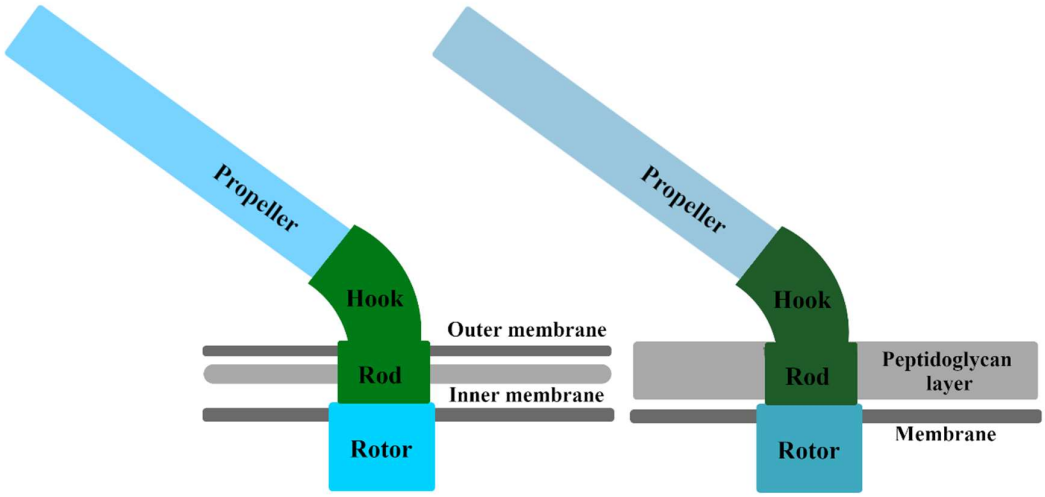



***Figure 1.2*. Parts of bacterial flagellum.** Main parts of a bacterial flagellum are the propeller outside the cell, rotor and joint consisting of hook and the rod. Left: In gram-negative bacteria rod is placed in the periplasmic space and outer membrane. Right: In gram positive bacteria the rod is in the peptidoglycan layer.

The propeller or filament is the outer part of the flagella that rotates, built of protein flagellin or its subforms (Morgan et al., 1995; Terashima et al., 2008). In Epsilonproteobacteria FlaA, major flagellin and FlaB, minor flagellin are examples of subforms (Josenhans and Suerbaum, 2002; Wassenaar et al., 1994). In other species, flagellin is referred to with different names, such as FliC in *E.coli* (Reid et al., 1999) and *S. enterica* (McQuiston et al., 2004) and the hag gene in *B. subtilis* (Mirel and Chamberlin, 1989). How bacteria ensure that the flagellum reaches its finite length is not clear and there are multiple competing models (Hughes, 2017). Two main mechanisms investigated are rates of polymerization, that occurs at the tip of the

flagellum (Evans et al., 2013) and the relationship between flagellin secretion and the *pmf* (proton motive force) driving the export (Renault et al., 2017). In *C. jejuni* FlaG has been identified as a length regulator due to its deletion increasing flagellar median length and interaction with FliA, the sigma 28 factor that controls flagellin expression (Inoue et al., 2018).

The flagellin filament is capped with FliD (Ikeda et al., 1996; Song et al., 2017). The presence of FliD is crucial for FliC assembly into the filament, and its deletion results in a complete lack of motility in *S. enterica* (Ikeda et al., 1996). This self-assembling protein can have different stoichiometries forming a pentameric structure in *S. enterica* and a hexameric one in *E. coli* (Song et al., 2017).

The hook and the rod traverse the cell wall. The hook, built mainly of FlgE, FlgK and FlgL lies between the bottom of the filament and the rod (Figure 1.3). The latter structure is a cylindrical structure with the L-ring (FlgH) above, followed by the distal rod consisting of FlgG, FlgF, then a P-ring (FlgI), and the proximal rod (FlgC, FlgB) (Terashima et al., 2008). Finally, the rod structure connects to the motor via FliE (Minamino et al., 2008; Terashima et al., 2008).

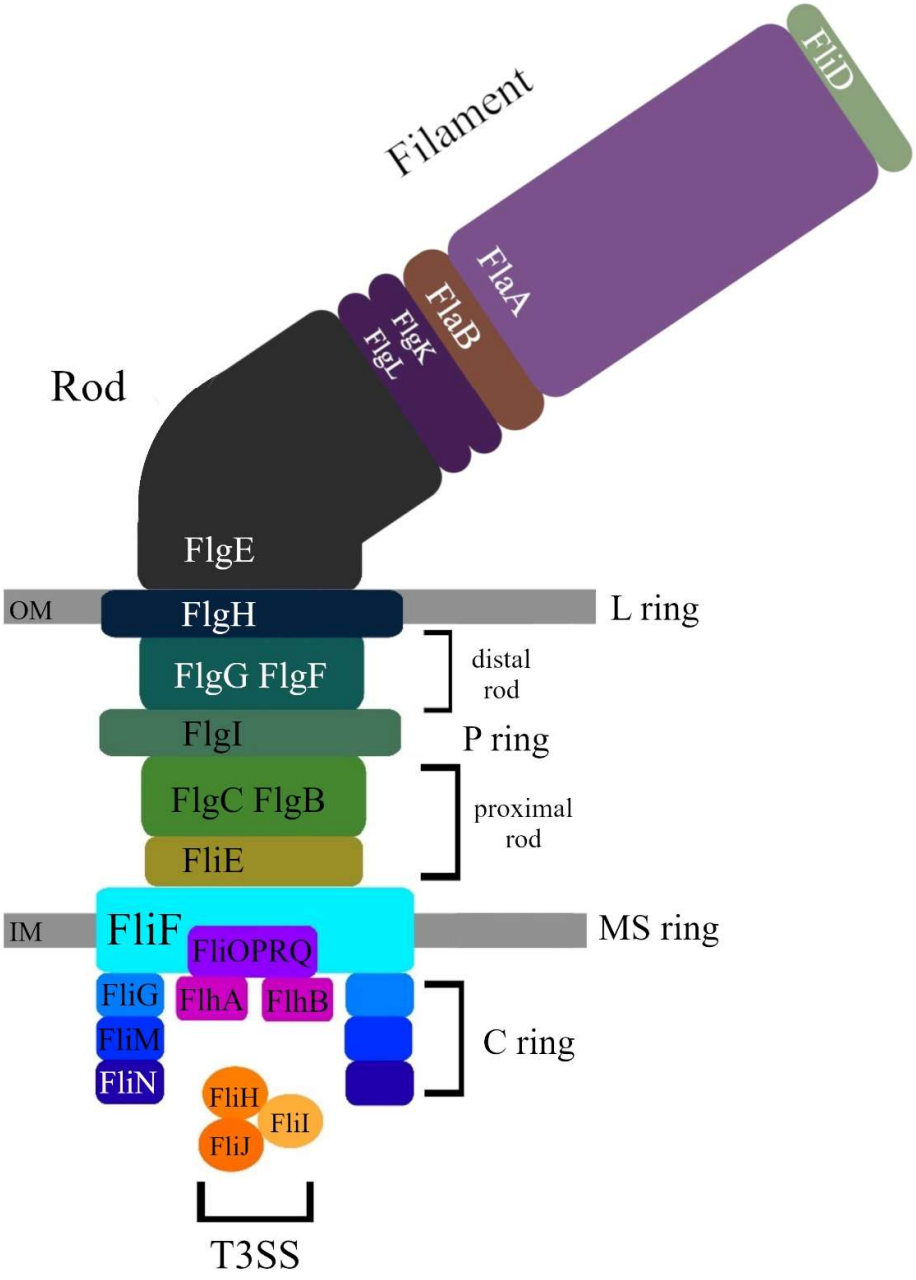



***Figure 1.3*. Flagellum structure in gram-negative bacteria**. The intracellular part of flagella is composed of the motor (MS ring, C ring and T3SS) and rod (proximal rod, P ring, distal rod, L ring), which connects to the hook (FlgE). Long horizontal grey lines denote approximate location of bacterial membranes. T3SS has membrane and MS ring embedded parts consisting of FliOPRQ, FlhA and FlhB and cytoplasmic components FliH, FliI, FliJ (not represented in actual stoichiometry). Top right: FlgK and FlgL join the filament, built of FlaB and FlaA (in *H. pylori* and *C. jejuni* or FliC in other species) with the hook (FlgE). Filament is capped by FliD.

The molecular motor can be divided in stator and rotor parts. The rotor is the fixed structure composed of multiple rings embedded in the inner membrane. Stators are small protein complexes that dynamically associate and disassociate from the rotor while delivering ion translocation energy driving the rotation (Terashima et al., 2008; Wadhwa et al., 2021; Zhou et al., 1998). Stators are formed of MotA and MotB proteins in a 5:2 ratio with the MotB protein surrounded by MotA (Santiveri et al., 2020). Translocation of a $H^+$ ion through the $MotA_5B_2$ channel induces a conformational change that serves as an energy source for driving the motor. At least three species have been shown to have a 5:2 stoichiometry (*C. jejuni*, *V. alginolyticus* and *Shewanella oneidensis*) and it is likely conserved (Santiveri et al., 2020). The number of stators associated with a rotor at a given time can change depending on the rotation speed required (Wadhwa et al., 2021) and need to increase force; e.g., encountering more viscous environment is met by recruitment of more stator units (Lele et al., 2013). The number of stators recruited to the rotor is also different among species (Beeby et al., 2016).

The rotor is built of the MS-ring, C-ring and a Type 3 Secretion System (T3SS)-like complex, also called flagellar T3SS (fT3SS), embedded in the MS-ring (Chevance and Hughes, 2008; Minamino et al., 2008). The MS-ring is built of a single protein FliF (Ueno et al., 1992) and mostly located on the outside of the inner membrane with two transmembrane domains traversing the membrane and a small portion of the protein interacting directly with the C-ring (Bergeron, 2016). The upper part of FliF faces the FliE protein in the joint (Kubori et al., 1997; Macnab, 2003). The fT3SS embedded in the MS-ring is built of FlhA, FlhB, FliO, FliP, FliQ and FliR proteins. It transiently interacts with a cytoplasmic complex built of FliI, FliJ and FliH proteins (Terashima et al., 2008) that ensure the transport of proteins secreted via the fT3SS that include building blocks of the flagellum itself (Minamino, 2014) as well as various other proteins, such as virulence factors and invasion antigen proteins (Chaban et al., 2015). For instance, while bacteria have multiple secretion systems, some species, like *C. jejuni*, do not have a separate T3SS system and its flagellar equivalent fulfils other secretion tasks, such as secretion of virulence factors (Konkel et al., 2004). A recent FliPQR structure shows a complex in a 5:4:1 ratio, although FliO was expressed but not observed in this complex (Kuhlen et al., 2018). FliH-C binds FliI-N in a 2:1 ratio and serves as a negative regulator of FliI ATPase activity (Imada et al., 2016; Lane et al.,

2006; Minamino and Macnab, 2000). FliH also interacts with FliN to guide the FliH-FliI complex towards the C-ring in *S. enterica* (McMurry et al., 2006), docking it at FlhA and FlhB in the T3SS-like complex (Minamino et al., 2009). FliJ binds FliI in a 1:6 ratio, resulting in a $FliH_{12}$-$FliI_6$-FliJ cytoplasmic complex (Minamino, 2014). Partial deletion of FliH, however, allows it to retain some transport and does not eliminate it completely (González-Pedrajo et al., 2002).

The C-ring is built of FliG, FliM and FliN (Noreen et al., 1994), and in some cases FliY, a subform of FliN found, for example, in *C. jejuni* (Henderson et al., 2020) *H. pylori, Clostridia* and various *Bacillus* species (Lowenthal et al., 2009). FliG interacts with the stator and transmits the torque to the upper flagellar structures (Lloyd et al., 1996; Minamino et al., 2011). FliM receives a signal for rotation-switch through binding of CheY, and transfers it to FliG (Minamino et al., 2019). CheY is a chemotaxis protein and the key component in transducing signals from positive attractants to changing flagella rotation direction and thus movement of the cell (Eisenbach, 1996). CheY activity is regulated by phosphorylation, performed by kinase CheA. The phosphorylated form (CheY-P) is binding FliM (Barak and Eisenbach, 1992; Welch et al., 1993).

FliN in the C-ring is involved in docking the cytoplasmic fT3SS components (McMurry et al., 2006) with MS-ring housed fT3SS, and also participates in binding of CheY (Sarkar et al., 2010).

In addition, polar bacterial motors contain additional anchoring structures that surround the stators and the rotor in the shape of periplasmic disks or scaffolds (Zhou and Roujeinikova, 2021). These structures differ between species much more significantly than motors themselves, being associated with the outer membrane, inner membrane, or both, and have different or as yet unidentified molecular components (Zhou and Roujeinikova, 2021).

## 1.3. Flagella biosynthesis: transcription and translation

The entire complex of the flagellum is resource costly for a bacterium to produce, and therefore flagella transcription initiation and downstream processes are tightly regulated.

*C. jejuni* and *H. pylori* both represent an interesting exception of the trend of motility-relevant genes being clustered in operons and instead these genes are interspersed throughout the chromosome (Josenhans and Suerbaum, 2002). In *C. jejuni,* the flagella and motor assemble quickly after division (Muller et al., 2014).

While the flagellar structure itself is fairly universal, flagellar gene transcriptional regulation among species is highly divergent. Mechanisms that guide the localization of flagellar components are also variable, from the landmark protein systems like TipN and TipF in *Caulobacter* to FlhF and FlhG that may have different functions depending on species (Aihara et al., 2014).

Temporal regulation of flagellar component production is achieved by transcribing them in three sequential groups termed class I, class II and class III, that are dependent on each other in a downstream manner, to ensure that the production of flagella building blocks occurs only once the structures they assemble on are completed (Chevance and Hughes, 2008). Prokaryotic RNA polymerases achieve specificity with σ factors, components recruited to control transcription of groups of genes, for example in response to stress (Kazmierczak et al., 2005).

In well studied model organisms like *S. enterica,* class I genes encode the master regulator FlhDC which in turn initiates class II gene transcription that is mediated by the σ70 transcription factor (Figure 1.4). Class II genes are structural genes that build the rotor, the T3SS-like system, the rod and the hook as well as regulators of class III genes, σ28 and FlgM (and anti-σ28 factor (Chevance and Hughes, 2008; Kazmierczak et al., 2005). Transcription of class III structural genes that build the propeller only initiates when these structures are completely assembled and fT3SS secretes FlgM out of the cell, releasing σ28 to transcribe flagellin and other class III proteins (Aldridge and Hughes, 2002). It is difficult, however, to draw any generalizations regarding flagellar gene transcription regulation in bacteria, as other widely used model organisms like *V. alginolytocus* and *C. crescentus* have vast differences across the entire process, including using different σ factors for different classes (Aldridge and Hughes, 2002).

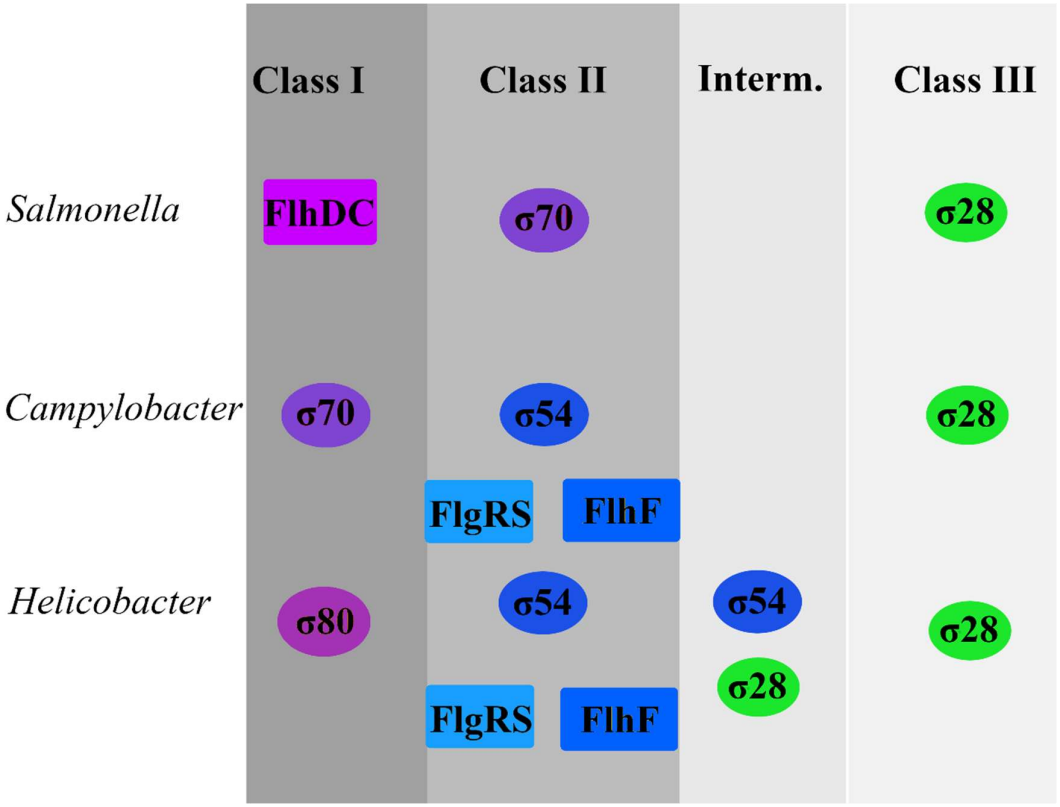



***Figure 1.4.*** **Transcriptional regulators and flagellar component classes.** Transcription of flagellar components is typically divided in three temporally separate classes, governed by transcriptional regulators and sigma factors, with some species having a distinct additional class.

*Campylobacter* and *Helicobacter* lack a class I master regulator and instead class I proteins are constitutively transcribed by σ70 or σ80, respectively (Kazmierczak et al., 2005; Niehus et al., 2004; Wösten et al., 2008) (Figure 1.4). In *C. jejuni*, the class II regulators are σ54 (RopN) and the FlgR/FlgS two component system, while the class III regulator is σ28, also known as FliA (Wösten et al., 2004). The FlgR/FlgS two-component system consists of a sensory kinase (FlgS) and response regulator (FlgR)) (Beier and Frank, 2000; Brahmachary et al., 2004; Spohn and Scarlato, 1999). FlgM, the anti-σ28 factors, is influenced by both class II and III regulators (Wösten et al., 2004). *C. jejuni* also has a regulator with an unknown mechanism, FlhF, that appears to be independent of the FlgR/FlgS pathway (Balaban et al., 2009; Ren et al., 2018). Deletion of *flhF* results in a lack of flagella and widely altered transcription landscape (Ren et al., 2018; Stoakes, 2017). This protein has been shown to be constitutively expressed, has a detectable GTPase activity and dimerizes (Balaban et al., 2009; Bange et al., 2007; Ren et al., 2018).

In *H. pylori* class II genes are transcribed by σ54 and regulated by the FlgRS system, and class III genes by σ28, although there is a wide group of genes that falls under the influence of both class II and III forming a separate intermediate class between the two, including FlhF itself. FlhA, the structural component of the T3SS-like system, and FlhF appear to be the two major regulators in *H. pylori*, where an FlgM/σ28 feedback loop is connected to FlhA and the positive regulation of flagella components, while FlhF has an uncharacterized parallel mechanism. Assembly of the MS-ring, C-ring and fT3SS is a checkpoint that leads to activation of RpoN (σ54) dependent genes, for example FlhA of the fT3SS has been shown to bind FlgS that can subsequently regulate the RpoN-dependent promoters (Tsang et al., 2015). A unique detail in *H. pylori* process is the specific involvement of FliH in the signal transduction and its link to the FlgRS system (Tsang and Hoover, 2015). Deletion of *Helicobacter flhF* also completely abolishes motility and flagella production (Niehus et al., 2004).

FlhF has been proposed to have not only functions in flagella assembly regulation, but also in directing FliF towards assembly in poles in *Vibrio cholerae* (Green et al., 2009), and in *P. aeruginosa*, where removal of this protein causes deviation from the usual flagella assembly at the cell pole and motility defects (Murray and Kazmierczak, 2006).

Besides spatiotemporally regulated transcription, an additional level of post-transcriptional control of flagellar components is present as well. FliW modulates the activity of CrsA, a translational regulator that binds mRNA and blocks translation of the FlaA protein, flagellin (Altegoer et al., 2016). This system in *C. jejuni* has been shown to be involved in ensuring flagellin is directed towards the cell poles and controls many other flagellar proteins besides flagellin itself (Dugar et al., 2016; Li et al., 2018).

In the light of continuously worsening antimicrobial resistance issue anti-virulence therapies as an alternative are a rapidly emerging field, and connection between motility and virulence comes into consideration (Martínez et al., 2019). Flagellar formation as a process has been considered as a drug target to prevent biofilm formation. In *Pseudomonas aeruginosa* and *Proteus mirabilis*, mupirocin has been shown to decrease flagellar formation and flagellin expression (Horii et al., 2003). From known flagellar structural components, FliD assembly has been considered as a potential drug target in *C. jejuni* (Chintoan-Uta et al., 2016) and in *H. pylori* (Mohammad et al., 2016). Swarming is a specific type of motility that refers

to collective movement of cells under stimulating conditions. To inhibit swarming behaviour multiple approaches have been considered, mainly around signalling, however targeting of stators as part of the motor is another promising approach (Rütschlin and Böttcher, 2020). In this context additional understanding of the FliG may facilitate targeting motor as source of motility with greater precision. A recent study screening for novel drug targets in *E. coli* pathogenic strains identified flagellar components FliN, FliH and MotB among proteins that co-evolve with virulence genes and could be considered as drug targets (Kaur et al., 2021), pointing to possibility that motor component regulation and assembly could be in the spot-light of novel antimicrobial design.

## 1.4. Assembly of the bacterial rotor

Assembly of the complex structure of flagellar rotor has been a long-time study target. Evidence differs regarding which component is the first to assemble: in *S. enterica* it is thought that the FliF ring serves as nucleus for other structures to assemble onto it (Morimoto et al., 2014), but in *E. coli* it has been found that fT3SS assembles first, followed by FliF (Li and Sourjik, 2011). Assembly of FliG and the C-ring closely follows FliF and T3SS, as FliF and FliG are known to bind each other in 1:1 ratio(Levenson et al., 2012; Lynch et al., 2017; Xue et al., 2018). In *H. pylori*, deletion of FliF results in reduced amount of FlhA associated with the membrane (Tsang and Hoover, 2015).

Until recently, evidence that FliG has a 33-34 unit stoichiometry whereas FliF has a 25-26 unit stoichiometry could not be reconciled (Thomas et al., 2001, 1999) with available binding data, until the solution of FliF ring structures in 2020 using cryo-electron microscopy (cryoEM) (Johnson et al., 2020; Kawamoto et al., 2021). The new data explained that the FliF ring has a matching sub-stoichiometry on the outside to that of FliG (33-34) as suggested by 1:1 binding data and a 21-23 unit sub-stoichiometry on the inside of the ring structure (Johnson et al., 2020; Kawamoto et al., 2021). It is worth noting that in both studies, there were particle populations with stoichiometries above and below these numbers, and the most represented particle population was used as the basis for the final structure. Whether stoichiometry variations occur within species *in vivo* or are an effect of sample preparation remains unclear, as it has been suggested that

construct truncation (Kawamoto et al., 2021) and lack of interactions with fT3SS (Johnson et al., 2020) may lead to assembly artifacts.

Folding of FliG has been shown to be co-dependent with FliF, so it is one of the earliest parts of the rotor to assemble (Morimoto et al., 2014). As of this moment there are two *S. enterica* FliF ring structures clearly showing either 33 or 34 subunit stoichiometry, and both of these are high resolution cryoEM structures (Johnson et al., 2020; Kawamoto et al., 2021). Several cryo-electron tomography (cryoET) structures of flagellar rotors from different species show different FliG stoichiometries, including 33/34 in *S. enterice* (Thomas et al., 2001, 1999), 34 in *Vibrio alginolyticus* (Carroll et al., 2020) and 46 in *Borrelia burgdorferi* (Chang et al., 2020). This shows that such stoichiometries can widely differ between species assuming that in the case of a larger FliG ring there would be a matching FliF ring. Early EM studies of *S. enterica* showed a distribution of 32-36 subunits with a Gaussian curve distribution making 34 the most commonly found variant (Young et al., 2003). Low resolution cryoET data for different species show varying diameters of the density matching the C-ring, implying these wider diameters are likely constructed from more units (Beeby et al., 2016; Qin et al., 2017). FliG structures are conserved and similar in length (with the difference of a short additional N-terminal helix) (Figure 4.2 (4.2.1)) (Xue et al., 2018) that further supports the likelihood that different diameters in species with as yet undetermined stoichiometry of the C-ring are likely achieved via variation in the number of subunits (Beeby et al., 2016) rather than differences in the FliG structure itself.

Point mutations and deletions in FliF of *Vibrio alginolyticus* and *Caulobacter crescentus* interfere with flagella formation through abolishing interaction with FliG (Grunenfelder et al., 2003; Ogawa et al., 2014). At the same time, while FliF-FliG interaction is crucial for the assembly, FliF-FliG fusion mutants in *S. enterica* that lack the FliG N-terminus and FliF C-terminus are still able to assemble functional rotors with a strong clockwise rotation bias, reduced diameter, and presumably reduced number of subunits (estimated to be 28 to 33 fold) (Sakai et al., 2019).

## 1.5. FliG and FliF – central components of the rotor

FliG is an entirely α-helical protein consisting of three distinct domains: the N-terminal domain (FliG-N), middle domain (FliG-M) and C-terminal domain (FliG-C) (Lee et al., 2010). Each of these domains contribute to the crucial functions of FliG in the rotor (Figure 1.5). The N-terminal domain binds FliF and thus bridges the structures above and below the inner membrane (Levenson et al., 2012; Lynch et al., 2017; Xue et al., 2018). The C-terminal domain interacts with stators and transfers the torque ensuring the energy for movement of the entire flagellar structure (Kojima and Blair, 2001; Lloyd et al., 1996; Minamino et al., 2019; Zhou et al., 1998). The FliG middle domain binds FliM (Lam et al., 2013; Vartanian et al., 2012), which binds FliN, thus linking FliG into the C-ring structure. The FliG middle domain is also crucial for rotation direction switching – mutations in the EHPQR motif in *Vibrio alginolyticus* create rotation-bias and rotation-lock mutants, and based on NMR and MD studies the EHPQR motif transfers the signal from CheY binding FliM to the conformational change inducing a directional switch (Nishikino et al., 2018).

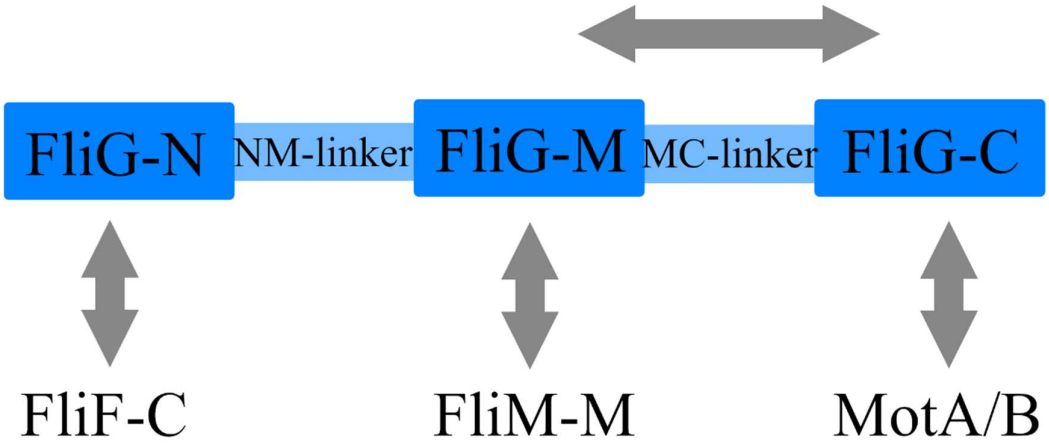



***Figure 1.5.*** **FliG domains and their involvement in interactions**. FliG-N binds FliF-C, FliG-M binds FliM-M and FliG-C binds the stator. FliG-M and FliG-C domains have intraprotein interactions.

FliF and FliG interactions have been captured in two crystal structures of FliG-N and FliF-C from two different species – *H. pylori* and *T. maritima* (PDB-ID: 5WUJ and PDB-ID: 5TDY respectively) (Lynch et al., 2017; Xue et al., 2018).

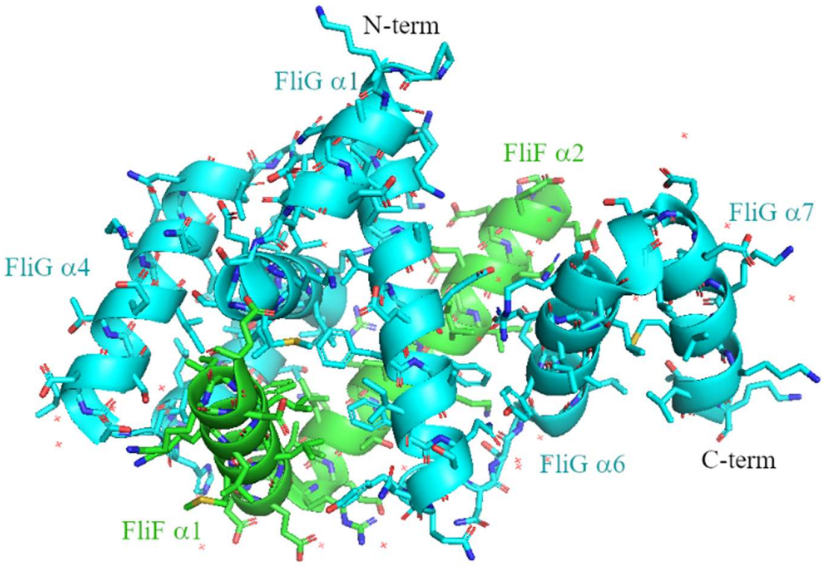



***Figure 1.6.*** **Crystal structure of *H.pylori* FliG N-terminus (cyan) co crystallized with FliF-C helices (green)**. FliG consists of 7 helices α1 to α7 and FliF has two resolved helices.

There is detailed structural information on all FliG domains available – full length unbound FliG in *A. aeolicus* (Lee et al., 2010), and FliG-C and FliG-M, as well as co-crystallized FliG-M and FliM-M in *T. maritima* and *H. pylori* from different authors (Brown et al., 2002; Lam et al., 2012, 2013; Minamino et al., 2011; Paul et al., 2011b; Sircar et al., 2015; Vartanian et al., 2012). All structured domains of FliG are entirely α-helical.

All proteins building the rotor and the flagellum itself are involved in interactions with ring structures above and below the ring in which the protein is in. FliF, the protein that builds all of the MS-ring, positioned above the inner membrane, has recently been shown to be the transition point of different stoichiometries within the system via high-resolution cryoET (Johnson et al., 2020; Kawamoto et al., 2021). The ring is built from isoforms of the same FliF chain that have different conformations, ensuring that the outer part of the ring has one stoichiometry (e.g. 34), compatible with the C-ring. The inner surface has a different stoichiometry that matches structures above it and houses the T3SS-like complex (fT3SS) inside it (Johnson et al., 2020; Kawamoto et al., 2021) (Figure 1.7). Serving as a scaffold for this secretion machinery is another crucial FliF function, which ensures that after rotor is assembled, components needed to build the remaining parts of the flagellum can be exported.

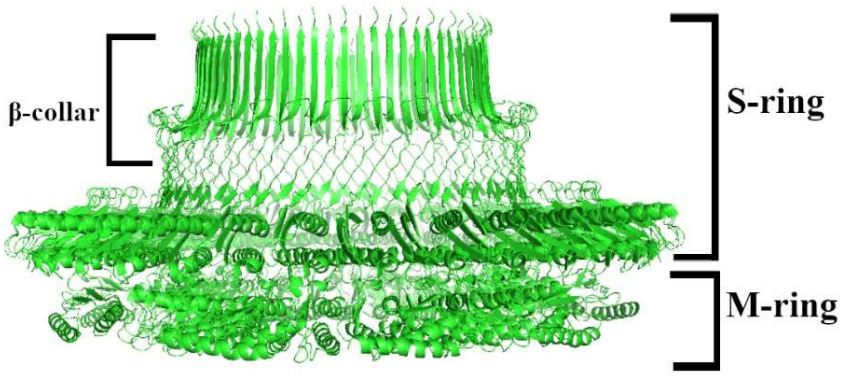



***Figure 1.7.*** **Structure of the FliF (MS) ring**. S-ring is above the inner membrane. M-ring faces the cytoplasm. Protein is mostly α-helical except for β-sheet containing ring towards the hook structures.

FliF has both α-helical and β-sheet domains with two transmembrane domains. In the N-terminus there is a part speculated to be secretion signal, while the C-terminal cytosolic domain contains the FliG-binding domain (Bergeron, 2016; Johnson et al., 2020). The regions in-between the transmembrane domains are comprised of three α-helical RBM (Ring binding motif) domains, RBM1, RBM2 and RBM3. The latter domain is split in two parts and intersected by a long β-collar (Johnson et al., 2020; Kawamoto et al., 2021) (Figures 1.7, 1.8).

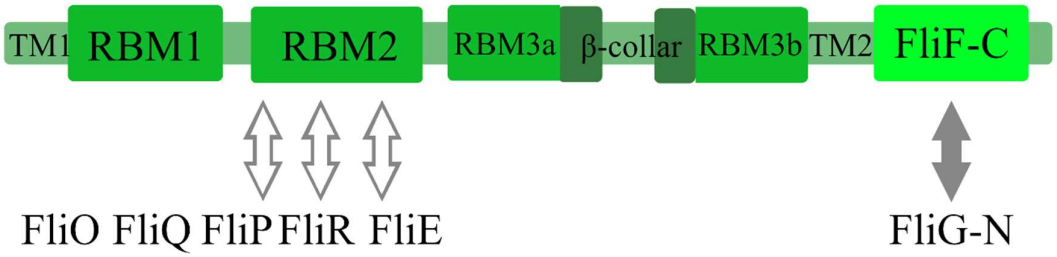



***Figure 1.8.*** **FliF domains and interactions.** Grey arrow – established interaction between FliG-N and FliF-C, empty arrows, hypothetical interactions.

While it is clear that FliF houses the fT3SS, the exact residues that interact with FlhA or FlhB remain unknown and the interaction has only been proved indirectly, through intergenic suppression (Kihara et al., 2001). Little was known about FliE until its structure was recently solved (Johnson et al., 2021), and the interaction between FliE

and FliF was proposed based on suppressor mutation in a study where the FliF N209D mutation restored motility to a non-motile FliE mutant strain (Hendriksen et al., 2021). That would place the interaction in RBM2, however a suppressor mutation does not necessarily occur in an interacting pair of residues and may occur in another site and e.g. compensate for lack of function by inducing a conformational change.

The flagellum with its fT3SS shares many similarities with the T3SS injectisome found for instance in *Salmonella*, *Shigella* and *Yersinia*. FliF RBM domains share homology with the SctD and SctJ injectisome components (Abrusci et al., 2014; Bergeron, 2016). The existence of a FliG homologue in the injectisome is currently speculative (Abrusci et al., 2014). Cross linking studies between FliP and FliR and the T3SS component SctJ E138 imply that RBM2 in FliF may have similar interactions with FliP and FliR, however they are yet to be directly proven (Kuhlen et al., 2018).

## 1.6. C-ring, domain placement and rotation switch

Bacterial flagella are bi-directional and can switch rotation direction in response to environmental stimuli; however, such a rotation switch does not result in two diametrally opposite directional movements, but rather CCW and CW rotation corresponds to linear swimming movement towards an attractant and chaotic tumbling movement as a response to a repellent , at least in peritrichous bacteria like *S. enterica* and *E. coli* (Berg and Brown, 1972; Darnton et al., 2007; Eisenbach et al., 1990). Switching events in amphitrichous bacteria seem to involve much more complex coordinated macrophenotypic processes, as in *C. jejuni* when both motors rotate CCW and the flagellum facing the direction of movement is wrapped around the cell, while the opposite, propelling motor determines the direction. Switching to CW direction results in tumbling movements and unwrapping of the wrapped flagellum (Cohen et al., 2020).

Despite numerous available crystal structures of FliG domains and their interaction partners, exact placement and arrangement of FliG domains within the ring assembly remains a subject to debate. Cross-linking has shown interaction between adjacent FliG-C domains and among adjacent FliG-M domains (Lowder et al., 2005) as well as between FliG-C and FliG-M (Baker et al., 2016; Kim et al., 2017). Transition of FliG-C from interacting with the FliG-M of the same unit to interacting with the next

one has been proposed to occur during FliG ring assembly (Baker et al., 2016) and to be the key conformational difference between clockwise (CW) and counter-clockwise (CCW) states of the C-ring (Carroll et al., 2020). These sates will be referred to as "$C^0$" and "$C^{-1}$" in the remainder of this thesis (Figure 1.9).

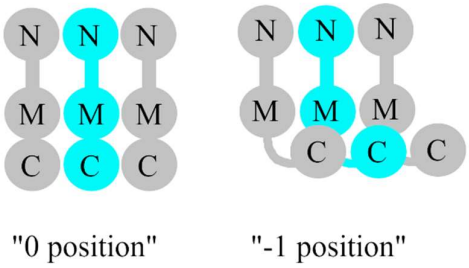



***Figure 1.9*. FliG-M and FliG-C modes of interaction. Three adjacent copies of protein are depicted.** FliG domains are designated with N, M and C. Left: position where FliG-C packs against FliG-M of the same subunit. Right: position where FliG-C packs against FliG-M of neighbouring subunit.

The need for the C-terminal domain of FliG (FliG-C) to interact with MotA/B has universally placed FliG-C in all predictions on the outer side of the C-ring (Figure 1.10), but placement and relationship between the middle domain and N-terminus has been subject to speculation (Paul et al., 2011b; Stock et al., 2012). It has been proposed that all three domains could be separate or that the N-terminus and middle domain could be condensed into an "NM-domain", but most propositions are also affected by the need to solve the then existing conundrum stoichiometry incompatibility between FliG and FliF (Paul et al., 2011b; Stock et al., 2012), with the hypotheses about the latter having proven to be false to date. The NM-domain hypothesis is supported by a pair of crosslinked residues in an experiment in *T. maritima* (Paul et al., 2011b). An intermediate state with a partially bent linker conformation has been proposed based on co-evolution data (Khan et al., 2018). FliG middle domain (FliG-M) co-crystallization with FliM middle domain (FliM-M) firmly places this domain facing downward, towards the cytoplasm (Vartanian et al., 2012).

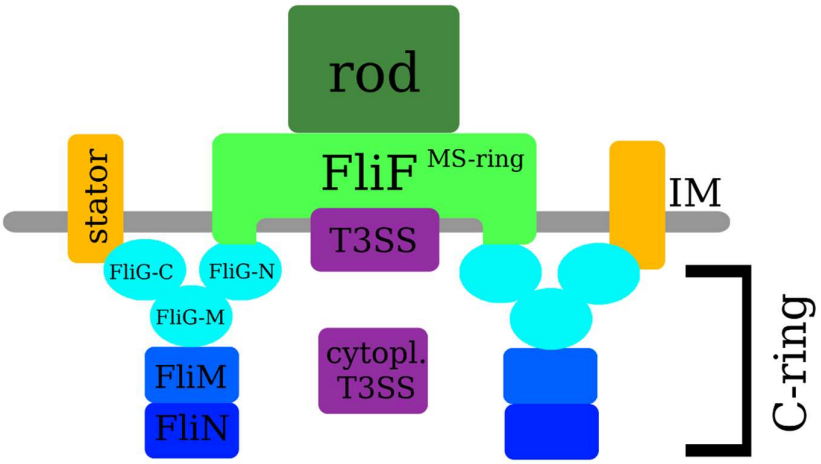



***Figure 1.10*. Flagellar motor components.** Overall structure of the flagellar motor. On the cytoplasmic side, the C-ring is composed of FliM, FliN and FliG that interacts with MS-ring, built of FliF proteins, residing in the periplasmic space and traversing the inner membrane (IM). The rotor (C-ring and MS-ring) interacts with stators traversing the inner membrane. FliG is central in joining the rotor rings and the stator. The T3SS-like system transports proteins through the inner membrane. FliG consists of N-terminal (FliG-N), middle (FliG-M) and C-terminal (FliG-C) domains. Rotor further transmits the movement upward in the flagellum, driving the rod.

*S. enterica* cryoEM studies show that the CW to CCW switch affects the overall diameter of the C-ring with the CCW conformation being wider (Sakai et al., 2019). Cross-linking experiments in *E. coli* propose that FliG-M and FliM-M interactions remain stable during switching and the FliG-C domain makes a rotation (Paul et al., 2011a). CheY-P (phosphorylated form of CheY) binds FliM and subsequently also an N-terminal segment of FliN (Sarkar et al., 2010).

## 1.7. *Campylobacter jejuni* rotor protein pseudorevertants

Pseudorevertants are mutants that arise in a strain with a compromised function by spontaneous occurrence of mutation in a component different than one initially affected that results in restoration of the function. Deletion of the *flhF* locus in *C. jejuni* abolishes motility and formation of flagella. However it is restored by appearance of pseudorevertants with reversions not in any flagella regulatory or transcription proteins, but directly in the flagellar structural components such as motor components FliF, FliG, FliH and ChePep (*Cj1178c*), an Epsilonproteobacteria specific protein (Table 1.1.) (Stoakes, 2017). ChePep is a chemotaxis regulator that is involved in switching

regulation and is only found in Epsilonproteobacteria, localized at the same cell pole(s) where flagella are present (Howitt et al., 2011). The mechanism behind these mutations restoring the flagella biosynthesis is currently unknown.

***Table 1.1*. Mutations found in *ΔflhF* pseudorevertants (Stoakes, 2017).**

| Protein | Location | Mutation | Revertant (Stoakes, 2017) |
|---|---|---|---|
| ***FliF*** | MS ring | V177L | PR1 |
| ***FliF*** | MS ring | M99I | PR2 |
| ***FliF*** | MS ring | A102T | PR3 |
| ***FliF*** | MS ring | E104K | PR4 |
| ***FliF*** | MS ring | S214T | PR5a, PR5b |
| ***FliF*** | MS ring | S103R | PR6a |
| ***FliF*** | MS ring | P220L | PR6b |
| ***FliG*** | C ring | T33A | PR4 |
| ***FliG*** | C ring | T33A | PR6c |
| ***FliH*** | T3SS cytoplasmic | S151N | PR6a, PR6b |
| ***Cj1178c*** | Chemotaxis | A475P | PR3 |
| ***Cj1178c*** | Chemotaxis | A475P | PR5a, PR5b |
| ***Cj1178c*** | Chemotaxis | A475P | PR6c |

Mutation T33A in *C. jejuni* FliG is localized in the N-terminus and equivalent residue is resolved in available crystal structures from other species. From pseudorevertant mutations found in FliF only three are found in resolved structures and only one is reasonably conserved to make a comparison between species. Various revertant strains likely work by different mechanisms as their gene expression patterns are different (Stoakes, 2017). The exact effects by which mutations in structural protein compensate for lack of the regulatory protein, FlhF, remain unknown.

Despite the motor being a fundamentally conserved structure, there are previous examples of appearance of pseudorevertants with mutations in these structural proteins to restore the crucial function of motility. In *S. enterica*, compromising the function of FliG and FliM results in mutations in MotA and MotB, the stator proteins (Togashi et al., 1997). Eliminating FliF-C and FliG-N and directly fusing the proteins in *S. enterica* prompts appearance of pseudorevertants with mutations in FliG, FliM and FliN that alleviate the strong CW bias, but not the reduced motility (Sakai et al., 2019). A two-amino acid deletion in FliF prompts appearance of mutation in FlhA (also in *S. enterica*), and another set of FliF mutants mutants in this species have given rise to pseudorevertants with mutations in FliG, FliM, FlgC, FlgF, MotA and MotB (Kihara et

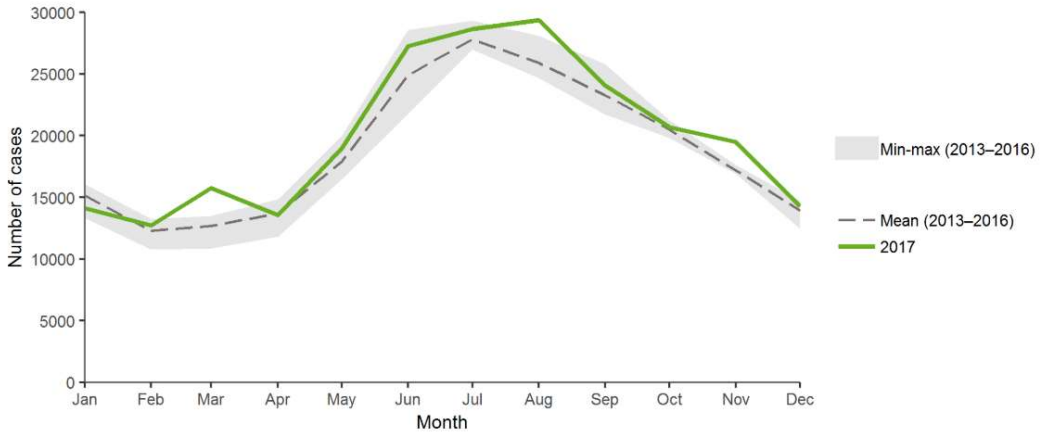
al., 2001; Komatsu et al., 2016). Examples in other species are also found, such as *C. crescentus*, where disruption of FliF-FliG interaction in non-motile strains eventually leads to appearance of motility restoring mutations in FliF and FliG that restores the phenotype (Grunenfelder et al., 2003). The common denominator between these pseudorevertant examples across species is restoration of function of a structural protein with a mutation in another structural protein. In case of *flhF* it is unknow if direct binding between FlhF and the motor components occurs.

## 1.8. *Campylobacter* and *Helicobacter* - pathogens and model organisms

### 1.8.1. Pathogenicity

*C. jejuni* is the leading cause of food poisoning in developed countries, mainly acquired from undercooked poultry (Butzler, 2004), but may it be found in food of various animal origin, contaminated water or transmitted through contact with farm animals and occasionally pets (Horrocks et al., 2009; Kaakoush et al., 2015). Infection mainly leads to diarrhoea, but in rare cases can instigate Guillain-Barré syndrome and Miller Fisher syndrome (Kaakoush et al., 2015). In addition to being a common pathogen, *C. jejuni* is also highly infections – it is predicted that ingesting as little as 360 cells is sufficient for the infection to occur (Hara-Kudo and Takatori, 2011).

*Campylobacter* infections soar in summer months, with a total of 250,161 cases reported in the EU/EAA in 2017 (Figure 1.11).




*Source: Country reports from Austria, Cyprus, the Czech Republic, Denmark, Estonia, Finland, France, Germany, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, the Netherlands, Norway, Poland, Romania, Slovakia, Slovenia, Spain, Sweden and the United Kingdom.*

***Figure 1.11*. Distribution of confirmed campylobacteriosis cases by month, EU/EEA, 2013–2016 and 2017.** An increase in infection occurs in the summer months. From ECDC report (ECDC, 2019).

The most affected group are children under 4, but of the 72 confirmed deaths, the majority occurs in the age group of 65 years and above (ECDC, 2019). Campylobacteriosis has been the most widely reported zoonosis in Europe in 2019 (220682 cases in EU in 2019), followed by salmonellosis (87923 cases in 2019), and has remained in the leading position since 2005 (EFSA, 2021). The average incidence of campylobacteriosis in Europe is 64.9 cases per 100,000 people (ECDC, 2019). Similarly to the EU/EAA, in England, Wales and Scotland the peak of campylobacteriosis is during summer, however with much higher occurrence rates of 89.7 in England (year 2016) and 109.0 in Scotland (year 2019) per 100,000 people (Health Protection Scotland, 2020; Public Health England, 2018).

Data from a number of European countries also suggests some correlation between increased temperature and precipitation (leading to increase in humidity) and campylobacteriosis incidence (Lake et al., 2019). Likewise in England and Wales association studies show some degree of link between infections, temperature and rainfall, however it is not a straightforward correlation that accounts for the entire infection dynamic and more factors remain to be identified (Djennad et al., 2019).

In Singapore, a highly developed country with equatorial climate and fairly constant temperature throughout the year, the average incidence of campylobacteriosis in 2015 was 7.6 per 100,000 with a gradually increasing trend since from 2002 (Zwe

and Yuk, 2017), but a month-by-month data has not been reported. Asymptomatic infections and flawed detection and reporting protocols make estimating the impact on campylobacteriosis in developing countries difficult (Platts-Mills and Kosek, 2014). While one study suggests that the pathogen is also harmful to the poultry birds it primarily resides in (Humphrey et al., 2014), potentially decreasing efficiency of poultry production, it is contradicting with other data (Bailey et al., 2018). Therefore the main economic impact *C. jejuni* has remains on burdening the healthcare system and its impact on animal welfare remains under debate and further study (Awad et al., 2018).

*C. jejuni* invades its host organism via multiple mechanisms that include adhesion, chemotaxis, toxin production as well as motility (Bolton, 2015; Guerry, 2007). Cia proteins (*Campylobacter* invasion antigens) are a group of proteins found to be secreted through the flagellar T3SS-like system and are connected to colonization of epithelial cells (Konkel et al., 2004). The T3SS-like secretion system is the only one of its type in *C. jejuni* and motility and secretion of Cia proteins have both been shown to be integral in invasion (Konkel et al., 2004). *C. jejuni* flagellin is glycosylated and this modification has been implicated in virulence in this species (Guerry et al., 2006). It is also among the most glycosylated bacterial proteins known (Szymanski et al., 2003).

*H. pylori* is an organism that resides in the human stomach and numerically dominates human stomach microflora (Andersson et al., 2008; Bik et al., 2006). *H. pylori* infection is present in almost half of the human population, but is mostly asymptomatic (Correa and Houghton, 2007; Perry et al., 2006). The natural reservoir and transmission sources of *Helicobacter* are human to human and it has co-evolved with *Homo sapiens* as its host (Cover and Blaser, 2009; Payão and Rasmussen, 2016). In combination with other irritants this bacterium can cause gastric lesions that evolve into chronic gastritis and can further progress into gastric cancer (Correa and Houghton, 2007). It is the only bacterium designated as a carcinogen and its role in gastric cancer has been extensively studied (Parkin et al., 2005; Wroblewski et al., 2010). *H. pylori* can also cause conditions like duodenal ulcer disease and Mucosa associated lymphoid tissue (MALT) lymphoma (Peek and Crabtree, 2006; Wroblewski et al., 2010). The extent of involvement of *H. pylori* in gastro-oesophageal reflux disease remains unclear as contradicting roles have been suggested (Sharma and Vakil, 2003). Association studies also link *H. pylori* even to extragastric diseases, like iron deficiency anaemia, liver cirrhosis, autoimmune diseases (Idiopathic thrombocytopenic purpura (ITP)) and

acute autoimmune responses (Guillain-Barré syndrome), atherosclerosis and others (Suzuki et al., 2006). While the exact relationship remains unclear, *H. pylori* has also been suggested to confer protection to some diseases and due to its complex and long co-evolution with its human host it may be more than just a pathogen to be eradicated (Cover and Blaser, 2009).

One of the central aspects of *H. pylori* pathogenicity is the cytotoxin-associated genes pathogenicity island (cagPAI), an operon that encodes a type IV secretion system (T4SS) and CagA protein, that when delivered into host cell by this system interferes with cellular signalling. It is not present in all strains (Backert et al., 2010).

### 1.8.2. Morphology

*Campylobacter* and *Helicobacter* are both Epsilonproteobacteria (Rosenberg, 2014), but have a distinctly different flagella placement. *C. jejuni* has one flagellum on each pole, while *H. pylori* has multiple flagella concentrated on one pole. These organisms thus represent an amphitrichous and a lopotrichous flagella placement type, while more commonly experimentally used *E. coli* and *S. enterica* are peritrichous. Therefore, these organisms, while more difficult to cultivate, provide crucial tools in studying flagellar transcription, placement and assembly mechanisms and differences among them.

*C. jejuni* cells are long and spiral shaped (Figure 1.12), with medial length of 2μm and median width of 0.35μm. The flagella, one located at each pole, are from 2 to 5μm long and with a 24nm diameter (Inoue et al., 2018; Muller et al., 2014). Chemoreceptor arrays are placed close to the poles, but it is not clear if this is due to proximity of the flagella that would ultimately respond to chemical signals by changing rotation direction (Muller et al., 2014).

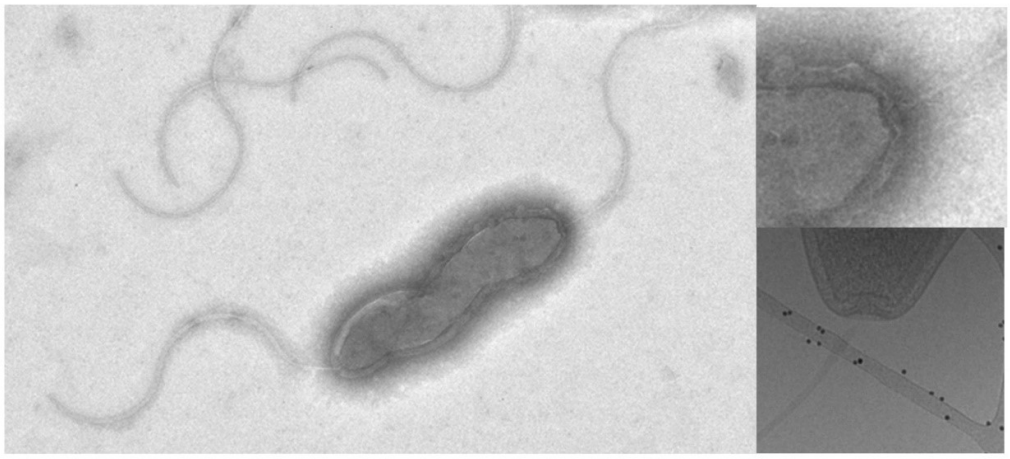


***Figure 1.12*. Electron micrograph of *Campylobacter jejuni* NCTC11168** (by E.A.Stoakes) (left and right top) and a cryoEM image on the bottom right. The characteristic membrane crater-like structure is visible at the poles of a *C. jejuni* cell.

*Helicobacter pylori* cells are spiral shaped, 5µm long and 0.5µm wide. Five to seven sheathed flagella are clustered in one pole of the cell (O'Rourke and Bode, 2001) (Figure 1.13). The sheath provides additional protection from the acidic environment of human stomach (Geis et al., 1993; Suerbaum et al., 1993). The morphology of the cells can change and rod-shaped, filamentous and coccoid forms have been observed (Krzyzek and Gościniak, 2018).

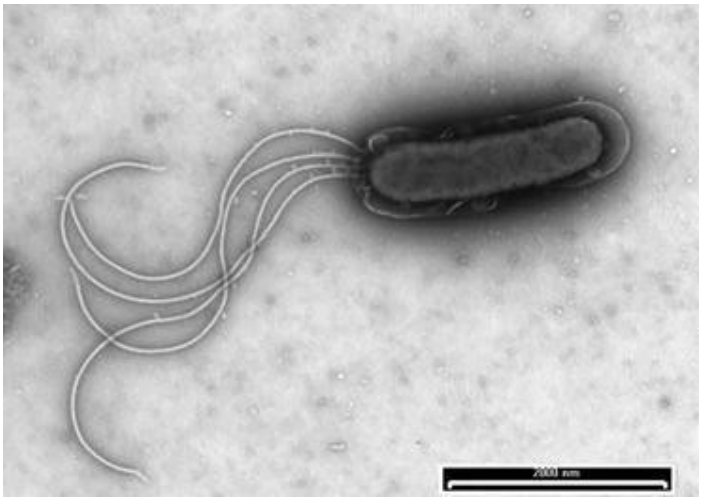


***Figure 1.13*. Electron micrograph of *Helicobacter pylori* from** (Constantinescu et al., 2016.)**.** Multiple flagella originate from a single pole of the cell.

## 1.9. Biophysical approaches for studying protein structures

### 1.9.1. MD simulations

Molecular dynamics (MD) simulation is a versatile tool for conducting computational experiments on timescales/lengthscales that cannot be detected in laboratory experiments. The utility of MD lies in its ability to both generate predictions and hypotheses for verification in the laboratory that may otherwise not be uncovered, as well as extracting additional information and expanding and complementing results of practical experiments (Hollingsworth and Dror, 2018).

MD simulations may be divided into atomistic and coarse-grained (CG) resolutions (Figure 1.14). Atomistic simulations represent each atom in the structure and solvent with appropriate chemical size, charge, bond length etc. Atomistic simulations provide an accurate representation of the system, but are costly to run and may not be feasible for multi-protein complexes. In cases of large systems, to lessen the computational cost and time, coarse-graining is performed. This groups nearby atoms together into a representative bead and while some information may be lost this way, generally CG models can perform as well as their atomistic counterparts if carefully parameterized and when used to ask the right questions, such as protein-lipid interactions (Periole and Marrink, 2013) or studying very large protein systems or dynamics of domains in multi-domain proteins (Takada, 2012). CG models are not suitable for studying folding processes and may fail to register events that rely on finer changes, such as side chain conformational change (Periole and Marrink, 2013; Takada, 2012). MD simulations can be used to simulate not only proteins in solution, but also lipid bilayers with various composition and membrane proteins embedded in them (Monticelli et al., 2008; Shelley et al., 2001; West et al., 2009), nucleic acids (Cheatham and Case, 2013) and post-translational protein modifications (Petrov et al., 2013) like glycans (Allen et al., 2021). Ever growing computational abilities and development of force-fields for different types of molecules enable accurate representation of more and more complex systems.

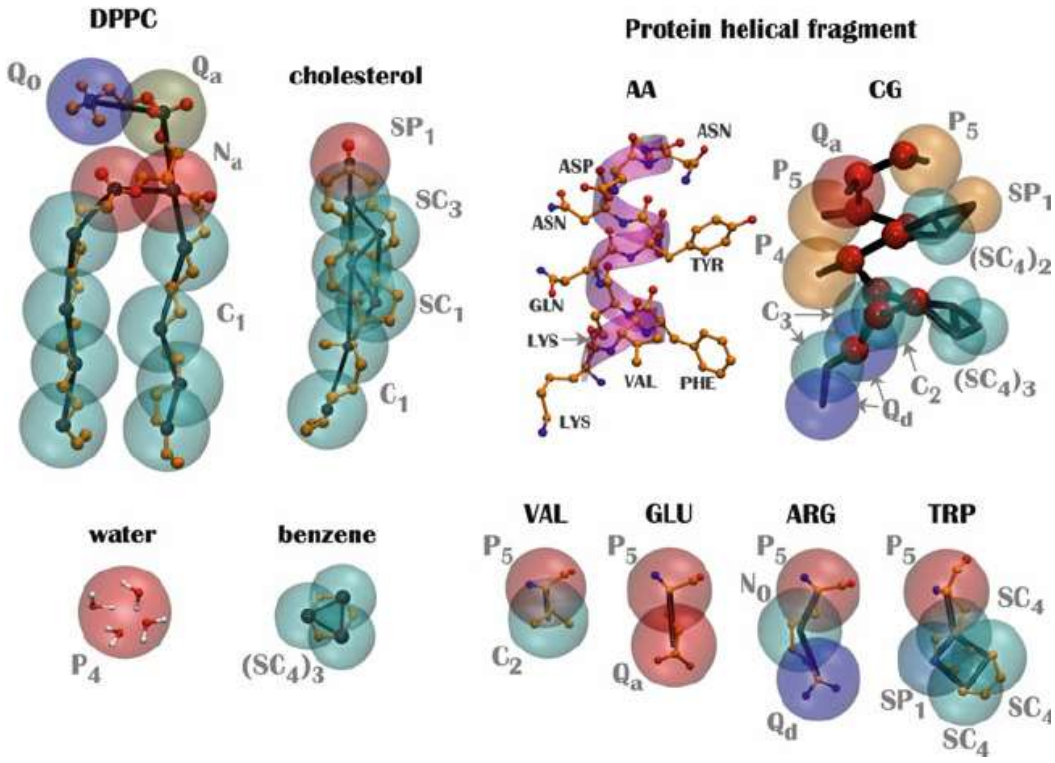



***Figure 1.14*. Principle of coarse-graining in Martini forcefield from** (Periole and Marrink, 2013). Examples of molecular structures in atomistic representations and their CG counterparts.

There are multiple molecular simulation programs available, most commonly used being GROMACS (Abraham et al., 2015) and NAMD, but examples also include CHARMM (Brooks et al., 2010), AMBER (Salomon-Ferrer et al., 2013), LAMMPS (Thompson et al., 2022) and Desmond (Bowers et al., 2006).

The force field is a combination of equations that describe forces and energies acting on each element in a system at a given time point. These are used to represent covalent bond length vibration, angle change between any three covalently linked atoms, torsion angles of dihedrals, along with non-bonded interactions, including van der Waals forces for all atoms and electrostatic interactions for polar and charged atoms (Figure 1.15) (Kukol, 2008). Covalent links are intramolecular terms while nonbonded interactions are intermolecular terms. The parameters for each term are either obtained from experimental knowledge based on other biophysical and biochemical methods or from quantum mechanics calculations (Gonzalez, 2011). For example for calibration of a forcefield structural data provide bond length and angles, thermodynamic experiments

provide solvation energies while electron densities are derived quantum chemical calculations (van Gunsteren et al., 2006).

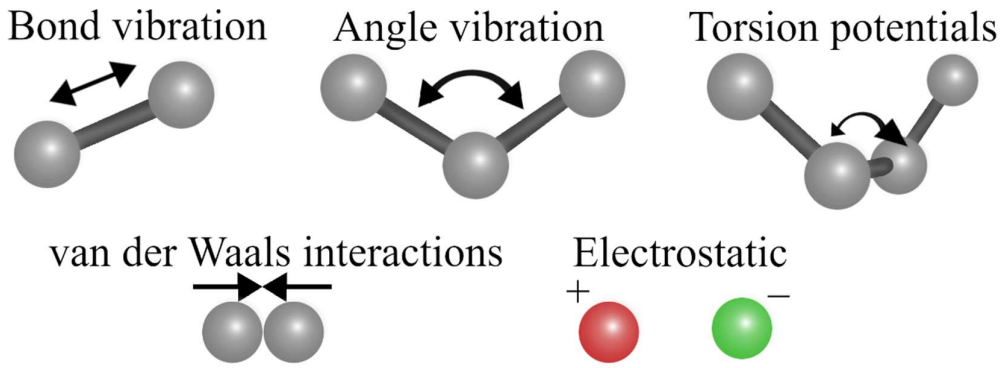



***Figure 1.15*. Visual representation of types of forces and interactions included in a force-field**. Basic types of bond movements and non-bonded interactions that can be characterized between atoms. Adapted from - (Kukol, 2008)

Validation of MD simulation with experimental data is not always straight-forward, because simulation and experimental data suffer each from a different set of imprecisions. Simulation quality is influenced by degrees of freedom modelled in the system, sampling, how the force-field has been parametrized and the boundary conditions used in the system (van Gunsteren et al., 2018). Experimental data values are often averages over different time scale from those typically encountered in simulations or some data e.g. NMR values are also derived, not directly observed (van Gunsteren et al., 2018). Simulation of well characterized small molecules and peptides with accurately measurable data like solvation of cyclohexane and folding of small peptides are used to validate force-fields (van Gunsteren et al., 2006).

An MD simulation run is associated with sequential time steps, and the position of each atom will be calculated after each step taking into account all forces that were acting on it in the previous step. A typical simulation occurs in a periodic box that is infinitely replicated in all directions to prevent artifacts due to artificial boundaries. While infinite replication of a periodic box removes the serious distortions that simulation in a vacuum would have and allows use of explicit water molecules or other solvents along with improved long range interaction evaluation, it may also introduce an increase in attractive and repulsive forces and induce cooperative effects between the copies (van Gunsteren et al., 2018, 2006). Distorting interactions between periodic box

copies can typically be avoided by increasing the box size such that the solute of interest, such as a protein molecule, does not interact significantly with its periodic image. If the system replicates biological conditions, it also typically has to maintain a constant pressure and temperature and that is achieved with thermostats and barostats, respectively. Alternatively, other simulation ensembles may be enforced, for example to simulate a crystallographic unit cell, in which constant temperature but fixed volume is desired.

The mathematical basis for each simulation time step is derived from Newton's second law of motion, where the force *F* on each particle in a system with mass *m* can be related to acceleration *a*: $F = m \times a$ , and this can be expressed as a change in coordinates (*r*) over time:

$$F = \frac{\Delta^2 r}{\Delta t^2} \times m$$

***Equation 1.1.***

The force exerted on each particle can also be evaluated by calculating the negative gradient of the underlying potential energy (*V*), which is calculated using a forcefield:

$$F_i = -\frac{\Delta V}{\Delta r}$$

***Equation 1.2.***

While the principles behind describing forces as illustrated in Figure 1.12 is the basis for most classical force-fields, individual mathematical expressions between them may differ. An example of terms behind the CHARMM family of forcefields (Brooks et al., 2010; MacKerell et al., 2000; Ramos Sasselli et al., 2016) is shown below:

$$\sum_{ele}(r) =$$

$$\sum_{bonds} k_b(r - r_{0,b})^2 +$$

$$\sum_{UB} k_{UB}(S - S_{0,UB})^2 +$$

$$\sum_{angle} k_a(\theta - \theta_{0,a})^2 +$$

$$\sum_{dihedrals} k_{d,n}(1 + \cos(n\chi - \Delta_{d,n})) +$$

$$\sum_{impropers} k_i(\psi - \psi_{0.i})^2 +$$

$$\sum_{vdw} \varepsilon_{d,n} \left[ \left( \frac{R_{min,ij}}{r_{ij}} \right)^{12} - 2 \left( \frac{R_{min,ij}}{r_{ij}} \right)^{6} \right] +$$

$$\sum_{electrostatic} \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}}$$

***Equation 1.3.***

where UB stands for Urey-Bradley potential, vdw for van der Waals forces, electrostatic for electrostatic forces, bonds describe bond length, angle describes angle between three atoms, dihedrals describe special placement of four atoms, improper describe improper dihedrals.

The system before simulation is energy minimized to remove any steric clashes and placed in an appropriate solution and/or conditions corresponding to a relevant experiment (Dokholyan, 2012).

During simulation, each step involves calculating all forces described by the forcefield acting on each atom and moves them accordingly from the starting position, then repeats the process multiple times. A smaller time step ensures higher precision (and stability), but increases computational cost.

In a biological system a water model is used and physiological conditions are usually represented by adding monovalent atoms (typically sodium and chloride) to maintain total neutral charge in the system (Patodia, 2014) and represent the ionic strength of the environment (Zhang et al., 2010). Water models are tested and continuously improved (Chatterjee et al., 2008; Medina et al., 2011). Divalent ion models exist, and can be used when studying a particular biological question, such as divalent ion interactions with nucleic acids or lipids (Böckmann and Grubmüller, 2004; Mukherjee and Bhattacharyya, 2013). However, generally solvent is maximally simplified (Patodia, 2014) and not all ion types found in physiological solution are represented, as that would introduce unnecessary complexity to the system and sodium chloride solutions are well parametrized. With regards to coarse-grained systems,

simulations can suffer from solvent freezing due to the simplified solvent models, and modified antifreeze may be added, if necessary (Marrink et al., 2007).

Equilibration is performed to ensure stability of the system before the actual production run is commenced; the system is typically heated up and kinetic energy distributed within it (Patodia, 2014), often while limiting distortion of the solute of interest (such as a protein) via e.g. position restraints. The simulation run time is selected depending on the goal, approximation level of the system (atomistic vs coarse-grained) and how quickly multiple runs of the same simulation converge. The length must be selected such that it would be long enough for the molecular motion of interest (for example movement of domains of molecules for biological systems) to occur within the size and complexity level of the system simulated (Boldon et al., 2015; Dokholyan, 2012; Perilla et al., 2015).

MD simulation data, particularly for biomolecules, can be analysed with multiple methods, like RMSD (root mean square deviation), RMSF (root mean square fluctuation) (Kukol, 2008), residue pairwise proximity matrix and many others.

### 1.9.2. Protein structure modelling

Some proteins are difficult to crystallize due to lack of stabilizing interactions with interaction partners, membrane proteins etc, and therefore they may not have solved structures available. Homology modelling makes it possible to derive models based on available structures (or fragments of structures) from homologs. While these models have potential drawbacks, they often provide a necessary starting material for further hypothesis generation, and can serve as initial structures for MD simulations. Homology modelling uses known crystal structures as templates to build a structure of a given sequence, for example using the long standing standard program in the field, MODELLER (Webb and Sali, 2016).

Alternative approaches are structure prediction from sequence without using a specific template. Recent developments in structure generation include neural-network based AlphaFold and its derivatives, capable of predicting protein structures (Jumper et al., 2021). *De novo* prediction of a structure solely from sequence information is developing and notable and a well-established standard in the field is Rosetta (Rohl et al., 2004). *De novo* prediction of multimeric states has also been reported with Rosetta (Andre et al., 2007).

### 1.9.3. Cryo-electron tomography

Electron microscopy-based structure solving methods like cryoEM and cryoET have recently made a huge contribution to visualization of structures that were previously difficult to study. CryoET is a method that reconstructs a protein from imaging a 3D object from multiple angles. Crystallography relies on the homogeneity of a protein sample or complex while cryoET allows the study of structures that are difficult to purify or isolate their stable complexes, without the need to remove them from their *in situ* environment (Lučić et al., 2013). Added gold particles or prominent features are tracked throughout every image in the tilt series leading to reconstruction of a sub-tomogram. Multiple sub-tomograms can be classified and averaged resulting in a higher resolution image of a structure, complex or feature.

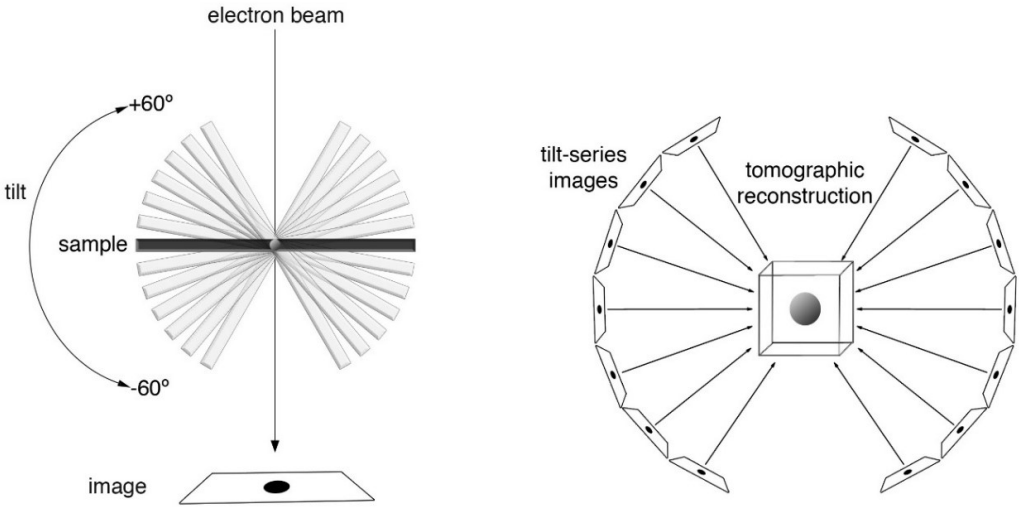



***Figure 1.16*. Sample rotation in the electron beam leads to a collection of tilt series of images.** Tracking a feature in every image allows retrieval of 3D reconstruction from a series of 2D images. Image from Wikipedia under Creative Commons licence (https://commons.wikimedia.org/wiki/File:Electron_Tomography.tif)

The ultrastructure of *C. jejuni* due to its polarized structure has been successfully studied with cryoET before (Muller et al., 2014). The flagellar motors in cryoET images have clearly visible ring structures (Muller et al., 2014) making it a very attractive model for motor reconstruction (Figure 1.16). This technique has been used in flagella component deletion studies in different species (Beeby et al., 2016) and even conventional electron microscopy can detect large scale differences in fusion mutants (Thomas et al., 2001). With very large datasets cryoET can reconstruct structures to a

resolution comparable with crystallography and NMR as evident from recent breakthroughs in solving flagellar motor component structures (Johnson et al., 2020; Kawamoto et al., 2021; Kuhlen et al., 2018). Whether conformational changes can be detected in *C. jejuni* pseudorevertant cryoET reconstructions of a small dataset remains to be seen.

## 1.10. Aims of the thesis

FliG of the bacterial motor has not been studied in detail with MD and this study sets out to broaden our understanding of this crucial to bacterial motility protein. Conformation of linker between FliG-N and FliG-M needs to be additionally addressed to explore conflicting hypotheses proposed in the literature so far. Currently here is no high resolution FliG ring model available unlike recently solve FliF so this study bridges the gap by proposing a possible model that can serve for further experiment design and hypothesis generation.

There is no known *C. jejuni* motor ring stoichiometry or structures available, but recent data implies that it could be different from 34 seen in *S. enterica* and *V. alginolyticus*, because *B. burgdorferi* has a very different stoichiometry and C-ring diameter. This study attempts to elucidate more about *C. jejuni* motor with cryoET.

As continuation of previous *C. jejuni* deletion strain and pseudoreverant studies, additional deletion strain and complementation strains are created.

1. Investigate conformational dynamics of FliG N-terminus
2. Model the conformation of FliF-C bound FliG in *H. pylori*
3. Model a FliG ring assembly as part of the C-ring
4. Elucidate the stoichiometry of *Campylobacter* motor and density differences in a FliF* pseudorevertant
5. Investigate complementation of *ΔfliF* from external locus

# Chapter 2 – Methods

## 2.1. Molecular dynamics simulations and homology modelling

Molecular dynamics simulations and homology modelling was performed as described in (Tupiņa et al., 2022).

### 2.1.1. System setup

Ionizable amino acid side chains can adopt different protonation states depending on the pH of the environment. Each side chain can have its protonation equilibrium constant $pK_a$ calculated similarly as titration curve of a compound. The H++ server (Gordon et al., 2005) was used to predict protonation state of ionizable residues, using $\varepsilon^{int} = 10$, $\varepsilon^{out} = 80$ at pH = 7.5, representing the physiological environment inside the bacterial cytoplasm (Kobayashi et al., 2000; Krulwich et al., 2011; Slonczewski et al., 2009). Two crystal structures PDB ID:5WUJ (Xue et al., 2018) and PDB ID:5TDY (Lynch et al., 2017) were simulated both in solution and under crystallographic unit cell conditions, with the pH representing the crystal solution (pH 5.6 for 5WUJ and pH 7.5 for 5TDY) (Table 2.1). After identification of E91 in 5WUJ as having potentially variable protonation state in pH 7.5, systems were set up in both protonated/deprotonate E91 variants (Table 2.1). To assess reproducibility, we utilized both the CHARMM36m (Huang et al., 2016) and AMBER14SB (Maier et al., 2015) forcefields in separate simulations for each structure.

For simulations in solution beginning from each structure, we used physiologically representative ion concentrations (150 mM NaCl), while the salinity in simulated crystallographic conditions was adjusted by removing sodium atoms that would be bound by sodium citrate in the crystallization solutions (Lynch et al., 2017; Xue et al., 2018). In addition, a variant of PDB ID 5WUJ with residues expressed and purified, but unresolved in the electron density (residues 1 to 6 and 112 to 115 of FliG, and residues 518 to 523 and 559 to 567 of FliF respectively) were built with the ModLoop (Fiser and Sali, 2003) server. Additionally, the crystal structure of full length FliG (PDB ID:3HJL) (Lee et al., 2010) was simulated in solution using the CHARMM36m force field (Huang et al., 2016).

### 2.1.2. MD simulations

All MD simulations were performed with the GROMACS 2018 package (Abraham et al., 2015). Equations of motion were integrated using the leap-frog algorithm with a 2 fs time step. We performed energy minimization using the steepest descents algorithm for ≤100,000 steps, followed by a 10 ns NPT equilibration with position restraints on the protein backbone atoms (using a force constant of 1000 kJ $mol^{-1}$ $nm^{-2}$) for all structures. The unrestrained production runs in the NPT ensemble were run for 100, 200 or 500 ns depending on the system (Table 2.1). The TIP3P (Jorgensen et al., 1983) water model was used for all systems. Temperature for the solvated proteins under effective physiological conditions was maintained at 300 K using the v-rescale thermostat (Bussi et al., 2007), while a temperature mimicking the crystallization conditions was set to 289 K and 298 K for 5WUJ and 5TDY, respectively. The Parrinello-Rahman barostat (Parrinello and Rahman, 1981) was used for isotropic pressure coupling with a reference pressure of 1 bar. Electrostatic interactions were calculated according to the Particle Mesh Ewald (Darden et al., 1993) algorithm with a real space cut-off of 1.2 nm for CHARMM36m and 1.0 for AMBER14SB. For the calculation of van der Waals interactions, we applied a cut-off at 1.2 nm, switching the potential after 1.0 nm for CHARMM36m and 1.0 cut-off for AMBER14SB. The LINCS (Hess et al., 1997) algorithm was used for constraining all covalent bonds involving hydrogen atoms. The forcefield, simulation sampling, and relevant conditions for each system are summarized in Table 2. VMD (Humphrey et al., 1996) and PyMOL (DeLano, 2002) were used for trajectory and structure visualization.

***Table 2.1*. Simulated systems in this study.**

| *System name* | *Structure (PDB ID)* | *No of proteins* | *No of atoms* | *Forcefield* | *Duration (ns)* | *Protonation state of E91* | *Number of replicas* | *Conditions* |
|---|---|---|---|---|---|---|---|---|
| FliG-N | FliG only (5WUJ) | 1 | 49581 | CHARMM36m | 200 | deprot | 3 | physiol. |
| FliG-N | FliG only (5WUJ) | 1 | 49575 | CHARMM36m | 200 | prot | 3 | physiol. |
| FliG-N/ FliF-C | 5WUJ | 1+1 | 49592 | CHARMM36m | 500 | deprot | 3 | physiol. |
| FliG-N/ FliF-C | 5WUJ | 1+1 | 49598 | CHARMM36m | 500 | prot | 3 | physiol. |
| FliG-N/ $FliF\text{-}C^{tet}$ | $5WUJ_{tet}{}^{\Pi}$ | 4+4 | 230810 | CHARMM36m | 100 | deprot | 1 | physiol. |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| FliG-N/ FliF-$C^{tet}$ | $5WUJ_{tet}{}^{\Pi}$ | 4+4 | 230822 | CHARMM36m | 100 | prot | 1 | physiol. |
| FliG-N/ FliF-$C^{tet}$ | $5WUJ_{tet}{}^{\Pi}$ | 4+4 | 230810 | AMBER14SB | 100 | deprot | 1 | physiol. |
| FliG-N/ FliF-$C^{tet}$ | $5WUJ_{tet}{}^{\Pi}$ | 4+4 | 230822 | AMBER14SB | 100 | prot | 1 | physiol. |
| FliG-N/ FliF-$C^{cryst}$ | 5WUJ | 6+6 | 28643 | CHARMM36m | 100 | deprot | 1 | crystal |
| FliG-N/ FliF-$C^{cryst}$ | 5WUJ | 6+6 | 28673 | CHARMM36m | 100 | prot | 1 | crystal |
| FliG-N/ FliF-$C^{cryst}$ | 5WUJ | 6+6 | 28625 | AMBER14SB | 100 | deprot | 1 | crystal |
| FliG-N/ FliF-$C^{cryst}$ | 5WUJ | 6+6 | 28643 | AMBER14SB | 100 | prot | 1 | crystal |
| FliG-N/ FliF-$C^{expr}$ | $5WUJ_{expr}$ * | 6+6 | 28865 | CHARMM36m | 100 | deprot | 1 | crystal |
| FliG-N/ FliF-$C^{expr}$ | $5WUJ_{expr}$ * | 6+6 | 28883 | CHARMM36m | 100 | prot | 1 | crystal |
| FliG-N/ FliF-$C^{expr}$ | $5WUJ_{expr}$ * | 6+6 | 28865 | AMBER14SB | 100 | deprot | 1 | crystal |
| FliG-N /FliF-$C^{expr}$ | $5WUJ_{expr}$ * | 6+6 | 28883 | AMBER14SB | 100 | prot | 1 | crystal |
| | 5TDY | 2+2 | 175879 | CHARMM36m | 100 | - | 3 | physiol. |
| | 5TDY | 4+4 | 14017 | CHARMM36m | 100 | - | 3 | crystal |
| | $5TDY_{bent}$ | 1+1 | 68590 | CHARMM36m | 100 | - | 3 | physiol. |
| | $5TDY_{eln}$ | 1+1 | 75677 | CHARMM36m | 100 | - | 3 | physiol. |
| | 3HJL | 1 | 203600 | CHARMM36m | 100 | - | 3 | physiol. |

*$5WUJ_{expr}$ – system with unresolved, but expressed residues added.
$^{\Pi}5WUJ_{tet}$ – system with tetramer made from FliG-N/FliF-C complexes

Computational experiments were run on Bioinformatics Institute in-house Linux cluster and ASPIRE1 of Singapore National Supercomputing Centre. BII cluster uses 8 nodes of 2 GPUs (Nvidia GeForce RTX 2080 Ti) and 24 CPUs (Intel® Xeon® Gold 5118 CPU @ 2.3 GHz) each, while ASPIRE1 employs 4 nodes per simulation, with each node using 1 GPU (Nvidia Tesla K40t) and 24CPUs (Intel® Xeon® CPU E5-2690 v3 @ 2.6 GHz).

### 2.1.3. Model building

Sequences were aligned with MUSCLE software (Edgar, 2004; McWilliam et al., 2013). Homology models of full length *H. pylori* FliG were built in MODELLER (Webb and Sali, 2016). Initially a homology model of the N-terminus and the middle domain was built generating using the *H. pylori* structures (PDB-IDs: 5WUJ and 3USY) (Lam et al., 2013; Xue et al., 2018) and *A. aeolicus* full-length structure (PDB-ID: 3HJL) (Lee et al., 2010). In addition, the linker region was defined and constrained as α-helical in MODELLER. From a total of 2,000 models, the 10 with the best DOPE (Shen and Sali, 2006) and molpdf (Shen and Sali, 2006) scores were submitted to the SWISS-MODEL webserver (Schwede et al., 2003; Waterhouse et al., 2018) Structure Assessment and QMEAN (Benkert et al., 2008) evaluation tools. The model that was in both the MODELLER score top 10 groups, had QMEAN score of 0.77 ± 0.06 and had no Ramachandran outliers was selected (QMEAN scores were very close for all the tested models, but 0.77 was among the best). To create the full-length model, the crystal structure of the C-terminal and middle domains of *H. pylori* (PDB-ID: 3USY) (Lam et al., 2013) was overlaid onto the model of the middle domain that both structures share using VMD (Humphrey et al., 1996).

Whole FliG rings were built by fitting three copies of the FliG bound to FliF C-terminal region (FliF-C/FliG) and FliM middle domain (FliM-M) complex in a segment of the *V. alginolyticus* cryoET densities (Carroll et al., 2020), via manual placement and subsequent energy minimization. The system was then subjected to Molecular Dynamics Flexible Fitting (MDFF) (see section 2.1.4 below). For the *H. pylori* models, FliM-M was added by overlaying the model's middle domain with the crystal structure of the *H. pylori* middle domain and FliM middle domain (PDB-ID: 4FQ0 (Lam et al., 2013)) in Chimera (Pettersen et al., 2004). For the *A. aeolicus* ring, a homology model of FliM-M was built using the *A. aeolicus* sequence (UniProt entry: A0A7C5QKN5) and *H. pylori* FliM-M PDB-ID: 4FQ0 structure as template. The middle copy of the three monomers simulated in MDFF was used to build an entire ring from 34 copies of the monomer.

The ring was placed in the cryoET density with the *Fit in Map* function in Chimera (Pettersen et al., 2004) using 20 Å resolution for the map created from the atomistic model, and used to obtain a final correlation value.

### 2.1.4. Flexible fitting

MDFF is used to dynamically fit atomic coordinates into experimental density maps (Trabuco et al., 2009). An additional external potential derived from an experimental map is defined on a grid to drive gradual fitting of coordinates. All MDFF simulations were performed on a trimeric unit of the *H. pylori* FliF-C/FliG/FliM-M complex for the clockwise (CW) or counterclockwise (CCW) cryoET density maps from *V. alginolyticus* obtained from *V. alginolyticus* mutants that are locked in one of the two directions of flagella rotation (Carroll et al., 2020). All such simulations were run using NAMD2.9 (Trabuco et al., 2009) with the CHARMM36 FF (Huang and Mackerell, 2013) with implicit solvent ($\varepsilon = 80$). The nonbonded interactions were switched off between 1.0 and 1.2 nm using a force-switching method (Steinbach and Brooks, 1994). Langevin dynamics was used to maintain the temperature at 300 K with a coupling coefficient of 5 $ps^{-1}$. The flexible fitting was run in 10-15 separate simulations, with increased scaling factors from 0.1 to 15. The bias was applied to all non-hydrogen atoms with atom-dependent mass weighting. Additional restraints in MDFF were used to maintain correct chiral centres, peptide bond conformations, and secondary structures of each protein. All simulation systems were generated using CHARMM-GUI through Map Utilizer (Qi et al., 2017).

### 2.1.5. Analysis

MD simulation data was initially analysed using RMSD (root mean square deviation) measurement in GROMACS (Abraham et al., 2015) and plotted in xmGrace. Index files defining the elements were also created in GROMACS. RMSD measurement in MD effectively assesses the structural drift of an element (groups of atoms or amino acids) across the time of simulation with reference to the starting coordinates, usually the energy minimized starting structure, after least-squares fitting. In the analysis performed here, helix boundaries were defined using energy minimized structures instead of crystal structure, because they have water and additional solvents from crystallization removed and minimized structured were the direct input used in simulation. RMSD measurement as comparison between two different structural conformations as in Results 4.2.2 or two different species as in Results 3.2.5 is defined

as distance between atom in the reference structure and the same atom (or nearest equivalent by homology) in the other structure. RMSD is defined as follows, where δ is the distance between the atoms.

$$RMSD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\delta_i^2}$$

***Equation 2.1.***

Crossing angle measurements between helix pairs to observe dynamic flexibility were performed in VMD (Humphrey et al., 1996). To perform this, 4 points were defined, namely the Cα atoms of the first and last residues of each of the two helices in the crossed pair. A visual representation of this is shown in section 3.2.2.

The fitting between the simulation coordinates and experimental density map was calculated in Chimera (Pettersen et al., 2004). The *Fit in Map* tool locally maximizes the overlap between atomic coordinates and the density map. The correlation coefficient was calculated as the minimum mean cosine angle between the vectors derived from the experimental and simulation maps calculated on the grid. This value can vary from -1 to 1, i.e., from anti-correlated to identical, respectively.

## 2.2. *Campylobacter* strain construction and sequencing

### 2.2.2. Growth and storage conditions of microorganisms

The *E. coli* strains used for cloning (NEB® 5α or DH5α) were grown at 37ºC in standard LB (Lysogeny Broth) medium or LB-agar plates (1% agar) in aerobic conditions. *C. jejuni* NCTC11168 and its derivatives were grown on Campylobacter blood-free selective agar supplemented with CCDA (containing Cefoperazone and Amphotericin B) (Oxoid), plates for approximately 24 hours following streaking on Mueller-Hinton (MH) (Oxoid)1% agar plates. Liquid cultures of *C. jejuni* were grown in Mueller Hinton broth and all steps of cultivating *C. jejuni* were performed in a variable atmosphere incubator (VAIN) (Whitley VA500 workstation cabinet) (90 % $N_2$ (v/v), 6 % $CO_2$ (v/v), 4 % $O_2$ (v/v)). All cell lines used in this project are listed in Table 2.2.

*E. coli* glycerol stocks were made using a 1:1 ratio of overnight culture and 50% glycerol in cryo-stock vials. *C. jejuni* stocks were made by scrapping cells of a plate and resuspending them in a Microbank$^{TM}$ bead stock vial (Fischer Scientific). All stocks were stored at -80ºC.

Apramycin (SIGMA-ALDRICH) was used in a final concentration of 60 µg/ml, chloramphenicol of 25-30 µg/ml and kanamycin of 25 µg/ml in plates and liquid cultures. Apramycin plates were covered with aluminium foil during preparation due to the light sensitivity of the antibiotic with every effort made to minimise exposure to light during all procedures.

***Table 2.2*. Bacterial strains used in this study.**

| *Bacterial strain* | *Genotype* | *Source* |
|---|---|---|
| *E. coli DH5α* | fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17 | NEB |
| *E. coli NEB 5α* | fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17 | NEB Gibson Assembly kit |
| *C. jejuni NCTC11168* | Sequenced laboratory strain, differences from original isolate identified by Emily Stoakes (Stoakes, 2017) | Prof. Charles Penn, University of Birmingham |
| *C. jejuni NCTC11168 ΔflhF fliF$_{M99I}$* | *ΔflhF fliF$_{M99I}$* (pseudorevertant PR2) | (Stoakes, 2017) |
| *C. jejuni NCTC11168 ΔflhF ΔfliF* | NCTC11168 *ΔflhF*::Kan$^{r}$ *ΔfliF*::Cat$^{r}$, promoterless Kan$^{r}$ cassette inserted in *flhF* locus, promoterless Cat$^{r}$ cassette inserted in *fliF* locus | Created by Emily Stoakes and Charlotte Dixon |
| *C. jejuni NCTC11168 ΔfliF* | NCTC11168 *ΔfliF*::Cat$^{r}$, promoterless Cat$^{r}$ cassette inserted in *fliF* locus | This study |
| *C. jejuni NCTC11168 ΔflhF ΔfliF:fliF* | NCTC11168 *ΔflhF*::Kan$^{r}$ *ΔfliF*::Cat$^{r}$, *Cj0046*::*natfliF,* promoterless Kan$^{r}$ cassette inserted in *flhF* locus, promoterless Cat$^{r}$ cassette inserted in *fliF* locus, *fliF* under | This study |

| | | |
|---|---|---|
| | native promoter inserted in *Cj0046* locus | |
| *C. jejuni NCTC11168 ΔflhF ΔfliF:fliF $_{M99I}$* | NCTC11168 *ΔflhF*::Kan$^r$ *ΔfliF*::Cat$^r$, *Cj0046*::*natFliF $_{M99I}$,* promoterless Kan$^r$ cassette inserted in *flhF* locus, promoterless Cat$^r$ cassette inserted in *fliF* locus, *fliF$_{M99I}$* under native promoter inserted in *Cj0046* locus | This study |
| *C. jejuni NCTC11168 ΔfliF:fliF* | NCTC11168 *ΔfliF*::Cat$^r$, *Cj0046*::*natfliF, fliF* under native promoter inserted in *Cj0046* locus | This study |
| *C. jejuni NCTC11168 ΔfliF:fliF $_{M99I}$* | NCTC11168 *ΔfliF*::Cat$^r$, *Cj0046*::*natfliF $_{M99I}$, fliF$_{M99I}$* under native promoter inserted in *Cj0046* locus | This study |
| *C. jejuni NCTC11168 ΔflhF ΔfliF:porA-fliF* | NCTC11168 *ΔflhF*::Kan$^r$ *ΔfliF*::Cat$^r$,*Cj0046*::*natfliF,* promoterless Kan$^r$ cassette inserted in *flhF* locus, promoterless Cat$^r$ cassette inserted in *fliF* locus, *fliF* under porA promoter inserted in *Cj0046* locus | This study |
| *C. jejuni NCTC11168 ΔflhF ΔfliF:porA-fliF $_{M99I}$* | NCTC11168 *ΔflhF*::Kan$^r$ *ΔfliF*::Cat$^r$, *Cj0046*::*natfliF $_{M99I}$,* promoterless Kan$^r$ cassette inserted in *flhF* locus, promoterless Cat$^r$ cassette inserted in *fliF* locus, *fliF$_{M99I}$* under porA promoter inserted in *Cj0046* locus | This study |
| *C. jejuni NCTC11168 ΔfliF:porA-fliF* | NCTC11168 *ΔfliF*::Cat$^r$, *Cj0046*::*natfliF, fliF* under porA promoter inserted in *Cj0046* locus | This study |
| *C. jejuni NCTC11168 ΔfliF:porA-fliF $_{M99I}$* | NCTC11168 *ΔfliF*::Cat$^r$, *Cj0046*::*natfliF $_{M99I}$, fliF$_{M99I}$* under porA promoter inserted in *Cj0046* locus | This study |

### 2.2.3. PCR and cloning

Inserts and vectors for cloning were amplified using high fidelity Q5 polymerase (NEB) according to manufacturer's instructions using 0.5 µM final primer concentration and extension times based on manufacturers' reported polymerase amplification speed, length of target amplicon and rounded up to 5-10% additional time. Screening colonies for successful clones was carried out by whole cell colony PCR using the GoTaq Green Master Mix (Promega) according to manufacturer's instructions. Where PCR products were amplified from a plasmid template, NEB *Dpn*I reaction was used subsequently to eliminate plasmid template DNA before transformation.

PCR products were run in 1% agarose TAE gels with NEB 1kb plus, NEB 1kb extend, 1kb Life Sciences Ladder or 100bp Hyper Ladder according to expected product size. PCR products were purified using Wizard SV Gel and PCR Clean-up System (Promega). SYBR green and GelPilot dye was used for visualization of DNA in the gel.

Primers were designed using the NEB Tm calculator for reactions using NEB reagents and the OligoCalc (Kibbe, 2007) Tm calculator for all other primers. Primer pairs were designed with the aim of Tm 60˚C and 50% GC content where possible. Lyophilised primers were resuspended to 100 µM concentration in nuclease free water according to the volume indicated in the delivery sheet and afterwards used to prepare 10 µM stocks (PCR working concentration).

Suicide vectors from pC46-fdxA backbone were assembled using Gibson Assembly Cloning kit (NEB, UK) according to manufacturer's instructions using 100 ng vector and insert in a 3:1 molar ratio:

| **Reaction component** |
| --- |
| **10µl 2X Gibson assembly Master Mix** |
| **100ng vector** |
| **3x fold molar excess insert to vector** |
| **Nuclease free water up to 20µl** |

A 20 µl Gibson reaction was incubated at 50°C for 15 minutes and transformed in NEB 5α *E. coli* cells provided with the Gibson kit. In the negative control reaction, no insert was added. NEBiocalculator was used to calculate the appropriate insert amount corresponding to 100 ng of vector. Original pC46 vector that had been used to

create pC46-fdxA (Table 2.4) was obtained for University of Leicester (Elgamoudi, 2016; Elgamoudi and Ketley, 2018).

***Table 2.3*. Primers used in this study.** For Gibson assembly steps see section 5.2 (Figure 5.2). Gibson assembly pairs per step are red-green colour coded. There are four primers per step, a pair red-green pair for insert and a red-green pair for amplifying vector. Green and red show each side for Gibson assembly reaction, where insert and vector will anneal.

| *Primer description* | *Sequence* | *Purpose* | *Origin* |
|---|---|---|---|
| *apramycin_fwd* | actttgctaaagataaattaaaacTCAGCCAATCGACTGGC | vector construction, anneals to apramycin, overhang for Gibson reaction with pC46-fdxA vector | Emily Stoakes |
| *apramycin_rev* | atcttaatattaaaaaggagaaaacATGCAATACGAATGGCGAAAAG | vector construction overhang for Gibson reaction with pC46-fdxA vector | Emily Stoakes |
| *pC46-fdxA_rev* | GTTTTAATTTATCTTTAGCAAAAGTAG | vector construction, amplifies pC46-fdxA vector, step1 | Emily Stoakes |
| *pC46-fdxA_fwd* | GTTTTCTCCTTTTTAATATTAAGATTTAAAAG | vector construction, amplifies pC46-fdxA vector, step1 | Emily Stoakes |
| *fliF_+_native_promoter_fwd* | acattatttaaagctaggccggccgTTAAGATAGATAAATCAAATAAAAATATGTAAAATTC | vector construction, anneals to *fliF* with native promoter, overhang for Gibson reaction with vector step2 | Emily Stoakes |
| *fliF_+_native_promoter_rev* | tttcgattggcgcgcctgagggattTCATCACATATCCTTTTCG | vector construction, anneals to *fliF*, overhang for Gibson reaction with vector step2 | Emily Stoakes |
| *pC46_fdxA_apramycin_rev* | CGGCCGGCCTAGCTTTAAATAATG | vector construction, vector amplification for Gibson reaction, step2 | Emily Stoakes |
| *pC46_fdxA_apramycin_fwd* | AATCCCTCAGGCGCGCCA | vector construction, vector amplification for | Emily Stoakes |

| | | | |
|---|---|---|---|
| | | Gibson reaction, step2 | |
| *porA_fwdII* | acattatttaaagctaggccggccTTAAAACAACTATATATTACTTCTCGTTTAC | vector construction, anneals to porA promoter, overhang for Gibson reaction with vector step 3 | Modified from Emily Stoakes |
| *porA_rev* | gatgaagcatatttttaaaatccatGAGAATTCTCCTTGTCAAAAATTAATAAAAC | vector construction, anneals to porA promoter overhang for Gibson reaction with vector step 3 | Emily Stoakes |
| *pC46aprafliFnative_rev* | AGGCCGGCCTAGCTTTAAATAATG | vector construction, vector amplification for Gibson reaction, step3 | Emily Stoakes |
| *pC46aprafliFnative_fwd* | ATGGATTTTAAAAATATGCTTCATC | vector construction, vector amplification for Gibson reaction, step3 | Emily Stoakes |
| *natFliF_fw1* | TTAAGATAGATAAATCAAATAAAAATATGTAAAATTC | Sanger sequencing for vector verification | Emily Stoakes |
| *natFliF_rev1* | TCATCACATATCCTTTTCG | Sanger sequencing for vector verification | Emily Stoakes |
| *pC46 F* | GCTCCAAATCCTTTGGTG | Sanger sequencing for vector verification | This study |
| *pC46 R* | GCCCAAAGCTCCAG | Sanger sequencing for vector verification | This study |
| *PCR F Cat* | CCTTCAAAGCTTGTCCACGG | Colony PCR, anneals to $Cat^R$ casette | Charlotte Dixon |
| *PCR R Cat* | ACGGACAGGTAGGCGTTTT | Colony PCR, anneals to $Cat^R$ casette | Charlotte Dixon |
| *Apramycin46 out* | GACAGGTGGCTCAAGGAG | Sanger sequencing of apramycin insertion site | This study |

| | | | |
|---|---|---|---|
| *Apramycin46 in* | GCTGACCGATGAGCTCG | Sanger sequencing of apramycin insertion site | This study |
| *FliF46 out* | CTTTAGAGAAATTTAATG CAGCAA | Sanger sequencing of *fliF* insertion site | This study |
| *FliF46 in* | TAAAGTTGCCCAATTTGA TGAAG | Sanger sequencing of *fliF* insertion site | This study |
| *27F* | AGAGTTTGATCMTGGCTC AG | Screening for 16S rRNA contamination | (Jiang et al., 2006) |
| *1492 R* | CGGTTACCTTGTTACGAC TT | Screening for 16S rRNA contamination | (Jiang et al., 2006) |
| *Downstream FliF 2 R* | CTACTGCTCTTGGGCCG | Sanger sequencing and insert detection in *Cj0046* | This study |
| *Cj0046 F* | GAGCCAATCCTATTT CATCAG | Sanger sequencing and insert detection in *Cj0046* | This study |
| *Cj0046 R* | GTTAAAACTCCACTAATC AACACC | Sanger sequencing and insert detection in *Cj0046* | This study |
| *Apra screen F* | TCAGCCAATCGACTGGCG | Sanger sequencing and insert detection in *Cj0046* | This study |
| *Apra screen R* | ATGCAATACGAATGGCGA AAAG | Sanger sequencing and insert detection in *Cj0046* | This study |
| *porAseqR* | GAGAATTCTCCTTGTCAA AAATTAAT | Sanger sequencing and porA insert detection | This study |
| *prom_seqR* | CATATCCATTATTTGCATT ACTACAC | Sanger sequencing and porA insert detection | This study |

Plasmids were extracted with the QIAprep Spin Miniprep Kit (Qiagen) or Monarch Plasmid Miniprep Kit (NEB) according to manufacturers' instructions. Vectors for amplification in *E. coli* were transformed as follows: 5-10 ng of plasmid were gently mixed with 50µl competent cells (pre-thawed on ice). Transformation mix kept for 30 min on ice, then heat-shocked for 30s in 42˚C. 950µl RT LB medium added,

incubated at 37˚C for 1h while rotating. Plated on LB plated (50µl, 100µl and 150µl to achieve different colony densities).

***Table 2.4*. Plasmids used in this study.**

| *Plasmid* | *Description* | *Origin* |
|---|---|---|
| *pC46-fdxA* | Vector containing *Cj0046* sequence for homologous recombination, $Cat^R$ cassette under native promoter, fdxA promoter with no ORF | Created by Emily Stoakes from pC46 (APPENDIX E) |
| *pMK-RQ del_FliF (GeneArt construct)* | Deletion of *fliF* in *C. jejuni* with homologous recombination, $Cat^R$ cassette | Created by Emily Stoakes |
| *pC46-fdxA-AprR* | Vector containing *Cj0046* sequence for homologous recombination, $Cat^R$ cassette under native promoter, fdxA promoter with $Apr^R$ casette | This study |
| *pC46-fdxA-AprR-natfliF* | Vector containing *Cj0046* sequence for homologous recombination, fdxA promoter with $Apr^R$ cassette, $Cat^R$ cassette replaced with *fliF* under native promoter | This study |
| *pC46-fdxA-AprR-porA-fliF* | Vector containing *Cj0046* sequence for homologous recombination, fdxA promoter with $Apr^R$ cassette, $Cat^R$ cassette replaced with *fliF* under porA promoter | This study |
| *pC46-fdxA-AprR-natfliF*$_{M99I}$ | Vector containing *Cj0046* sequence for homologous recombination, fdxA promoter with $Apr^R$ cassette, $Cat^R$ cassette replaced with *fliF*$_{M99I}$ under native promoter | This study |

| | | |
|---|---|---|
| *pC46-fdxA-AprR-porA-fliF$_{M99I}$* | Vector containing *Cj0046* sequence for homologous recombination, fdxA promoter with *Apr$^R$* cassette, *Cat$^R$* cassette replaced with *fliF$_{M99I}$* under porA promoter | This study |

### 2.2.4. Transformation of *Campylobacter jejuni*

*C. jejuni* cells of the necessary genotype (NTCTC11168, *ΔflhF, ΔfliF* or *ΔflhFΔfliF*) were grown on an MH plate overnight and scraped off the plate with a loop, dislodging in 100 µl of ice-cold wash buffer (272mM sucrose, 15% glycerol (v/v)). The cell suspension volume was made-up to 1 ml with wash buffer and cells were washed by 20 min 4000 g centrifugation and gentle resuspension 3 times with the same buffer. The final suspension was aliquoted in 100 µl and frozen in liquid nitrogen or used fresh for the following steps.

One cell-aliquot was mixed with 700 to 900 ng of plasmid in ice-cold electroporation cuvette (0.2 cm) and electroporated at 2.5 kV, 200 Ω, 25 µF (Bio-Rad Gene pulser 0.2). The cells were plated out on antibiotic-free MH plates using either 90 µl of cells or 10 µl mixed with 90 µl of MH broth. After overnight incubation at 37˚C under microaerophilic conditions (VAIN), plates were washed with 2 ml MH broth and cells were dislodged with a loop. 100 µl of the cell slurry (or a 10x dilution: 10 µl+ 90 µl MH broth) was plated on reduced antibiotic plates (10 µg/ml chloramphenicol or 26 µg/ml apramycin). Incubation in microaerophilic cabinet (VAIN) at 37˚C for 3-5 days was followed by picking single colonies and streaking them on standard antibiotic concentration plates. Clones were screened for the presence of the insert with GoTaq whole cell PCR, targeting an amplicon within each one of the antibiotic resistance genes.

### 2.2.5. Screening and verification of *C. jejuni* strains

Genomic DNA (gDNA) was extracted with QIAgen Blood and Tissue Kit (Qiagen) with manufacturer's modifications for use on Gram-negative bacteria. The concentration and purity were measured with both NanoDrop (Thermo Scientific) and as well as the Qubit dsDNA broadrange assay kit (ThermoFisher Scientific). *ΔfliF* and

*ΔflhFΔfliF* strains were verified by amplifying *fliF* and flanking regions encompassing the entire homologous recombination region followed by Sanger sequencing to verify that upstream and downstream genes were not affected. The *Cj0046* region encompassing the integration region could be amplified, but could not be reliably Sanger sequenced. After gDNA extraction, samples with a low 260/230 absorbance ratio that indicates organic compound contamination that can potentially interfere with amplification reactions were further purified performed using QIAamp Mini spin columns and QIAgen clean-up protocol for QIAamp MinElute columns (Qiagen). DNA was eluted typically in 30 µl Elution Buffer (EB) even in protocols suggesting 50 µl in the instructions, to increase the DNA concentration.

Whole genome sequencing was performed to verify correct integration of sequences in the chromosome and identify any additional mutations that may affect the phenotype. Uniquely barcoded libraries for Illumina short-read sequencing were prepared with the Nextera XT library prep kit (Illumina) according to manufacturer's instructions. Library quality was checked on select samples with Agilent Bioanalyzer high sensitivity DNA kit and concentration of all libraries was determined using QuBit 2.0 broad-kit range. Concentration of all libraries were normalise and pooled at equimolar concentrations and sequenced on MiSeq V2 using 2x250bp PE protocol (Illumina).

MultiQC (FastQC analysis for multiple samples) was used for sequencing data quality checks (Ewels et al., 2016). Bowtie2 was used to map the fastq files to the reference genome using the "very sensitive" parameter (Langmead and Salzberg, 2012). Sequencing data was then analysed with mpileup and VarScan2 SNP (single nucleotide polymorphism) and INDEL tools (Koboldt et al., 2009) . A minimum coverage of 15, minimum average quality of 30, and minimum frequency 0.9 was selected in VarScan2 search. The Artemis Genome viewer was used to find deletions (Carver et al., 2012).

## 2.3. Cryo-electron microscopy and electron microscopy data collection

### 2.3.1. Cryo-electron microscopy grid preparation

Sample optimization for CryoEM was performed with fixed and live cells varying the type of grid and amount of gold particles. *C. jejuni* was cultured as in section 2.2.2 either until an $OD_{600}$ (Optical density) of 1.2 (early stationary) was reached or late stationary cells were used (collected 2 hours after reaching early stationary phase).

Fixed cells were prepared by fixing a cell amount equivalent to $OD_{600}$=1 in a final concentration of 2.5% glutaraldehyde/PBS (phosphate-buffered saline) and after incubation at RT (room temperature) for 30 min, washed and resuspended with cold PBS twice at 2000 g for 5 min. A varying amount of PBS was used for the final resuspension to change and assess cell density on grids. Wide bore tips were used for all steps to avoid breaking flagella structures. Live cells were grown to early stationary phase and either plunge frozen immediately in liquid ethane, or spun down at 2000 g for 5 min and resuspended in PBS prior to freezing to increase concentration.

Two different kinds of gold particles were tested – 10 nm gold nanoparticles with OD=1 (Sigma Aldrich) and protein-conjugated gold (Immunogold Conjugate protein A, particle size 10nm, BII Solutions) in variable ratios to sample volume.

For initial tests, Lacey carbon grids with irregular holes were used; for final samples regular hole (diameter - 3.5 µl) shaped carbon grids with 1 µm interspacing were used. All grids were glow-discharged for 1 minute before use. Samples were frozen with a with Leica GP2 rapidplunge-freezing device with 2.5 s blotting time.

Optimization was performed on an EOL 2100Plus microscope with a Gatan OneView Camera (Advanced Bioimaging Research Technology Platform (RTP), University of Warwick, and datasets for tomograms were collected at Midlands Regional Cryo-EM Facility (University of Leicaster) on Thermo Fisher Titan Krios G3. Tilt series were collected from 50.99 to -59.00 degrees (see APPENDIX B be for rawtlt values) with SerialEM software with pixel spacing of 3.462 Å. Nominal tilt axis angle 84.9, 300KV and 2.7 mm Cs (spehircal abberation). Magnification 42000.

### 2.3.2. Negative stain electron microscopy grid preparation

For flagella counting in negative stain electron microscopy, cells of each complementation strain created as described in section 2.2.6 were cultivated as in section 2.2.2 to late stationary phase. The amount of cells equivalent to $OD_{600}$=1 was taken from each sample and mixed with 600 µl 5% glutaraldehyde. These were incubated for 30 min at RT and spun down at 2000 g for 5 min. They were then washed with 1 ml PBS using wide bore tips, before being gently resuspended in 50 µl PBS. Copper grids were glow discharged before use. 10 µl of bacterial cell sample was applied on the grid with a wide bore tip and excess liquid was removed with filter paper after 1 min incubation. Uranyl acetate was added to the grid in a similar size drop to the sample and removed after 1 minute with filter paper. This was additionally dried by placing face down on another filter paper. All samples were prepared both in the original concentration and with 10 x dilution in PBS in case the sample would be overcrowded for counting.

A minimum of 100 cells were observed for each sample in negative stain microscopy using an EOL 2100Plus microscope with Gatan OneView Camera and recorded as having flagella or having no flagella. The results were then averaged to percentage of flagellated cells. One cell end was recorded as one count, but even before the observation of results, cells with two complete flagella were expected to be very rare.

### 2.3.3. Cryo-electron microscopy data analysis

Sub-tomograms were built using Etomo from IMOD 4.11.18 suite (Kremer et al., 1996). When building the seed model for tracking, 30 fiducials were selected were possible, however WT *C. jejuni* samples had low fiducial density, therefore most samples had a lower number of fiducials. The first seed model was selected automatically and then manually additional fiducials were added after visual examination of the tilt series. Positioning tomograms were built with a thickess selection 1200 voxels. A boundary model was created on compressed (binned by 2) sample tomograms. The final tomogram thickness was around 1100 voxels. No post-processing cropping was performed.

The completed sub-tomograms were to be used as input for model building in Relion. CTF estimation was run with CTFfind 4 (Rohou and Grigorieff, 2015).

In addition sub-tomogram averaging and model building was attempted with EMAN2 2.31. Intial runs did not give a result with no symmetry (C1) imposed, therefore (C39) was selected (see Chapter 5).

The tilt series were imported with apix 3.462Å.

The settings for automated tomogram reconstruction were as follows: npk=20 tltkeep=0.9 outsize=1k niter=2,1,1,1 bytile (on) notmp (on) pkkeep=0.9 clipz=-1 bxsz=32 pk_mindist=0.125 filterto=0.45. The rawtlt file was specified instead of selecting tltstep. Boxsize here refers to box used to search for fiducials.

CTF estimation settings were: dfrange=0.2,2,0.1 psrange=60,120,2 tilesize=256 voltage=300 cs=2.7 nref=15 stepx=20 stepy=40

Particles to reconstruct were selected by manually placing the selection circle over the centre of the motor, with size 168 (later determined to be too small).

Particles were extracted with: boxsz_unbin=168 maxtilt=100 padtwod=2.0 wiener (on) shrink=1 tltkeep=1.0 alioffset=0,0,0.

Intial model generation used: sym=c39 gaussz=-1.0 filterto=0.02 fourier (on) learnrate=0.1 niter=5 applysym (on) shrink=2 (applied because of computational resource limitations). The obtained model was used as input for 3D reconstruction.

3D model generation used: niter=5 sym=c39 mass=11000000.0 (very rough estimate) localfilter (on) goldstandard=-1 pkeep=0.8 maxtilt=90.0. The 3D model was used as input for subtilt refiemenent.

Subtilt refinement used: niters=4 padby=2.0 keep=0.5 maxalt=90.0

Upon specifying symmetry C33 or C39, low resolution models were obtained. Model building was limited by computational limitatations (RAM size). C33 models were reconstructed with apix 0.3462Å instead of 3.462Å and resulted in wrong size models, not directly comparable with other data and are thus not further presented. The boxsize used for particle selection was later determined to be too small.

Reconstruction in newer version (2.91) of EMAN2 was performed similarly, but some options are change dor updated. Symmetry was varied (C37, C39, C42).

The tilt series were imported with apix 3.462Å.

npk=20 tltkeep=0.9 outsize=1k niter=2,1,1,1 bytile (on) notmp (on) corretctrot (on) pkkeep=0.9 clipz=-1 bxsz=32 patchtrack (on) filterres=40.0. The rawtlt file was

specified instead of selecting tltstep. Boxsize here refers to box used to search for fiducials.

CTF estimation was first run with handedness check on, that indicated that correction is necessary and tomogram reconstruction was re-run specifying tltax= -275.0.

Final CTF estimation settings after handedness correction were: dfrange=0.2,2,0.1 psrange=60,120,2 tilesize=256 voltage=300 cs=2.7 nref=15 stepx=20 stepy=40

Particles to reconstruct were selected by manually placing the selection circle over the centre of the motor, with size 336 (larger than before)

Particles were extracted with: boxsz_unbin=336 maxtilt=100 padtwod=2.0 curves=-1, curves_overlap=0.5 shrink=1 tltkeep=1.0

Intial model generation used: sym=c39 (or c37, c42) filterto=0.02 fourier (on) learnrate=0.1 niter=5 applysym (on) shrink=1. The obtained model was used as input for 3D reconstruction.

3D model generation used: niter=5 sym=c39 (or c37, c42) mass=11000000.0 (very rough estimate) goldstandard=-1 pkeep=0.8 maxtilt=90.0. The 3D model was used as input for subtilt refiemenent. Always received warning in afterwards that nomalization to mass is not possible, but solution proceeded better when providing mass.

Subtilt refinement used: niters=4 padby=2.0 keep=0.8 maxalt=45.0

# Chapter 3 - MD simulations of FliG-N/FliF-C complex

## 3.1. Introduction

FliG and FliF protein rings are central to the flagellar rotor structure and assembly, as the MS-ring (built of FliF) is the earliest assembling structure alongside fT3SS (Li and Sourjik, 2011; Morimoto et al., 2014). The binding between FliG-N and FliF-C joins the C-ring (built of FliG, FliM and FliN) below the inner membrane and MS-ring above it (Lynch et al., 2017; Noreen et al., 1994). FliG also has the crucial function in the C-ring of transmitting the torque from stator to rotor (Lloyd et al., 1996; Minamino et al., 2011). FliG has three domains FliG-N, FliG-M and FliG-C (Lee et al., 2010; Levenson et al., 2012) Interactions between FliG-N and FliF-C have been extensively studied with mutational studies (Grunenfelder et al., 2003), biochemical methods (Ogawa et al., 2014), NMR (Levenson et al., 2012) and crystallization (Lynch et al., 2017; Xue et al., 2018) (Introduction, 1.5). These studies have characterized residues important for binding in *H. pylori*, *T. maritima*, *V. alginolyticus* and *C. crescentus* (also described in sections 3.2.3., 1.5). There are two crystal structures of FliG-N binding FliF-C, one from *H. pylori* (PDB ID: 5WUJ) (Xue et al., 2018) and one from *T. maritima* (PDB ID: 5TDY) (Lynch et al., 2017).

For the *H. pylori* structure 5WUJ, the unit cell contains six identical copies of FliG-N/FliF-C complex (Figure 3.1). A single copy of FliG-N/FliF-C complex is entirely α-helical and *H. pylori* FliG consists of seven helices α1 to α7 while FliF-C consists of 2 helices (Figure 3.2).

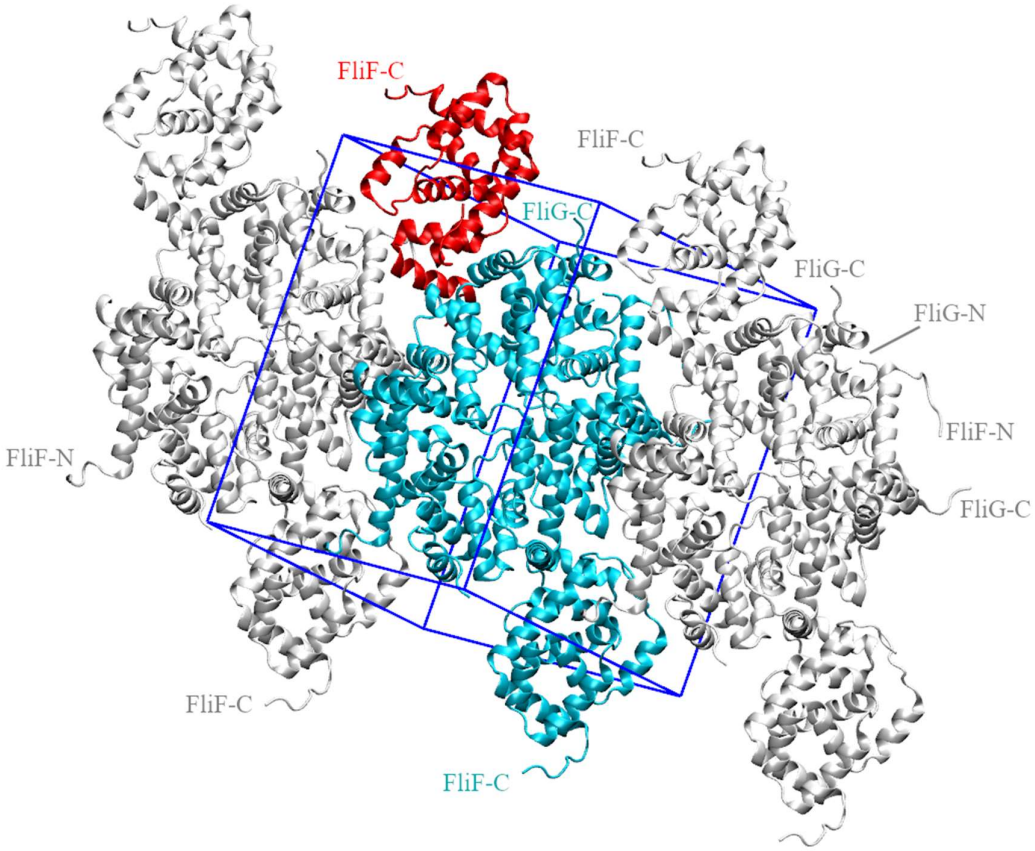



***Figure 3.1.*** **Unit cell of crystal structure 5WUJ.** Unit cell box is shown with 6 FliG-N/FliF-C complexes, one of which is coloured in red; molecules in neighbouring unit cells are shown in grey. Unit cell copies and boundary generated in VMD.

MD simulation of known crystal structures can generate otherwise unobservable new hypotheses and mechanistic insights. MD simulations have so far not been used in study of FliG-N and FliF-C and could reveal novel insights about these crucial for motility proteins. FliF unbound FliG structure (Lee et al., 2010) differs in the linker conformation between FliG-N and FliG-M from those seen in FliG-N/FliF-C structures (Lynch et al., 2017; Xue et al., 2018) as well as those proposed for alternative models (Khan et al., 2018; Paul et al., 2011b). In this study MD simulations are used to explore the conformation of FliG-N and helices that link it to FliG-M in physiologically-representative and crystal lattice conditions.

## 3.2. Molecular dynamics simulations analysis of FliG-N

Prior to simulations, analysis of residues in the 5WUJ structure which may have variable protonation state was performed using the H++ server (Gordon et al., 2005). E91 was identified as a candidate for which the protonation may change, and potentially influence the protein structure. E91 is surrounded by other negative charges in the interface between FliG α6 and α5 and FliF α2, raising the question on whether the protonation state of this residue could have potential structural effects. To explore this possibility, multiple MD simulations of the FliG-N/FliF-C complex presented throughout this chapter were performed while testing both possible E91 protonation states (Method details in sections 2.1.1, 2.1.2).

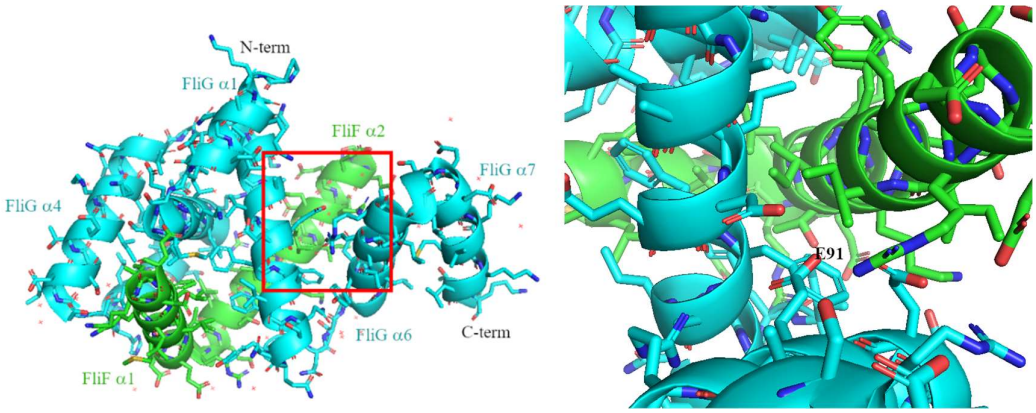



***Figure 3.2*. E91 of FliG, surrounded by nearby negative charges.** Left - 5WUJ crystal structure. Squared area shown in detail on the right. E91 could have a variable protonation state according to H++ server test. FliG (cyan), FliF (green). Protein is shown in cartoon representation while side chains as sticks.

Using visual observation and preliminary RMSD analysis of the proteins simulated in water, the protein core was defined as consisting of α2 to α5, while α6 and α7 were distinctly separate. The small α1 was too flexible and not found in sequences of many species of model organisms and pathogens (Figure 3.3) and therefore not included in the core for measurement definition purposes.

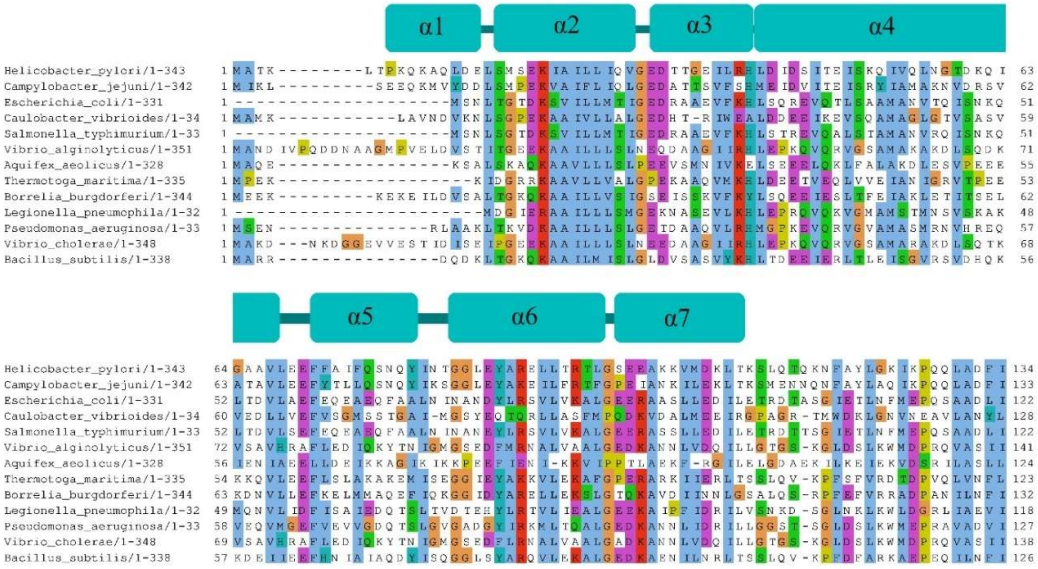



***Figure 3.3*. FliG sequence alignment and helix positions for *H. pylori* FliG-N.** Helices resolved in 5WUJ crystal structure are shown above the alignment. The figure was created using Jalview (Waterhouse et al., 2009).

After examination of the structure, the helix boundaries for purpose of RMSD analysis were set as follows (Table 3.1). Definition of helices is similar to that used in (Xue et al., 2018), but more stringent in counting a residue as part of the helix as opposed to being a linker, taking into account behaviour of the protein in simulation, where residues at the end of helices may appear part of it in the crystal structure, but may be unstable during simulations. The residue numbering matches both the actual *H. pylori* sequence (strain 26695) and the numbering in the structure deposited in the PDB (Xue et al., 2018):

***Table 3.1*. Helix boundaries used in analysis of 5WUJ.**

| *Protein* | *Helix number* | *Starting residue* | *Ending residue* |
|---|---|---|---|
| *FliG* | α-helix 1 | 7 | 15 |
| *FliG* | α-helix 2 | 17 | 29 |
| *FliG* | α-helix 3 | 31 | 41 |
| *FliG* | α-helix 4 | 42 | 56 |
| *FliG* | α-helix 5 | 60 | 80 |
| *FliG* | α-helix 6 | 84 | 97 |
| *FliG* | α-helix 7 | 99 | 111 |
| *FliF* | α-helix 1 | 524 | 540 |
| *FliF* | α-helix 2 | 544 | 559 |

These boundaries were used throughout subsequent analyses.

### 3.2.1. FliG-N conformation in unbound state

A conformation of FliG-N unbound to FliF can be seen in the FliG full-length crystal structure of *A. aeolicus* (Lee et al., 2010). FliF-C has been reported to co-fold with FliG-N in *T. maritima* (Levenson et al., 2012; Lynch et al., 2017) therefore firstly a simulation with the *H. pylori* FliF-C removed from the complex was set up for a simulation of 200 ns in triplicate to observe, if FliG-N in *H. pylori* is stable without the presence of FliF-C. A high overall RMSD (details in Methods 2.1.5.) reflecting conformational instability was observed in all three runs, with the FliG core collapsing and helices 6 and 7 being highly flexible. The protonation state of E91 did not have any effect on the structural stability. Conformations adopted by the protein core at the end of the simulation were different and did not converge to a single state (Figure 3.4).

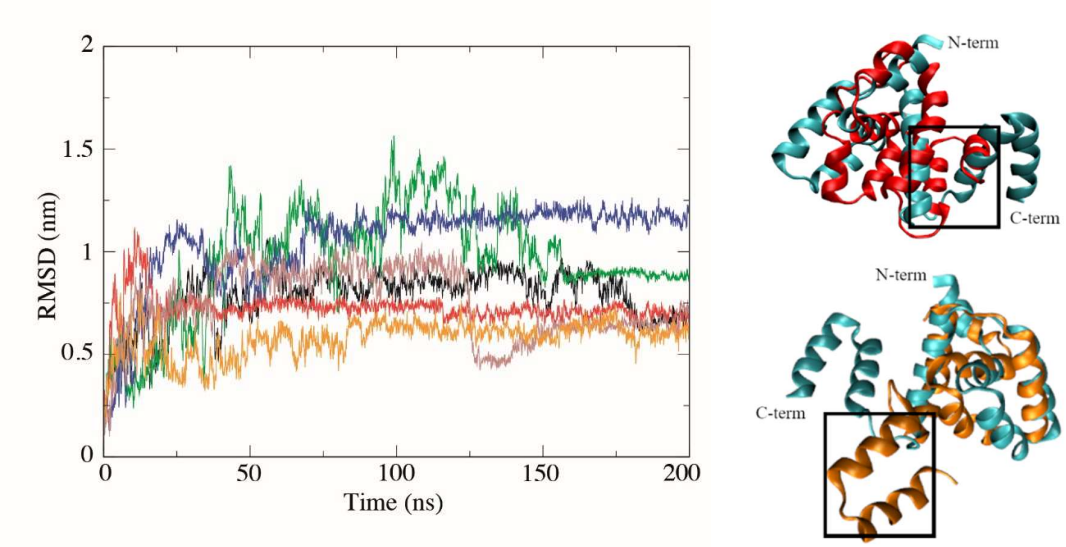



***Figure 3.4*. Structural drift of FliG in solution.** Backbone RMSD of FliG-N in the absence of FliF after fitting on protein backbone. System used CHARMM36m force field; different colours represent three independent runs of E91 deprotonated (black, green, red) or protonated (blue, orange, brown). On the right, snapshots are shown for the two of the runs (red – E91 deprotonated, orange – E91 protonated) after 200 ns versus the initial structure (cyan). Box indicates position of helices 6 and 7. Protein is shown in cartoon representation.

In addition, a 100 ns simulation of *A. aeolicus* FliG full length protein was performed in triplicate (Figure 3.5) to test the stability of the only available FliF-C unbound FliG-N structure. Here the N-terminus in the context of the full-length protein

retained its conformation even in the absence of FliF, unlike the behaviour observed in FliG-N only simulations (Figure 3.4).

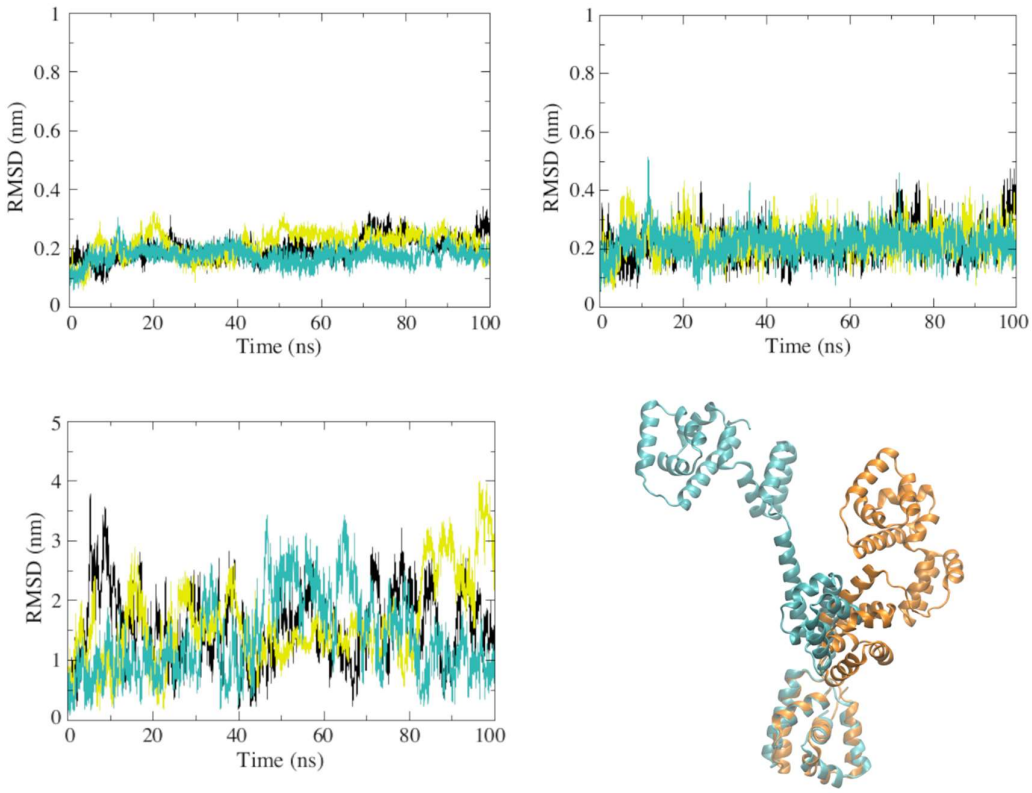



***Figure 3.5*. Structural drift of full-length *A. aeolicus* FliG (3HJL) in solution.** Backbone RMSD after fitting on backbone of protein core for: core (top left), FliG helix 2 (top right), elongated helix. System used the CHARMM36m force field, different colours represent three independent runs. 3HJL structure before (cyan) and after (orange) 100 ns simulation, aligned over residues 5 to 90 in VMD.

### 3.2.2. *H. pylori* FliG-N/FliF-C complex dynamics in solution

Next, simulations of the FliG-N/FliF-C complex were extended to 500 ns as convergence was not observed in the first 100 ns. During this time frame, the FliG-N core (helices 2 to 5) and both FliF helices showed consistent stability, while helices 6 and 7 remained flexible and did not converge into a fixed position during any of the three replicas (Figure 3.6). The high RMSDs of helices 6 and 7 contributed to the overall high RMSD of the backbone, while the core remained stable (Figure A, Appendix A), showing a distinctly different behaviour between the FliG-N globular part and the helices that are to serve as linker with FliG-M domain.

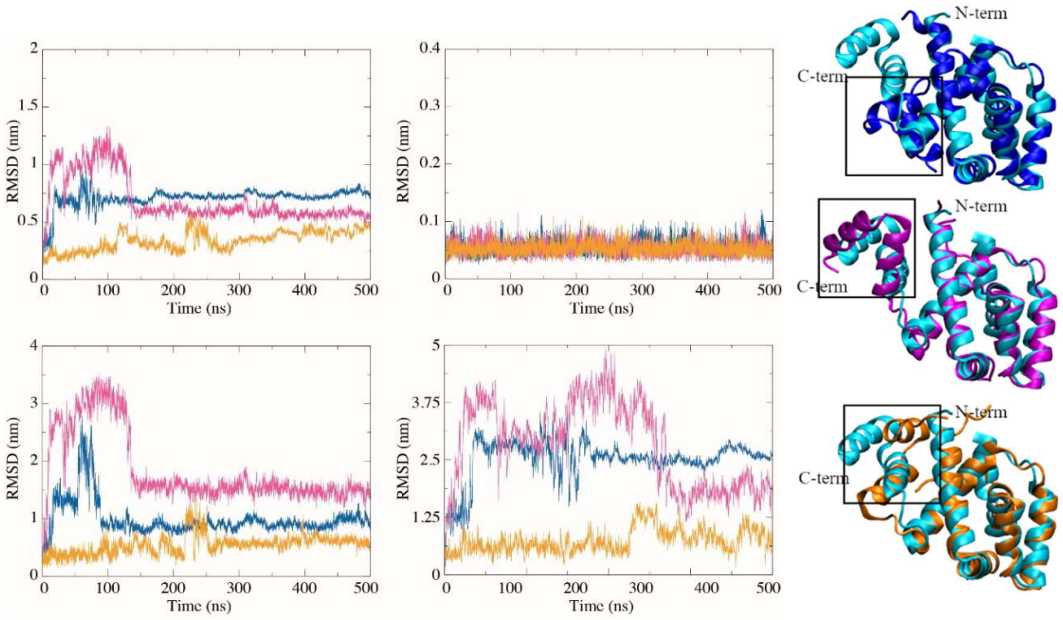



***Figure 3.6*. Structural drift of 5WUJ complex in solution.** Backbone RMSD after fitting against backbone of protein core for: whole protein (top left), FliG helix 2 (top right), FliG helix 6 (bottom left), and FliG helix 7 (bottom right). Different colours represent three independent runs. On the right, snapshots are shown for the three runs (blue, magenta, orange) after 500 ns versus the initial structure (cyan). Box indicates position of helices 6 and 7. CHARMM36m forcefield, E91 deprotonated.

In addition to RMSD analysis, a crossing angle between helices 6 and 7 was defined using the first and last Cα atom of each helix as an additional measure to observe the dynamics of these elements. For comparison, a crossing angle for helices 3 and 4 was defined in the same way (Figure 3.7). Angles in the crystal structure for the helical pairs were -49° and -44°, respectively. In the three runs of simulation, the average angle over the whole simulation for α6 to α7 was -54±10°, -51±14° and -50±19° and for α3 to α4 it was -42±5°, -41±6° and -42±5°. In the case of α6-α7, the larger standard deviation reflects higher flexibility, as opposed to α3-α4 that as part of the FliG-N core maintain the same position.

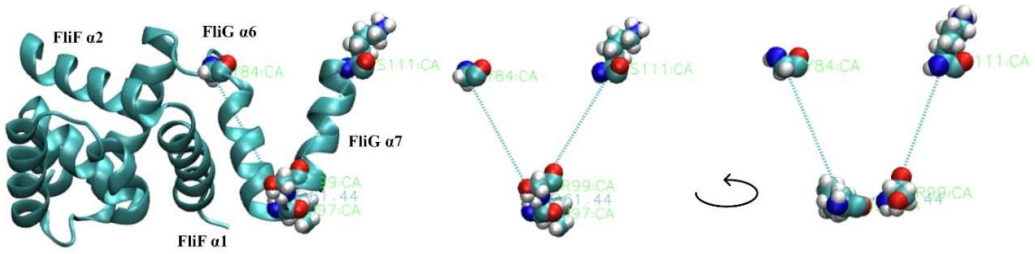



***Figure 3.7*. Example of dihedral selection and measurement in VMD**. Left FliG-N/FliF-C complex after first 100 ns (CHARMM36m, E91 deprotonated) in cartoon representation and first and last amino acid of each helix in VdW sphere representation. Middle – only amino acids selected shown. Right - rotated view of the same selection to show all four points of the dihedral measured. Cα of each amino acid used as measurement point.

The same setup with E91 protonated in three runs produced three different behaviours and helices 6 and 7 again contributes significantly to the overall backbone RMSD (Figure 3.8).

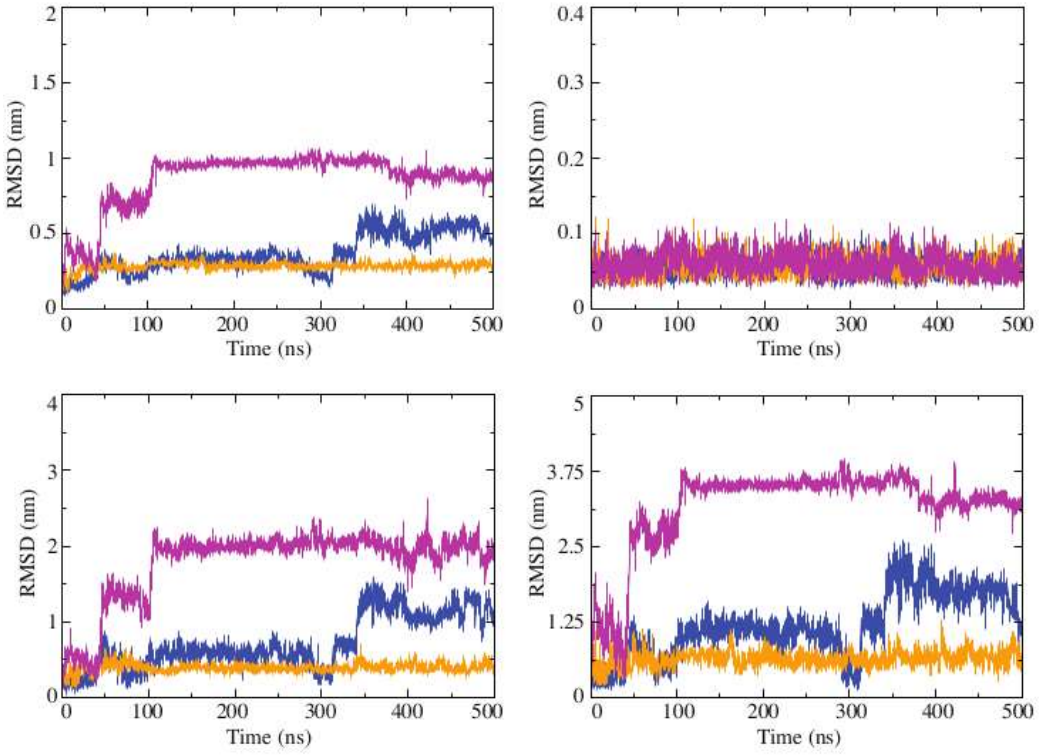



***Figure 3.8*. Structural drift of 5WUJ complex in solution.** Backbone RMSD after fitting against backbone of protein core for: whole protein (top left), FliG helix 2 (top right), FliG helix 6 (bottom left), and FliG helix 7 (bottom right). Different colours represent three independent runs. On the right, snapshots are shown for the three runs (blue, magenta, orange) after 500 ns versus the initial structure (cyan). Box indicates position of helices 6 and 7. CHARMM36m forcefield, E91 protonated.

### 3.2.3. *H. pylori* FliG-N/FliF-C complex dynamics in tetramer

Investigation of crystal contacts of the 5WUJ crystal structure within the unit cell revealed extensive hydrophobic contacts between helices 6 and 7 of one molecule and helices 3 and 4 of the neighbouring one (Figure 3.9). The interface is mainly formed of hydrophobic contacts with a tightly packing hydrophobic core made of T34, L38, T47, I53, V54, A89, L93, L97, V105, L109, flanked by two salt bridges K108-D44, R39-T96 and possibly E31-R95 given that side chains can rotate in solution (Xue et al., 2018). In conjunction with the instability of helices 6 and 7 in solution during simulation as described above, this revealed the possibility that this conformation may be enforced by crystal contacts.

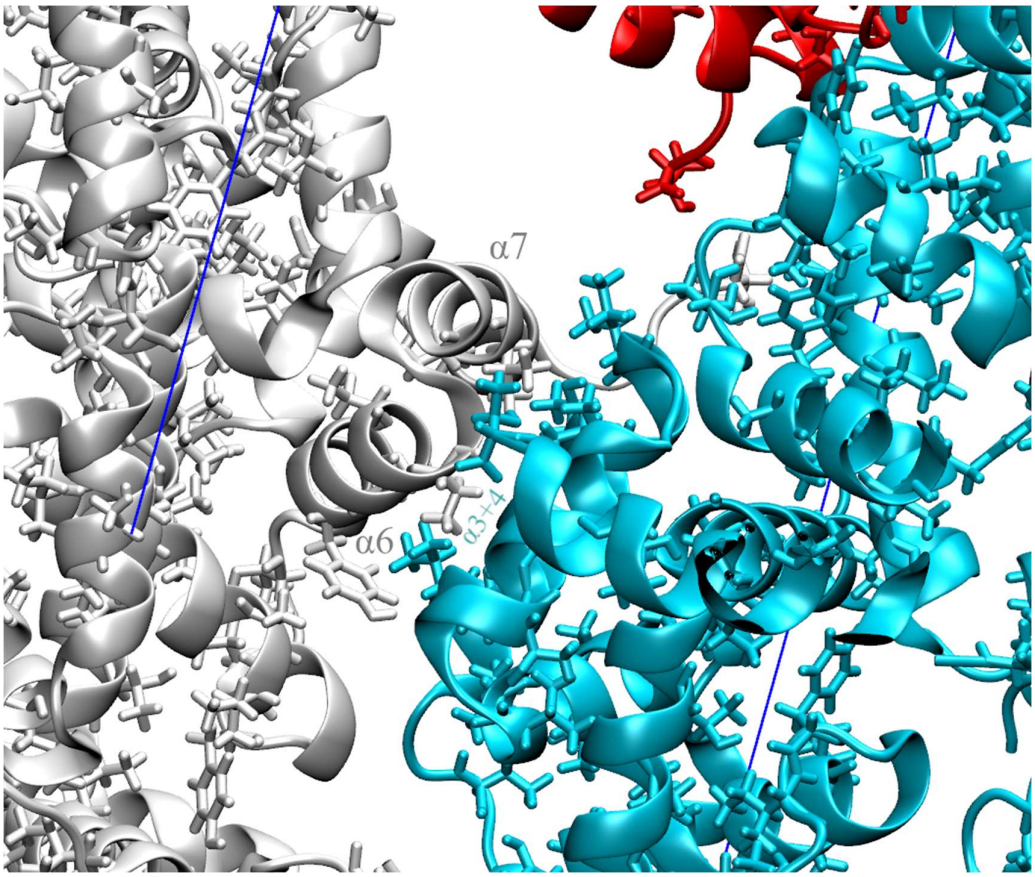



***Figure 3.9*. Hydrophobic crystal contacts between neighbouring crystallographic unit cells in 5WUJ.** Helices 6 and 7 of a grey FliG molecule pack against helices 3 and 4 of cyan FliG molecule. Blue lines are unit cell boundaries as in Figure 3.1. Protein is shown in cartoon representation while hydrophobic residues found in the interface shown in stick representation.

Initially, to test this theory a tetramer system was set up from the crystallographic asymmetric unit assembly (FliG-N/FliF-C$^{tetr}$) (Figure 3.10). This system contained three copies of α6-α7 against α3-α4 interfaces of mainly hydrophobic interactions as seen in Figure 3.9, but would not directly replicate unit cell periodicity and was performed in a solution. The outer tetramers of the 4 complex array increasingly destabilized during a 200 ns simulation while the middle interface remained more stable (Figure 3.10). There was no clear pattern regarding E91 protonation state or the influence of force field (both CHARMM36m and AMBER14SB were tested) (Figure 3.11 to 3.14). Since all systems showed an increasing RMSD trend during the runs, this suggested that the tetrameric array is not stable and is likely to further destabilize over long time scales. Thus, a new setup was sought in order to test stability in the context of the crystallographic environment, as described below.

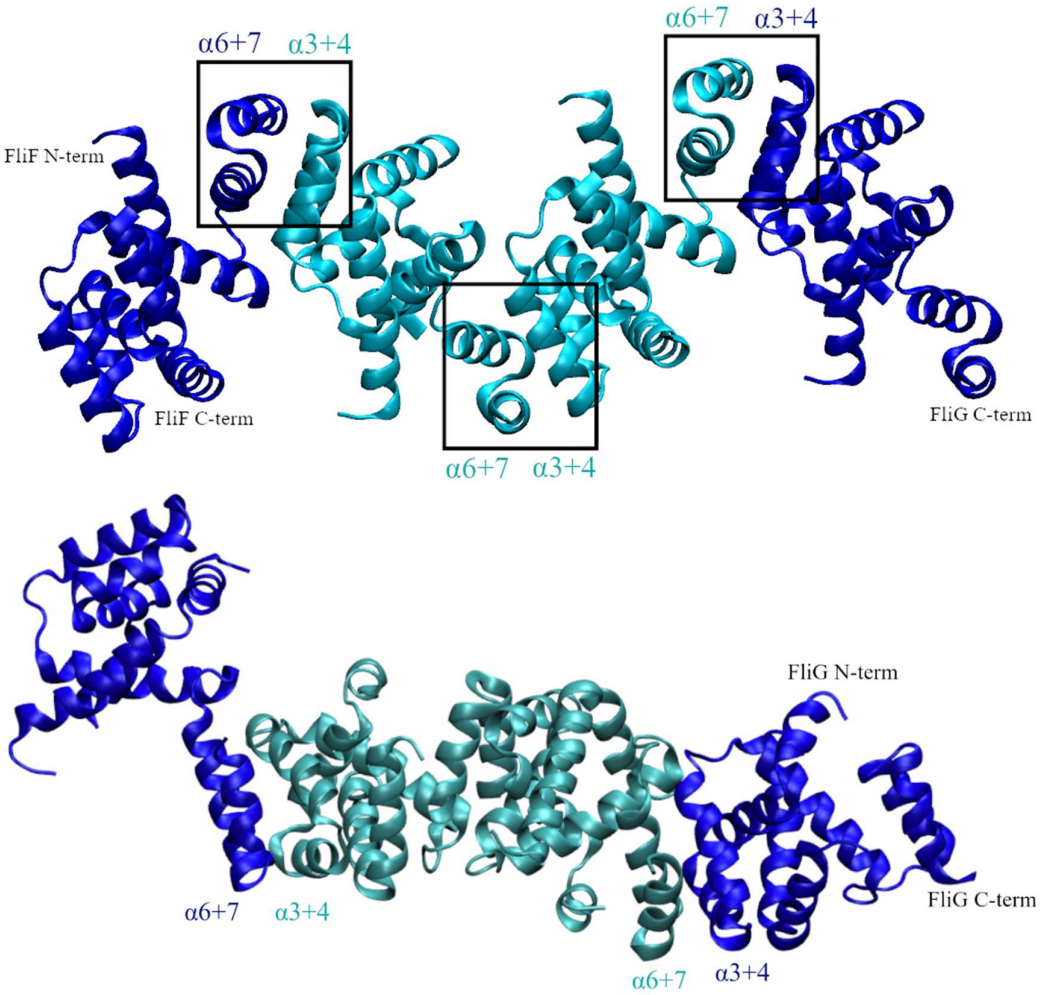



***Figure 3.10*. FliG-FliF crystal contact tetramer.** Top – FliG-N/FliF-C$^{tet}$ before simulation with packing areas between complexes indicated in boxes. Bottom –

simulation after 200 ns, using the CHARMM36m forcefield, with E91 deprotonated. Outer two monomers dark blue, inner two monomers cyan. The core of the outermost left monomer became destabilized. Protein is shown in cartoon representation.

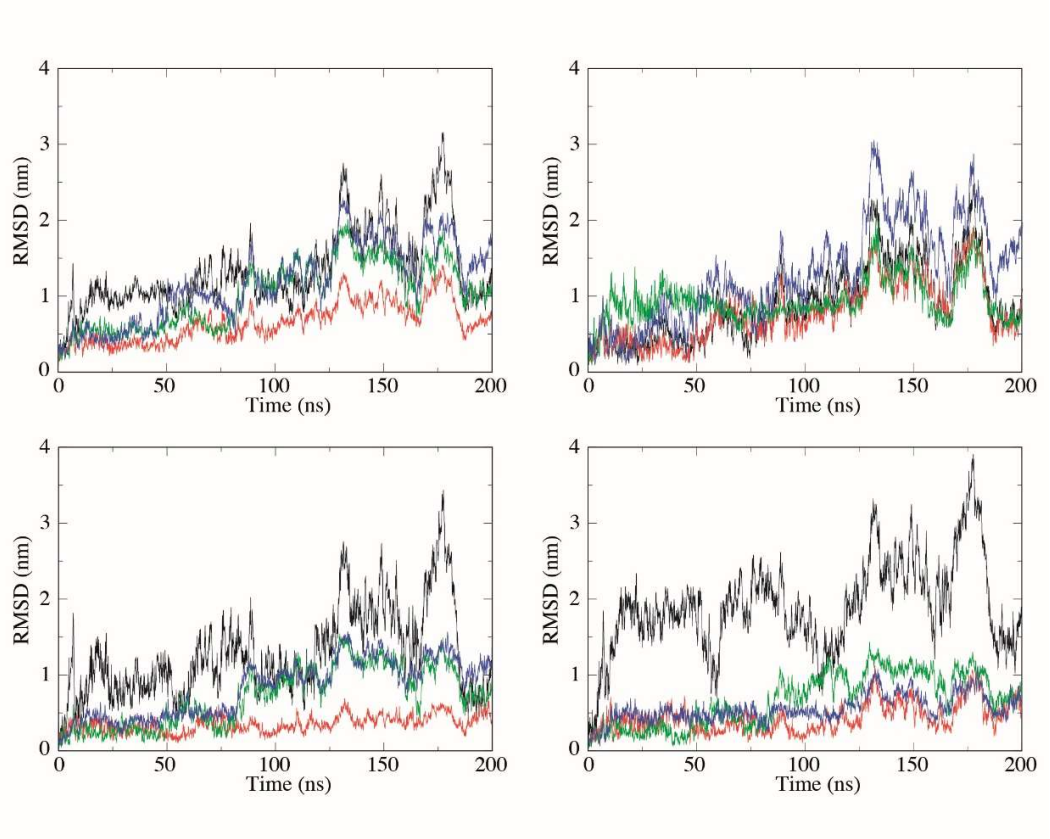



***Figure 3.11*. Structural drift of FliG-N/FliF-C$^{tet}$.** Backbone RMSD after fitting on backbone of protein core for: whole protein (top left), FliG helix2 (top right), FliG helix6 (bottom left), and FliG helix 7 (bottom right). System used the CHARMM36m force field with E91 deprotonated; different colours represent each FliG-FliF complex in the system.

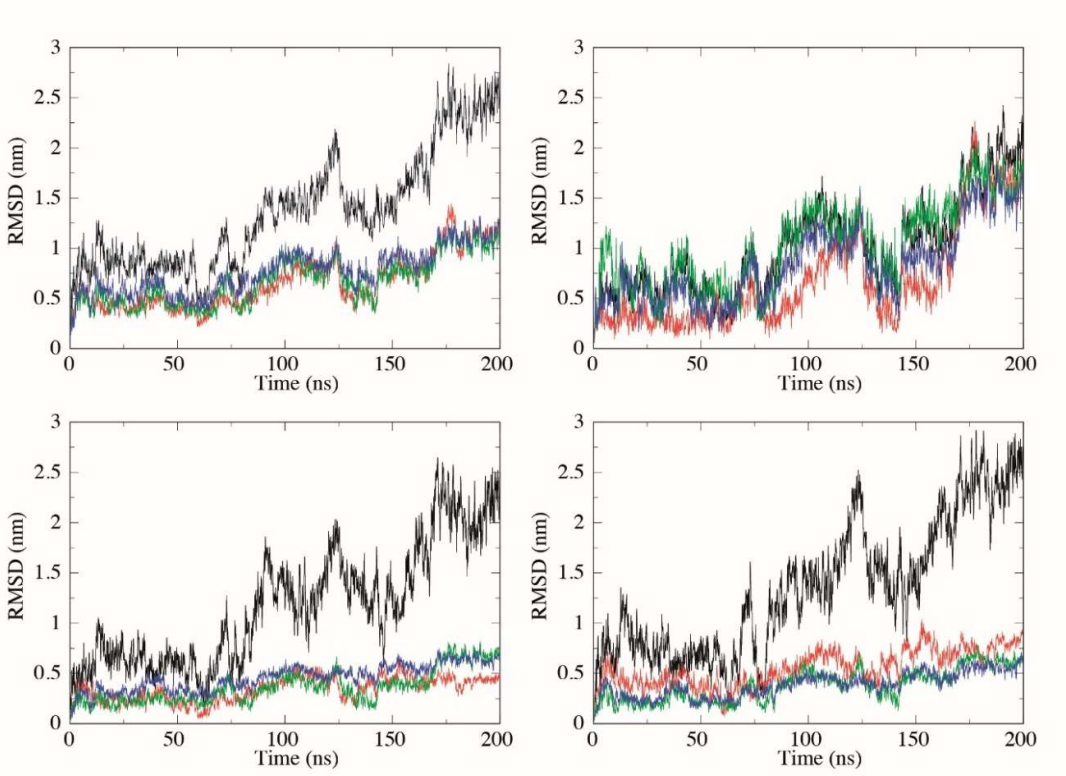



***Figure 3.12*. Structural drift of FliG-N/FliF-C$^{tet}$.** Backbone RMSD after fitting on backbone of protein core for: whole protein (top left), FliG helix2 (top right), FliG helix6 (bottom left), and FliG helix 7 (bottom right). System used the CHARMM36m force field with E91 protonated; different colours represent each FliG-FliF complex in the system.

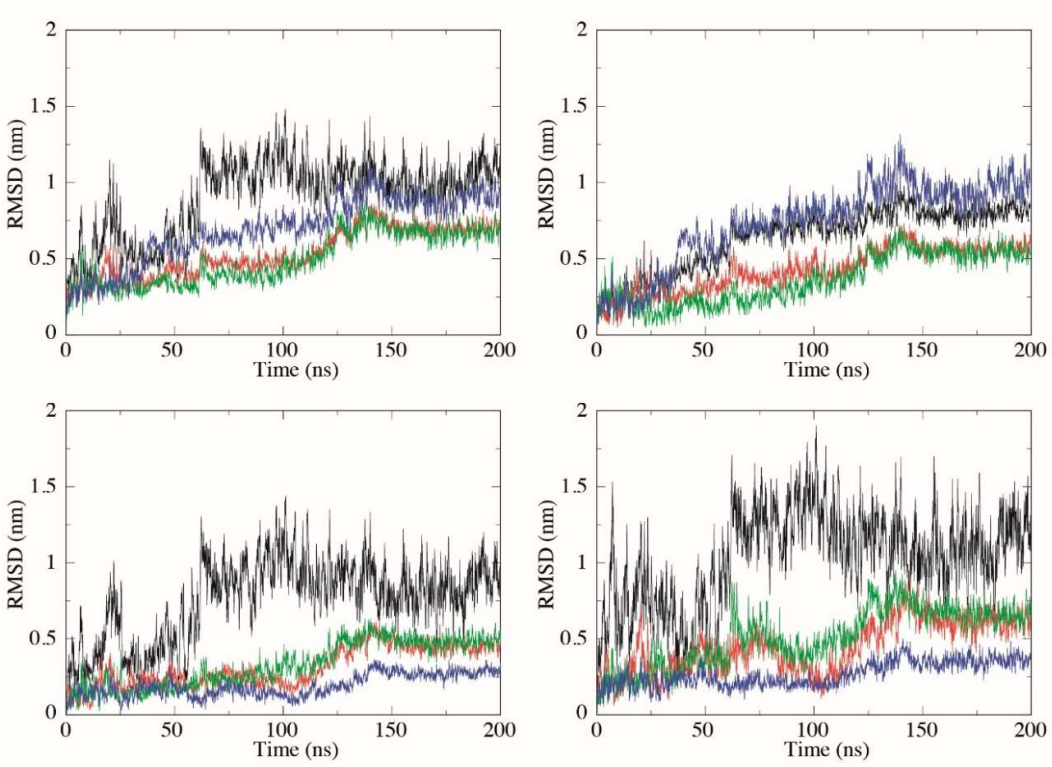



***Figure 3.13*. Structural drift of FliG-N/FliF-C$^{tet}$.** Backbone RMSD after fitting on backbone of protein core for: whole protein (top left), FliG helix2 (top right), FliG helix6 (bottom left), and FliG helix 7 (bottom right). System used the AMBER14SB force field with E91 deprotonated; different colours represent each FliG-FliF complex in the system.

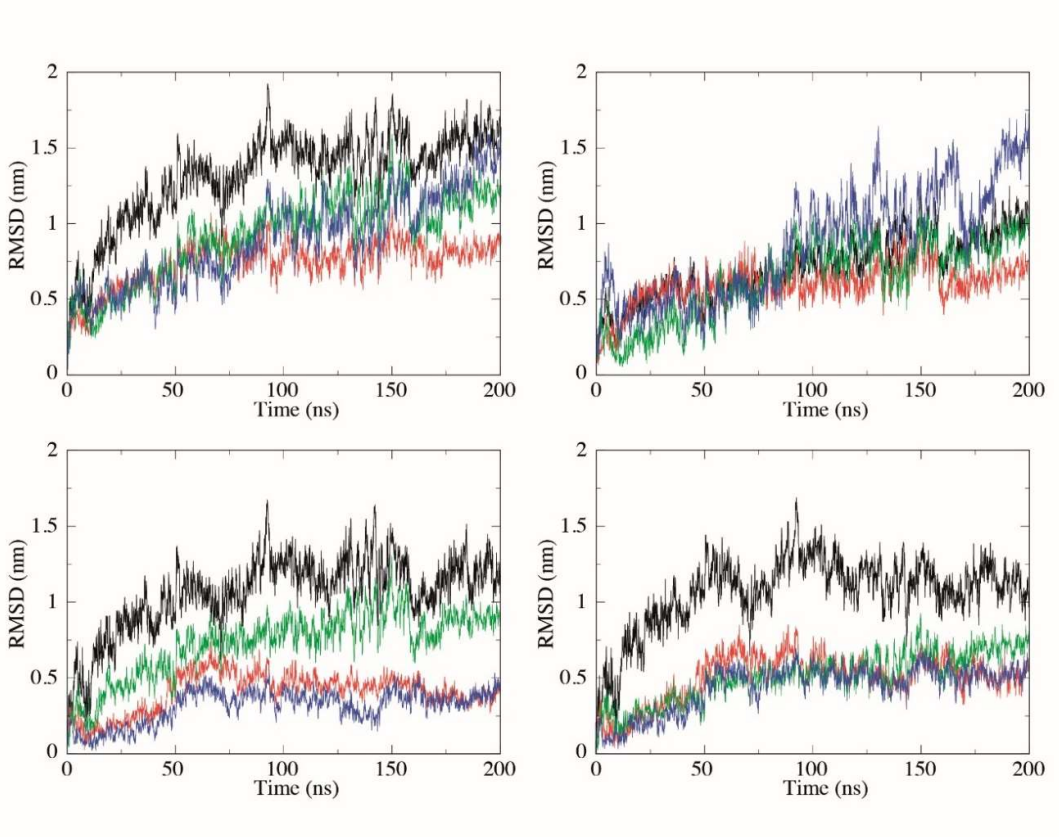



***Figure 3.14*. Structural drift of FliG-N/FliF-C$^{tet}$.** Backbone RMSD after fitting on backbone of protein core for: whole protein (top left), FliG helix2 (top right), FliG helix6 (bottom left), and FliG helix 7 (bottom right). System used the AMBER14SB force field with E91 protonated; different colors represent each FliG-FliF complex in the system.

### 3.2.4. *H. pylori* FliG-N/FliF-C complex dynamics under crystal conditions

A simulation replicating unit cell conditions was set up (in E91 protonated/ deprotonated variations). Unit cell periodicity was fully reproduced and contained six copies of FliG-N/FliF-C. Stability was significantly increased overall in the unit cell setup as compared to the single FliG-N/FliF-C copy in solution and the tetramer in solution and particularly in the previously flexible helices 6 and 7 confirmed this (RMSD $\leq$ 0.3 nm) (Figure 3.15).

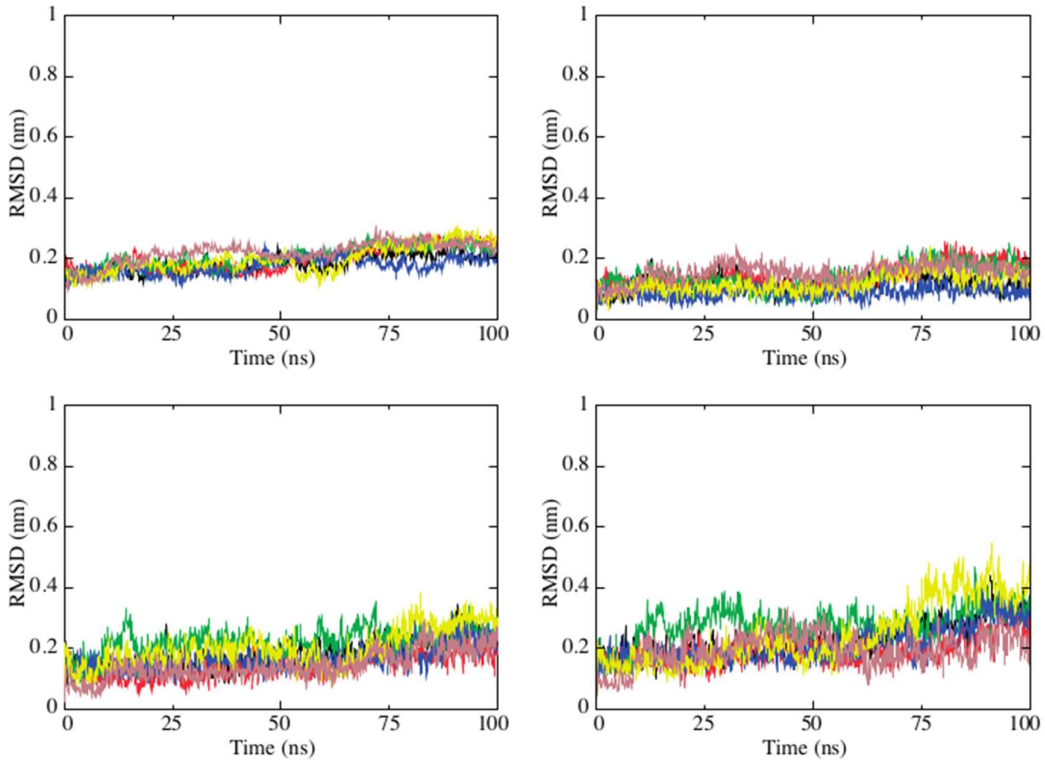



***Figure 3.15*. Structural drift of FliG-N/FliF-C$^{cryst}$ in the unit cell.** Backbone RMSD after fitting against backbone of protein against whole protein (top left) and backbone of protein core for: FliG helix2 (top right), FliG helix6 (bottom left), and FliG helix 7 (bottom right). System used the CHARMM36 force field with E91 deprotonated; different colours represent each FliG-FliF complex in the unit cell.

Using the same approach of crossing angle measurement of helix pairs (as for FliG-N/FliF-C complex in solution) under simulated crystallographic conditions, α6 to α7 had average values of -53±3°, -48±4°, -57±7, -49±3°,-54±4°,-54±6° in each copy of the unit cell, while the α3-α4 pair had average values of -46±4°, -46±5°, -47±4°, -40±4°, -45±4°, and -48±4° for each copy. Standard deviations were comparable among all helices, and were minimal, confirming the comparable stability of these structural elements is due to crystal contacts.

Finally, it is noteworthy that the protonation state of E91 did not have an influence on stability and the trend remained the same (RMSD ≤ 0.3 nm) (Figure 3.16).

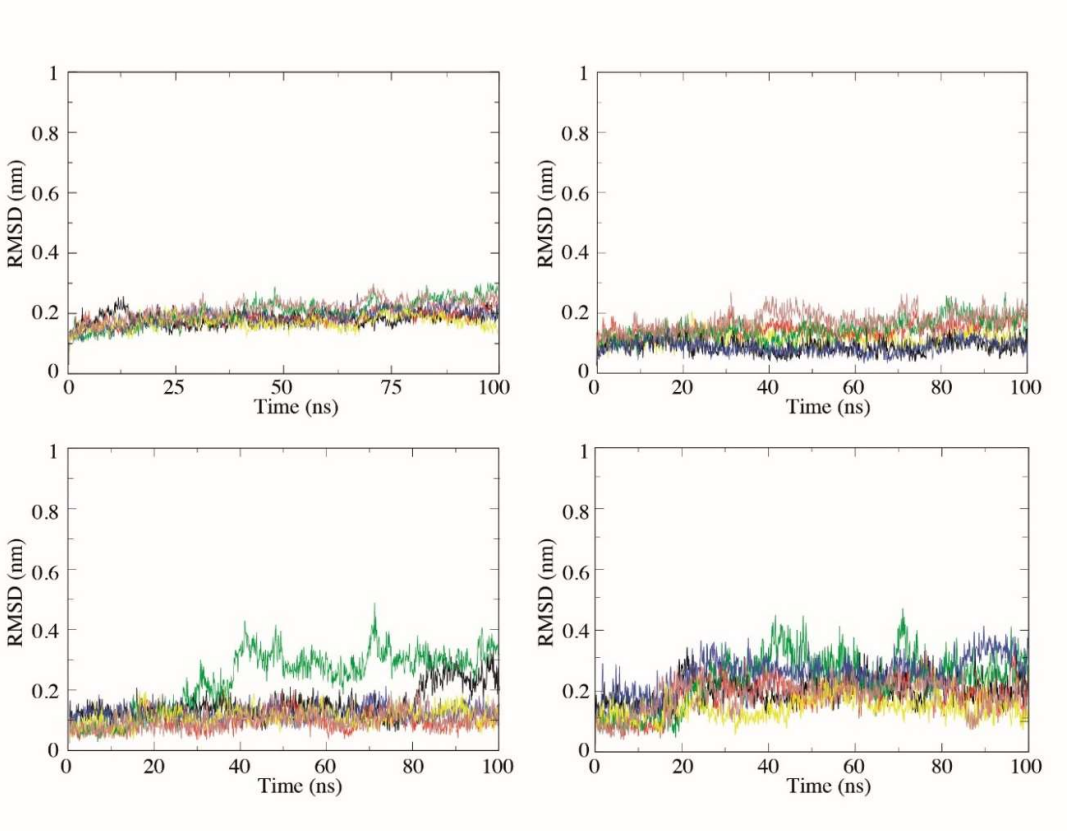



***Figure 3.16*. Structural drift of FliG-N/FliF-C$^{cryst}$.** Backbone RMSD after fitting on backbone of protein core for: whole protein (top left), FliG helix2 (top right), FliG helix6 (bottom left), and FliG helix 7 (bottom right). System used the CHARMM36m force field with E91 protonated; different colours represent each FliG-FliF complex in the unit cell.

Furthermore, to confirm that the force field was not influencing the complex stability, additional simulations were run for the same system using AMBER14SB in both E91 protonation states; these again showed very similar trends (Figure 3.17, 3.18).

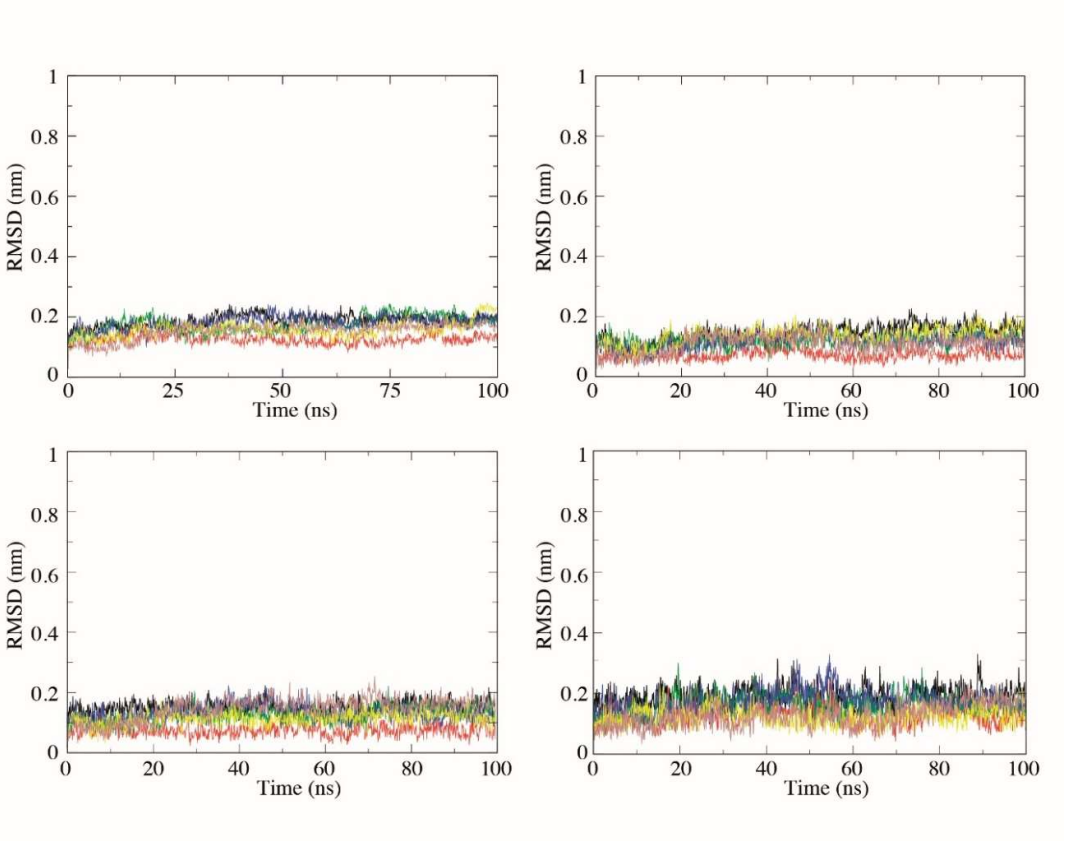



***Figure 3.17.*** **Structural drift of FliG-N/FliF-C$^{cryst}$.** Backbone RMSD after fitting against backbone of protein core for: whole protein (top left), FliG helix2 (top right), FliG helix6 (bottom left), and FliG helix 7 (bottom right). System used the AMBER14SB force field with E91 protonated; different colours represent each FliG-FliF complex in the unit cell.

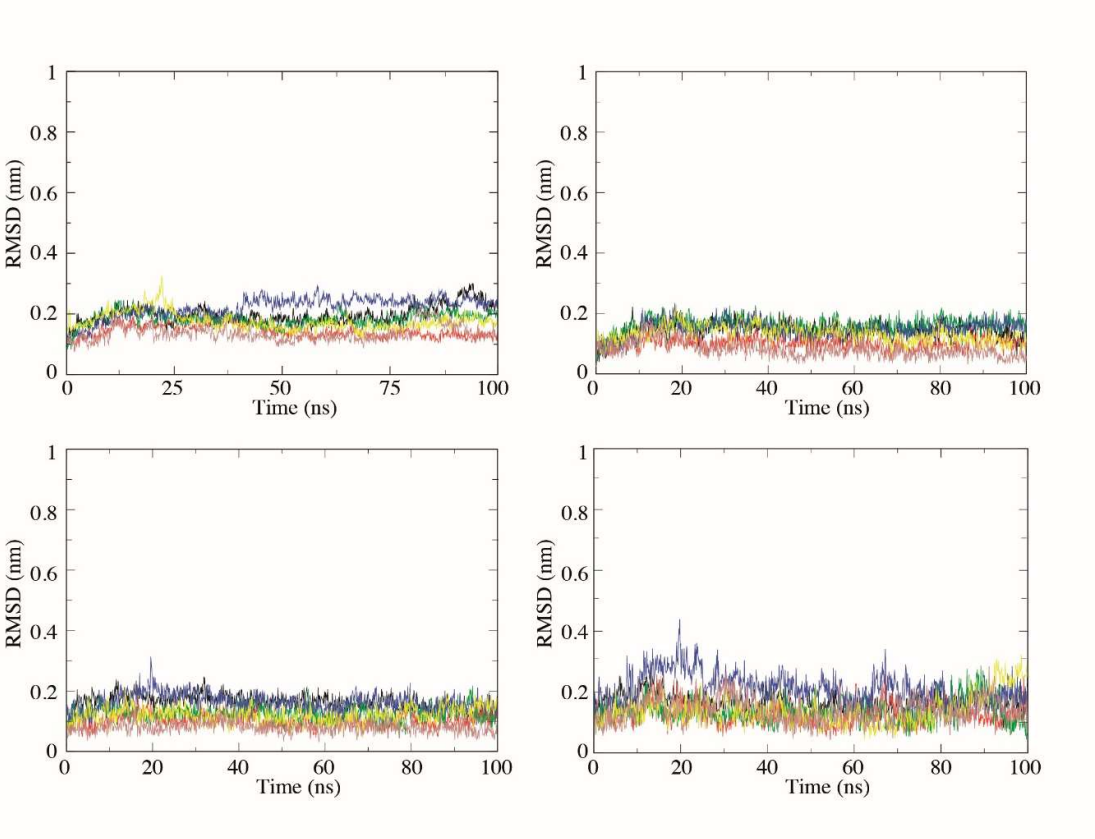



***Figure 3.18*. Structural drift of FliG-N/FliF-C$^{cryst}$.** Backbone RMSD after fitting on backbone of protein core for: whole protein (top left), FliG helix2 (top right), FliG helix6 (bottom left), and FliG helix 7 (bottom right). System used the AMBER14SB force field with E91 deprotonated; different colours represent each FliG-FliF complex in the unit cell.

Not all residues that are expressed in a construct appear resolved in the crystal structure, and often flexible ends that are present in a protein sample are not visible in the structure. To verify if expressed, but unresolved residues at both ends of the FliG-N/FliF-C complexes may have any influence on the unit cell stability, residues 1 to 6 and 112 to 115 of FliG, and residues 518 to 523 and 559 to 567 of FliF respectively (Xue et al., 2018) were built with the ModLoop web server (Fiser and Sali, 2003) (Figure 3.19. This system performed comparably with the original unit cell system as generated from data deposited in the PDB:5WUJ entry, in both forcefields (CHARMM36m and AMBER14SB and protonation states of E91 (Figures 3.20 to 3.23), to check for force-field dependency. The added residues were flexible, but the rest of the structure remained very similar to the starting structure (Figure 3.19).

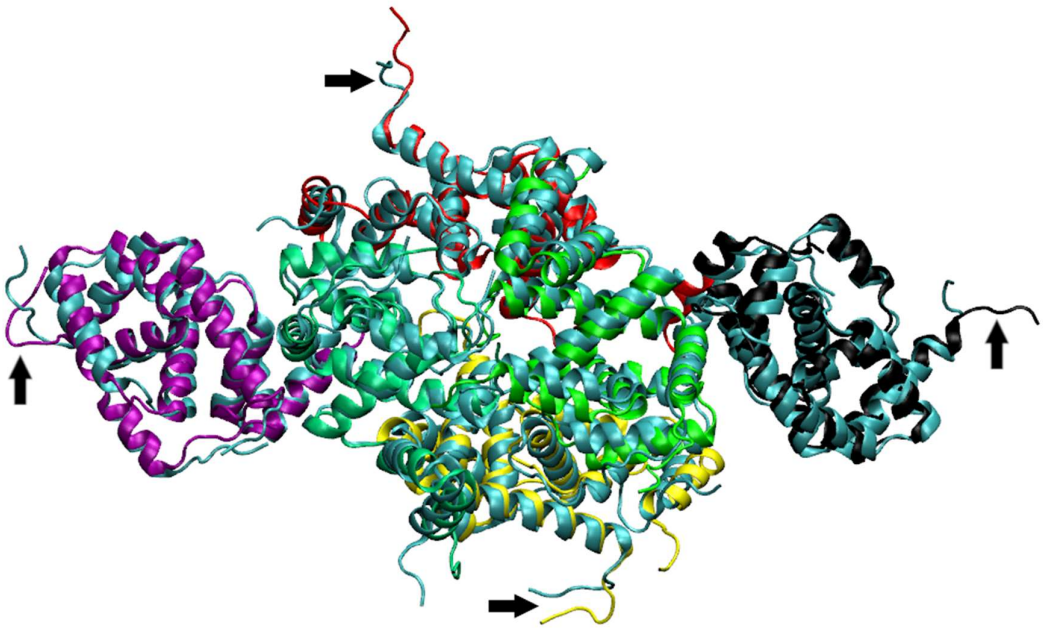


***Figure 3.19*. FliG-N/FliF-C$^{expr}$ after 100ns**. Starting structure – cyan. Copies of FliG-N/FliF-C$^{expr}$ in unit cell after simulation – purple, pale green, red, green, yellow, black. Arrows point to examples of the added residues at the end of both chains. E91 deprotonated, CHARMM36m forcefield.

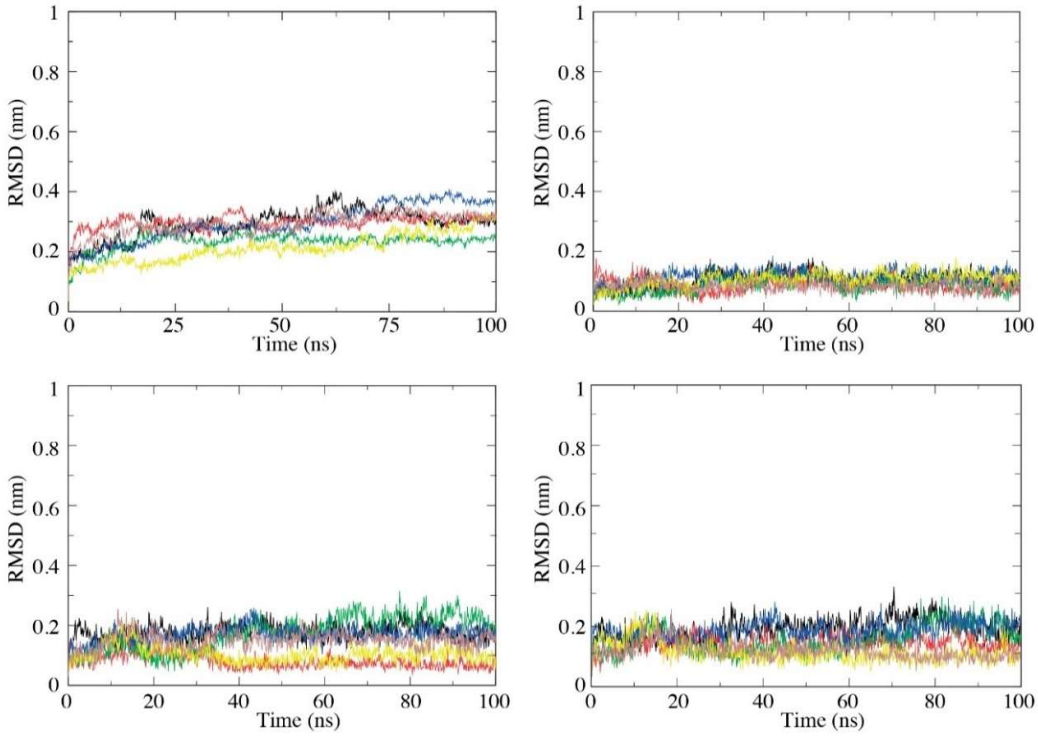



***Figure 3.20*. Structural drift of FliG-N/FliF-C$^{expr}$.** Backbone RMSD after fitting on backbone of protein core for: whole protein (top left), FliG helix2 (top right), FliG helix6 (bottom left), and FliG helix 7 (bottom right). System used the AMBER14SB force field with E91 deprotonated; different colours represent each FliG-FliF complex in the unit cell.

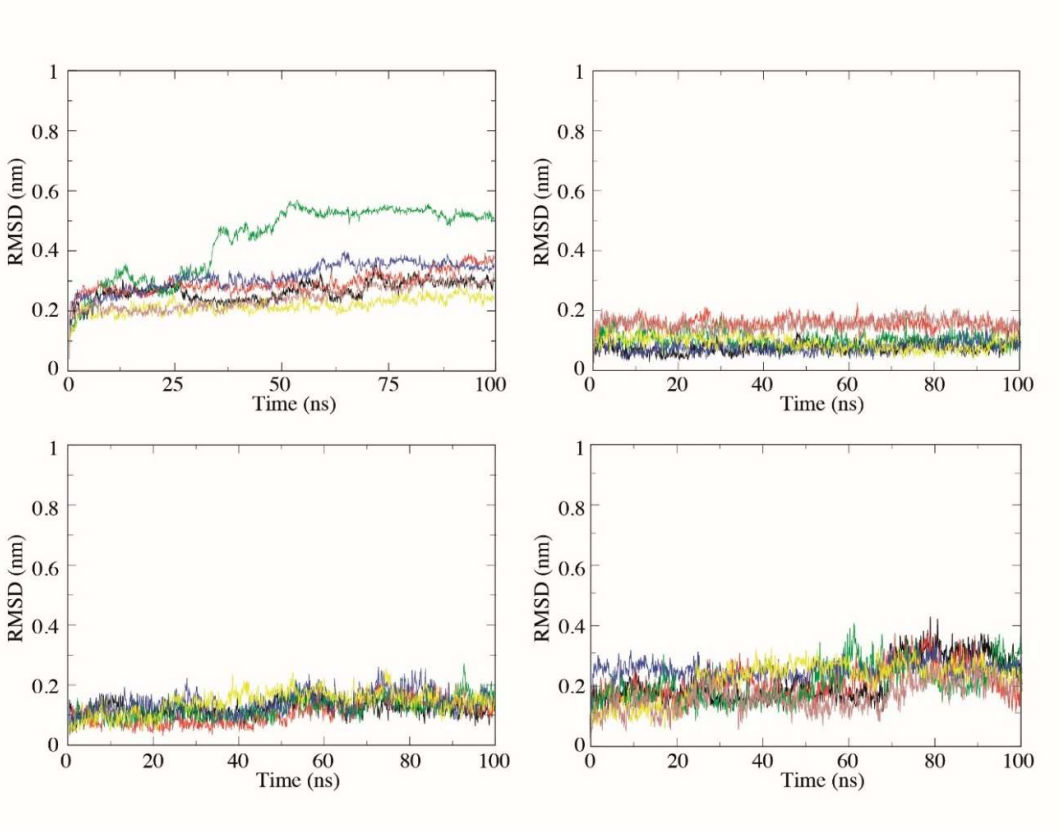



***Figure 3.21*. Structural drift of FliG-N/FliF-C$^{expr}$.** Backbone RMSD after fitting on backbone of protein core for: whole protein (top left), FliG helix2 (top right), FliG helix6 (bottom left), and FliG helix 7 (bottom right). System used the CHARMM36m force field with E91 protonated; different colours represent each FliG-FliF complex in the unit cell.

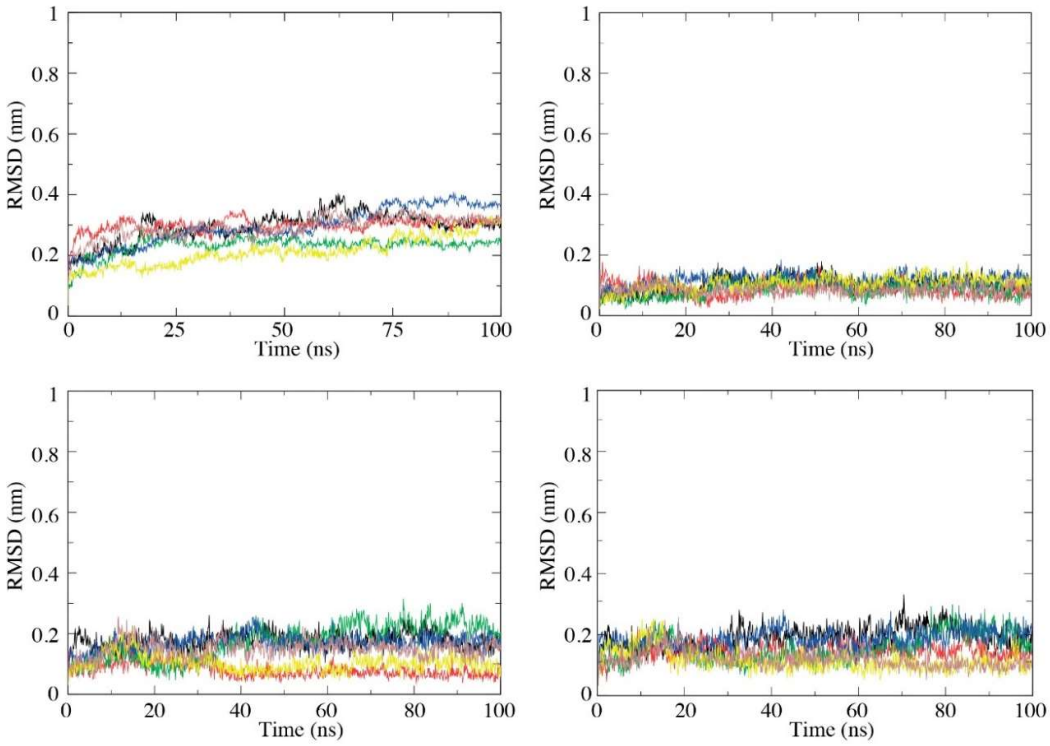



***Figure 3.22*. Structural drift of FliG-N/FliF-C$^{expr}$.** Backbone RMSD after fitting against backbone of protein core for: whole protein (top left), FliG helix2 (top right), FliG helix6 (bottom left), and FliG helix 7 (bottom right). System used the AMBER14SB force field with E91 deprotonated; different colours represent each FliG-FliF complex in the unit cell.

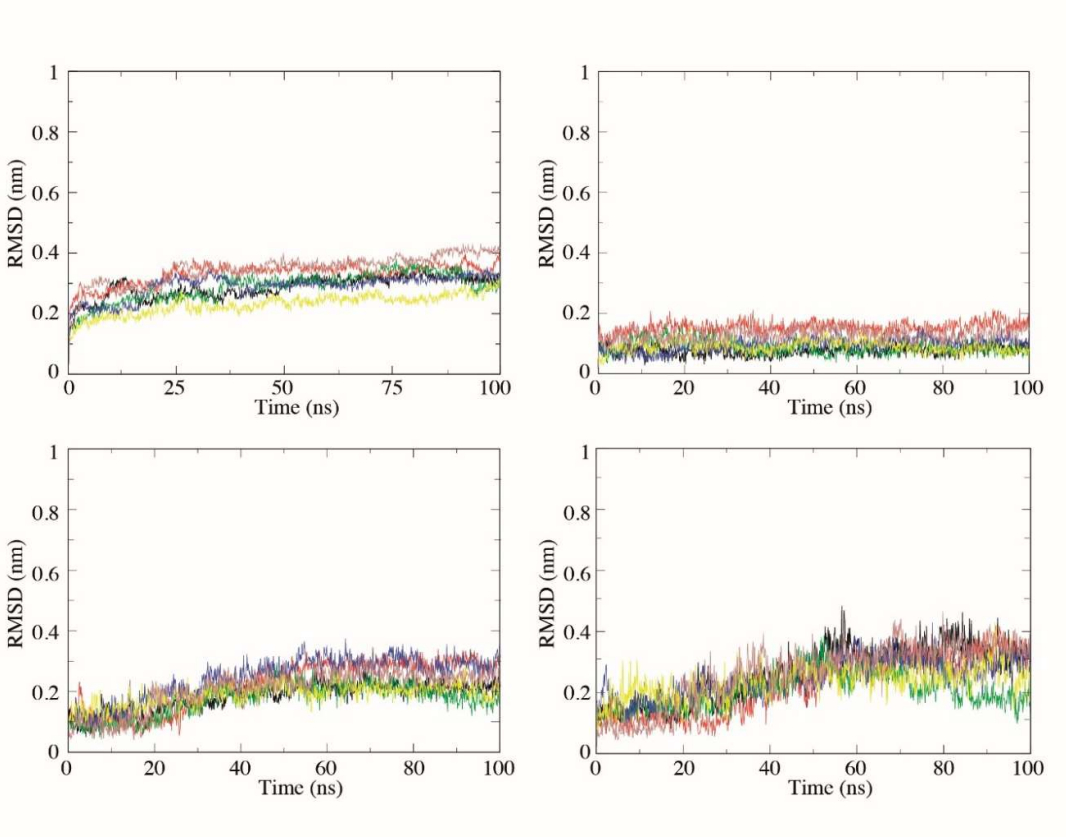



***Figure 3.23*. Structural drift of FliG-N/FliF-C$^{expr}$.** Backbone RMSD after fitting against backbone of protein core for: whole protein (top left), FliG helix2 (top right), FliG helix6 (bottom left), and FliG helix 7 (bottom right). System used the AMBER14SB force field with E91 protonated; different colours represent each FliG-FliF complex in the unit cell.

### 3.2.5. *T. maritima* FliG-N/FliF-C complex dynamics in solution and crystal

*H. pylori* is one of the two presently available FliG-N/FliF-C structures, the other coming from *T. maritima*, a thermophilic bacterium often used in crystallization studies due to proteins of thermophilic organisms being more robust (Dalhus et al., 2002; Razvi and Scholtz, 2006). The *T. maritima* unit cell of 5TDY contains four copies of FliG-N/FliF-C complex molecules in two conformations (Figure 3.25), one of which is highly similar to *H. pylori* ($5TDY_{bent}$) (Figure 3.26), but the other one adopts a different conformation, forming an extended helix from helix 5 onwards ($5TDY_{eln}$), which is different from the extended linker conformation in *A. aeolicus* (Lynch et al., 2017). Unlike in $5TDY_{bent}$ in $5TDY_{eln}$ a small third helix is seen in the FliF chain (Figure 3.24).

Testing behaviour of the same protein in different species is an important comparison to establish whether the results seen previously are species specific.

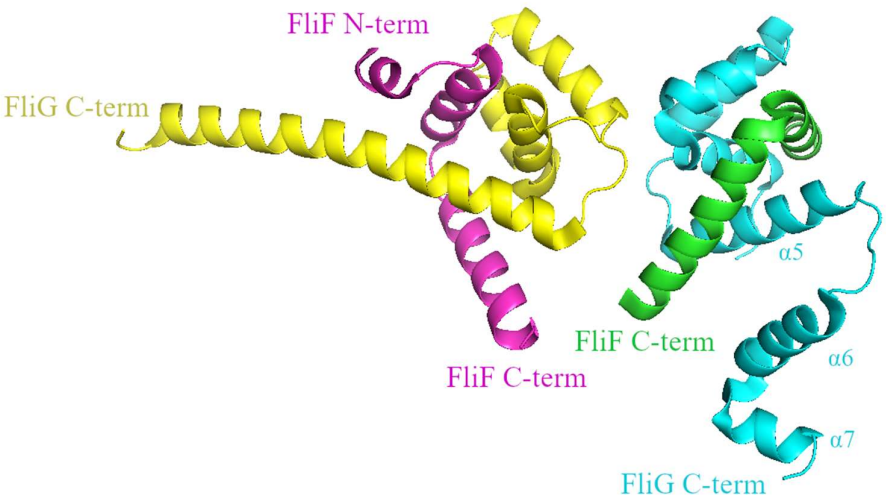



***Figure 3.24*. Asymmetric unit of 5TDY**. Left - $5TDY_{eln}$ conformation (FliG – yellow, FliF-magenta), right - $5TDY_{bent}$ conformation (FliG-cyan, FliF-green).

For analysis, helical boundaries were defined as follows (Table 3.2):

***Table 3.2.* Helix boundaries used in analysis of 5TDY**

| *Protein* | *Helix number* | *Starting residue* | *Ending residue* |
|---|---|---|---|
| *$5TDY_{bent}$ conformation* | | | |
| *FliF* | α-helix 1 | 3 | 20 |
| *FliF* | α-helix 2 | 23 | 35 |
| *FliG* | α-helix 2 | 41 | 54 |
| *FliG* | α-helix 3 | 56 | 65 |
| *FliG* | α-helix 4 | 68 | 80 |
| *FliG* | α-helix 5 | 85 | 101 |
| *FliG* | α-helix 6 | 109 | 122 |
| *FliG* | α-helix 7 | 124 | 132 |
| *$5TDY_{eln}$ conformation* | | | |
| *FliF* | α-helix 1 | 136 | 153 |
| *FliF* | α-helix 2 | 156 | 168 |
| *FliF* | α-helix 3 | 171 | 174 |
| *FliG* | α-helix 2 | 183 | 196 |
| *FliG* | α-helix 3 | 198 | 207 |
| *FliG* | α-helix 4 | 210 | 222 |
| *FliG* | elongated α-helix | 227 | 265 |

This numbering corresponds to that found in the crystal structure (Lynch et al., 2017), but is not reflective of the residue numbering in the *T. maritima* sequence, as it is merely sequential from the first residue in the asymmetric unit and proceeds through all chains.

Individual residues equivalent to those found in *H. pylori* may be deduced from the sequence alignment (Figure 4.2., in Chapter 4.2.). For RMSD measurements, $5TDY_{bent}$ (5TDY with 5WUJ-like conformation) core was defined as α2-α5, exactly the same as in 5WUJ (Figures 3.2, and 3.3., in Chapter 3.2., Figure 3.24). In $5TDY_{eln}$ (5TDY with elongated conformation) the core was defined as α2-α4, because part of α5 included in the elongated helix was subject to different dynamics and could thus artificially alter the apparent measured RMSD of the core (Table 3.2) (Figure 3.24).

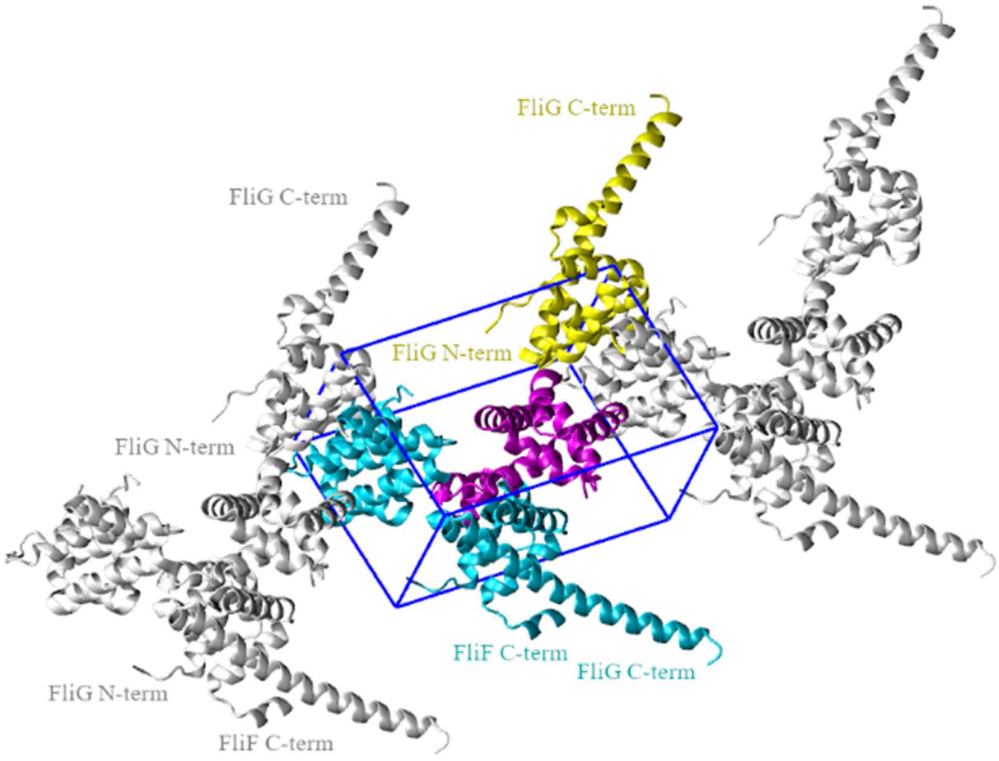



***Figure 3.25*. Unit cell of crystal structure 5TDY**. Unit cell box is shown, containing 4 FliG-N/FliF-C complexes; 2 of 4 are depicted in magenta (5WUJ-like conformation) and yellow (elongated conformation), with remaining ones in cyan. Molecules in neighbouring unit cells are shown in grey. Protein is shown in cartoon representation.

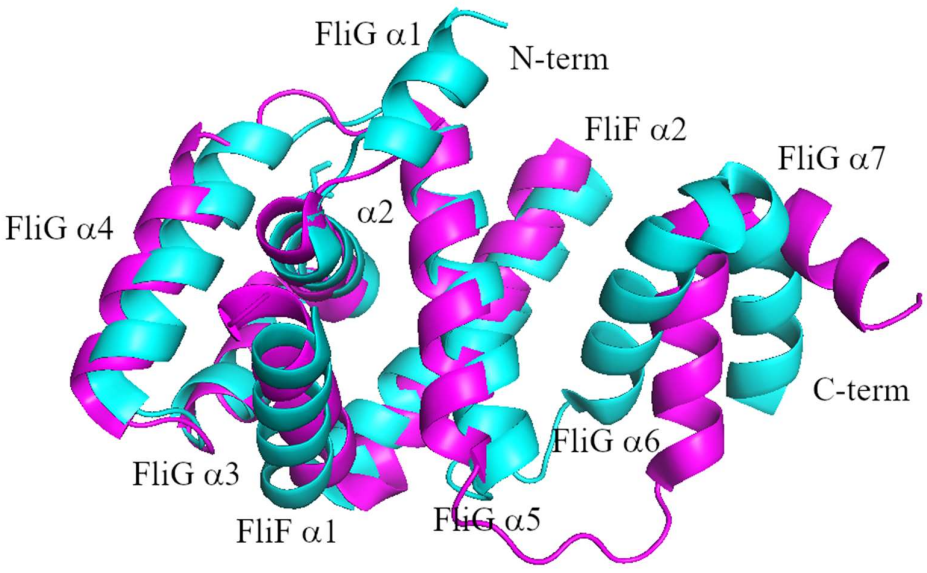



***Figure 3.26.*** **Comparison of crystal structure conformations for different organisms.** Superposition is shown for the FliG and FliF fragments from *H. pylori* (PDB-ID: 5WUJ, cyan) and *T. maritima* (PDB-ID: 5TDY, magenta) in their bent conformations). The conformations are highly similar; RMSD = 2.44 Å over 427 atoms (all atom RMSD).

To verify the behaviour of FliG helices 6 and 7 and confirm it is species-independent, performed simulations of both isolated conformations as well as the asymmetric unit in solution and in crystallographic conditions were performed. The conformation of 5TDY that is close to 5WUJ ($5TDY_{bent}$) showed identical behaviour, with a decreased stability of the extended helix in the solution and increased stability in the presence of crystal contacts, indicating that both variants are artificially stabilized by crystal contacts and unlikely to be reflective of the conformation in solution (Figures 3.27 to 3.32).

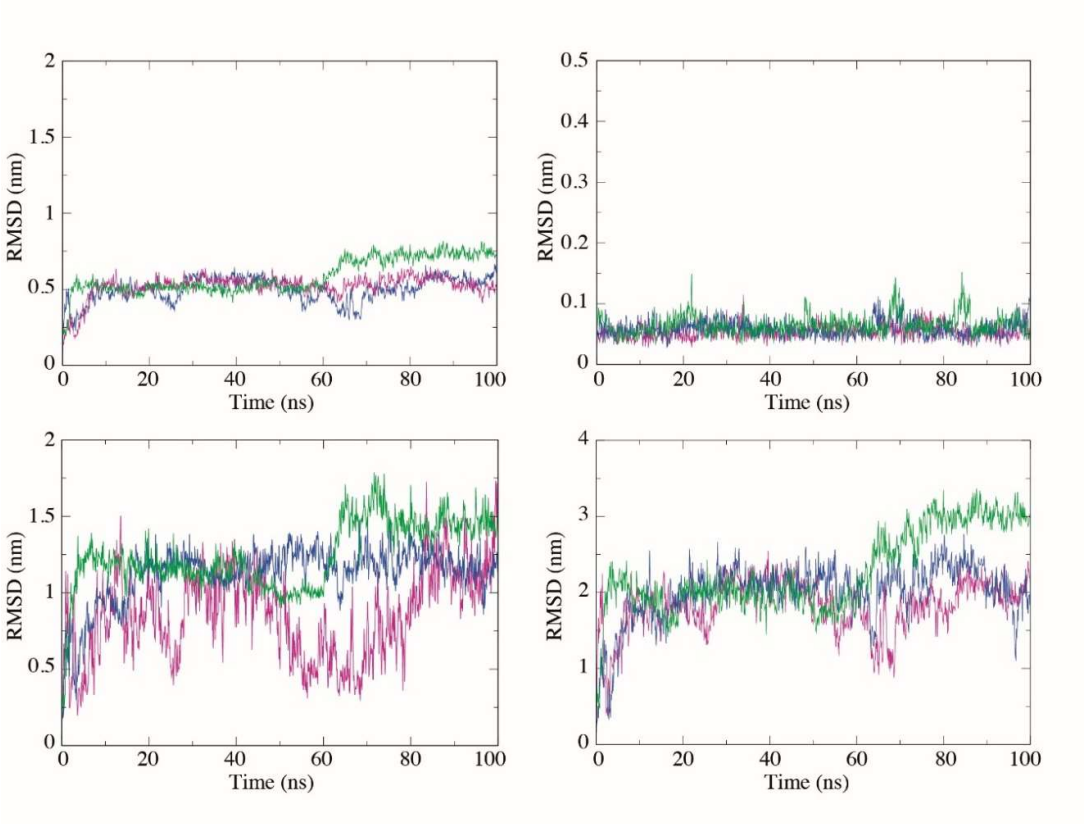



***Figure 3.27*. Structural drift of 5TDY complex (5WUJ like conformation) in solution in asymmetric unit.** Backbone RMSD after fitting against backbone of protein against whole protein (top left) and backbone of protein core for: FliG helix 2 (top right), FliG helix 6 (bottom left), and FliG helix 7 (bottom right). System used the CHARMM36m force field, different colours represent three independent runs.

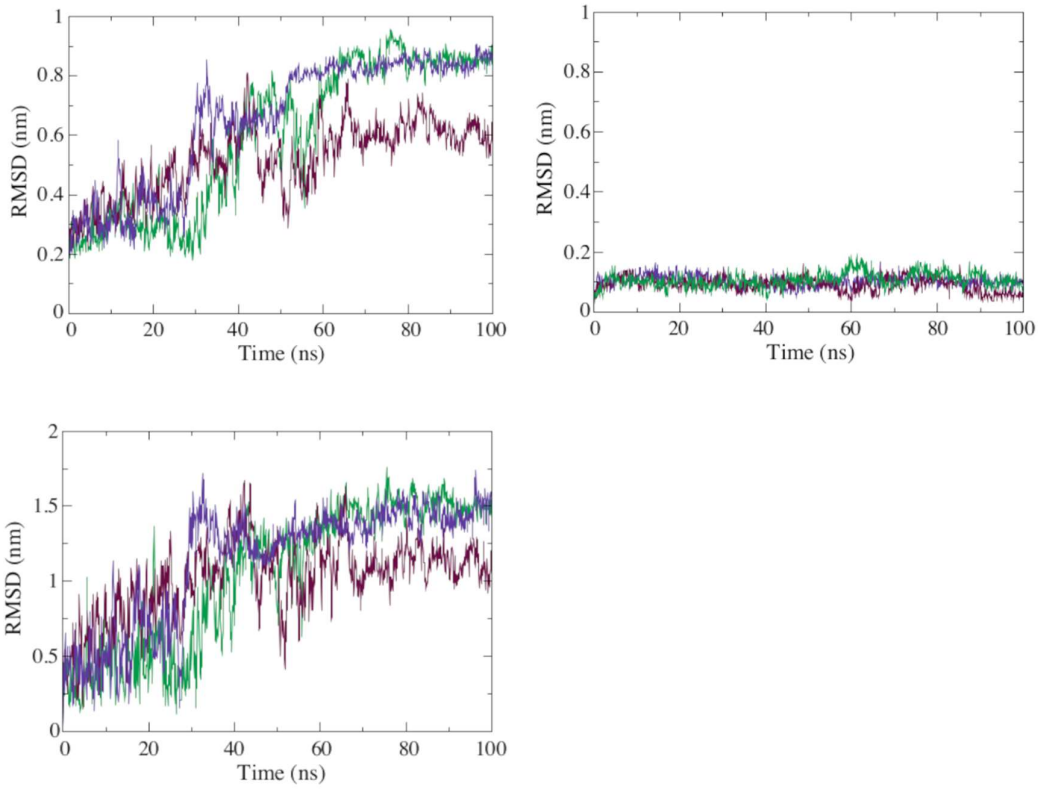



***Figure 3.28*. Structural drift of 5TDY complex (elongated conformation) in solution in asymmetric unit.** Backbone RMSD after fitting against backbone of protein against whole protein (top left) and backbone of protein core for: FliG helix 2 (top right), elongated helix (bottom left). System used the CHARMM36m force field, different colours represent three independent runs.

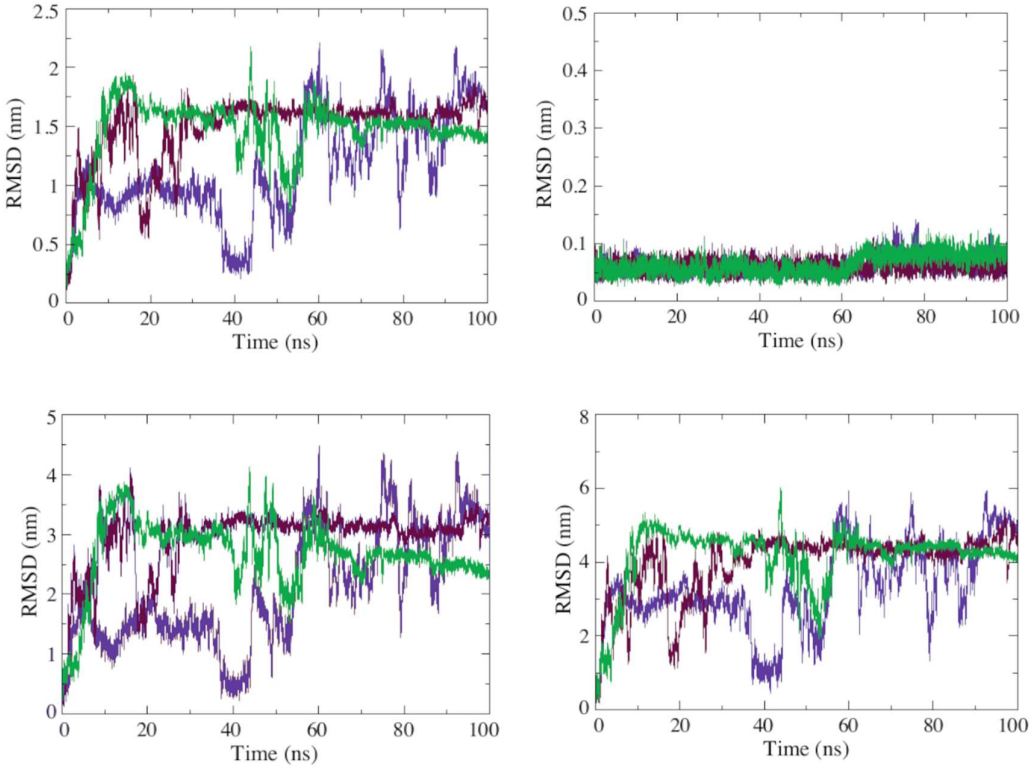



***Figure 3.29*. Structural drift of isolated 5TDY complex (5WUJ like conformation) in solution.** Backbone RMSD after fitting on backbone of protein core for: whole protein (top left), FliG helix 2 (top right), FliG helix 6 (bottom left), and FliG helix 7 (bottom right). System used the CHARMM36m force field, different colours represent three independent runs.

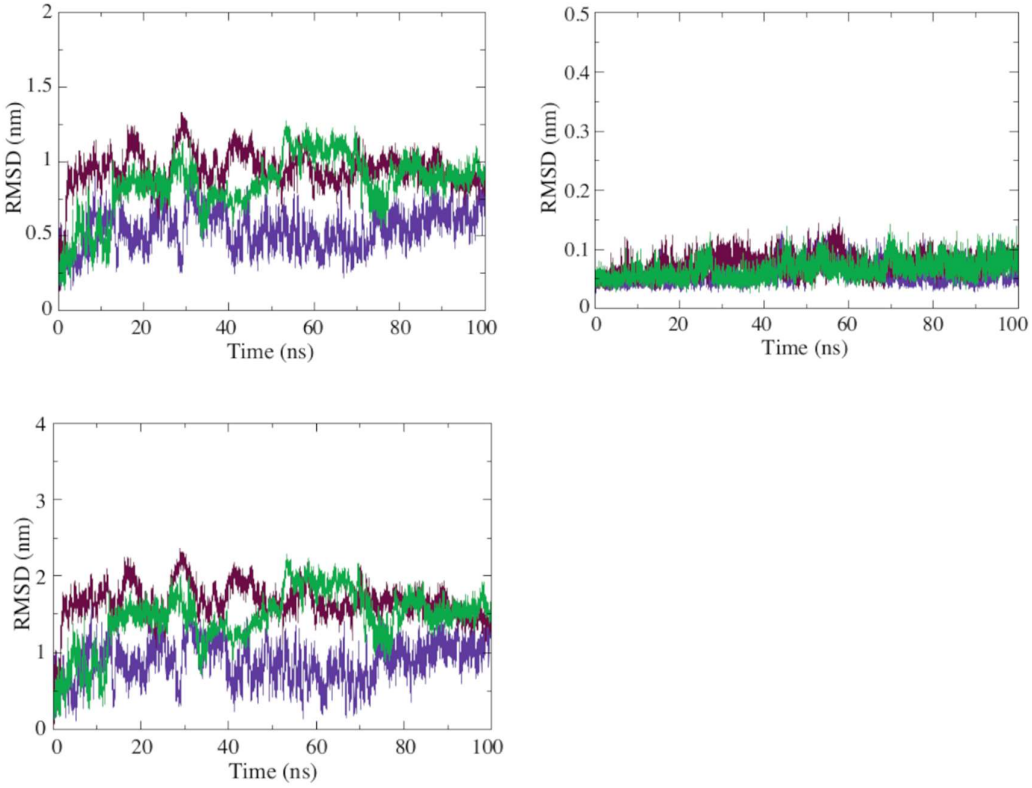



***Figure 3.30*. Structural drift of isolated 5TDY complex (elongated conformation) in solution.** Backbone RMSD after fitting on backbone of protein core for: whole protein (top left), FliG helix 2 (top right), elongated helix (bottom left). System used the CHARMM36m force field, different colours represent three independent runs.

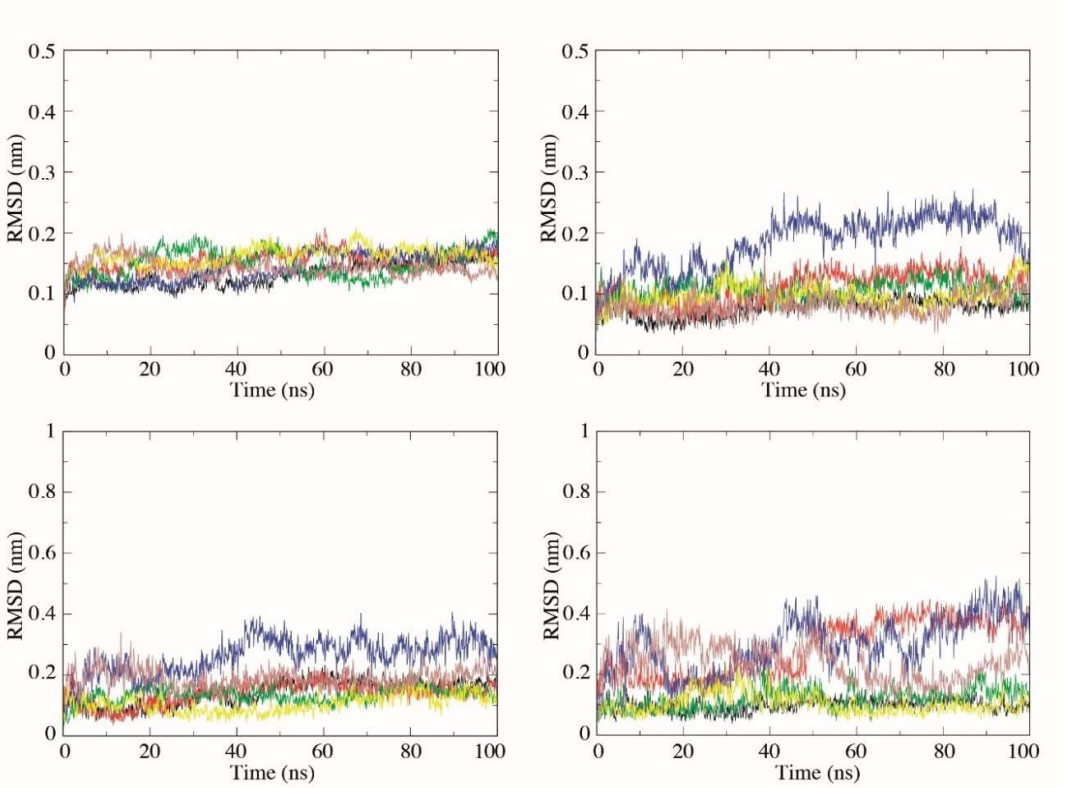



***Figure 3.31*. Structural drift of 5TDY in unit cell (5WUJ like conformation).** Backbone RMSD after fitting on backbone of protein core for: whole protein (top left), FliG helix 2 (top right), FliG helix 6 (bottom left), and FliG helix 7 (bottom right). System used the CHARMM36m force field, different colours represent three independent runs with two molecules in each unit cell.

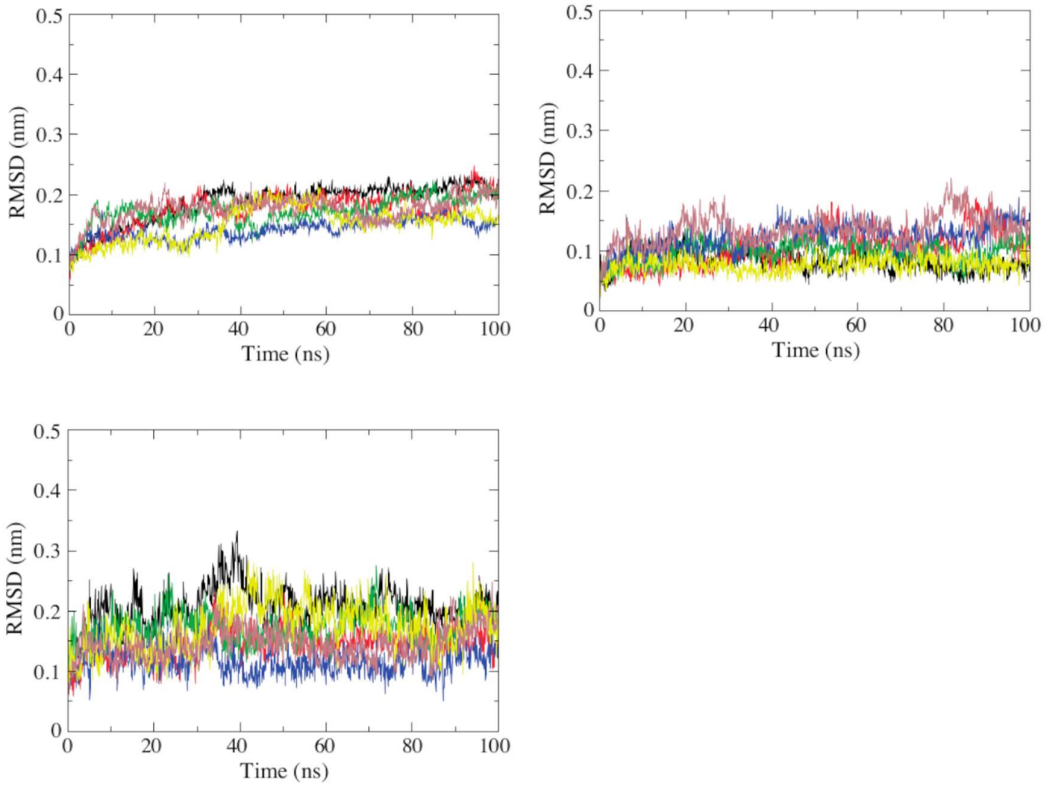



***Figure 3.32*. Structural drift of 5TDY in unit cell (elongated conformation).** Backbone RMSD after fitting on backbone of protein core for: whole protein (top left), FliG helix 2 (top right), extended helix (bottom right). System used the CHARMM36m force field, different colours represent three independent runs with two molecules in each unit cell.

## 3.3. Discussion

Crystal structures are static snapshots of a protein. However, in the cellular environment proteins can undergo many changes, including individual conformational changes and changes induced once assembled into larger-scale complexes. MD simulations can reveal this behaviour in solution or other environments. Here FliG-N of *H. pylori*, *T. maritima* and *A. aeolicus* were studied to gain new insights about the FliG protein.

FliG-N simulations with FliF-C removed were consistent with previous data which shows that FliG and FliF co-fold and are required for mutual structural stability (Levenson et al., 2012; Lynch et al., 2017), since they did not converge to a single conformation and the core part of the domain appeared to collapse. There was also no

obvious similarity between the structures after simulation of independent replicas or of the N-terminus of *A. aeolicus* from the full-length FliG structure. Simulations of full-length FliG structure in the absence of bound FliF maintained stability in each domain and collapsing was not observed, but instead a structural drift across the entire molecule occurred.

Despite the high sequence conservation locally, there is a higher number of negatively charged residues in the *A. aeolicus* N-terminus that may contribute to this different behaviour. Further experiments and varied setups would be required to investigate the conformation and stability of N-terminus in the absence of FliF in these two species and possibly *T. maritima* N-terminus as well.

FliG-N/FliF-C complex simulations showed distinct instability of α6 and α7 in solution as opposed to the protein core (helices α2-α5). The initial hypothesis that the E91 protonation state may have a role in stability of the conformation was not confirmed. A single replica simulation shown in Figure 3.7 (blue colour) may be seen as an exception, but it appears to be stochastic as there was still no coherent reproducible behaviour across extended simulation runs. This was confirmed by the similar behaviour of $5TDY_{bent}$ simulations when treated in both the asymmetric unit in solution or when both conformations were separated in solution (Figures 3.30, 3.30) in *T. maritima*. Therefore, behaviour of the FliG-N helices 6 and 7 is species independent.

Interpretation of a protein crystal structure into biological properties of the protein involves many considerations as crystallographic conditions and artefacts created during crystallization can alter potential conclusions made. For example, the crystal environment can change binding properties of a ligand (Søndergaard et al., 2009), bias sidechain conformation (Arpino et al., 2012) and influence position of ions (Krah and Takada, 2016). Care must be taken when verifying, if crystal packing contacts represent a biologically relevant oligomerization interface (Carugo and Argos, 1997; Dey et al., 2018; Luo et al., 2015; Sardis and Economou, 2010).

The interface between neighbouring FliG-N α3-4 and α6-7 was previously proposed to be representative of the ring assembly (Xue et al., 2018). The observation that crystal conditions reduced the structural drift for all elements, but in particular for α6 and α7 in both *H. pylori* and *T. maritima* prompted consideration that this conformation needed further investigation. In *H. pylori,* additionally, neither forcefield nor the E91 protonation state led to any different observations. Simulations of

FliG-/FliF-C tetramer arrays in the absence of crystallographic periodicity were indicative, but not conclusive given potential forcefield specific behaviour or E91 protonation state influence, necessitating simulation in the crystal unit cell. While mutation of some residues from α3-4 and α6-7 crystal interface, initially proposed to be biologically relevant, did affect motility of *E. coli*, (Xue et al., 2018), they could easily compromise motility through disrupting FliG function in another way leading to the same macrophenotype.

Additionally, the $5TDY_{eln}$ conformation, which is different from *A. aeolicus* likewise was destabilized in solution and shown to be stabilized by crystal contacts, indicating both conformations, $5TDY_{bent}$ and $5TDY_{eln}$ may not represent a conformation found physiologically during ring assembly (Figure 3.25).

Collectively, the simulation data presented here suggest that based on the available FliG-N/FliF-C structures, part of the protein that has to join FliG-M is found in a conformation that may not correspond to that needed in a ring assembly context and further modelling is needed to address the question of how the different parts of the proteins that are enforced in their conformations by crystal contacts that may be found *in vivo*.

# Chapter 4 - Integrative modelling of full-length *Helicobacter pylori* FliG and complete ring assembly

## 4.1. Introduction

FliG consists of three distinct domains, FliG-N, FliG-M and FliG-C (Lee et al., 2010). Investigation of the dynamics of *H. pylori* and *T. maritima* FliF-C bound FliG N-terminal domain in solution and crystal structure in Chapter 3 showed that conformation of helices that link the N-terminal domain with middle domain is enforced by crystal contacts. This opened question about the possible conformation under biological conditions.

When modelling FliG structures, the linker between FliG-N and FliG-M (NM-linker) has often been omitted and typically indicated with a dashed line (Carroll et al., 2020; Chang et al., 2020), while multiple proposed models in the literature do not generally agree with one another (Paul et al., 2011b). The only crystallized NM-linker in a full length structure is that of the *A. aeolicus* FliG (Lee et al., 2010), while in other species no solved structures of the NM-linker are available. Docking of structures in the original 5WUJ crystal structure paper did not account for the distance that is needed to join FliG-N and FliG-M, but instead it simply treats these domains separately (Xue et al., 2018). Another study suggests a that linker is not present at all, and that FliG-N and FliG-M are in a tight interaction, practically forming one domain (Paul et al., 2011b). An intermediate conformation of FliG-N is proposed in a co-evolution study, but only in the context of the separate FliG-N domain (Khan et al., 2018).

The FliG ring assembly has been reported to be a 34 mer (with 33-35 variation possible) in *S. enterica* (Thomas et al., 2006, 1999), 34 in *V. alginolyticus* (Carroll et al., 2020) and more recently a suprising 46 in *B. burgdorferi* (Chang et al., 2020). The FliF ring has been resolved in high resolution as a 33 and a 34-mer in *S. enterica* (Johnson et al., 2020; Kawamoto et al., 2021; Takekawa et al., 2021) and due to the 1:1 binding ratio and co-folding (Levenson et al., 2012; Lynch et al., 2017), it can be reasonably expected that the FliG and FliF stoichiometry in a species should match.

There is no clear agreement about placement of the FliG domains in the low reslution C-ring densities (Paul et al., 2011b).

Therefore MD results were followed up with homology modelling using known structural data to create a full lenght FliF-C bound FliG model, and expanded to a whole FliG ring modelling with addition of the FliM-M, using fitting into cryoET density and applying MDFF techniques (Figure 4.1).

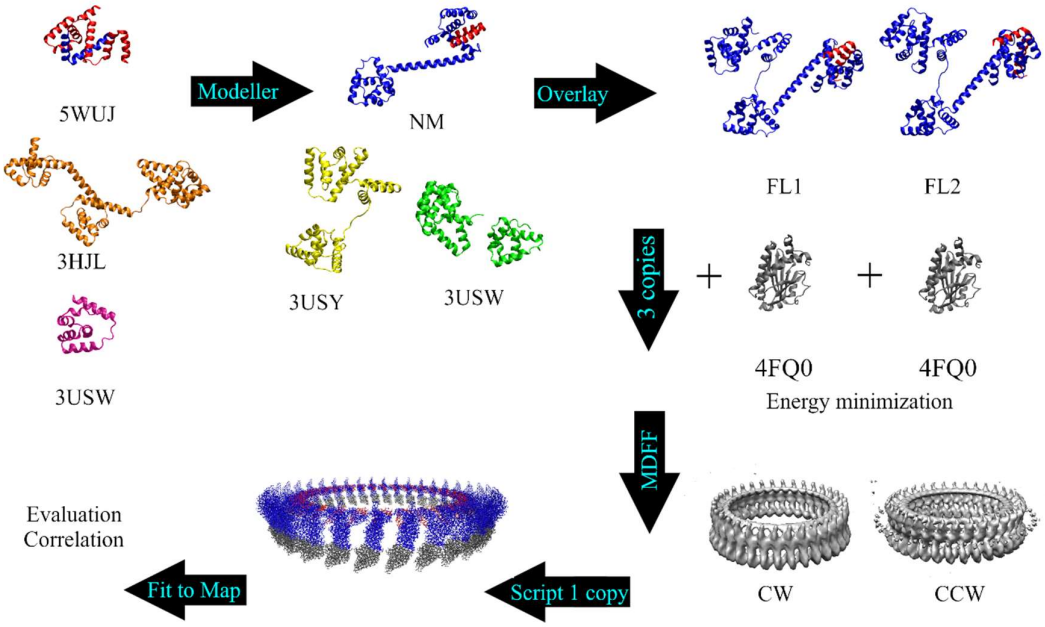



***Figure 4.1*. Model building workflow from a single FliG domain to a ring assembly**. Homology model of FliG-NM was built based on three crystal structures, and subsequently used to construct two full-length FliG models (FL1 and FL2). FliM-M was added to both models. After building rings in all possible combinations, best models were picked by visual evaluation and subjected to energy minimization in GROMACS

FliG is mostly conserved and therefore suitable for homology modelling of the *H. pylori* FliG/FliF-C crystal structures (Figure 4.2).

## 4.2. Modelling of *H. pylori* FliG/FliF-C

### 4.2.1. Homology modelling of *H. pylori* FliG/FliF-C from crystal structures

Simulation data in Chapter 3 indicated that the protein conformation with helices 6 and 7 oriented in the bent conformation (as found in crystal structures 5WUJ and 5TDY) is unlikely to be maintained in solution, but are instead artefactually stabilized by crystal contacts. To address the question of how the *H. pylori* NM-linker may be structured in a ring assembly, integrative and homology modelling approaches had to

be employed, using structural data for separately crystallized domains and a full length FliG structure from *A. aeolicus* to build a full length FliG/FliF-C model.


***Figure 4.2*. Sequence alignment of FliG from various flagellated bacterial species representing model organisms and pathogens.** FliG sequence is highly conserved throughout the protein length and of similar total length. Figure created in Jalview (Waterhouse et al., 2009).

The NM-linker of *A. aeolicus* full-length FliG structure (in its unbound state, i.e., not bound to FliF) adopts an extended state, instead of the two separate helices observed in other structures. The middle domains from *H. pylori*, corresponding to

PDBs 5WUJ and 3USW (Lam et al., 2012) were used as additional input for MODELLER (Webb and Sali, 2016). Structure of *H. pylori* N-terminus (5WUJ) and one of the two *T. maritima* crystallized conformations with a very similar fold ($5TDY_{bent}$) (Section 3.2.5, Figure 3.26) showed almost identical behaviour and were the initial motivation to extend FliG modelling to more than one domain. Simulation results also showed that the other *T. maritima* conformation ($5TDY_{eln}$) that ends with an elongated helix (Section 3.2.5, Figure 3.24) that is different from *A. aeolicus* is also stabilized by crystal contacts. Therefore, $5TDY_{eln}$ was used as an input to make an additional model for comparison (Figure 4.3).

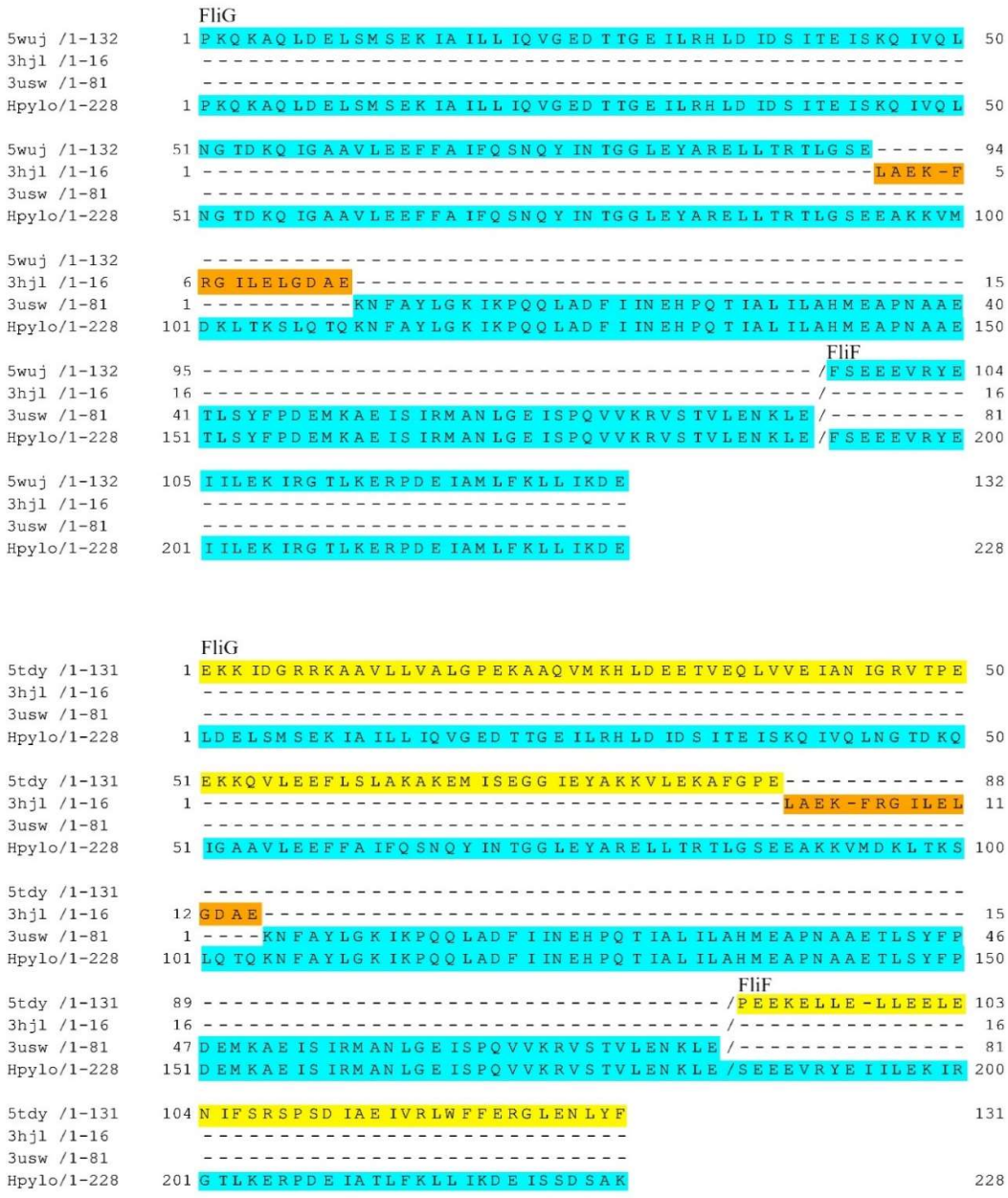



***Figure 4.3*. Alignments of *H. pylori* sequences (FliG/FliF-C) from structures used in the model building, coloured by species,** including 5WUJ: *H. pylori* (cyan) FliG-N/FliF-C, 3HJL: *Aquifex aeolicus* (orange) FliG and 5TDY: *Thermotoga maritima* (yellow) FliG-N/FliF-C, to demonstrate the contribution of each structure and species to the models.

The 5WUJ based models selected by molpdf (Sali, A; Blundell, 1993) and DOPE (Shen and Sali, 2006) scores were overall structurally similar, as also confirmed via the QMEAN score (Benkert et al., 2008) from the SWISS-MODEL webserver tool

(Schwede et al., 2003; Waterhouse et al., 2018). Visual evaluation also came into play in selecting models where the NM-linker and α5 were joined by a short loop rather than a rigid, but sharply bent helix turn that would be less likely present in an actual structure. The SWISS-MODEL webserver tool indicated that the helical region has a lower confidence than the rest of the structure, which is not surprising as FliG-N/FliF-C and FliG-M were built with *H. pylori* sequence based on the *H. pylori* structure while NM-linker is based on a different species. $5TDY_{eln}$ (*T. maritima* FliG-N/FliF-C that ends with an elongated helix) based models were selected by the same criteria, but had lower QMEAN scores and more Ramachandran outliers (Hooft et al., 1997).

After obtaining these two-domain models, the C-terminal domain of *H. pylori* was next integrated. There are two known conformations of *H. pylori* C-terminus with part of domain rotating 180˚ as captured by X-ray structures with PDB IDs: 3USW and 3USY (Lam et al., 2012). Both conformations were added to 5WUJ and $5TDY_{eln}$ based two-domain models (Figure 4.4, 4.5) and are further referred to as C1 and C2 in the context of the homology models here. This is because, although their potential biological relevance for the clockwise (CW) and counter-clockwise (CCW) states is implied, there is not presently enough information to assign either conformation to a specific state (Lam et al., 2012).

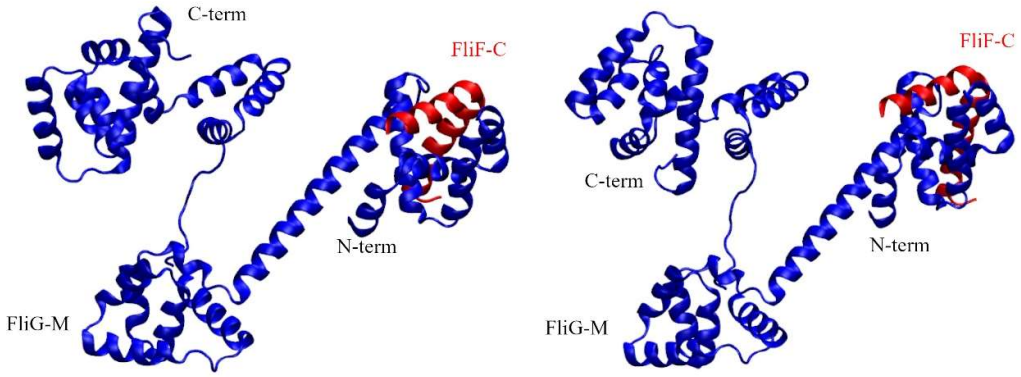



***Figure 4.4*. Two *H. pylori* models based on 5WUJ N-terminus with two different C-terminal domains.** Red: FliF-C, dark blue: FL FliG. Left: C1 from 3USW. Right: C2 from 3USY

Since the middle domain and C-terminus of $5TDY_{eln}$ based models are identical to the 5WUJ based models, certain key differences arise from inclusion of *T. maritima*

N-terminal domain. In particular, the orientation of the N-terminal domain and the extended helix is opposite when comparing models. The orientation of the FliF-C domain is inverted in both $5TDY_{eln}$ based models compared to the 5WUJ-based models. FliF consists of three helices instead of two in the $5TDY_{eln}$-based models, reflecting the crystal structure.

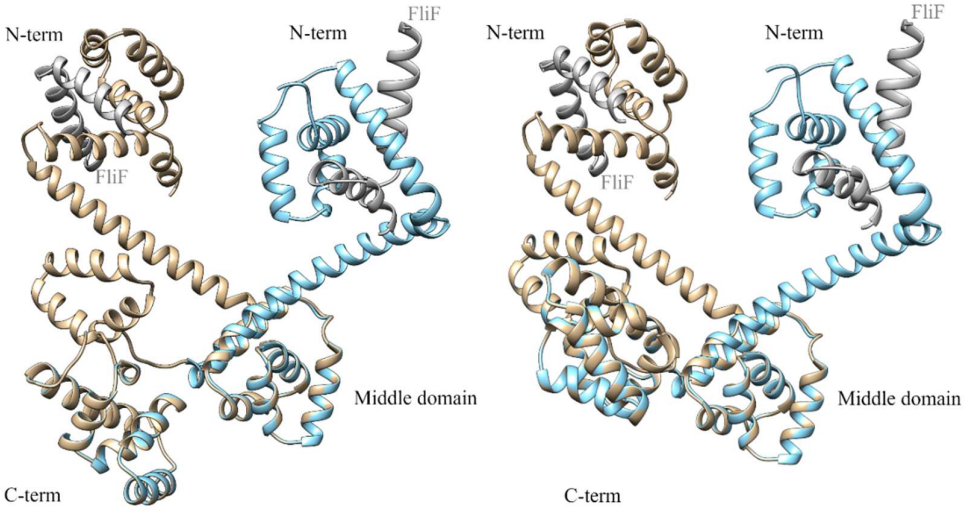



***Figure 4.5*. Comparison of 5WUJ and $5TDY_{eln}$ based models**. Left: FliG/FliF-C, C1. Right FliG/FliF-C, C2. Light brown – 5WUJ based models, light blue – $5TDY_{eln}$ based models. FliF-C coloured in grey in all models.

### 4.2.2. Modelling of FliG with Alphafold

Finally for comparison, a full length model of FliG was predicted using Alphafold (Jumper et al., 2021). While manual overseeing of the modelling process as in the case of MODELLER is preferable to most automated modelling servers due to "black box" issues, Alphafold excels in its field with its recent breakthroughs. The Alpahfold FliG conformation (disregarding the absence of FliF-C) partially contradicts our model and displays a conformation akin to that seen in 5WUJ and $5TDY_{bent}$ (*T. maritima* conformation that is similar to 5WUJ) (Figure 4.6).

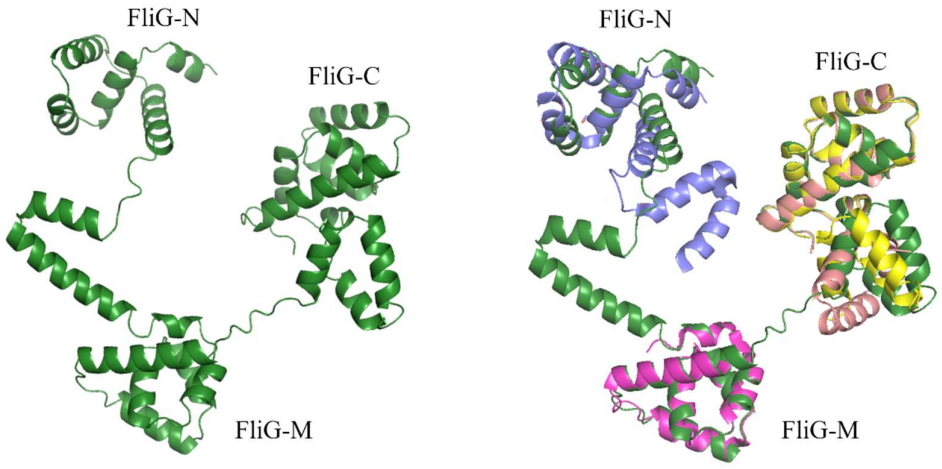



***Figure 4.6*. Alphafold prediction of *H. pylori*.** Prediction model overlaid with crystal structures of individual domains. FliG-N from 5WUJ (cyan), FliG-M from 3USW (pink) and two conformations of FliG-C from 3USW (salmon) and 3USY (yellow).

This prediction is not unlikely *per se* as Alphafold uses known crystal structures as part of its prediction algorithm. The similarity of the *H. pylori* prediction with actual crystallographic structures for *H. pylori* is reflected by RMSD values of 7.4 Å (over 670 fitted all atoms) for FliG-N and 0.55 Å (over 497 atoms) for FliG-M. For FliG-C, equivalent RMSD values were 0.80 Å (over 701 atoms) for 3USW and 0.53 Å (over 655 atoms) for 3USY. On the individual domain level, most of the domains are thus very close to the crystal structures (Figure 4.6). For FliG-N, helices 1 to 5 adopt similar fold, and the higher RMSD arises from helix 6 which has a completely different crossing angle relative to α1-α5, but could still potentially accommodate FliF-C binding (which was not included in the prediction). The predicted and actual middle domain structures are almost identical. The predicted C-terminal domain is closer to the 3USY conformation than 3USW, as assessed both visually and quantitatively. The overall model of the complete FliG, however, has two very flexible and open regions compared to the MODELLER-derived models, where there is only one flexible region between FliG-M and FliG-C that comes directly from crystal structure.

The Alphafold modelling approach gives a very different result for the area of interest between FliG-N and FliG-M than the step-by-step MODELLER approach. While Alphafold models individual domains match with the know crystal information very well on the separate domain level, currently the Alphafold model does not provide

a better answer on how to link the FliG-N and FliG-M domains with certainty than the classic homology modelling approach.

## 4.3. Modelling of FliG rings

The next aim was to validate the FliG/FliF-C models by fitting them into known cryoET densities of the C-ring and modelling an entire FliG ring assembly using an approach that combines MDFF and fitting into cryoET density data. To date there is no high-resolution structural data of FliG ring assembly available.

Initially, the different models of FliG/FliF-C or FliG only were manually placed into cryoET density in Chimera and correlations were calculated with Fit to Map function to assess agreement of the ring structures. The EM structure density from *S. enterica* for the C-ring and MS-ring solved in 2006 (Thomas et al., 2006) was used as at the start of the project it was the best available density of such type at the time and it suffers from non-matching inner and outer "lobe" stoichiometry, where the outer lobe was refined to 34 proteins of FliG/FliM/FliN and inner to 25 proteins of FliF (Thomas et al., 2006), depicting FliG and FliF stoichiometry conundrum (Figure 4.7), which was not resolved until the recent release of high resolution FliF ring structures (Johnson et al., 2020; Kawamoto et al., 2021). It was therefore only possible to work with placing monomers in the density until *V. alginolyticus* CW and CCW cryoET ring densities were resolved (Carroll et al., 2020) opening up the possibility for better ring building. In those structures, both the inner and outer lobes have a matching 34-mer stoichiometry and there are clear links between each monomer like shape in the density in the inner and outer lobes (Carroll et al., 2020), unlike the *S. enterica* structure (Thomas et al., 2006), indicating domains of FliG could be joined via these "density bridges" (Figure 4.7). Notably, before the release of the *Vibrio* structures other modelling attempts also suffered from the assumed 34 vs. 25 mismatch, as seen in the papers of which the FliG-N structures used in the simulation study come from (Lynch et al., 2017; Xue et al., 2018) as well as in theoretical models of *E. coli* and *T. maritima* (Kim et al., 2017; Paul et al., 2011a, 2011b).

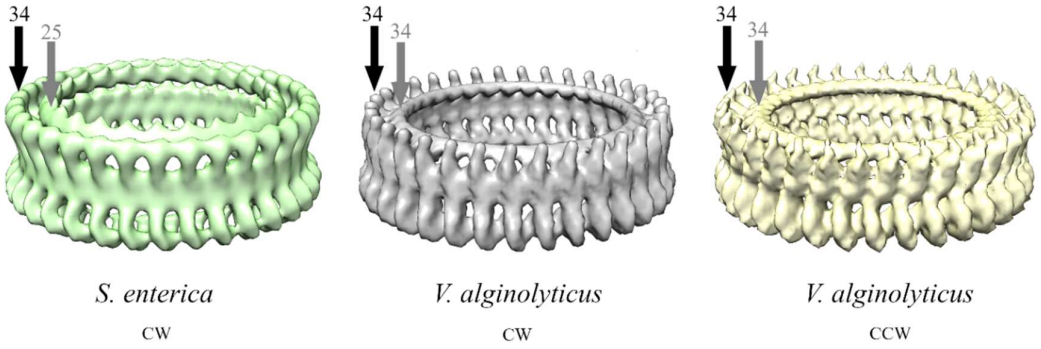



***Figure 4.7.*** **Densities of *S. enterica* and *V. alginolyticus* C-rings.** *V. alginolyticus* has CW and CCW fixed states resolved. Black arrow shows outer "lobe" and grey arrow inner "lobe" of each density and indicates stoichiometry that they were resolved with.

Homology models or the crystal structure (in case of *A. aeolicus*) were both used as input for the ring generating script and rings were placed in *V. alginolyticus* densities. The script takes an input structure and rotates it for a fixed angle and places it a defined distance from a centre of a space. It then creates a defined number of copies of the structure around that centre in the same plane. The ring generating script was repeatedly used with different input rotation value against the centre until visually reasonable output was obtained. Angles and corresponding rings were picked based on visual evaluation (in particular, there were no clashes or overlaps of domains, and the N-terminal domain was facing the centre of the ring). Upon placing some of these rings into the cryoET density, an issue was identified whereby the angle of the FliG NM-helix was very different from the observed "density bridge", and so these ring models were rejected as unlikely. While a notably better fit of the FliG homology models over those from *A. aeolicus* initially seemed promising, several issues were spotted after careful consideration that led to a complete re-work of the ring design process. Among them, the most crucial was the positioning of FliG-M in the density usually attributed to FliM-M that would make the model irreconcilable with FliM and FliN placement. The flexible appearance of the full-length *H. pylori* linker between FliG-M and FliG-C indicated that the domain likely could be positioned to better match the density and prompted to seek approaches beyond a "simple static" combination of models in a ring.

Therefore, FliM-M was next added to all *H. pylori* sequence containing models based on the FliM-M/FliG-M co-crystal structure, namely PDB ID: 4FQ0 (Lam et al., 2013). For *A.aeolicus* FliM, no crystal structure was available and therefore a homology model was built using a the *A. aeolicus* FliM sequence and the *H. pylori* FliM-M

structure as a template. Integration of FliM-M to the crystal structure and model-based rings constructed previously resulted in it being placed partially outside the density and at an angle that would make FliM interaction with FliN impossible, again proving the need for different methodology to reconcile the model with information known for other C-ring proteins.

## 4.4. MDFF based FliG/FliF-C/FliM-M ring modelling

MDFF (Molecular Dynamics Flexible Fitting) is a method that used molecular dynamics simulation to fit a structure into a reference space - an EM, cryoEM, cryoET density or SAXS envelope. Using this method can help to fit a structural model into target density, that otherwise, if manually placed, would be partially outside due a particular linker conformation between domains for instance. Simulation would allow the linker to adopt a different conformation and guide the structure into density.

Three copies of the FliG/FliF-C/FliM-M models were next manually placed side by side into the *V. alginolyticus* cryoET density (Carroll et al., 2020). A trimmed density matching to three monomers of 34 stoichiometry seen in the density with a trimer of the FliG/FliF-C/FliM-M model was subsequently used to perform Molecular Dynamics Flexible Fitting (MDFF) of 10 to 15 steps. Both clockwise (CW) and counterclockwise (CCW) motor states representing cryoET densities were used. After the run the middle copy of the three was chosen for further work, because the copies on the outside of a multimer can behave differently as well as influence each other. The final middle copy was finally used as input for the previously described ring-generation process, in order to generate a full 34-mer ring that was again fitted into the *V. alginolyticus* cryoET density (Figure 4.10). The fit to map function from Chimera was used to obtain a final measure of correlation (Table 4.1).

***Table 4.1*. Summary of ring models and correlation with Fit-to-Map function in Chimera.**

| Model | C-terminus type | cryoET density type | Correlation with cryoET map |
|---|---|---|---|
| **5WUJ model N-terminus based** | 3USY | CW | 0.954 |
| **5WUJ model N-terminus based** | 3USY | CCW | 0.996 |
| **5WUJ model N-terminus based** | 3USW | CW | 0.935 |
| **5WUJ model N-terminus based** | 3USW | CCW | 0.959 |
| **5TDY model N-terminus based** | 3USY | CW | 0.923 |
| **5TDY model N-terminus based** | 3USY | CCW | 0.885 |
| **5TDY model N-terminus based** | 3USW | CW | 0.897 |
| **5TDY model N-terminus based** | 3USW | CCW | 0.923 |
| ***Aquifex aeolicus* FL structure** | - | CW | 0.851 |
| ***Aquifex aeolicus* FL structure** | - | CCW | 0.844 |
| **Alphafold FliG model** | - | CW | 0.942 |
| **Alphafold FliG model** | - | CCW | 0.952 |

While correlation below 0.9 could be used as a "first checkpoint" parameter in rejecting ring models, individual assessment of each post-MDFF monomer was performed to ascertain which models are structurally most reasonable and make sense in the context of what is previously known from biochemical and structural studies. Structures were assessed for properties including: if they maintained structural integrity (unfolding, particularly of FliF-C was sometimes observed), domain-domain packing, and orientation of the domains.

The *A. aeolicus* models had the lowest correlation with the respective densities, and in addition, the C-terminal domain completely lost its shape and aggregated with

the neighbouring molecule in a manner that is biologically implausible. The $5TDY_{eln}$ conformation-based model with the C-terminal region from the 3USY structure showed a loss of domain tertiary structure during MDFF in the CW and CCW densities and the same occurred in the case of the 5TDY extended conformation model with the C-terminal region based on the 3USW structure in CCW density. All four $5TDY_{eln}$ conformation based models showed partial unravelling of the two FliF helices from the ends exposed to the solvent, likely due to their position being opposite of those in models where the N-terminal region was based on 5WUJ (Figure 4.5).

The Alphafold starting model has a conformation reminiscent of the 5WUJ crystal structure (i.e. bent conformation of helices 6 and 7 in *H. pylori* FliG-N) but during the MDFF run this opened up into an intermediate conformation (Figure 4.8). The molecule significantly compacts its middle and C-terminal domains from their starting positions, which is expected. This is likely indicative of the necessity for linker to be long between the FliG-M and FliG-C domains. However, as this model does not have FliF helices bound, it was only used as a comparative model.

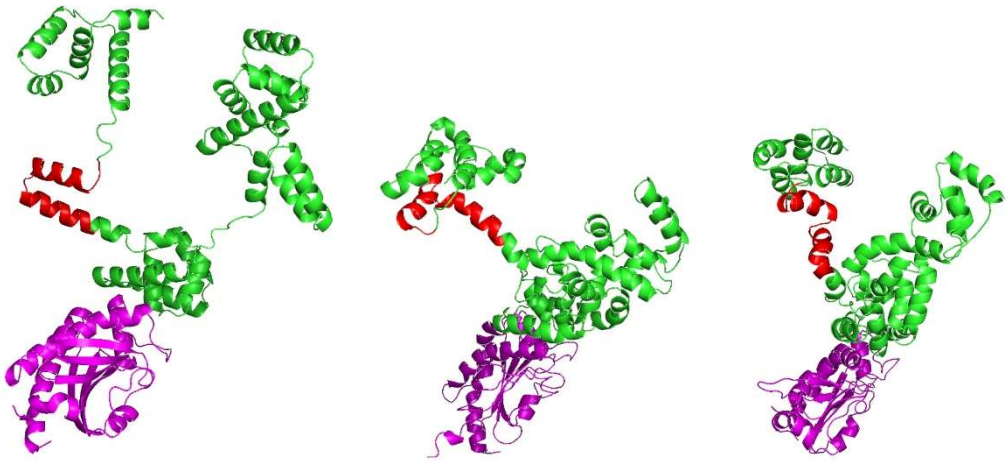


***Figure 4.8*. Alphafold prediction model and MDFF outcomes.** Alphafold predicted FliG (green) with crystal structure of FliM-M (magenta) added, before MDFF (left), after 3 copy MDFF run in CCW trimer density (middle) and after 3 copy MDFF run in CW trimer density (right). The red part corresponds to helices 6 and 7 as seen in 5WUJ crystal structure.

5WUJ-based models that demonstrated the highest correlation were subsequently evaluated via additional features for final selection of the best integrative models. For both CW variants the helix between the N-terminal and middle domains is

fully extended, but for CCW variants slightly bent, but individual parts were not packed against each other (Figure 4.9). For the CW variant, the 3USY based C-terminal region was selected over the 3USW based model as there were interactions between the C-terminal and N-terminal regions that have not been experimentally reported and are likely not biologically relevant. For CCW models the same was true for the 3USW based C-terminal region, and therefore the 3USY based model was selected, supported by the fact that it has the highest correlation score.

The resulting C-ring models are in agreement with cryoET densities (Figure 4.10).

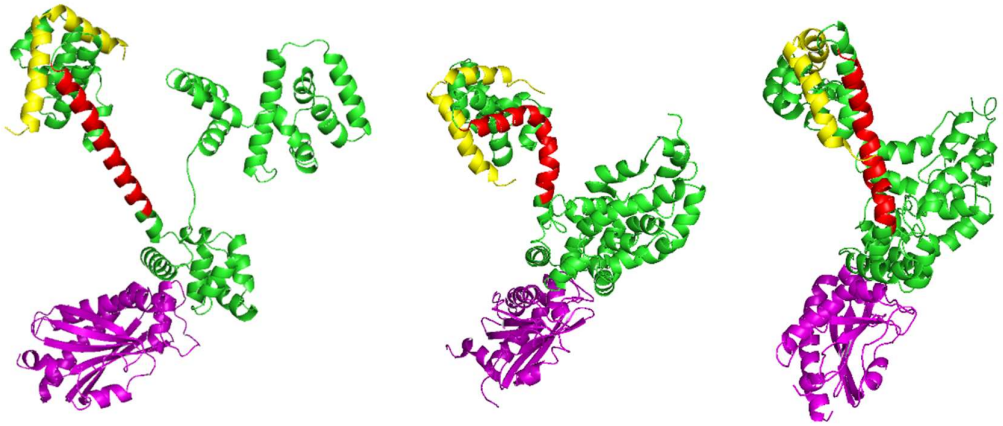


***Figure 4.9*. Monomers used in building FliG/FliM ring.** FliF-C/FliG/FliM-M complex before MDFF, based on homology modelling and crystal structures (left), after 3 copy MDFF run in CCW trimer density (middle) and after 3 copy MDFF run in CW trimer density (right). All models have 3USY based C-terminal domain conformation. After MDFF, individual models were used to build 34 copy ring models. FliG (green), FliM-M (magenta), FliF-C (yellow). The red part corresponds to helices 6 and 7 as seen in the 5WUJ crystal structure.

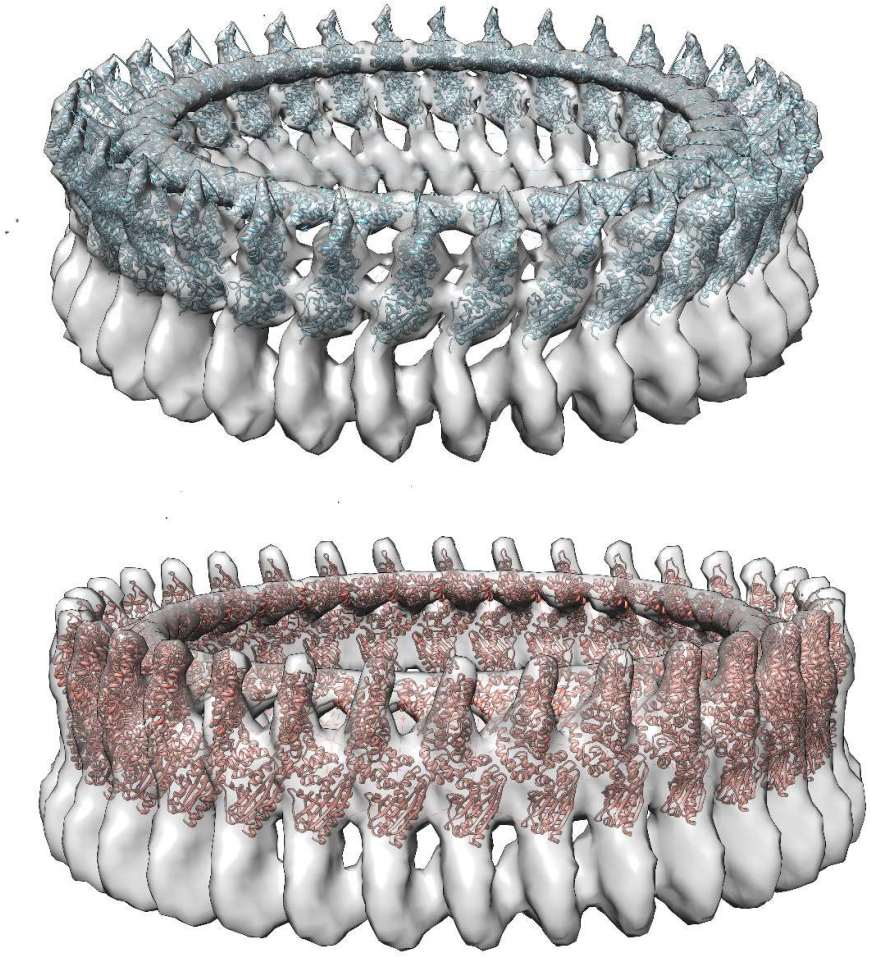


***Figure 4.10*. Full length *H. pylori* FliF-C/FliG/FliM-M 34-mer model docking into CCW and CW rotor cryoET density of *V. alginolyticus***. Top – full length FliG/ FliM-M model in CCW density, bottom – full length FliG model in CW density. The density is oriented with the FliN layer at the bottom of the image, and the FliF/membrane layer at the top.

## 4.5. Discussion

The question about finding a physiologically relevant linker conformation between FliG-N and FliG-M that arose from results presented in chapter 3 lead to the creation of a workflow that combines, homology modelling, fitting into cryoET density and MDFF (Figure 4.1). Combing existing crystal structure information via homology modelling resulted in a reasonable FliG/FliF-C model (Figure 4.4), however it was not fully reconcilable with known cryoET data. Therefore, MDFF was employed to reach

the final MODELLER derived FliG/FliF-C/FliM-M models (Figure 4.9). Furthermore, the final *H. pylori* models obtained after MDFF are discussed (Figure 4.9).

The compacting of *H. pylori* C-terminal domain observed after MDFF is expected since in previous "static" monomer fitting attempts previously this domain resided partially outside of the density and MDFF provided the density as a restraint (Figure 4.9). However, it is also likely to be biologically relevant, because in an assembled state of such complex machinery, it would be unlikely for FliG-C to remain as a flexible linker as seen in the crystal structure. It is, though, imaginable that this linker is flexible under certain conditions, e.g., before or during the assembly steps.

The observation that in the final MODELLER-based FliG models the CCW state retains an extended helix and the CW state is in a slightly bent state provides an opportunity to speculate whether the helix may have a functionality in the switching between the two states (Figure 4.9). However, further investigation would be needed to support that.

Observing that the comparative Alphafold model after MDFF "opened up" compared to the crystal-structure like conformation is encouraging for the hypothesis that the helix has to be extended or partially extended to bridge FliG-N and FliG-M domains across the space between the inner and outer lobes of the C-ring density (Figure 4.8).

Based on both the final MODELLER-based integrative homology model and the Alphafold-based model, following MDFF, it is clear that FliG-N and FliG-M are likely linked by an extended linker. The precise behaviour of the two models during MDFF, however, differs. Thus, in the CCW state fitting of the homology model the linking helix is slightly bent, but fully extended in the CW state. In contrast, in both the Alphafold CCW and CW models, after opening up the bent state, helix 6 is observed to pack against FliG-N due to being flanked by flexible linkers and likely unable to assume an extended helical state because of that. It should be noted that a search of the Alphafold database showed the recent addition of an independent *H. pylori* model for the protein, which is very similar to the Alphafold model created in this study, and interestingly, low-confidence scores were present in areas surrounding the FliG helices 6 and 7 show; thus, while individual domains are indeed able to reproduce their crystallographic counterparts, the linkers between them should be subject to caution and/or may adopt different structures under different states of function or assembly (Figure 4.11).

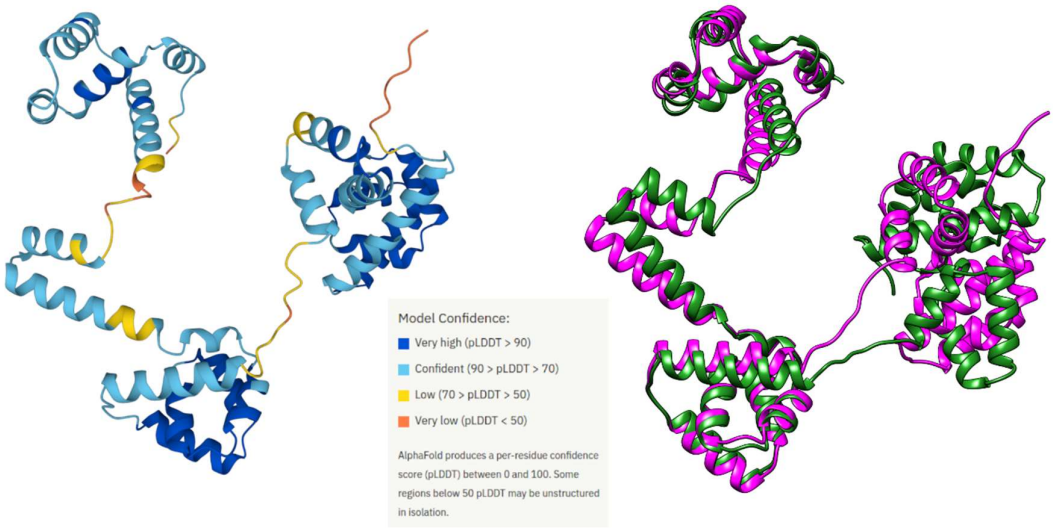



***Figure 4.11*. Alphafold *H. pylori* FliG model confidence score.** Left *H. pylori* FliG Alphafold model from database, coloured by confidence score from Alphafold webpage. Comparison between Alphafold model crated with Alphafold script (dark green) and model currently available in Alphafold database (magenta), that has confidence scores available for comparison.

FliM is the protein that receives signal for rotation-switch and passes it upwards to the FliG (Minamino et al., 2019), which will result in flagellar rotation changing direction from clockwise (CW) to counter-clockwise (CCW). This process involves major structural rearrangements as seen from differences in cryoET densities (Carroll et al., 2020). FliM-M was added to all FliG/FliF-C models as part of the MDFF based modelling approach.

While *A. aeolicus* does contain a sequence that by homology is identified as FliM in the Uniprot database (Bairoch et al., 2004; Bateman, 2019) (entry A0A7C5QKN5), it should be considered putative only. Annotation of the original genome sequence of *A. aeolicus* did not identify a *fliM* (Deckert et al., 1998) and this has therefore been accepted in schematic depictions of the *A. aeolicus* flagellum (Takekawa et al., 2015). The possible absence of FliM is not addressed in the rotational switching model in the context of the *A. aeolicus* FliG crystal structure, where analysis of FliG and FliM interaction is part of the discussion (Lee et al., 2010). It is nevertheless likely that it must at least have a functional homologue, since the entire rotational

switching mechanism is difficult to imagine without any corresponding component in place. Indeed, search of SMART database (Letunic et al., 2015) also identifies the typical domains in *A. aeolicus* FliM. Due to some remaining uncertainty of FliM status, homology modelling based on *H. pylori* with the most similar *A. aeolicus* sequence available may be a sub-optimal choice. The *A. aeolicus* homology model also had lower quality in all scores than other homology models built in this study. The use of the FliM-M domain here, however, had a secondary importance in preventing FliG-M from falling into the FliM-M "space" during MDFF simulation runs. The *A. aeolicus* FliM-M model therefore should not be used to draw any conclusions about the FliM ring in this species or interactions with FliG that have a high chance to differ from those of the *H. pylori* motor. The *A. aeolicus* FliG/FliM-M ring has thus served in this study as a comparative model.

The final CCW monomer model still has part of the of the C-terminal domain in proximity of the extended helix (corresponding to helices 6 and 7) (Figure 4.9), although the C-terminal domain is not directly interacting with the N-terminal domain like in the other rejected models (APPENDIX D). This interaction is likely artificial and simply a by-product of the fitting of the structure into the density. Extended runs of MDFF and/or more extensive sampling to equilibrate the structures may reveal a more optimized CCW monomer model. The MODELLER-based MDFF models (Figure 4.9) seem slightly more physiologically feasible than MDFF processed Alphafold models (Figure 4.8), because there are no structural "breaks"- loops even, if the helical linker bends unlike in the Alphafold derived model.

None of the FliG models in this study, before or after MDFF, bring the two residues that cross-link in a study that argues there is no extended linker between FliG-N and FliG-M (Paul et al., 2011b). FliG protein in the cross-linking study was overexpressed without presence of FliF, but 1:1 binding and folding is important for both proteins (Levenson et al., 2012; Lynch et al., 2017), therefore the state FliG is found in the cross-linking experiment may represent transport, assembly or other intermediate state in a process involving FliG, other than final functional ring assembly.

Both final ring assemblies (Figure 4.10) have high correlation coefficients with the cryoET densities (>95). The CCW density model in this regard demonstrates the possibility that the loops on the outside of FliM-M may interact with neighbouring FliM-M copies in a dynamic setting. MD simulations of the entire ring complex may be

needed to demonstrate this. The conformation of FliM-M in the CW ring, although fitting well, may not necessarily be reflective of the most optimal variant as the subunits are oriented more vertically leading to a greater distance between them compared to the CCW counterpart.

It should be noted that while an optimized process was developed to generate the best models, there are potential drawbacks of the work, and additional work that could be done, as outlined below:

Firstly, the issue of initial FliG placement in the cryoET density prior MDFF must be addressed. FliG C-terminus can either interact with FliG-M of the same monomer (termed "$C^0$ position" in this study), or the FliG-M next to it ("$C^{-1}$ position") (1.6., Figure 4.9.) The placement of the FliG/FliF-C/FliM-M trimer during MDFF was by default in $C^0$ position and the effects of trimming a different $C^{-1}$ starting position could also be explored. In this regard, a bias has been created in the modelling process.

In addition, the results that favour an extended helix joining FliG-N and FliG-M do not exclude the possibility that the bent conformation may be present in some other conditions, such as during intermediate states of assembly.

Finally, the *H. pylori* ring was modelled here using the best available cryoET density that comes from *V. alginolyticus* and has a 34-mer stoichiometry. However, low resolution data shows that different species can have a variable C-ring diameter (Beeby et al., 2016) while newly released data shows the *B. burgdorferi* C-ring ring has 46-fold symmetry (Chang et al., 2020). Since there are no large differences in FliG length (Figure 4.2) it is more likely that different diameters are achieved with varying subunit number. A low resolution cryoET study of *H. pylori* shows it has a wider diameter (Qin et al., 2017) than the *S. enterica* motor resolved before, therefore it is possible that the *H. pylori* ring may have a stoichiometry larger than 34-fold that was modelled in here. Importantly, the *B. burgdorferi* flagellum is unusual as it is a periplasmic flagellum (Chang and Liu, 2019) instead of a flagellum where the propeller is in contact with the outside environment, and therefore vastly different stoichiometries may be a functional adaptation that sets *B. burgdorferi* apart from other types of flagella. The full density of *B. burgdorferi* was not available in the database at the time of model building. Interestingly, however, the figures in the paper indicate that in the authors' attempt at modelling FliG, FliM and FliN in the C-ring density, the N-terminal domain is placed a considerable distance apart from FliG-M and FliG-C, and this distance could only be

conceivably achieved by an extended, straight connecting helix (Chang et al., 2020). Future availability of this density and the authors' model could possibly deliver additional support to the *H. pylori* extended helix model.

While the modelling using this methodology creates a theoretical possibility that may reflect how the FliG ring looks, it can, however, be used for generation of other experiments and hypotheses; these range from further use in MD experiments, to cross-linking experiments *in vitro*.

As a direct continuation of the results in Chapter 3, here a FliG model compatible with the MD data obtained has been constructed. To fill the void of a FliG ring structure that is not currently solved, unlike that of its binding partner FliF (Johnson et al., 2020; Kawamoto et al., 2021), two possible C-ring models corresponding to CW and CCW states using an improved FliG/FliF-C/FliM-M model have been created. Further studies of FliG domain interactions are needed, while many aspects remain uncovered to be explored in the future, such as the influence of the C-ring stoichiometry which likely differs between species.

# Chapter 5 - CryoET and EM studies of *Campylobacter jejuni* flagellar motor

## 5.1. Introduction

*C. jejuni* is a model organism in flagella research that represents the amphitrichous mode of flagella placement, and it is used to explore flagella transcription regulation and mechanisms that govern flagella biosynthesis and placement (Balaban and Hendrixson, 2011). FlhF is a regulatory protein, primarily associated with bacteria characterised with polar flagella placement. Evidence suggests that it has different functions in different bacterial species; deletion of *flhF* makes cells aflagellate, but the mechanism of how it is involved in flagella regulation and assembly remains unclear (Ren et al., 2018; Stoakes, 2017). In *C. jejuni* in an *ΔflhF* background pseudorevertants arise to restore motility by SNP mutations directly in the genes coding for the flagella motor structural components *fliF*, *fliG* and *fliH* (Stoakes, 2017). In some pseudorevertants, a single non-synonymous SNP in *fliF* was sufficient to restore motility. Among the best characterized mutants is a strain containing $fliF_{M99I}$ (Table 1.1.); characterization of this pseudorevertant ($FliF_{M99I}$) revealed vast changes across the transcriptome in a growth-phase dependent manner (Stoakes, 2017). One of the aims of this project was to further confirm that the single *fliF* mutation is responsible for the phenotypes observed by Stoakes (2017). We will therefore aim to characterise *C. jejuni* constructs in which a gene expressing $FliF_{M99I}$ at varying levels is re-introduced back into the chromosome (see Methods Chapter, Section 2.2.2., Table 2.3, and also Introduction Chapter, Section 1.7., Table 1.1) in both *ΔflhFΔfliF* background and a *ΔfliF* background. Effects of $fliF_{M99I}$ in the presence of *flhF* have not so far been observed.

For a long time, an *S. enterica* low resolution EM map (Thomas et al., 2006) was the single source of information of the C-ring, until the recent release of *V. alginolyticus* maps (Carroll et al., 2020). No structure at any resolution is available for the *C. jejuni* C-ring, nor is there an established stoichiometry for the C-ring in this species, making it hard to ascertain the effect of the SNPs observed in the pseudorevertants. In this part of the project we aimed for a a low resolution cryoET reconstruction of WT and pseudorevertant (*ΔflhF* $fliF_{M99I}$) *C. jejuni* flagellar motors (see Methods Chapter, Section 2.2.2., Table 2.3, and also Introduction Chapter, Section 1.7., Table 1.1) to

determine the stoichiometry and size of the *C. jejuni* motor as well as check for large scale structural rearrangements in the pseudorevertant macromolecular structure.

There are two different approaches explored in this chapter. A microbiology approach is used to create a set of complementation strains to explore the $fliF_{M99I}$ phenotype. Sequencing is performed to verify the deletion and integration process. Subsequently, cryoET datasets are collected of WT *C. jejuni* and $fliF_{M99I}$ containing pseudorevertant to gain additional structural insights. Results will be presented in separate subsections.

## 5.2. *Campylobacter* complementation strain construction

*C. jejuni* NCTC11168 strain is a laboratory strain obtained from Prof. Charles Penn, University of Birmingham. NCTC11168 was first sequenced in a published study (Parkhill et al., 2000), but the isolate was later found to be non-motile. A motile isolate was sequenced later (Gaynor et al., 2004), but the NCTC11168 variant used in this study comes from the original clinical isolate and is different from both of published isolates. Differences between the NCTC11168 strain used in this study and the sequence published have been identified by Emily Stoakes (Stoakes, 2017). *C.jejuni* NCTC11168 strain cannot express heterologous vectors, and therefore, only suicide vectors can be used for complementation and overexpression studies.

Construction of the *C.jejuni* NCTC11168 *ΔflhFΔfliF* strain previously carried out in our laboratory (Emily Stoakes and Charlotte Dixon) utilized the promoter-less antibiotic cassettes expressing kanamycin and chloramphenicol resistance respectively. In this project, the *C. jejuni* NCTC11168 *ΔfliF* strain was constructed using the same suicide vector (GeneArt Gene Synthesis, Thermo Fisher Scientific), enabling the deletion of the *fliF* locus in the wild-type background (Methods, 2.2.3., Table 2.4) that replaces the *fliF* locus with a chloramphenicol resistance cassette via homologous recombination (Figure 5.1). The plasmid construct enabled the deletion of most of the *fliF* locus except 111bp at the 5'end and 49 bp at the 3'end. Homologous recombination was achieved via 427 bp- and 498 bp-long regions targeting the upstream and downstream regions of the *fliF* locus that include parts of the *hisC* and *fliG* genes respectively.

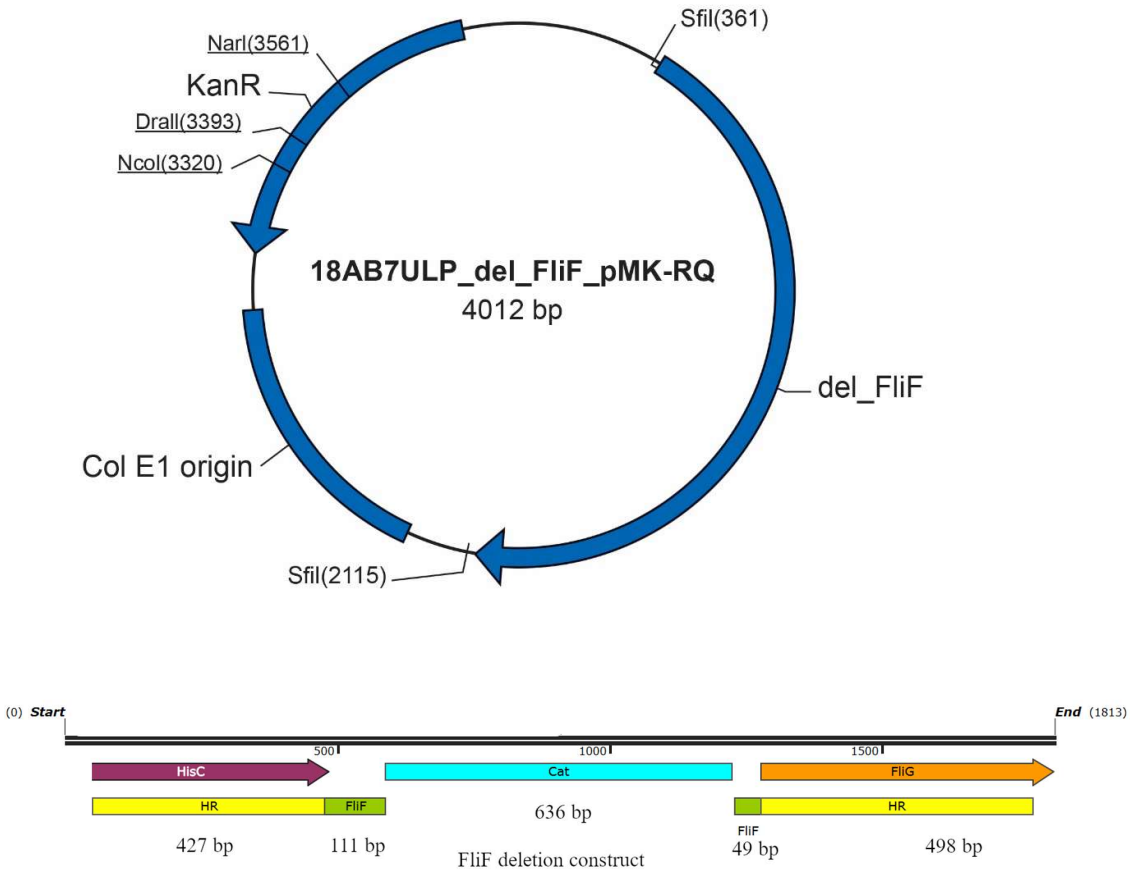



***Figure 5.1. fliF* deletion vector scheme.** Top: Vector contains kanamycin resistance and del_FliF contains full chloramphenicol cassette that integrates in *fliF* locus. Below: detailed del_FliF sequence. HR denotes regions that engage in homologous recombination. FliF – remaining fragments of FliF, Cat – chloramphenicol resistance cassette, HisC upstream gene fragment in HR region, FliG downstream gene fragment in HR region.

Putative mutants were subjected to Sanger sequencing using primers that annealed outside the recombination sites to check for any mutations and scar-less recombination sites. Those clones that were correct were carried forward for further analysis. The same analysis was also applied to the already available *C. jejuni* NCTC11168 *ΔflhFΔfliF* for which the recombination sites were not checked before. Analysis focused on verifying that the flanking genes were not compromised and the downstream *fliG* RBS (ribosome binding site) (AGGA) remained intact, ensuring that FliG transcription would not be compromised in the newly created *ΔfliF* strain. A single *ΔfliF* clone was selected for further work.

The *C. jejuni* vector pC46 (APPENDIX E) is widely used for complementation studies and was also utilised for the construction of strains expressing FliF and FliF$_{M99I}$. The vector allows for the integration of the gene of interest in the *Cj0046* pseudogene locus (Reuter and van Vliet, 2013; Thomas et al., 2011). However, selection for a successful insertion is based on a chloramphenicol resistance cassette. Therefore, a modified pC46 had to be created that allowed for selection of successful complementation with an alternative antibiotic cassette. Apramycin is one of the few used and characterized resistance cassettes used for *C. jejuni* genetic manipulation (Cameron and Gaynor, 2014). Using Gibson assembly, a pC46 plasmid was first constructed in which aparamycin under the control of the native *C. jejuni* promoter *fdxA* was inserted divergently from the chloramphenicol cassette (Shaw et al., 2012) (Figure 5.2, Step 1). Successful constructs were selected on agar plates with both chloramphenicol and apramycin. The existing chloramphenicol cassette was then replaced, using Gibson assembly, with either wild-type *fliF* or *fliF*$_{M99I}$ under the control of the wild-type *fliF* native promoter or the *C. jejuni porA* prmoter. The latter is a strong promoter that is expected to constitutively overexpress *fliF* (Jervis et al., 2015) (Figure 5.2, Steps 2-4)). This time successful constructs were first selected for their resistance to apramycin, and then screened for their sensitivity to chloramphenicol. Putative clones were screened by whole cell PCR; plasmid DNA was extracted from positive clones, followed by Sanger sequencing (Figure 5.2).

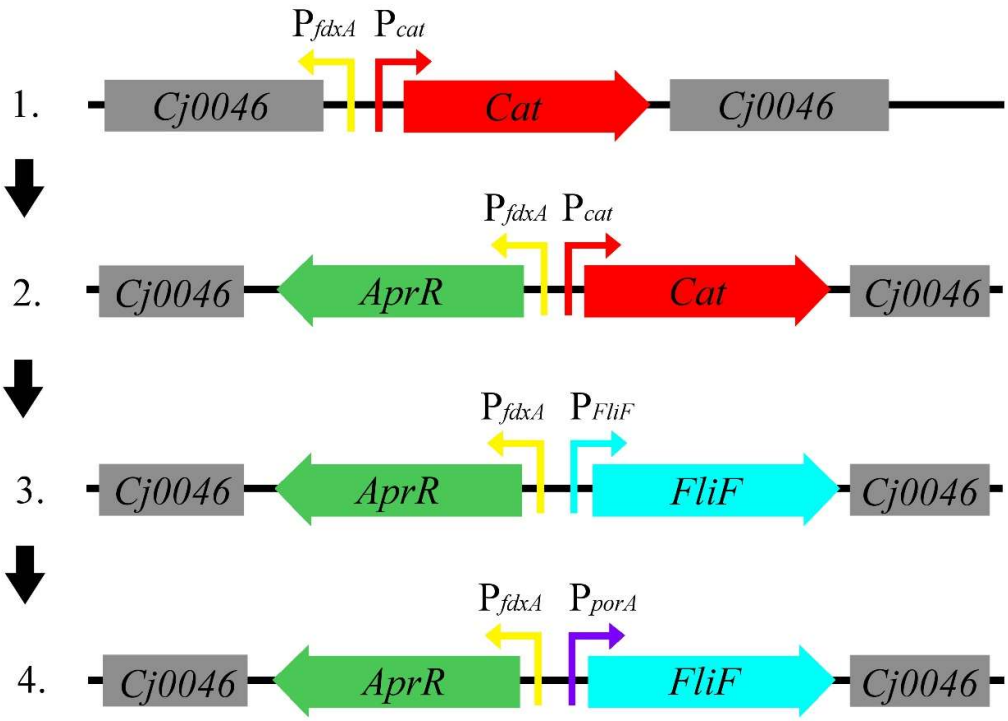



***Figure 5.2.*** **Cloning scheme to create pC46 integration vector with apramycin resistance and FliF ORF**. Step 1. pC46-fdxA (modified pC46 with fdxA promoter inserted (created previously, Table 2.4., APPENDIX E). Step 2. Apramycin resistance inserted under fdxA promoter. Step 3. Chloramphenicol resistance is exchanged to FliF under its native promoter. Step 4. Native FliF promoter is exchanged to porA promoter.

*ΔfliF* or *ΔflhFΔfliF C.jejuni* cells were then transformed with the pC46-AprR plasmids and positive recombination clones were identified with colony PCR, using primers targeting upstream and downstream regions of the *Cj0046* insertion site, as well as primers targeting the inserted sequences.

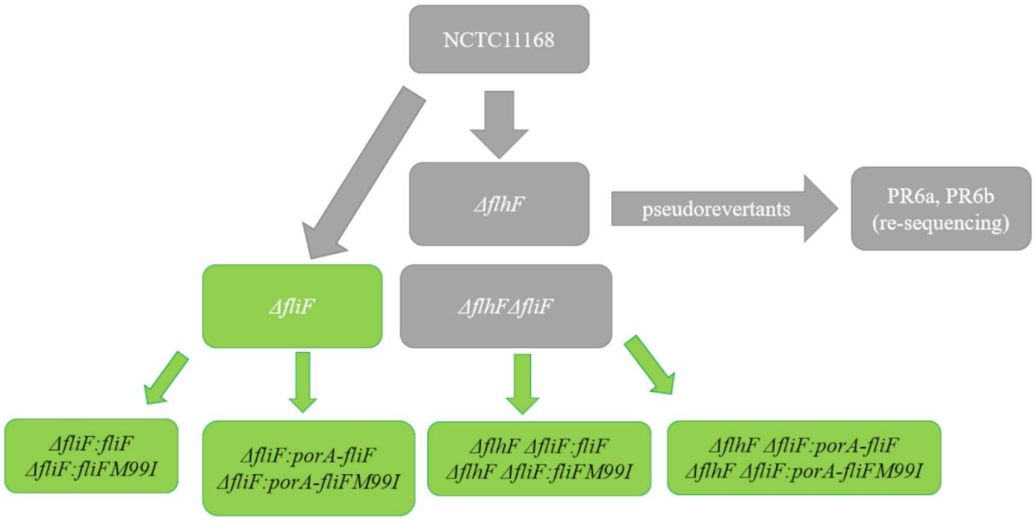



***Figure 5.3*. Lineage of *C. jejuni* strains created and sequenced.** Green – strains created in this study. Grey - previous origin (see Methods 2.2.2). NCTC11168 and *ΔflhF* were not sequenced and are shown to depict the lineage of strains only.

## 5.3. Sequencing of complementation strains

Although loci in the immediate vicinity of insertion and deletion sites in the chromosome were checked during the construction of the new strains through PCR and Sanger sequencing, whole genome sequencing was also carried out on all new and parental strains to eliminate possibility of other loci having been affected during the genetic manipulation steps and while passaging the strains afterwards (Figure 5.3). At least two different clones from each strain constructed in this project were sequenced (Table 5.1).

Genomic DNA was extracted, and sequencing libraries were prepared as described in Methods (2.2.5); all samples were pooled and sequenced using a single MiSeq V2 cartridge. Analysis of the fastq files using MultiQC (Ewels et al., 2016), revealed that although the duplication rates were 30-40% due to the high coverage per sample, the sequencing data was of good quality, with good distribution of reads amongst the pooled samples. All sequenced strains were mapped to the NCBI *C. jejuni* NCTC11168 (Accession Number: NC_002163.1); all samples had above 95% reads mapping to the *C. jejuni* chromosome, and above 170x coverage.

Analysis of the sequencing data showed that the deletion of *fliF* in the *ΔflhFΔfliF strain* created previously by (Stoakes, 2017) was successful. It also

confirmed that the integration of wild type *fliF* or *fliF*$_{M99I}$ in all complementation strains was at the expected location in the *Cj0046* pseudogene and all *ΔflhF* strains contained the *flhF* deletion.

***Table 5.1*. Whole genome sequencing**. Table includes information on GC content, reads and coverage. Reference genome - NC_002163.1

| *Strain and sample ID* | *GC content (%)* | *Reads per sample* | *Average Read length (bases)* | *Reads mapping to C. jejuni NCTC11168 (paired)* | *Coverage* | *Coverage standard deviation* | *Average Insert length (bases)* | *Average Insert length standard deviation (bases)* |
|---|---|---|---|---|---|---|---|---|
| *ΔflhFΔfliF _S21* | 36 | 1167363 | 147 | 1127089 (96.55%) | 200 | 209 | 159 | 2208 |
| *ΔflhFΔfliF:natfliF*$_{M99I}$*_S13* | 38 | 1331738 | 116 | 1291386 (96.97%) | 179 | 242 | 121 | 2252 |
| *ΔflhFΔfliF:natfliF _S10* | 35 | 1333815 | 168 | 1274060 (95.52%) | 260 | 256 | 195 | 3662 |
| *ΔflhFΔfliF:natfliF _S11* | 37 | 1026438 | 135 | 992566 (96.7%) | 162 | 207 | 148 | 2551 |
| *ΔflhFΔfliF:natfliF*$_{M99I}$*_S14* | 37 | 1052629 | 149 | 1009366 (95.89%) | 181 | 234 | 163 | 2521 |
| *ΔflhFΔfliF:natfliF _S12* | 37 | 1260776 | 125 | 1225348 (97.19%) | 184 | 230 | 130 | 1005 |
| *ΔflhFΔfliF:natfliF*$_{M99I}$*_S15* | 37 | 1365957 | 139 | 1312002 (96.05%) | 219 | 270 | 149 | 2716 |
| *ΔflhFΔfliF:porAfliF*$_{M99I}$*_S18* | 36 | 897874 | 164 | 857200 (95.47%) | 170 | 182 | 182 | 2867 |
| *ΔflhFΔfliF:porAfliF _S16* | 36 | 1197731 | 135 | 1157008 (96.6%) | 188 | 190 | 148 | 2626 |
| *ΔflhFΔfliF:porAfliF _S17* | 37 | 1200815 | 135 | 1158907 (96.51%) | 188 | 252 | 141 | 1874 |
| *ΔflhFΔfliF:porAfliF*$_{M99I}$*_S19* | 36 | 1053132 | 155 | 1006584 (95.58%) | 189 | 189 | 174 | 2743 |
| *ΔfliF_S20* | 36 | 1197248 | 145 | 1160971 (96.97%) | 203 | 214 | 153 | 1121 |
| *ΔfliF:natfliF_S1* | 36 | 1963291 | 160 | 1882796 (95.9%) | 363 | 447 | 177 | 2246 |
| *ΔfliF:natfliF_S2* | 37 | 1955322 | 141 | 1890014 (96.66%) | 321 | 410 | 152 | 2306 |
| *ΔfliF:natfliF*$_{M99I}$*_S3* | 36 | 2475400 | 146 | 2389504 (96.53%) | 423 | 439 | 160 | 2449 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *ΔfliF:natfliF$_{M99I}$_S4* | 37 | 2348536 | 138 | 2272913 (96.78%) | 378 | 467 | 147 | 1561 |
| *ΔfliF:porAfliF_S5* | 37 | 1816964 | 152 | 1745012 (96.04%) | 319 | 412 | 170 | 2961 |
| *ΔfliF:porAfliF_S6* | 37 | 1313886 | 157 | 1260674 (95.95%) | 239 | 299 | 179 | 3092 |
| *ΔfliF:porAfliF_S7* | 36 | 1048123 | 178 | 995926 (95.02%) | 214 | 260 | 212 | 3668 |
| *ΔfliF:porAfliF $_{M99I}$_S8* | 37 | 1022903 | 176 | 971349 (94.96%) | 206 | 276 | 202 | 3732 |
| *ΔfliF:porAfliF $_{M99I}$_S9* | 36 | 873008 | 174 | 829969 (95.07%) | 175 | 186 | 204 | 2813 |

SNP and INDEL analysis was carried out on all strains with respect to the NCBI *C. jejuni* NCTC11168 (Accession Number: NC_002163.1) sequence. In comparison with the reference strain seven point mutations were identified in all samples. The same mutations had been identified in the previous study in the NCTC11168 *C. jejuni* strain used in the laboratory (Stoakes, 2017) and most of them have been found to arise commonly during the domestication of this strain. Exception to this are the SNPs in cipA, coding for an invasion protein, and hisG coding for an ATP phosphoribosyltransferase, both of which are unique to this laboratory strain (Pascoe et al., 2019).

INDEL (insertion and deletion) analysis, which was not carried out previously by Stoakes (2017), revealed two deletions and one insertion present in all strains and parent strains. The insertion is present in an intergenic region between two divergent genes, *nalD* (a nicotinate-nucleotide adenylyltransferase) and *gapA* (a glyceraldehyde 3-phosphate dehydrogenase). The region is characterised by a string of six A nucleotides, that with the insertion becomes a string of seven A nucleotides. Although it may potentially have an influence on the expression of the surrounding genes, it is difficult to speculate.

***Table 5.2*. Mutations and insertions/deletions present in all strains created in this study in comparison with the reference genome (**NC_002163.1). PR6a and PR6b are not included.

| *Gene* | *Position in reference* | *Change* | *Amino acid change* | *Deletion and complementation strains* |
|---|---|---|---|---|
| | | ***Point mutations*** | | |
| *mreB* | 253191 | A to G | D to G | all |
| *cheA* | 262345 | A to G | I to T | all |
| *Cj0431* | 393542 | T to A | * to K | all |
| *Cj0455c* | 420550 | A to G | | all |
| *cipA* | 638796 | A to G | V to A | all |
| *Cj0807* | 760188 | A to G | K to E | all |
| *hisG* | 1525525 | G to A | G to D | all |
| | | ***Insertions/deletions*** | | |
| *Cj0031* | 48995 | AG to A | poly G 10 to 9 | all[1] |
| *Cj0184* | 180705 | TAC to T | V to I | all[1] |
| *dcuA* | 97896 | TTTCA to T | S to * | *ΔflhFΔfliF* and derivatives |
| | | | **Intergenic** | |
| | 1338383 | C to CA | | all[1] |

[1]in some samples reads of both alleles are present, but due to cut-off do not appear in VarScan2 analysis. This means that cell population was heterogeneous.

The other two deletions were found in the *Cj0031* and *Cj0184* genes. The *Cj0031* deletion is found in a poly-G region and is present in all strains (Table 5.2). It encodes a putative Type IIG restriction-modification system that encodes endonuclease and methyltransferase activities in one protein and is among a series of phase-variable genes in NCTC11168 (Aidley et al., 2017; Anjum et al., 2016). These genes are activated by stochastic replication errors and selective pressure that creates a heterogenous population (Aidley et al., 2017). Activation of *Cj0031* changes methylation and expression pattern and confers resistance to some, but not all *C. jejuni* phages (Aidley

et al., 2017; Anjum et al., 2016). A polyG tract of 9 G nucleotides is thought to be in the ON (activated) state for this *Cj0031* (Anjum et al., 2016).

One additional deletion was identified *ΔfliF* and *ΔflhFΔfliF* strains. The *dcuA* gene deletion creates a premature stop codon in this gene, encoding a C4-dicarboxylate transporter (Table 5.2) creating a truncated product of 134 aa instead of 445 aa. *C. jejuni* NCTC11168 has five different C4-dicarboxylate transporters (Wösten et al., 2017). Expression of *dcuA* is repressed under limited oxygen conditions and in the case of its deletion *dcuB* can partially compensate its function (Wösten et al., 2017). No growth impairment in Heart Infusion broth (a rich growth medium) between *ΔdcuA* and WT *C. jejuni* 81116 has been reported before (Wösten et al., 2017), which explains why despite an effective loss of a functional protein, normal growth was seen during cultivation of the *ΔflhFΔfliF* and *ΔflhFΔfliF* based complementation strains, although MH (standard medium for *C. jejuni*) was used. It is reasonably safe to assume that this genome difference should not impact the flagellation phenotype.

Analysis of the complementation strains revealed that there are multiple mutations upstream and downstream the integration site in *Cj0046* pseudogene, however it is not a coding sequence therefore they are not predicted to have a direct consequence on the phenotype. Surprisingly, for 10 of 19 strains sequenced, the mutations extend into the gene, *Cj0053c – mnmA*(*trmU*), which is downstream of *Cj0046* (Table 5.3) but the sequence of which did not form part of the targeted recombination sequence. The product of this gene is not associated with flagella or motility and encodes a tRNA-specific 2-thiouridylase. In nine of the strains the *mnmA* has acquired the same non-synonymous mutation – I309T (Table 5.3).

***Table 5.3*. Sequencing overview of unique mutations in *C. jejuni* complementation strains and parental strains.**

| *Strain* | *ID* | *Locus with mutation* | *Mutated residue* | *Reference sequence* | *Change* | *Selected for negative stain EM* | *Other synonymous mutations in Cj0053c* |
|---|---|---|---|---|---|---|---|
| *ΔfliF:natfliF* | D/S1 | | | | | yes | 2 |
| *ΔfliF:natfliF* | F/S2 | *Cj0053c* | 68623 | A | G | no | 9 |
| *ΔfliF:natfliF(M99I)* | J*/S3 | | | | | yes | |
| *ΔfliF:natfliF(M99I)* | L*/S4 | | | | | no | |
| *ΔfliF:porAfliF* | A/S5 | | | | | yes | |
| *ΔfliF:porAfliF* | B/S6 | | | | | no | |
| *ΔfliF:porAfliF* | D/S7 | | | | | no | 2 |
| *ΔfliF:porAfliF(M99I)* | B*/S8 | *Cj0053c*<br>*fliF* | 68623<br>288482 | A<br>A | G<br>C | no | 4 |
| *ΔfliF:porAfliF(M99I)* | C*/S9 | *Cj0053c*<br>*fliF* | 68623<br>288482 | A<br>A | G<br>C | yes | 9 |
| *ΔflhFΔfliF:natfliF* | 3/S10 | | | | | yes | |
| *ΔflhFΔfliF:natfliF* | 4/S11 | | | | | no | |
| *ΔflhFΔfliF:natfliF* | 7/S12 | | | | | no | |
| *ΔflhFΔfliF:natfliF(M99I)* | 2*/S13 | *Cj0053c* | 68623 | A | G | yes | 6 |
| *ΔflhFΔfliF:natfliF(M99I)* | 6*/S14 | *Cj0053c* | 68623 | A | G | no | 9 |
| *ΔflhFΔfliF:natfliF(M99I)* | 7*/S15 | *Cj0053c* | 68623 | A | G | no | 6 |
| *ΔflhFΔfliF:porAfliF* | 3/S16 | *Cj0053c* | 68623 | A | G | no | 8 |
| *ΔflhFΔfliF:porAfliF* | 4/S17 | | | | | yes | |
| *ΔflhFΔfliF:porAfliF(M99I)* | 1*/S18 | *Cj0053c* | 68623 | A | G | yes | 9 |
| *ΔflhFΔfliF:porAfliF(M99I)* | 5*/S19 | *Cj0053c* | 68623 | A | G | no | 9 |
| *ΔfliF* | βB/S20 | | | | | no | |
| *ΔflhFΔfliF* | E1/S21 | | | | | no | |

Despite not being found in parental strains and having appeared in independent batches of complementation strains, blast searches indicated that the I309T mutation in the *mnmA* gene is also found in other *C. jejuni* strain sequences suggesting that it is probably a natural variant. Presence of other synonymous mutations in the *mnmA* part closer to *Cj0046*, albeit still outside the targeted recombination area that would have been directly involved in homologous recombination during integration, indicates that it is still likely a side effect of the integration (Table 5.3). In pC46-Apr the target

recombination sequences are 765 and 644 nucleotides long on either side of the *Cj0046* integration site and practically involve almost the entire *Cj0046* locus. *Cj0053c* is outside this direct homologous recombination site.

Interestingly, there is an additional non-synonymous mutation in *fliF* in both *ΔfliF:porAfliF(M99I)* clones sequenced (Table 5.3). This involves changing residue Q9 to proline, which is a non-conservative mutation (a mutation that replaces the previous amino acid with different physicochemical properties). Analysis was carried out to determine how conserved this amino acid is in FliF sequences from other bacterial species. Using FliF sequences from the same motile bacterial species used in Chapter 4 (4.2.1., Figure 4.2) for the FliG analysis, sequence alignment showed that it is not conserved (Figure 5.4). Interestingly a sequence alignment carried out at an earlier stage of the project (APPENDIX C), that contained a few additional FliF sequences from species that are not known to be motile, but do encode a protein with homology to FliF indicate that this residue could be partially conserved, increasing the importance of the uncovered mutation in this site.

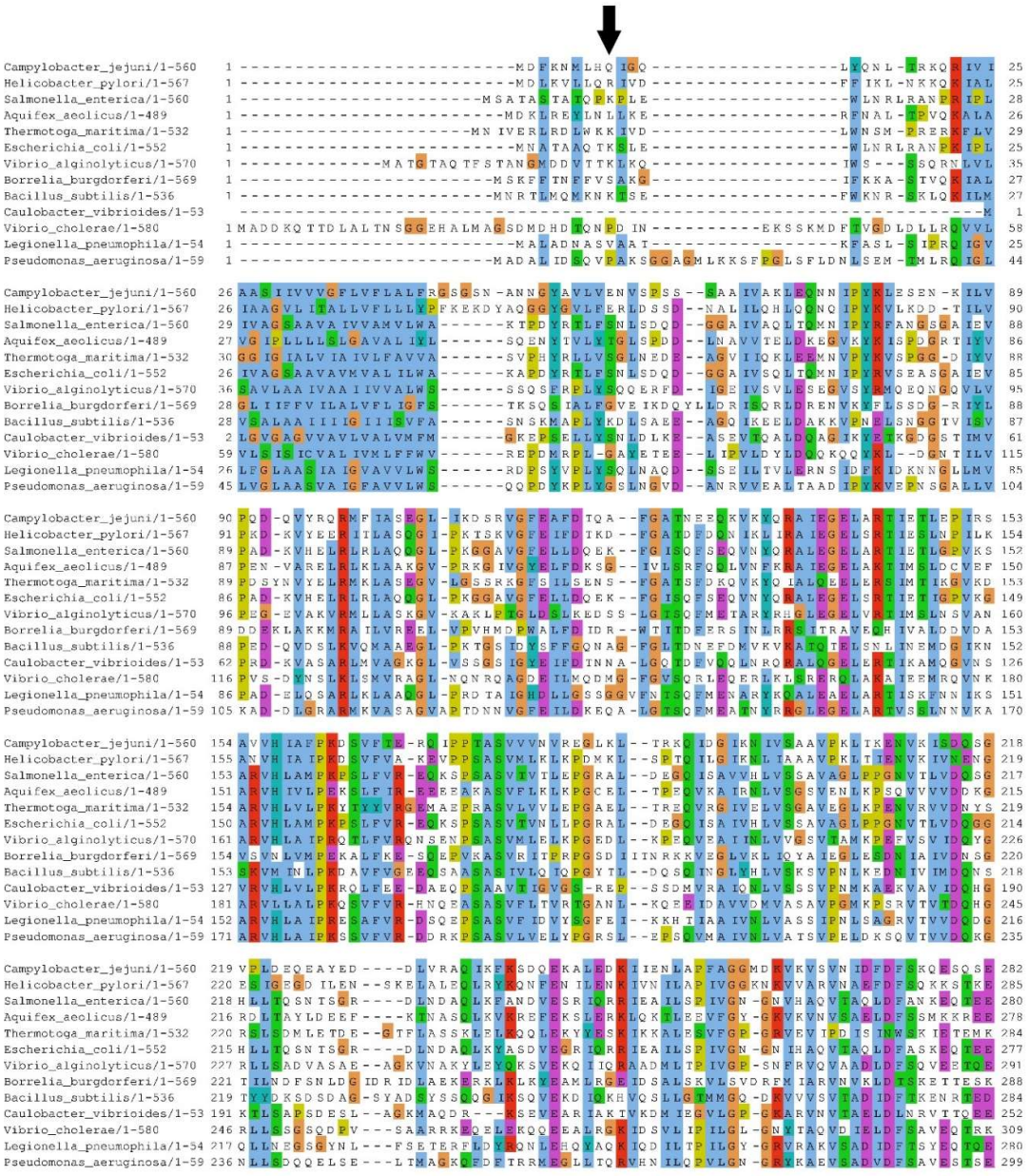



Figure continued on next page

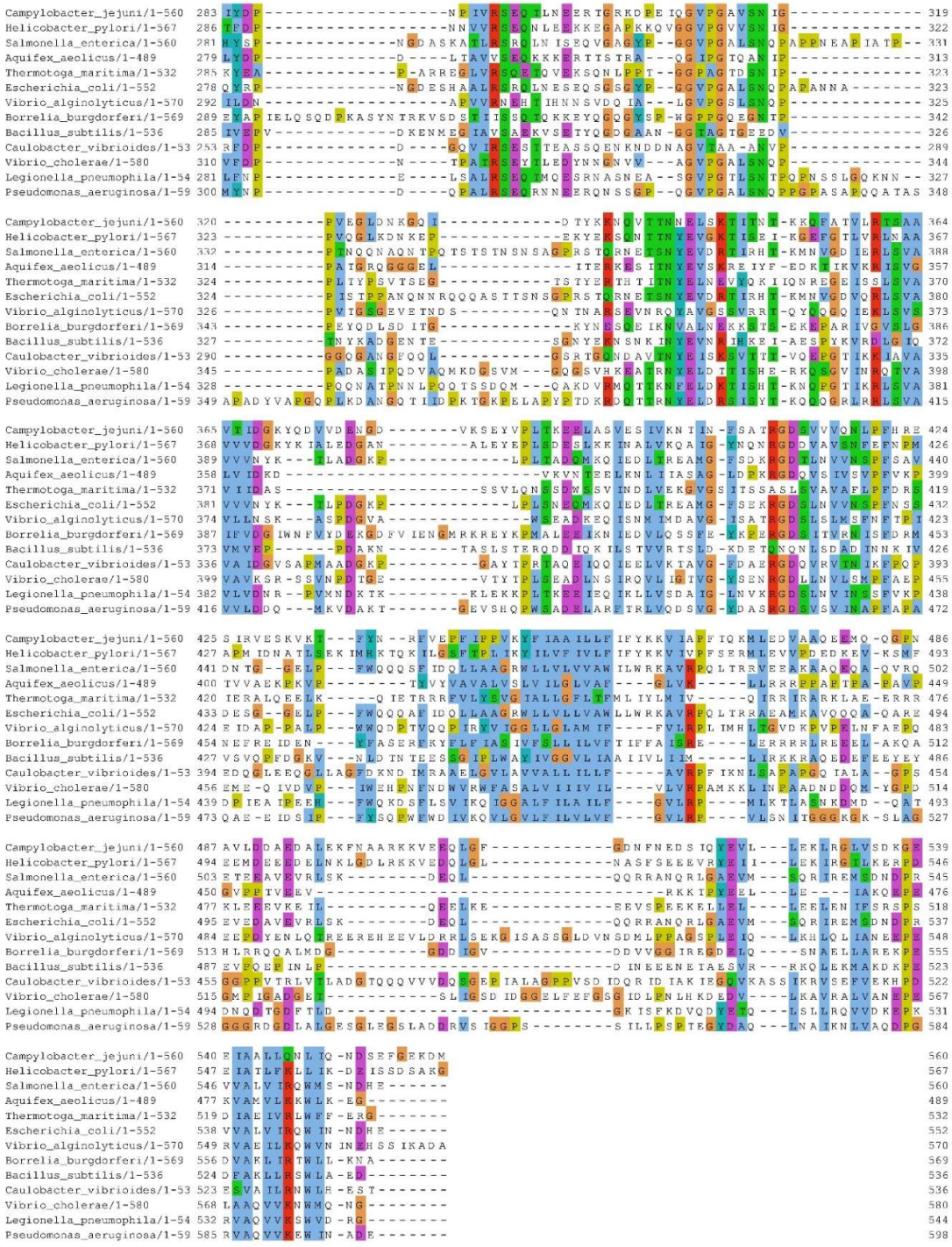



***Figure 5.4*. Sequence alignment of FliF from various flagellated bacterial species representing model organisms and pathogens.** The FliF sequence is highly conserved in the transmembrane domains (approximately residues 20-50 and 450 to 482 in *S. enterica*), moderately conserved in the RMB1 (approx. 5-103 in *S. enterica*) and RBM2 (approx. 125- 218 in *S. enterica*) domains, and has variable conservation in other parts (approx. 230 – 435 for split RBM3 with β-collar, 525 to 560 FliG binding domain in *S. enterica*).

According to Bergeron (2016), Q9 of the FliF N-terminus is predicted to be part of the protein's secretion signal (Bergeron, 2016) although a search of the Signal IP server 5.0 (Kihara, 2010) does not show that FliF would have a typical secretion signal and it has not been verified as a secretion signal in any molecular biology studies.

Despite the small mismatch between the expected genotype and sequencing results in several loci the complementation strains were otherwise found suitable for a preliminary phenotypic characterization with electron microscopy. Selection of the complementation strains characterised above was carried out as identified in Table 5.3.

## 5.4. Negative stain electron microscopy of complementation strains

Previous studies by Stoakes (2017) showed that *ΔflhF* is completely aflagellate in both exponential and stationary growth phases, while *ΔflhF* pseudorevertant with $fliF_{M99I}$ mutation has ~33% of unipolar flagellated cells. Only NCTC11168 strain has fully flagellated cells on both ends in exponential (~10%) and stationary (~45%) growth phases. In order to confirm the extent of regain of function in our complemented strains, in this case correct and timely assembly of flagella structure, the number of flagella and their placement was quantified. As mentioned above, several of our complemented strains contained additional SNPs. For the *ΔflhFΔfliF:natfliF(M99I)*, *ΔflhFΔfliF:porAfliF(M99I)* and *ΔFliF:porAfliF(M99I)* strains all sequenced clones contained the I309T mutation in *mnmA* (Table 5.3) and all *ΔflhFΔfliF* derived complementation strains have a *dcuA* truncation and potentially all observed strains have a phase-variable Cj0031 locus (Table 5.2). It is not known, if these SNPs have any influence on phenotype, although they do not seem to be directly connected to motility.

Prior results it was expected that *ΔfliF:natfliF* and *ΔfliF:porAfliF* would be flagellated and differences between these two would show if availability of FliF is a limiting factor for flagella assembly. Likwise *ΔfliF:natfliF(M99I)* and *ΔfliF:porAfliF(M99I)* would show, if the SNP has any effects in presence of FlhF and, if it is modulated by the amount of FliF. *ΔflhFΔfliF:natfliF* would not be flagellated as it should be equivalent to *ΔflhF* strain. If *ΔflhFΔfliF:porAfliF* was flagellated it would show that amount of *fliF* limits flagella assembly. *ΔflhFΔfliF:natAfliF(M99I)* should reproduce original pseudorevertant phenotype *ΔflhFΔfliF:porAfliF(M99I)* may have increased flagellation.

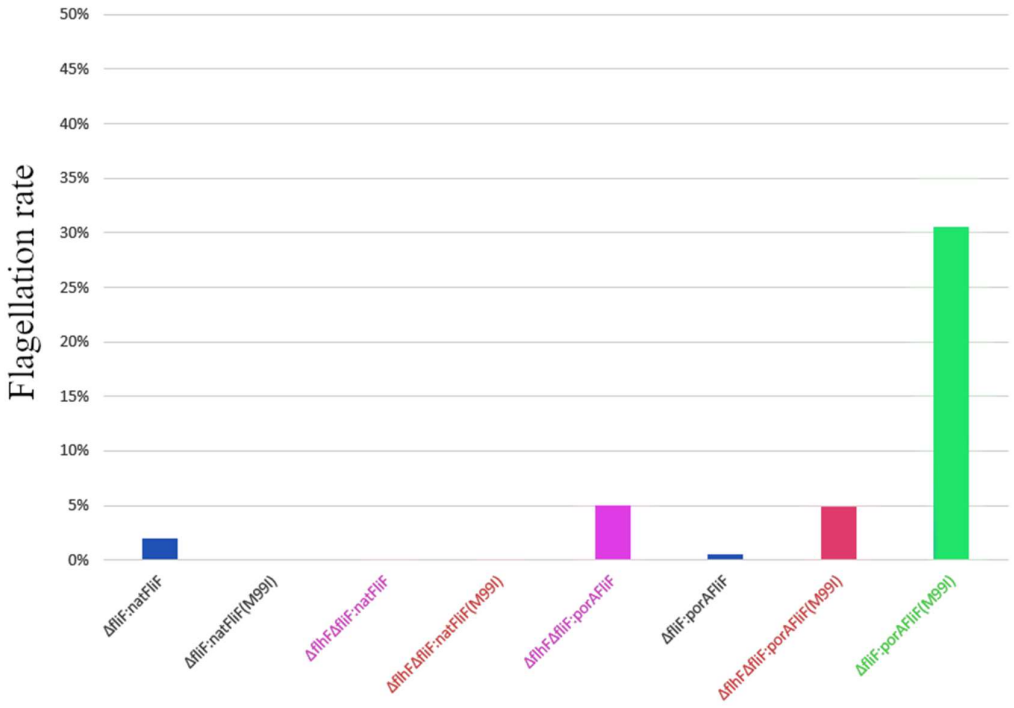



***Figure 5.5*. Flagellation rate of the *C. jejuni* complementation strains.** Colour coding denotes the strains where the sequenced genotype does not fully match the desirable genotype and is different between the compared strains. Red – $mnmA_{I309T}$ and truncated *dcuA* present; purple – only truncated *dcuA* present; green – $mnmA_{I309T}$ and $fliF_{Q9P}$ present. In addition, $fliF_{M99I}$ is present where integration was expected in strains where it is denoted in brackets in the strain name. Flagellation rate was calculated from n≥100 cell ends (in one technical replicate).

For each strain, a small amount of *C. jejuni* cells from late stationary phase was deposited onto an electron microscopy grid after being negatively stained using uranyl acetate. Grids were then analysed using electron microscopy where at least 100 individual cells were counted for each genotype/clone. All the strains examined, but one, had either no cells with flagella or a very small number. It was disappointing to see that complementation of the *fliF* wild-type deletion in the *Cj0046* locus did not work, regardless of the level of FliF expression. There was no clear trend, but strains in which FliF or FliF $_{M99I}$, were overexpressed in the absence of FlhF tended to have some cells with flagella. But numbers are too small to make any confident predictions. Interestingly, in the strain in which we identified an additional mutation in the N-terminal region of $FliF_{M99I}$, and where the protein was overexpressed, we observed that 30% of the cells were flagellated (Figure 5.5). In this strain, the $FliF_{M99I}$ protein is

overexpressed in the presence of FlhF. On the contrary, the respective strain with wild-type FliF overexpressed in the presence of FlhF, and in which we did not identify any spontaneous mutations, we very rarely observed any flagellated cells. A few cryoEM images collected from two strains (*ΔfliF:porAfliF* and *ΔfliF:porAfliF*$_{M99I}$) separately from negative stain EM show that non-flagellated cells not only do not express or assemble flagella, but the entire motor apparatus and the characteristic crater like shape at the cell pole is completely absent (Figure 5.6).

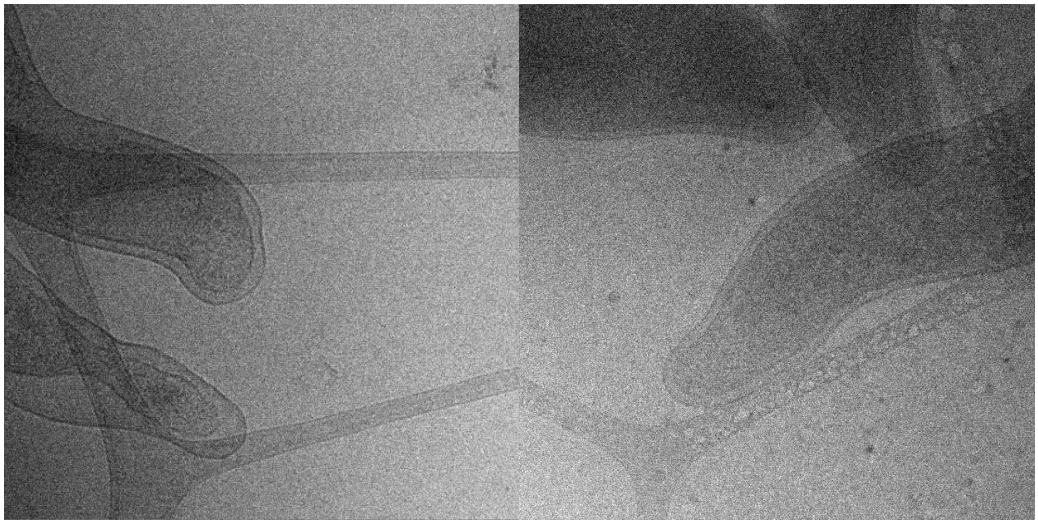


***Figure 5.6*. Cryo-electron microscopy images of *ΔfliF:porAfliF* and *ΔfliF:porAfliF*$_{M99I}$.** No motor insertion is visible at the cell pole and cell ends are completely rounded. Some flagellation is present in *ΔfliF:porAfliF*$_{M99I}$, in cell counting experiments, but the image shown has captured non-flagellated cells in this case.

## 5.5. *In silico* analysis of mutations in FliF structure

Although as described in the Introduction, there is no crystal structure available for *C. jejuni* FliF, *in silico* studies predict that it contains 3 RBM domains, RBM1, RBM2 and RBM3, intersected by the β-collar (Introduction 1.5., Figure 1.8). And even though, there are now new high resolution structures of FliF from other bacterial species, no structure has been predicted for the very N-terminal end of FliF (Johnson et al., 2020; Kawamoto et al., 2021). But recent expansion of the Alphafold database offers structural prediction of the entire *C. jejuni* FliF (although it should be noted that confidence score in many parts of the model is low (Figure 5.7)).

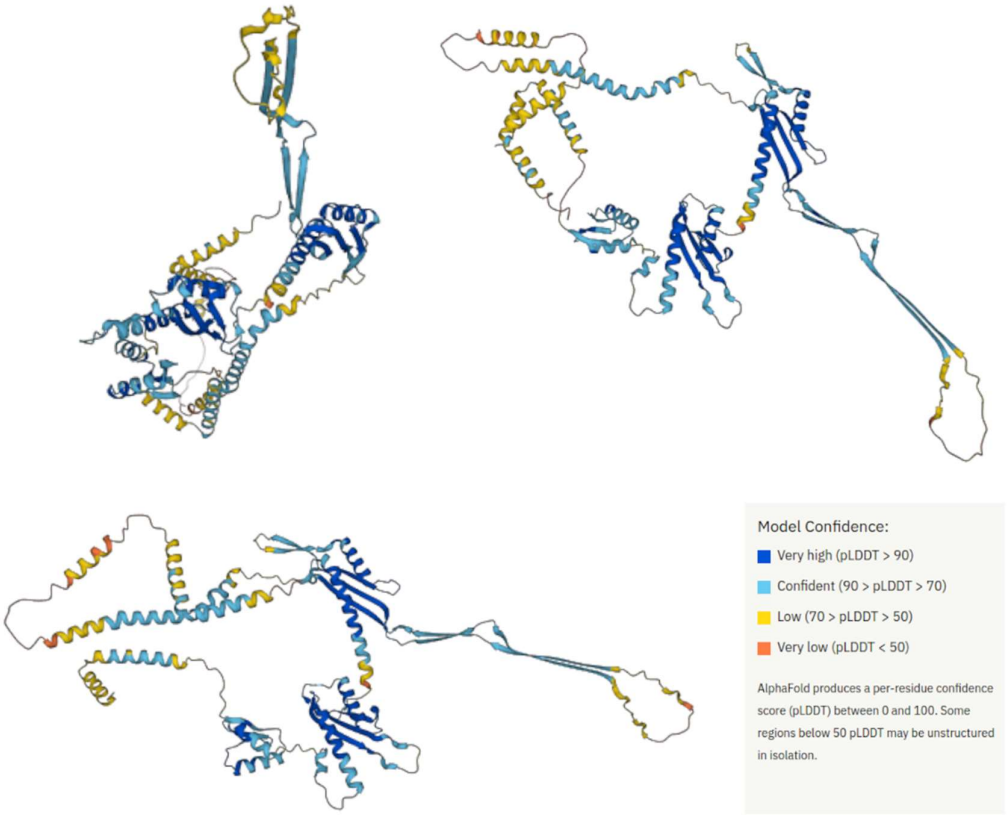



***Figure 5.7. C. jejuni* FliF Alphafold predictions from the database.** Three models were selected from database, that match the sequence used in this and previous FliF revertant study. The models have variable confidence across different areas.

Before the availability of these Alphafold structures, a Robetta server prediction (Kim et al., 2004) of *S. enterica* FliF N-terminus was made for the unresolved part of *S. enterica* FliF (because available FliF ring structures are of *S. enterica,* for easier comparison). Unlike Alphafold, the Robetta server prediction was carried out without the use of any templates and is solely a folding-based prediction from an input sequence (Figure 5.8).

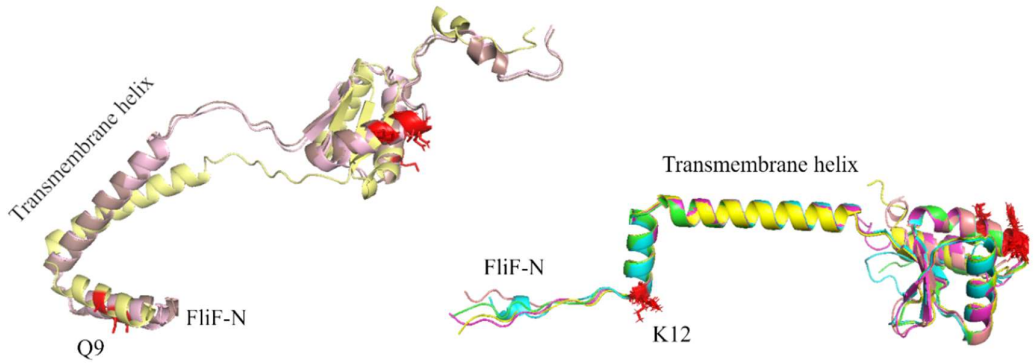



***Figure 5.8*. Predictions of FliF-N and transmembrane helix**. Left – Alphafold structural prediction fragments of first 125 amino acids in *C. jejuni* FliF sequence (3 models overlaid and aligned). Right – Robetta structural prediction of first 125 amino acids in *S. enterica* FliF sequence (5 models overlaid). Q9 in *C. jejuni* and K12 (its equivalent in *S. enterica*) are shown in red, in stick representation. Transmembrane helix is labelled. Residues known to be mutated in pseudorevertants from previous study M99, A102, S103, E104 (in *C. jejuni*), are also shown in red and in stick representation in the RBM1 domain above transmembrane helix. Equivalent residues in *S. enterica* structure are depicted the same way.

There are enough similarities between the between *C. jejuni* and *S. enterica* predictions from two different sources, although only the transmembrane helix and parts of the RBM1 domain have good confidence scores; the very N-terminal part, that would be cytoplasmic, is with low confidence scores. The area with the mutation of interest (Q9P) in both models is alpha helical and in the case of *S. enterica* unstructured right after the equivalent (K12) residue. *S. enterica* models found in Alphafold database also have an unstructured region after a short helix at the very N-terminus of FliF, similar to the one seen on Robetta prediction. In both cases the FliF fragment in the cytosol below the transmembrane domain could potentially interact with the membrane above it, serving as an anchor. How exchange of Q to P would affect this, it is difficult to estimate, but the amino acid in this position would lose its charge, and proline can introduce a kink in the helix, so it could be structurally significant.

Recently available structure of *A. aeolicus* FliF (Takekawa et al., 2021) was used to try and map the *C. jejuni* pseudorevertant mutations that are located in a hotspot region: M99, A102, S103, E104 (according to sequence alignment of FliF, Figure 5.4.). Highlighting these residues in the predicted structure shows that they are all in structured parts of the protein (Figure 5.8, 5.9). In the *A. aeolicus* FliF crystal structure

(Figure 5.9C) and in the *C. jejuni* predictions (Figure 5.9A) these mutations localize in the RBM1 domain, in α-helix.

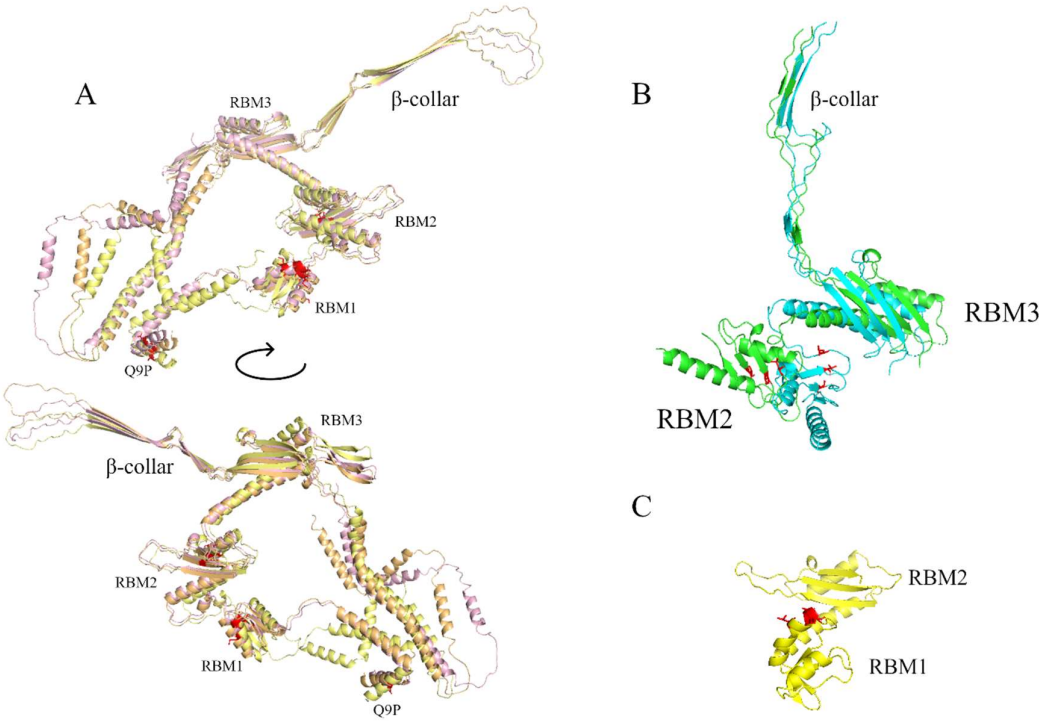



***Figure 5.9*. Overview of *C. jejuni* FliF predictions, known solved FliF structures and locations of pseudorevertant mutations**. A – three overlaid *C. jejuni* FliF predictions, from two opposite views. Novel Q9P mutation is denoted separately. B – two conformations of *S. enterica* 34-mer ring structure. C – crystal structure of RMB1 and RBM2 of *A. aeolicus*. All domains are labelled. In red are shown all previously identified *C. jejuni* pseudorevertant mutations or their equivalent residues in other species structures.

The other area of location of pseudorevertant mutations identified by Stoakes (2017) V177, S214 and P220 was are also mapped and were found to be located in RBM3, either in the β-sheet or adjacent linker. Placement of the MS-ring above the membrane and RMB2 placement shown in (Johnson et al., 2021) likely puts these residues in direct proximity of the inner membrane and likely to interact with T3SS components or the secreted proteins. In addition V177 (equivalent V176 in *S. enterica*) is part of the A-S-V-X-V motif, where A175 and V175 are shown to be crucial for interaction with FlhA (Kihara et al., 2001). Likewise in C. *jejuni* deletion of ΔA175-ΔS176 have the same effect (Boll and Hendrixson, 2013).

## 5.6. Sample creation and cryoET data collection

CryoET reconstruction could yield insight in *C. jejuni* motor appearance and stoichiometry as well as reveal if there are any large structural rearrangements in the $fliF_{M99I}$ carrying pseudorevertant. In order to do that a way to prepare samples had to be established.

Samples for cryoET data collection were optimized empirically testing two types of grids (lacey carbon irregularly shaped hole grids and regular hole shaped carbon grids), two types of gold particles (gold nanoparticles and Immunogold Conjugate protein A), use of wide bore tips to prevent mechanical damage to flagella and including additional washing steps resulting in the final protocol described in 2.3.1 to prevent clustering of cells and particles distribution seen in early samples (Figure 5.10).

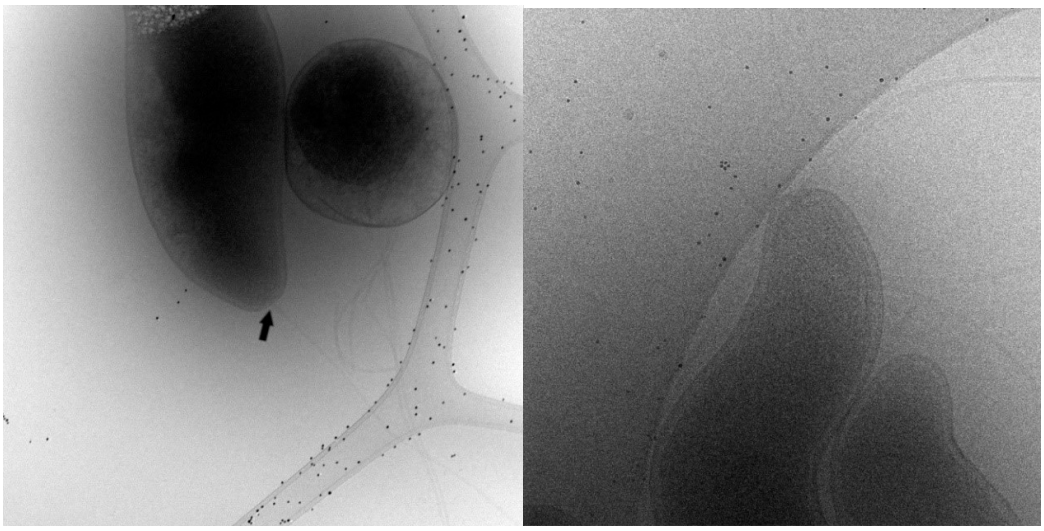


***Figure 5.10*. Examples of issues that had to be addressed during sample optimization**. Left – lacey carbon grid and gold particles sticking to the grid. Also, a coccoid cell with multiple flagella indicating stress to the cells, prompted change in culturing conditions. Arrow points to flagellum and a motor at the pole of the cell. Right – two normal shaped cells in even hole grid and some flagella potentially broken by mechanical force. Gold particles are sticking to the grid and some are clustered. In both images cells are adhering to each other as well.

First samples were prepared using lacey carbon grids. As gold particles tended to stick the grid and not be evenly distributed across the whole surface of the grid, we tested regular (even) hole grid. But even this resulted in the gold particles concentrating on the grid. A decision was therefore made to swap to immune-coated gold and this resulted in significantly improved fiducial distribution. Collecting bacterial cells at an

earlier time point in the growth curve, optimising the cell density and adding additional washing steps after collection to minimise cell clumping allowed us to prepare grids suitable for cryoET analysis. Sets of 8 grids with wild type and 8 grids of *ΔflhF fliF*$_{M99I}$ psueodrevertant were prepared. Data collection initially was done just for wild-type cells which were used to test our methodology and set up the analysis workflow. *ΔflhF fliF*$_{M99I}$ psueodrevertant data was collected in two later, separate sessions.

Collected tilt series images clearly showed that most bacteria have a clearly visible flagellum and ring in at least one of the poles, even before sub-tomogram reconstruction (Figure 5.11). The immuno-coated gold particle amount observed on the grids in the view in each image throughout the tilt series is variable and often below desirable number of 30 beads but still sufficient for sub-tomogram reconstruction in both IMOD and EMAN2. Not all samples have evenly distributed particles and in some there was a tendency for more particles on grid hole edges than in the ice surrounding the bacterium in the grid hole. Data was collected from the final optimized grids only and 44 tilt series were collected from these WT grids. In most tilt series there is only one cell end per view, except in three tilt series there are two cells/motors per imaging area, indicating that additional washing included in the final preparation protocol reduced cell clumping.

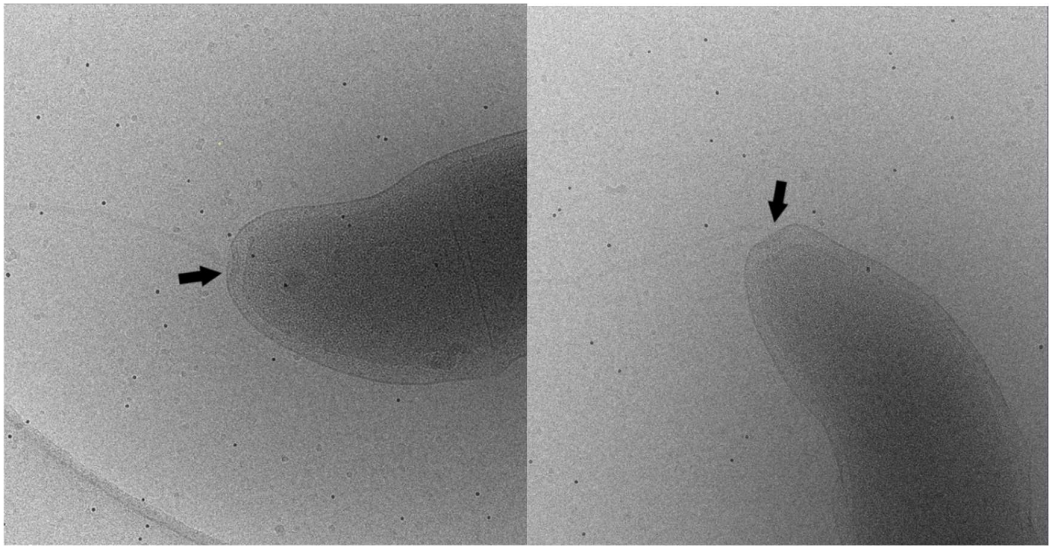


***Figure 5.11*. Examples of final tilt series images.** A single cell is found in most tilt series. A motor and flagellum are clearly visible, gold particles are found in ice and on grid, and are evenly distributed (≤ 30 beads). Arrows point toward a motor embedded in the cell pole and flagella attached to the motors.

It was decided that the images collected for the wild-type sample were of sufficient quality for processing and motor reconstruction. Even though, samples were prepared, processed and images were also collected for the pseudorevertant strains, time limitation meant that focus had to remain on the wild-type sample with the aim to use in the future for comparison to the motor structure of the pseudorevertant.

## 5.7. Analysis of cryoET data

Sub-tomogram reconstruction was done in IMOD (Kremer et al., 1996) with the intention to build the final model in Relion (Bharat and Scheres, 2016). In parallel, a unified approach of sub-tomogram reconstruction and model building was independently attempted in EMAN2 (Galaz-Montoya et al., 2015).

### 5.7.1. Sub-tomogram reconstruction in IMOD

44 tilt series were reconstructed in IMOD one by one. 30 fiducials were selected when available. The maximal number of fiducials was achieved by manually adding additional ones, that had not been identified in the automated search function. Fiducial tracking and alignment were carried out via multiple iterations of finding missing points in tracking or when misassigned to a different bead or feature, until no more gaps (in following each point throughout the tilt series) or large residuals in the tracked bead model could be found and the residual error could not be further reduced.

***Table 5.4*. High quality manual sub-tomogram reconstruction in IMOD statistics.** Many of the sub-tomograms have less than 30 fiducials.

| *Tilt series* | *# of points* | *Ratio of total measured values to all unknowns* | *Residual error mean and sd (nm):* | |
|---|---|---|---|---|
| *series4* | 19 | 4.91 | 0.209 | 0.127 |
| *series6* | 25 | 8.01 | 0.823 | 1.889 |
| *series7* | 14 | 5.34 | 0.566 | 0.429 |
| *series8* | 27 | 8.51 | 0.458 | 0.257 |
| *series9* | 22 | 7.44 | 0.468 | 0.364 |
| *series10* | 30 | 9.09 | 0.336 | 0.206 |
| *series11* | 30 | 9.09 | 0.449 | 0.257 |
| *series12* | 21 | 7.2 | 0.698 | 0.559 |
| *series13* | 30 | 9.09 | 0.476 | 0.287 |
| *series15* | 19 | 6.71 | 0.569 | 0.330 |
| *series16* | 24 | 7.75 | 0.473 | 0.280 |
| *series17* | 22 | 7.44 | 0.386 | 0.226 |

| | | | | |
|---|---|---|---|---|
| *series18* | 30 | 9.09 | 0.542 | 1.239 |
| *series19* | 30 | 9.09 | 0.405 | 0.259 |
| *series20* | 30 | 9.09 | 0.478 | 0.305 |
| *series21* | 18 | 6.46 | 0.461 | 0.293 |
| *series22* | 14 | 5.36 | 0.380 | 0.223 |
| *series23* | 14 | 5.36 | 0.469 | 0.282 |
| *series25* | 24 | 7.89 | 0.541 | 0.343 |
| *series26* | 30 | 9.09 | 0.414 | 0.239 |
| *series27* | 28 | 8.71 | 0.308 | 0.172 |
| *series28* | 16 | 5.79 | 0.354 | 0.211 |
| *series29* | 22 | 7.44 | 0.792 | 0.484 |
| *series31* | 16 | 5.93 | 0.379 | 0.214 |
| *series32* | 23 | 7.67 | 0.462 | 0.270 |
| *series33* | 19 | 6.66 | 0.505 | 0.426 |
| *series34* | 30 | 9.09 | 0.399 | 0.232 |
| *series35* | 17 | 6.2 | 0.435 | 0.395 |
| *series36* | 19 | 6.72 | 0.672 | 0.454 |
| *series37* | 18 | 6.46 | 0.778 | 0.660 |
| *series38* | 18 | 6.46 | 0.338 | 0.212 |
| *series39* | 19 | 6.72 | 0.409 | 0.231 |
| *series40* | 20 | 6.97 | 0.683 | 0.524 |
| *series41* | 30 | 9.09 | 0.472 | 0.271 |
| *series42* | 15 | 5.65 | 0.588 | 0.377 |
| *series43* | 29 | 8.65 | 0.495 | 0.321 |
| *series44* | 30 | 9.09 | 0.347 | 0.200 |
| *series45* | 29 | 8.9 | 0.428 | 0.251 |
| *series46* | 17 | 6.09 | 0.657 | 0.500 |
| *series47* | 21 | 7.21 | 0.309 | 0.193 |
| *series48* | 23 | 7.67 | 0.356 | 0.205 |
| *series49* | 25 | 8.1 | 0.342 | 0.204 |
| *series50* | 27 | 8.5 | 0.405 | 0.310 |
| *series51* | 14 | 5.31 | 1.363 | 1.644 |

### 5.7.2. Preliminary CryoET reconstructions of *C. jejuni* motor with EMAN2

Due to issues with installation of software packages and access to servers, parallel reconstruction of the same dataset was done in EMAN2 completely independently. This software cannot use IMOD created sub-tomograms for input due to format incompatibility. Sub-tomogram reconstruction is fully automated in EMAN2 and no manual improvement of tracing fiducials takes place like in IMOD. EMAN2 creates an initial model that is used to guide the 3D model reconstruction. After that, subtilt refinement takes place to create the final model. Two versions were used, EMAN2 2.31 and 2.91. Firstly, the reconstructions attempted with EMAN 2.31 (older version) are discussed. Results from EMAN 2.91 follow subsequently.

In case of a large dataset running reconstruction without symmetry (C1) can reveal the underlying symmetry, however this time it was not possible to get a result with C1 symmetry enforced, therefore different symmetries had to be pre-defined to obtain a result. As previous work in *S. enterica* suggested that motor ring had C34 symmetry with sub-populations of C33 and C35 (Thomas et al., 2006), the first attempts at building the model used C33 and C34 symmetries. Unfortunately for these a wrong pixel size input value (0.3462 instead of 3.462) was used and therefore had to be rescaled in Chimera by changing pixel size there. These models can therefore only be used for preliminary comparison only, because they cannot be used for building models in them. Several C39 reconstructions were attempted. C39 symmetry was selected based on a simple calculation, that the larger diameter of *C. jejuni* (52 nm), determined from low resolution cryoET, would require more subunits than *S. enterica* (44 nm) (Beeby et al., 2016). Currently there is no biological data that can definitively establish the *C. jejuni* motor stoichiometry.

C39 reconstructions obtained with EMAN 2.31 should be regarded as very preliminary and are potentially poorly reproducible as each run gives a different result (Figure 5.12). Reconstruction runs that did run until the end suffered from too small a box size that results in an edge cutting off part of the rings (Figure 5.12). Box size would need be optimized by trial and error and starting point was a smaller box due to computational memory limitations at the time. 3D reconstructions of two independent runs with the same symmetry appear vastly different (Figure 5.12) and it is not clear why. In addition, two earlier runs with C33 and C34 symmetry are shown for comparison. They do however show that C33 and C34 symmetry fits the dataset worse as the reconstructions come out less detailed, therefore C39 is closer to the true physiological stoichiometry than C33 or C34. The initial model generation step that comes before the 3D reconstruction shown in Figure 5.12 is obtained, is made based on random references in the EMAN2 2.31 reconstruction approach (used in results shown in Figure 5.12) and then used to search all particles, potentially leading to different outcomes.

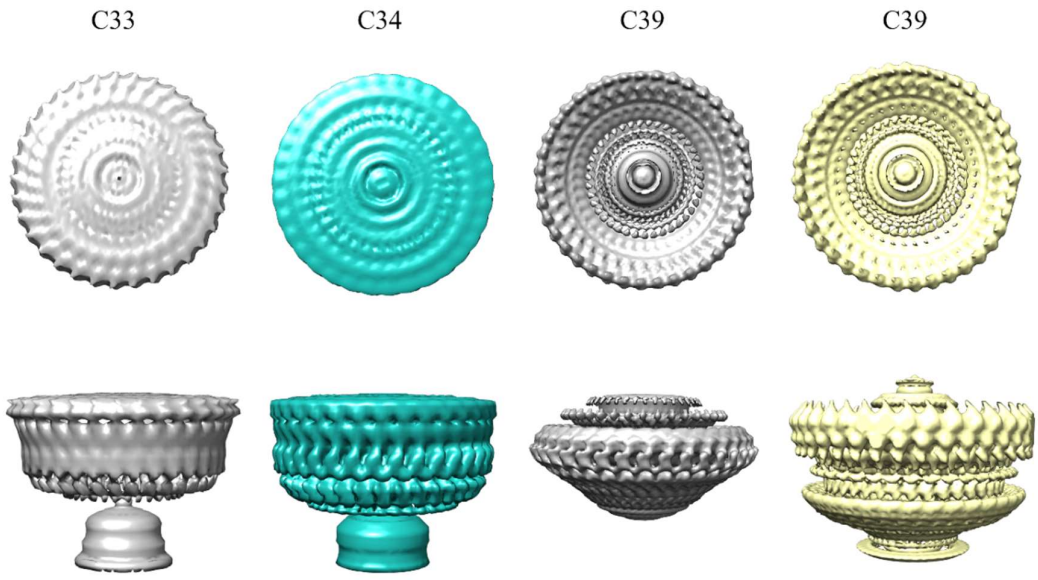



***Figure 5.12*. 3D reconstructions of WT *C. jejuni* motor using EMAN2 from four independent runs.** Top view and side view of respective models, processed with different pre-defined stoichiometries: C33, C34, C39 (two different runs).

Only one sub-tilt refinement of a 3D reconstruction was achieved with the EMAN2 2.31 version using the C39 symmetry model (Figure 5.13). It greatly elongates the model and shows multiple unidentified structures. The final sub-tilt refinement likewise suffers from too small a box size.

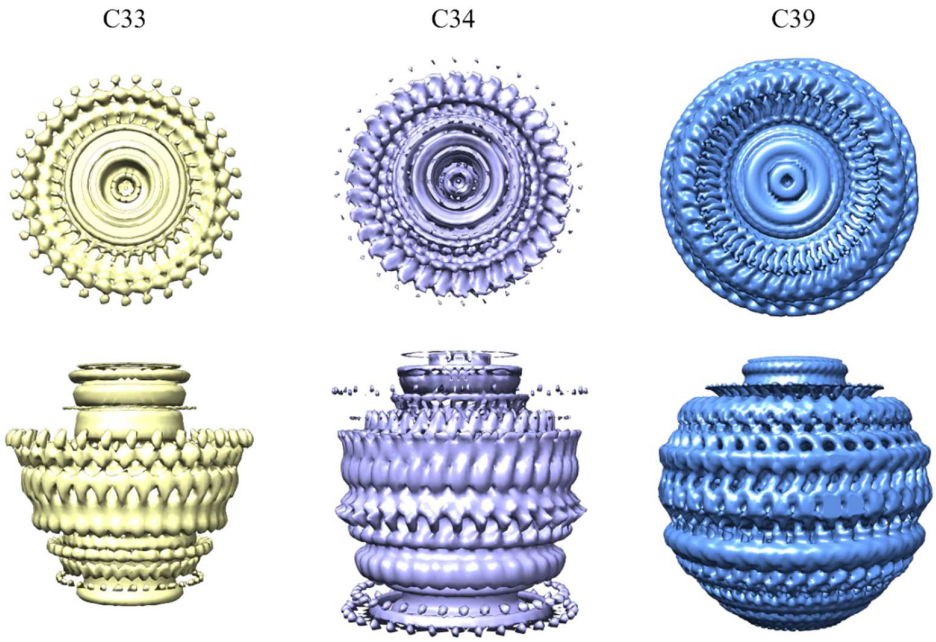



***Figure 5.13.*** **Sub-tilt refinement of a C33, C34 and C39 EMAN2 run.** Top view (left), side view (right).

A lack of resolution at the centre of the top view of the C39 reconstruction indicates that structures there have a stoichiometry very different from the C39 imposed. For comparison, C33 and C34 runs are added (Figure 5.13) that are very different. An exhaustive process employing multiple runs with different stoichiometries would be needed to employ this approach and find a stoichiometry best matching each area. It is impossible to assign proteins to the many rings seen. In addition, visual comparison depends a lot on the representation settings selected in Chimera.

Additional processing attempts were later attempted in EMAN2 2.91. Sub-tomogram reconstruction was started from scratch and unlike in version 2.31, where in each run a few sub-tomogram reconstructions would fail to due to memory limitations, all tilt series were successfully processed in version 2.91, so no unnecessary data loss occurred (Table 5.5).

***Table 5.5.*** **Sub-tomogram reconstruction statistics with EMAN2 2.91.**

| *Tilt series* | *Number of particles* | *loss* | *defocus* |
|---|---|---|---|
| *series6* | 1 | 0.5 | 1.3 |
| *series7* | 1 | 1.18 | 1.4 |
| *series8* | 1 | 0.51 | 1.4 |
| *series9* | 1 | 0.49 | 1.3 |

| | | | |
|---|---|---|---|
| *series10* | 1 | 0.49 | 1.4 |
| *series11* | 1 | 0.44 | 1.3 |
| *series12* | 2 | 0.64 | 1.4 |
| *series13* | 1 | 0.48 | 1.3 |
| *series15* | 1 | 0.69 | 1.4 |
| *series16* | 1 | 0.39 | 1.3 |
| *series17* | 1 | 0.73 | 1.2 |
| *series18* | 1 | 0.46 | 1.4 |
| *series19* | 1 | 0.55 | 1.2 |
| *series20* | 1 | 0.39 | 1.2 |
| *series21* | 1 | 0.91 | 1.2 |
| *series22* | 1 | 0.8 | 1.2 |
| *series23* | 2 | 0.92 | 1.3 |
| *series25* | 1 | 0.52 | 1.5 |
| *series26* | 1 | 0.5 | 1.1 |
| *series27* | 1 | 0.76 | 1.3 |
| *series28* | 1 | 0.87 | 1.3 |
| *series29* | 1 | 0.72 | 1.3 |
| *series31* | 1 | 0.99 | 1.6 |
| *series32* | 1 | 0.72 | 1.7 |
| *series33* | 1 | 0.81 | 1.4 |
| *series34* | 1 | 0.42 | 1.8 |
| *series35* | 1 | 0.84 | 1.4 |
| *series36* | 1 | 0.57 | 1.3 |
| *series37* | 1 | 0.86 | 1.4 |
| *series38* | 1 | 0.88 | 1.3 |
| *series39* | 1 | 0.58 | 1.3 |
| *series40* | 2 | 0.57 | 1.4 |
| *series41* | 1 | 0.55 | 1.3 |
| *series42* | 1 | 0.9 | 1.2 |
| *series43* | 1 | 0.41 | 1.3 |
| *series44* | 1 | 0.38 | 1.4 |
| *series45* | 1 | 0.56 | 1.4 |
| *series46* | 1 | 0.64 | 1.4 |
| *series47* | 1 | 0.62 | 1.5 |
| *series48* | 1 | 0.39 | 1.4 |
| *series49* | 1 | 0.42 | 1.2 |
| *series50* | 1 | 0.49 | 1.4 |
| *series51* | 1 | 0.75 | 1.2 |

Version 2.91 of EMAN2 incorporates new approaches for initial model reconstruction and 3D and subtilt-refinement are joined in a single step, but previous options are also available and upgraded since version 2.31. While the new approach did

not yield good results, running the older two step setup type (with separate 3D refinement followed by sub-tilt refinement) but in version 2.91 gave the best results. After varying input settings, several runs on the initial model appeared similar and therefore much more reproducible.

In addition, C39, C42 and C37 symmetries were applied, to explore the probable range of stoichiometries between known smaller (minimum 33)(Thomas et al., 2006) and larger (maximum 46)(Chang et al., 2020) diameters and stoichiometries currently known in other species. By careful examination of the structures obtained at various density representation cut-offs in Chimera and comparison with cryoET images from earlier studies, putative structures could be assigned to the inner architectures (Figure 5.14), such as the characteristic conical *C. jejuni* basal disk and the C-ring (Beeby et al., 2016). The rings below the basal disk likely correspond to the medial and proximal disks (Beeby et al., 2016), though they cannot be identified with absolute certainty. These disks are found in *C. jejuni* and are among additional structures that, unlike the C-ring and MS-ring, seem to have diversified during evolution between different species and are not found in all motors (Beeby et al., 2016; Chaban et al., 2018). In this case, these additional structures serve as landmarks to identify parts of the motor.

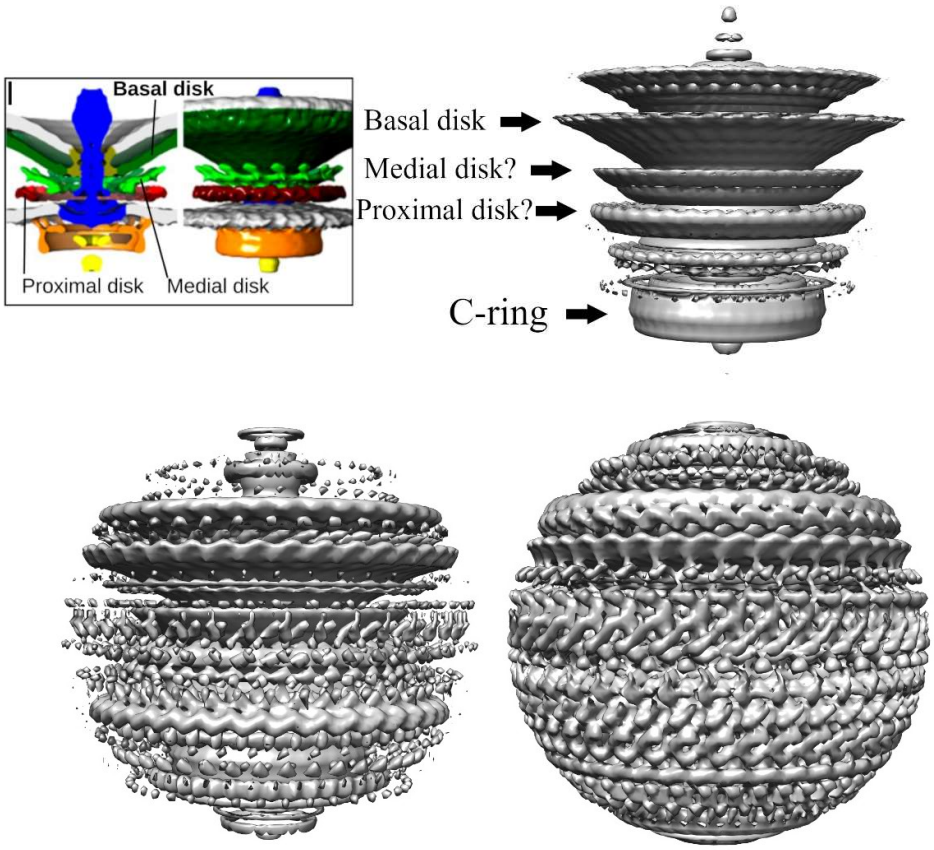



***Figure 5.14*. Different density representation cut-offs of 3D reconstruction of WT *C. jejuni* motor with C42 symmetry.** At maximal representation the reconstruction is limited by the spherical mask used (bottom right image). Deeper behind the outer circular structures, the motor structure is visible – in which the C-ring and basal disk, and possibly median and proximal disks – can be identified in comparison with previous low resolution data from *C. jejuni* (Figure 1I from (Beeby et al., 2016) shown upper left for comparison). C42 is presented as representative example.

The outer structures present in the density cannot at present be identified; they could be scaffold protein structures, or potentially even noise. Unlike the results from EMAN 2 version 2.31, version 2.91 produced similar results irrespective of the different stoichiometries (Figure 5.15).

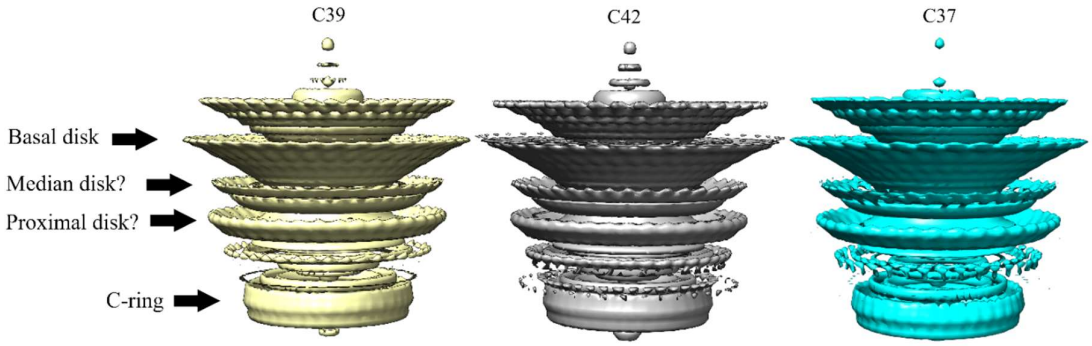



***Figure 5.15*. Comparison of *C. jejuni* motor solutions in C37, C39 and C42.** Identifiable and likely ring structures are denoted with arrows. Structures are similar between the two independent runs.

The diameter of what is highly likely to be the C-ring is 49.5 nm, which is smaller than the previously reported 52 nm derived from a low resolution cryoET study (Beeby et al., 2016) and is actually closer to that of *V. alignolyticus* (Carroll et al., 2020). It is noteworthy that the spherical mask applied seems to actually limit the density and could be changed in future runs for potential further improvement (Figure 5.14). There is no discernible symmetry at the C-ring area and imposed symmetries seem to rather fit the basal ring and other structures than the C-ring. Encouragingly, reconstructions of C39 and C42 with the same parameters seem very similar and show the same C-ring size (Figure 5.15).

## 5.8. Discussion

*C. jejuni* is an amphitrichous organism with a single large flagellar motors placed at each of the two ends of the cell (Muller et al., 2014) making it an attractive model organism for research into the flagella macromolecular assembly, structure and placement. On the other hand, unlike research in *E. coli* or *S. enterica* flagella, *C. jejuni* research is limited by the lack of genetic tools available for genetic manipulation and expression of heterologous proteins. A previous study of *C. jejuni* motility pseudorevertants opened intriguing questions about how mutations in motor components restore motility due to loss of a regulator protein FlhF (Stoakes, 2017).

FlhF is a protein with a signal recognition particle (SRP) domain and GTPase activity (Balaban et al., 2009) whose precise role and interactions with the structural

flagellar proteins (if any) remain unknown (Stoakes, 2017). The project focused on two different lines of investigation.

Firstly, we aimed to acquire confirmation that the FliF protein versions that spontaneously arose in the absence of FlhF were sufficient to restore motility. Stoakes (2017) observed that even though *C.jejuni* pseudorevertants were building flagella, the assembly was at a much slower pace than in the wild type cells. We therefore also aimed to test, if FliF availability was the limiting factor and whether the pseudorevertant FliF still had the ability to interact with FlhF and consequently conferring motility closer to wild-type levels. A microbiology approach was undertaken of creating a set of complementation strains expressing WT or mutant $FliF_{M99I}$ versions from an external locus at different levels, with or without the presence of FlhF in the cells.

Secondly, cryoET data was collected to further refine the available structures, and to elucidate the stoichiometry in wild-type *C. jejuni* C-ring and MS-ring components. Stoakes (2017) had speculated that the $FliF_{M99I}$ mutation might induce structural changes to the MS-ring, changing its interaction with the C-ring and other flagellar export components. Acquiring the wild-type structure would enable us to compare it to the pseudorevertant structures.

### 5.8.1. Complementation strain creation and phenotypic characterization in *C. jejuni*

Creation of complementation vectors and strains was verified by Sanger sequencing at each stage before proceeding with whole genome sequencing. QC analysis and high coverage obtained in all samples enabled calling of SNPs and INDELs with confidence (Table 5.1). Analysis of the whole genome data, confirmed the Sanger sequencing data, that the deletion of *fliF* from within its operon and insertions of the various *fliF* constructs in the *Cj0046* pseudogene were all successful. Unfortunately, a number of different mutations, either distal to the target sites on the chromosome, or close to the recombination regions were identified, arising at different times in the lineage (Figure 5.16).

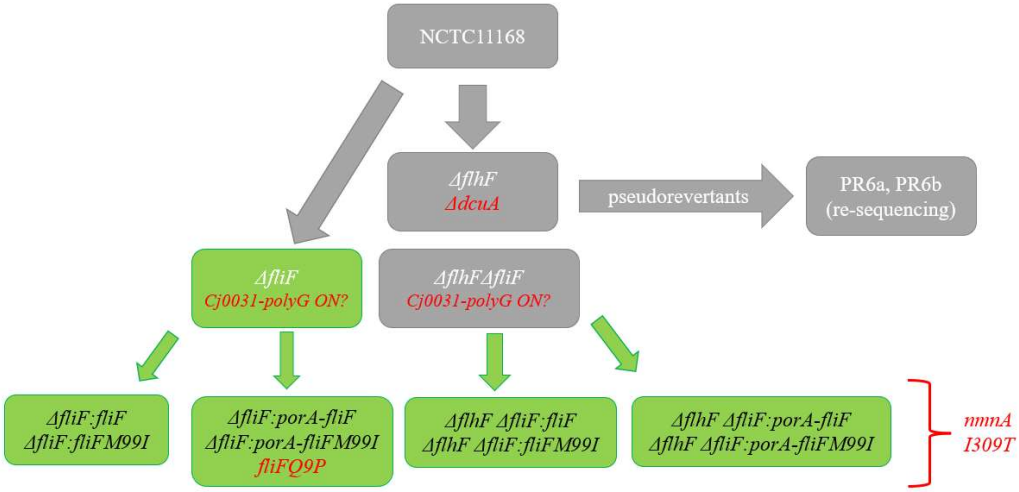



***Figure 5.16*. Timeline of *C. jejuni* mutations discovered.** In red novel mutations are marked, showing at what stage they would have likely appeared. *mnmA*$_{I309T}$ mutation is present in some, but not all clones of the complementation strains.

All the strains sequenced contain the *Cj0031* phase-variable gene which has likely been switched to the ON (activated) position (Anjum et al., 2016) due to a single nucleotide deletion in the homopolymer region (Figure 5.16). It would potentially affect the methylome of *C. jejuni* as the gene encodes a protein with putative endonuclease and methyltransferase activities; even though this is impossible to evaluate at present time, previous studies have suggested that the *Cj0031* regulon does not involve motility associated genes (Anjum et al., 2016). Re-analysis of this data (Stoakes, 2017) could yield additional insight on how often activation of this phase variable gene occurs in this wild-type *C. jejuni* strain and the associated mutant and pseudorevertant strains of that study.

All *ΔflhFΔfliF* strains have inherited a severe truncation of the *dcuA* gene, a C4-dixarboxylate transporter, that must have arisen due to a point mutation in *ΔflhF* creation or passaging even before creation of *ΔflhFΔfliF* (Figure 5.16). Known data however indicates that it is at least partially redundant in *C. jejuni* and loss of DcuA does not compromise growth (Wösten et al., 2017) which is consistent with the normal growth observed during my work with *ΔflhFΔfliF* and complementation strains derived from it. Going back to the Stoakes (2017) sequencing data, revealed that the *dcuA* truncation is present in the *ΔflhF* strain from which the pseudorevertants and all subsequent lineages were derived (Constantinidou, personal communication).

An extensive range of mutations in *Cj0046* upstream of the insertion site and in *Cj0053c* downstream of the insertion site shows that the homologous recombination process had more side effects than anticipated.

Additionally, in 9 out of 19 sequenced strains, non-synonymous mutation I309T was observed in the *mnmA* (*Cj0053c*) gene, downstream of the *Cj0046* pseudogene, where sequenced for the complementation were targeted (Figure 5.16). While a multitude of synonymous mutations in *mnmA* speaks to likely side effects of the integration (Table 5.3), the I309T mutation itself is found in other deposited *C. jejuni* sequences in NCBI, so likely it does not have a severe phenotype effect and may have originated stochastically. The synonymous and non-synonymous mutations in *mnmA* have arisen in complementation strains derived from both *ΔfliF* and *ΔflhFΔfliF* strains, while not being found in the parent strains themselves.

Last, but not least, a novel mutation Q9P was identified in FliF alongside the original pseudorevertant M99I mutation; the Q9P mutation only occurred when $FliF_{M99I}$ was expressed from the *porA* strong promoter, in *ΔfliF* genetic background. It is unlikely that this would be created by an integration defect, as the *fliF* part of pC46 constructs does not engage in homologous recombination itself and therefore it is likely an effect of the environmental pressure (stress and lack of flagella). Although interestingly, this mutation did not arise in the *ΔflhFΔfliF:porAfliF(M99I)* constructs, that used the same pC46-*porAfliF*$_{M99I}$ construct for recombination, but this time in the *ΔflhFΔfliF* genetic background. Identifying such SNPs highlight the importance of obtaining full sequencing data of the entire genome when creating a strain. Sanger sequencing of the area targeted with inserts or deletions will not reveal any distal SNPs arising due to the stress imposed on the organism during the genetic manipulation process or during subsequent subculturing. Whole genome sequencing is an invaluable tool to accompany any phenotypic characterization (Figure 5.16).

Complete absence of flagella in almost all complementation strains was initially an unexpected observation. Whole genome sequencing of the *fliF* locus showed that the promoter region upstream of *fliF* driving expression of the operon was not disturbed and neither was the ribosomal binding site upstream of *fliG* (the gene immediately downstream of *fliF*). But it strongly suggests that FliF function cannot be complemented with expression from a distal locus. Previous studies have shown that FliF is likely to co-fold with FliG for maximum stability (Levenson et al., 2012; Lynch et al., 2017).

Genes expressing both proteins belong in the same operon and are co-expressed, but our data suggests that they are probably co-translated and therefore co-targeted to the cell's poles for effective MS-, C-ring assembly and flagella production in *C. jejuni*. This inability to FliF and FliG to co-fold and efficiently assemble when expressed and translated distally from each other, is supported by the cryoEM images of *ΔfliF:porAfliF* and *ΔfliF:porAfliF*$_{M99I}$ (Figure 5.6) that were prepared alongside WT grid testing samples for the optimisation of the cryoET data collection (Section 5.5) that show that cells are not only aflagellate, but do not have motors inserted in the cell poles and the typical "crater-like" membrane indentation in the cell pole as a result of motor presence, as typically seen in WT cells (Figure 5.8). Because these grids were not intended for flagella counting, but qualitative observation only, cells used to prepare them were late-stationary phase, where we expect the majority of the cells to be bipolar with fully formed flagella (Stoakes, 2017).

Although randomly selected cryoEM images of *ΔfliF:porAfliF*$_{M99I}$ did not have flagella (Figure 5.6), in the negative stain EM cell counting experiments, this was the only strain with detectable motility restoration (to 30%) (Figure 5.5). This was also the only strain carrying the novel additional mutation in FliF, Q9P, in an area that is possibly the secretion signal (Bergeron, 2016), but there is no experimental proof. If that is the case, a mutation in this partially conserved residue could have an effect upon FliF secretion efficiency. The proline (P) residue, however, is found in other species in this position, and therefore further studies on this variant would be needed to make a definitive conclusion (Figure 5.4). Indeed, not all proteins use canonical secretion signals.

SRP pathway is a system that consists of Ffh protein and a small RNA complex, that targets transmembrane proteins during their translation, to protect their transmembrane domains from misfolding in the cytoplasm and by binding FtsY, guides the unfolded protein to SecYEG channel that is a part of general secretion system (Green and Mecsas, 2016). Coincidentally FlhF homodimer also contains a putative SRP function, as it shares a structural and sequence similarity the signal sequence binding protein Ffh bound to the SRP-receptor FtsY (Bange et al., 2007). Therefore, it is possible that Q9P mutation may have an effect on secretion efficiency of FliF indirectly through FlhF. A tentative speculation could be drawn that this mutation increases the secretion efficiency and together with an overexpression promoter porA, in the presence of

functional FlhF, can partially overcome the lack of motor assembly induced by lack of FliF-FliG co-folding. The flagellation rate is comparable to that of PR2 (containing the $FliF_{M99I}$ mutation) in the late stationary phase from an earlier study (Stoakes, 2017), therefore this Q9P containing strain could even be referred to as a novel pseudorevertant.

While our evidence that *C. jejuni* FlhF interacts with FliF directly is circumstantial, in a recent study Kojima et al (2021), demonstrated using FliF N-terminal deletions in *V. alginolyticus* that localization and assembly facilitated by FlhF is dependent on the presence of first 30 amino acids in FliF (Kojima et al., 2021). Since the Q9P mutation only arose in the presence of FlhF, which in fact is a protein with a putative SRP domain (Balaban et al., 2009; Bange et al., 2007), raises the possibility that this adaptation facilitates the interaction of the N-terminal FliF region with FlhF resulting in a more efficient targeting and translation of the FliF distally expressed and translated from FliG. This adaptation does not seem to be sufficient when lower levels of FliF are produced under the fliF native promoter.

The discovery of genetic changes during the creation of our *C. jejuni* strains demonstrates how difficult it is to maintain constant genetic background for comparable studies in this organism, especially for phenotypic assays, where multiple passaging rounds of a strain are necessary and where any genetic manipulation may induce unintended side effects. Appearance of a novel pseudorevertant mutation shows the high importance of motility in *C. jejuni* and high adaptability to restore it. Given its presence in the FliF part that has been associated with FlhF (Kojima et al., 2021), it may bring the elusive relationship of FlhF and *C. jejuni* motor proteins (Stoakes, 2017) one step closer to resolution.

### 5.8.2. CryoET reconstruction of *C. jejuni* WT motor

For a long time, the *S. enterica* EM map (Thomas et al., 2006) remained the only available input for C-ring modelling, until the recent structural resolution of *V. alginolyticus* (Carroll et al., 2020) and *B. burgdorferi* rings (Chang et al., 2020). Full *B. burgdorferi* ring is not deposited in publicly available database. The *V. alginolyticus* structures were the best available option and they were used in the C-ring modelling presented in Chapter 4; however, the question about stoichiometry differences of the MS- and C-rings between species and even within the species remains open.

Low resolution cryoET results show that the *C. jejuni* C-ring diameter is estimated to be 52 nm in diameter, while for *S. enterica* it is 44nm and for *Vibrio fischeri* 46 nm (Henderson et al., 2020). Thomas et al (2006) reported populations of different stoichiometries in the *S. enterica* C-ring with 33mer, 34mer and 35mer C-rings having maximum diameters of 46.5 nm, 47.8nm and 49.0 nm (Thomas et al., 2006). In the *V. alginolyticus* the 34mer C-ring in the CCW conformation is 46.2 nm at top, 46.6 nm at the middle, and 49 nm at the bottom, while the CW conformation is 49 nm at the top, 46.6 nm at the middle and 48.3 nm at the bottom; this demonstrates that the conformational state of the structure has a key influence on the diameter even throughout the ring (Carroll et al., 2020). The largest known stoichiometry to date of 46 subunits is found in the 62 nm wide *B. burgdorferi* C-ring (Chang et al., 2020). There might be a naturally varying stoichiometry in C-rings within the cell, but the distribution may be different from rings overexpressed and purified separately, so it is difficult to estimate.

*C. jejuni* C-ring stoichiometry is unknown, but during this project the first reconstructions carried out in EMAN2 were made with the known 33 and 34 stoichiometry from *S. enterica;* these were the only stoichiometries available at the time when the first reconstructions were run. It soon became clear that an approximation or an educated guess would have to be made for *C. jejuni* to accommodate the fact that it has a larger diameter. Simple proportional calculations based on the ring size of *S. enterica* and *C. jejuni*, number of known monomers in the *S. enterica* structure and assuming that the *C. jejuni* and *S. enterica* folded monomers were of the same size suggested that stoichiometries of 39 or 40 seemed like a reasonable starting point; therefore, further EMAN2 reconstruction attempts were made with C39 assumption. That did prove to deliver more detailed solutions (Figures 5.12, 5.13). However, later runs where recognizable *C. jejuni* structures were identified, with C39 and C42 stoichiometry, showed that it has an impact on the level of overall detail, but not on the now-identified C-ring part (Figures 5.14, 5.15). After the emergence of *V. alginolyticus* and *B. burgdorferi* structures, it is worth noting that the relationship between the number of subunits and the diameter is likely not entirely linear, so the actual C-ring stoichiometry in *C. jejuni* could be anywhere between 34 and 42 and multiple runs would be needed to obtain a more accurate model. Furthermore, not all parts of the motor share the C-ring stoichiometry, while MS-ring would be expected to match,

structures like stators associated with the motor or the periplasmic cage/scaffold structures can have an entirely different stoichiometry (Beeby et al., 2016; Qin et al., 2017), but only one symmetry can be enforced per run.

There could be also some particles with naturally varying stoichiometry in the dataset. Utilization of IMOD allows the reconstruction of high-quality sub-tomograms that could be used for model building in Relion. Features available in this software may allow to solve this and other issues issue by sorting motor particles into classes and utilizing the strength of Relion which is building the final reconstruction from all 2D images, not averaging of sub-tomograms.

*C. jejuni* like *H. pylori* is known to have additional scaffold structures surrounding the motor (Beeby et al., 2016; Qin et al., 2017; Zhou and Roujeinikova, 2021). What seems like the previously reported *C. jejuni* basal disk is seen in the best reconstructions as well as a part that matches the expected location of C-ring (Henderson et al., 2020). Its diameter is smaller than reported for *C. jejuni* before (Henderson et al., 2020). A lot of optimization options in Relion and EMAN2 reconstruction approaches remain, however, the preliminary reconstructions made in this study do show that the data collected is of sufficient quality to achieve a 3D reconstruction. Whether it will be successful with more powerful software (Relion) or by verifying stoichiometry with other experiments and then using that information in future to impose a stoichiometry on the EMAN2 reconstruction remains to be seen.

During this project we have also collected a significantly larger pseudorevertant $fliF_{M99I}$ dataset (44 for wild-type vs hundreds of tilt series for the pseudorevertant). That data is still to be analysed, but it should yield an even better outcome. It was not possible to engage with the pseudorevertant dataset reconstruction due to time constraints.

Further work with the cryoET dataset collected in this study should help with the identification of a more accurate stoichiometry for the *C. jejuni* structure, as well as creating a good reconstruction that could be used to model *C. jejuni* FliF and FliG rings with the methodology used in Chapter 4, to better understand the interactions between these two proteins. Complete ring reconstructions can also be used to explore the effects of the pseudorevertant mutations identified in FliF (Figure 5.9) and FliG (T33A) (Introduction 1.14, Table 1.1.) may finally be explained.

# Chapter 6 - Final discussion and Future work

## 6.1. Insights into *H. pylori* and *C. jejuni* flagellar motors

The bacterial flagellar motor is an intricate nano-machine, many finer aspects of which remain unknown; therefore, it continuously remains a fascinating study subject. From the differences in flagella numbers and placement between bacterial species, to stoichiometry differences in the composition of the flagellar rings of different species, to structures of individual proteins and conformational changes occurring when changing rotation direction – questions remain to be further investigated. In this study a combination of computational, structural, and molecular biology methods is used to address some of the missing pieces in the depiction of bacterial rotor we have today (Figure 6.1.).

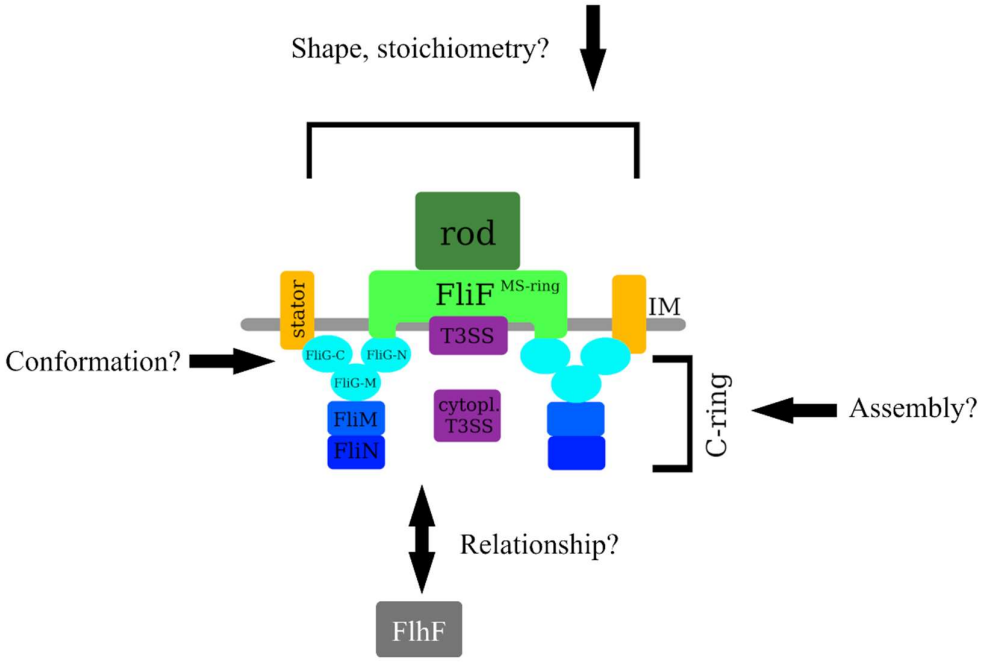



***Figure 6.1.*** **Summary of issues addressed in this study.** MD and integrative modelling were used to study conformation and ring assembly of FliG; genetic manipulation, sequencing, and EM were used to study the relationship between motor components and FlhF; and cryoET data was collected to study the stoichiometry and shape of the rotor.

Recent advances in cryo-electron microscopy are allowing an unprecedented insight into the structure of the flagellar motor that were previously unresolved, for example FliE and FliF (Johnson et al., 2021, 2020; Kawamoto et al., 2021). However, the structure of FliG in a ring assembly is currently unavailable. In this study, a combination of simulations, integrative/homology modelling, MD flexible fitting and cryoET densities was used to make the journey from MD studies of a single FliG domain in *H. pylori* to the prediction of the FliG/FliF-C/FliM-M ring arrangement.

The modelling process also illuminated the possible influence of stoichiometry composition differences between species (Carroll et al., 2020; Chang et al., 2020) that currently cannot be addressed due to a limited available number of cryoET maps of C-rings. CryoET datasets of the *C. jejuni* motor have been collected as part of this study, to attempt to elucidate the motor component stoichiometry in these species. The work in this study lays the groundwork to continue this line of investigation in future.

Motility is extremely important for bacteria, especially pathogens, and therefore targeted disruption of motility is often restored upon occurrence of pseudorevertants, in which mutations in another protein arise to compensate or bypasses the disruption. In *C. jejuni,* when motility is abolished through the deletion of the key motility gene *flhF*, a set of pseudorevertants with mutations in rotor components FliF, FliG and FliH have been found to arise (Stoakes, 2017). Collection of one of the pseudorevertants, *C. jejuni ΔflhF fliF$_{M99I}$* cryoET data alongside the WT was aimed at identifying any possible macroscopic conformational changes. Although, samples were prepared and cryoET images were collected, this goal remained out of reach due to complexity of the reconstruction process and time constraints.

A set of complementation strains were generated with the aim to confirm the compensatory effect of the M99I mutation in FliF in restoring motility to the *C. jejuni ΔflhF* strain and further explore how this mutation affects flagellar assembly and motility. The complementation strains created, demonstrated the dependency of FliG and FliF on being co-transcribed and co-translated and the proteins being spatially co-located for efficient motor assembly. A complementation strain expressing FliF$_{M99I}$, in this *ΔflhFΔfliF* background unexpectedly showed that the relationship of FliG and FliF also identified a new mutation in FliF Q9P that likely contributes to ring assembly efficiency increase alongside M99I again testifying to ability of this species to restore motility under pressure.

## 6.2. Improvement and validation of FliG/FliF-C/FliM-M ring models in *H. pylori*

Although the FliG/FliF-C/FliM-M ring structure models obtained in this study are a good starting point for further investigation in a number of directions, including those outlined below, there are improvements that could be explored.

Firstly, more repeats or alternative strategies for MDFF could be run, to sample a larger conformational space of models that fit into the cryoET density. For example, cascade MDFF is a promising refinement to standard MDFF that can help to overcome energy barriers and improve the fitting process (Dodd, 2017). More starting positions should be explored in MDFF setup as well, as discussed in 4.5., to explore effect on FliG C-terminus packing.

Secondly, it would be desirable to perform similar approaches using another cryoET density data set that has been solved (of *B. burgdorferi*) (Chang et al., 2020), but not as yet deposited in full in the EMDatabse (and was not been made available by the authors at the time of writing).

Thirdly, these ring models are based solely on the best fit in the map while retaining structural integrity. However, they differ in terms of the FliG-N/FliG-N interface prediction made with ClusPro (Comeau et al., 2004) server, using FliG-N/FliG-M/FliF-C model as input, that is not presented in the thesis (part of ongoing work). Output dimers or tetramers obtained were not reconcilable with the cryoET map, however they presented viable interaction interfaces. Perhaps applying MDFF to an alternative, interface focused trimer would yield an improved result (or alternative functional conformation).

In fact, initial coarse grained (CG) simulation runs based on the Martini force field of a FliG/FliF-C tetramer derived from the ClusPro prediction indicated possible alternative modes of packing between FliG-M and FliG-C domains and two different setups were used, representing both crystallized conformations of C-terminus (Lam et al., 2012). This line of enquiry was not completed due to technical difficulties in obtaining, exporting and analysing a matrix of all interacting residues efficiently; however, the information generated could be used to finetune future ring models.

In addition, the constructed ring models should also be compared with known information on interactions between FliG-N, FliG-M and FliG-C domains, mainly available from co-evolution and crosslinking experiments. They may show which of the previously known data represents clockwise (CW) or contraclockwise (CCW) rotation state of the motors and the C-ring, or which indicate shortcomings of the model.

There are multiple issues to consider when creating an all-encompassing model ring and joining together optimal fitting into the cryoET density, and uncovering detailed interaction interfaces between separate domains. It is a complicated task that requires novel approaches. A Martini-based CG study of FliG/FliF-C in the complete 34-mer assembly was designed to try to simulate the packing between domains similar to the CG tetramer work, but failed as the ring either deformed (instability of the CG assembly was presumed to be due to the absence of surrounding proteins in the C-ring complex), or remained too rigid in the case where restraints were introduced in a vertical plane in an attempt to account for missing components of the complex. In future, CG simulation could be applied to the rings derived from MDFF and more realistic interactions between domains could potentially be observed, but only if the issue with restraining the protein ring without compromising relevant domain movements could be solved.

## 6.3. Sequencing of *C. jejuni* complementation strains

Pseudorevertants in *C. jejuni ΔflhF* strain hold the key to the relationship between the obviously important, but mechanistically unclear function of FlhF along with the motor assembly and its positioning in the cell. To verify the phenotypes characterized before, it was necessary to complement *ΔflhFΔfliF* strains with the mutant $FliF_{M99I}$ variant to restore the phenotype as well as to observe the behaviour of $FliF_{M99I}$ in presence of FlhF by creating and complementing a new *ΔfliF* single deletion strain. Unexpectedly, it was discovered that complementation from an external locus does not allow for restoration of motility and flagella fail to assemble completely.

Additionally sequencing showed that insertion and deletion analysis should be performed on all previous pseudorevertant data, as analysis of the strains used in this project and simultaneous resequencing of two other pseudorevertants (not included in results, but performed in the same sample preparation and run) from Stoakes (2017)

study, suggests that a mutation resulting in the premature end of the C4-dicarboxylate transporter protein DcuA has originated during the construction of *ΔflhF* and carried over to all mutants and strains derived from it. While there is no direct link between this gene and motility, in general metabolism is linked to the ability of bacteria to utilize nutrients to synthesise the proteins they need; flagella synthesis is resource demanding. It is therefore important for future studies to further explore the effect of this mutation, particularly since transcriptome characterization of pseudo-revertants has revealed vast changes in metabolic landscape (Stoakes, 2017).

A novel mutation had occurred in FliF (Q9P) during the creation of the *ΔfliF:porAfliF(M99I)* strains and was found in both clones that were sequenced, further implying that it is giving the *C. jejuni* cells a selective advantage in a background in which FliF already has the M99I mutation and is overexpressed. One of these two clones that was used in negative stain EM was the only strain that showed some restoration of motility (30%). To discern the contributions of M99I and Q9P, further characterization and the creation of a strain with only Q9P would be needed.

Because complementation of wild type FliF did not work as planned from an external locus, the type of experiment that was originally intended to show restoration of phenotype would require devising a new strategy. NCTC11168 strain to express protein from heterologous plasmids, however it could be possible to delete both *fliF* and *fliG* together and perform the complementation in *Cj0046* with an operon containing both genes and variable (native or porA) promoters. Alternatively, genes could be targeted in their original operon via homologous recombination accompanied by high risk of additional mutation introduction in the process. Perhaps another *C. jejuni* strain that can express plasmids could be used; however, the overall phenotype differences between strains would introduce additional variability and difficulties in comparison with previous data. Discovery of the severe truncation of the *dcuA* gene in *ΔflhF* derived strains showed that for future work, in new lines of investigation, a new *ΔflhF* may have to be created.

## 6.4. MD simulations and pseudorevertant mutations in *C. jejuni*

MD simulation was initially considered as a tool to answer the questions that arose by the emergence of pseudorevertants in motor components FliF, FliH, and FliG in *C. jejuni* in the *ΔflhF* strain (Stoakes, 2017). However, the lack of structural information on FliF and low conservation level of FliH sequences at the start of the study led to alternative research directions. Instead, MD simulations were advantageously applied to study the conformation of FliG, as described in Chapter 3. The main difficulty of applying this method to studying pseudorevertant mutations is that – unlike disruptive mutations that most likely lead to destabilization of a structure or disruption of a binding interface – function-restoring mutations result in a WT (or WT-like) phenotypical outcome, and therefore their effects are not as easy to rationalize. For example, if the FliG T33A mutation in *C. jejuni* (T34 in *H. pylori*) confers a change to complex as a whole and has effects on multiple proteins, then this mutation could best be studied in a ring assembly context, not a single domain in solution. The ring assembly is now available (created in Chapter 4), however initial attempts at CG simulation of the whole FliG ring were not successful. Further optimization of restraints and setup is necessary.

Upon release of the FliF cryoEM structure (Johnson et al., 2020) one of the three mutations in the RMB2 domain (V177 in *C. jejuni*, V176 in *S. enterica*) was identified as conserved and suitable for simulation of an *S. enterica* structure. However, it was not certain how to define the environment around the domain, as it is likely in contact or partially embedded in the inner membrane. Therefore, trial CG simulations of the *S. enterica* FliF ring and the membrane were set up, but suffered from numerous shortcomings, like collapse of the ring and heavy influence of the starting setup. Rigorous optimization must be performed to create suitable CG simulation trials for the FliF ring, tackling the problems already identified. Together with the availability of the FliF RBM1 crystal structure and combination of prediction techniques for the transmembrane domains, it may now be possible to add missing information to anchor the FliF ring into the membrane. This would create a fairly large system that would be computationally costly even at the CG level. Some effects of single amino acid replacement could also be lost at the CG resolution (Introduction 1.9.1., Figure 1.14). Homology modelling of a *C. jejuni* sequence-based FliF ring remains challenging due to the stoichiometry issue – there are differences even between C33 and C34 solutions

in *S. enterica* (Johnson et al., 2020), which could be important due to *C. jejuni* larger diameter compared to other species (Beeby et al., 2016), depending on the role of the V177L mutation. In addition, binding of FlhA (fT3SS component housed in the MS-ring) with FliF (MS-ring) has been localized to the region spanning A175-S176 in FliF in *C. jejuni* (Boll and Hendrixson, 2013). Addition of FlhA would add additional complexity to simulations of such a system and interaction interfaces between FliF and FlhA are not structurally studied.

Most pseudorevertant residues between *C. jejuni* and *S. enterica* are conserved, which would be a pre-requisite to model them for *S. enterica*, given that introducing additional mutations *in silico* creates an additional level of uncertainty, which would mean FliF V177L and A102T and FliG T33A are the best candidates for simulation study (Table 6.1). However, all of these mutations are conservative – they replace an amino acid with one of similar charge and size. With the emergence of more structural information of FliF and optimization of simulation setups, MD could be utilized to look into the potential significance of the mutated residues in *C. jejuni* pseudorevertant strains.

***Table 6.1.*** **Conservation of known and novel pseudrevertant mutations in FliF and FliG.** In green is marked the mutation discovered in this study**.**

| Protein | Mutation | Conservation in sequence alignment | Has the same residue in a species with a structure available |
|---|---|---|---|
| **FliF** | V177L | Yes | *S. enterica* |
| **FliF** | M99I | Yes | no |
| **FliF** | A102T | Yes | *A. aeolicus* |
| **FliF** | E104K | Yes | no |
| **FliF** | S214T | Yes | no |
| **FliF** | S103R | No | no |
| **FliF** | P220L | Yes | no |
| **FliF** | Q9P | Partial* | no |
| **FliG** | T33A | Yes | *H. pylori* |

*-see Chapter 5.7.1

## 6.5. CryoET reconstructions of *C. jejuni* flagellar motor

Optimization of cryoET samples for plunge freezing led to collection of a cryoET dataset in WT *C. jejuni* and *ΔflhF fliF*$_{M99I}$ pseudorevertant. Reconstruction of the WT dataset was not accomplished to its full potential within the time allocated to this study, though key new structural insights were resolved. The test reconstructions show that the dataset is of sufficient quality to reach a 3D model. The most important future goal is to perform particle sorting without symmetry restraints and then reconstruct each class, if there are multiple, separately, in Relion. With the comparatively small WT dataset this may not be feasible, but certainly possible with the *ΔflhF fliF*$_{M99I}$ dataset.

While the FliF/FliG stoichiometry cannot be verified with the reconstructions currently obtained with EMAN2, some of the ring parts can be identified by similarity with previous structures; however, the C-ring part appears as a uniform ring with too little detail. In case the *ΔflhF fliF*$_{M99I}$ dataset, if a higher resolution reconstruction is obtained, it will be possible to model *C. jejuni* FliG and possibly FliF rings using homology modelling, fitting in map and the MDFF methodology as described in Chapter 4.

Results described in Chapter 3 and 4 have resulted in a peer-reviewed publication (Tupiņa et al., 2022).

# References


Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., Lindah, E., 2015. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX 1–2, 19–25. https://doi.org/10.1016/j.softx.2015.06.001

Abrusci, P., McDowell, M.A., Lea, S.M., Johnson, S., 2014. Building a secreting nanomachine: A structural overview of the T3SS. Curr. Opin. Struct. Biol. 25, 111–117. https://doi.org/10.1016/j.sbi.2013.11.001

Aidley, J., Holst Sørensen, M.C., Bayliss, C.D., Brøndsted, L., 2017. Phage exposure causes dynamic shifts in the expression states of specific phase-variable genes of Campylobacter jejuni. Microbiol. (United Kingdom) 163, 911–919. https://doi.org/10.1099/mic.0.000470

Aihara, E., Closson, C., Matthis, A.L., Schumacher, M.A., Engevik, A.C., Zavros, Y., Ottemann, K.M., Montrose, M.H., 2014. Motility and Chemotaxis Mediate the Preferential Colonization of Gastric Injury Sites by Helicobacter pylori. PLoS Pathog. 10, e1004275. https://doi.org/10.1371/journal.ppat.1004275

Aldridge, P., Hughes, K.T., 2002. Regulation of flagellar assembly. Curr. Opin. Microbiol. 5, 160–165. https://doi.org/10.1016/S1369-5274(02)00302-8

Allen, J.D., Chawla, H., Samsudin, F., Zuzic, L., Shivgan, A.T., Watanabe, Y., He, W.T., Callaghan, S., Song, G., Yong, P., Brouwer, P.J.M., Song, Y., Cai, Y., Duyvesteyn, H.M.E., Malinauskas, T., Kint, J., Pino, P., Wurm, M.J., Frank, M., Chen, B., Stuart, D.I., Sanders, R.W., Andrabi, R., Burton, D.R., Li, S., Bond, P.J., Crispin, M., 2021. Site-Specific Steric Control of SARS-CoV-2 Spike Glycosylation. Biochemistry 60, 2153–2169. https://doi.org/10.1021/acs.biochem.1c00279

Altegoer, F., Rensing, S.A., Bange, G., 2016. Structural basis for the CsrA-dependent modulation of translation initiation by an ancient regulatory protein. Proc. Natl. Acad. Sci. 113, 10168–10173.

Altegoer, F., Schuhmacher, J., Pausch, P., Bange, G., 2014. From molecular evolution

to biobricks and synthetic modules: A lesson by the bacterial flagellum. Biotechnol. Genet. Eng. Rev. 30, 49–64. https://doi.org/10.1080/02648725.2014.921500

Andersson, A.F., Lindberg, M., Jakobsson, H., Bäckhed, F., Nyrén, P., Engstrand, L., 2008. Comparative analysis of human gut microbiota by barcoded pyrosequencing. PLoS One 3, e2836. https://doi.org/10.1371/journal.pone.0002836

Andre, I., Bradley, P., Wang, C., Baker, D., 2007. Prediction of the structure of symmetrical protein assemblies. Proc. Natl. Acad. Sci. U. S. A. 104, 17656–17661. https://doi.org/10.1073/pnas.0702626104

Anjum, A., Brathwaite, K.J., Aidley, J., Connerton, P.L., Cummings, N.J., Parkhill, J., Connerton, I., Bayliss, C.D., 2016. Phase variation of a Type IIG restriction-modification enzyme alters site-specific methylation patterns and gene expression in Campylobacter jejuni strain NCTC11168. Nucleic Acids Res. 44, 4581–4594. https://doi.org/10.1093/nar/gkw019

Arpino, J.A.J., Rizkallah, P.J., Jones, D.D., 2012. Crystal Structure of Enhanced Green Fluorescent Protein to 1.35 Å Resolution Reveals Alternative Conformations for Glu222. PLoS One 7, e47132. https://doi.org/10.1371/journal.pone.0047132

Awad, W.A., Hess, C., Hess, M., 2018. Re-thinking the chicken–Campylobacter jejuni interaction: a review. Avian Pathol. 47, 352–363. https://doi.org/10.1080/03079457.2018.1475724

Baban, S.T., Kuehne, S.A., Barketi-Klai, A., Cartman, S.T., Kelly, M.L., Hardie, K.R., Kansau, I., Collignon, A., Minton, N.P., 2013. The Role of Flagella in Clostridium difficile Pathogenesis: Comparison between a Non-Epidemic and an Epidemic Strain. PLoS One 8, e73026. https://doi.org/10.1371/journal.pone.0073026

Backert, S., Tegtmeyer, N., Selbach, M., 2010. The versatility of helicobacter pylori caga effector protein functions: The master key hypothesis. Helicobacter 15, 163–176. https://doi.org/10.1111/j.1523-5378.2010.00759.x

Bailey, R.A., Kranis, A., Psifidi, A., Watson, K.A., Rothwell, L., Hocking, P.M.,

Kaiser, P., Stevens, M.P., Avendano, S., 2018. Colonization of a commercial broiler line by Campylobacter is under limited genetic control and does not significantly impair performance or intestinal health. Poult. Sci. 97, 4167–4176. https://doi.org/10.3382/ps/pey295

Bairoch, A., Boeckmann, B., Ferro, S., Gasteiger, E., Redaschi, N., Wu, C.H., Huang, H., Natale, D.A., Barker, W.C., Yeh, L.-S.L., Apweiler, R., Lopez, R., Magrane, M., Martin, M.J., O'Donovan, C., 2004. UniProt: The universal protein knowledgebase. Nucleic Acids Res. 32, D115–D119. https://doi.org/10.1093/nar/gkh131

Baker, M.A.B., Hynson, R.M.G., Ganuelas, L.A., Mohammadi, N.S., Liew, C.W., Rey, A.A., Duff, A.P., Whitten, A.E., Jeffries, C.M., Delalez, N.J., Morimoto, Y. V., Stock, D., Armitage, J.P., Turberfield, A.J., Namba, K., Berry, R.M., Lee, L.K., 2016. Domain-swap polymerization drives the self-assembly of the bacterial flagellar motor. Nat. Struct. Mol. Biol. 23, 197–203. https://doi.org/10.1038/nsmb.3172

Balaban, M., Hendrixson, D.R., 2011. Polar flagellar biosynthesis and a regulator of flagellar number influence spatial parameters of cell division in Campylobacter jejuni. PLoS Pathog 7, e1002420. https://doi.org/10.1371/journal.ppat.1002420

Balaban, M., Joslin, S.N., Hendrixson, D.R., 2009. FlhF and its GTPase activity are required for distinct processes in flagellar gene regulation and biosynthesis in Campylobacter jejuni. J. Bacteriol. 191, 6602–6611. https://doi.org/10.1128/JB.00884-09

Bange, G., Petzold, G., Wild, K., Parlitz, R.O., Sinning, I., 2007. The crystal structure of the third signal-recognition particle GTPase FlhF reveals a homodimer with bound GTP. Proc. Natl. Acad. Sci. 104, 13621–13625. https://doi.org/10.1073/pnas.0702570104

Barak, R., Eisenbach, M., 1992. Correlation between Phosphorylation of the Chemotaxis Protein Che-Y and its Activity at the Flagellar Motor. Biochemistry 31, 1821–1826. https://doi.org/10.1021/bi00121a034

Bateman, A., 2019. UniProt: A worldwide hub of protein knowledge. Nucleic Acids Res. 47, D506–D515. https://doi.org/10.1093/nar/gky1049

Beeby, M., Ribardo, D.A., Brennan, C.A., Ruby, E.G., Jensen, G.J., 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, E1917-26. https://doi.org/10.1073/pnas.1518952113

Beier, D., Frank, R., 2000. Molecular characterization of two-component systems of Helicobacter pylori. J. Bacteriol. 182, 2068–2076. https://doi.org/10.1128/JB.182.8.2068-2076.2000

Benkert, P., Tosatto, S.C.E., Schomburg, D., 2008. QMEAN: A comprehensive scoring function for model quality assessment. Proteins Struct. Funct. Genet. 71, 261–277. https://doi.org/10.1002/prot.21715

Berg, H.C., Brown, D.A., 1972. Chemotaxis in E coli analysed by 3D tracking. Nature 239, 500–504. https://doi.org/https://doi.org/10.1038/239500a0

Bergeron, J.R., 2016. Structural modeling of the flagellum MS ring protein FliF reveals similarities to the type III secretion system and sporulation complex. PeerJ 4, e1718. https://doi.org/10.7717/peerj.1718

Bharat, T.A.M., Scheres, S.H.W., 2016. Resolving macromolecular structures from electron cryo-Tomography data using subtomogram averaging in RELION. Nat. Protoc. 11, 2054–2065. https://doi.org/10.1038/nprot.2016.124

Bik, E.M., Eckburg, P.B., Gill, S.R., Nelson, K.E., Purdom, E.A., Francois, F., Perez-Perez, G., Blaser, M.J., Relman, D.A., 2006. Molecular analysis of the bacterial microbiota in the human stomach. Proc. Natl. Acad. Sci. U. S. A. 103, 732–737. https://doi.org/10.1073/pnas.0506655103

Böckmann, R.A., Grubmüller, H., 2004. Multistep binding of divalent cations to phospholipid bilayers: A molecular dynamics study. Angew. Chemie - Int. Ed. 43, 1021–1024. https://doi.org/10.1002/anie.200352784

Boldon, L., Laliberte, F., Liu, L., 2015. Review of the fundamental theories behind small angle X-ray scattering, molecular dynamics simulations, and relevant integrated application. Nano Rev. 6, 25661. https://doi.org/10.3402/nano.v6.25661

Boll, J.M., Hendrixson, D.R., 2013. A regulatory checkpoint during flagellar

biogenesis in campylobacter jejuni initiates signal transduction to activate transcription of flagellar genes. MBio 4, e00432-13. https://doi.org/10.1128/mBio.00432-13

Bolton, D.J., 2015. Campylobacter virulence and survival factors. Food Microbiol 48, 99–108. https://doi.org/10.1016/j.fm.2014.11.017

Bowers, K.J., Chow, E., Xu, H., Dror, R.O., Eastwood, M.P., Gregersen, B.A., Klepeis, J.L., Kolossvary, I., Moraes, M.A., Sacerdoti, F.D., Salmon, J.K., Shan, Y., Shaw, D.E., 2006. Scalable algorithms for molecular dynamics simulations on commodity clusters, in: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing, SC'06. Tampa, FL, USA, p. 43. https://doi.org/10.1145/1188455.1188544

Brahmachary, P., Dashti, M.G., Olson, J.W., Hoover, T.R., 2004. Helicobacter pylori FlgR is an enhancer-independent activator of σ54-RNA polymerase holoenzyme. J. Bacteriol. 186, 4535–4542. https://doi.org/10.1128/JB.186.14.4535-4542.2004

Brooks, B.R., III, C.L.B., A. D. Mackerell, J., Nillsonn, L., Petrella, R.J., Roux, B., Won, Y., Archontis, G., Bartels, C., Boresch, S., Caflish, A., Caves, L., Cui, Q., Dinner, A.R., Feig, M., Fischer, S., Gao, J., Hodoscek, M., Im, W., Kuzcera, K., Lazaridis, T., Ma, J., Ovichnikov, V., Paci, E., Pastor, R.W., Post, C.B., Pu, J.Z., Schaefer, M., Tidor, B., Venable, R.M., Woodcock, H.L., Wu, X., Yang, W., York, D.M., Karplus, M., 2010. CHARMM: The Biomolecular Simulation Program. J. Comput. Chem. 30, 1545–1614. https://doi.org/10.1002/jcc

Brown, P.N., Hill, C.P., Blair, D.F., 2002. Crystal structure of the middle and C-terminal domains of the flagellar rotor protein FliG. EMBO J. 21, 3225–3234. https://doi.org/10.1093/emboj/cdf332

Bussi, G., Donadio, D., Parrinello, M., 2007. Canonical sampling through velocity rescaling. J. Chem. Phys. 126, 014101. https://doi.org/10.1063/1.2408420

Butzler, J.-P., 2004. Campylobacter, from obscurity to celebrity. Clin. Microbiol. Infect. 10, 868–876. https://doi.org/10.1111/j.1469-0691.2004.00983.x

Cameron, A., Gaynor, E.C., 2014. Hygromycin B and apramycin antibiotic resistance cassettes for use in Campylobacter jejuni. PLoS One 9.

https://doi.org/10.1371/journal.pone.0095084

Carroll, B.L., Nishikino, T., Guo, W., Zhu, S., Kojima, S., Homma, M., Liu, J., 2020. The flagellar motor of Vibrio alginolyticus undergoes major structural remodeling during rotational switching. Elife 9, e61446. https://doi.org/10.7554/eLife.61446

Carugo, O., Argos, P., 1997. Protein-protein crystal-packing contacts. Protein Sci. 6, 2261–2263.

Carver, T., Harris, S.R., Berriman, M., Parkhill, J., McQuillan, J.A., 2012. Artemis: An integrated platform for visualization and analysis of high-throughput sequence-based experimental data. Bioinformatics 28, 464–469. https://doi.org/10.1093/bioinformatics/btr703

Chaban, B., Coleman, I., Beeby, M., 2018. Evolution of higher torque in Campylobacter-Type bacterial flagellar motors. Sci. Rep. 8, 1–11. https://doi.org/10.1038/s41598-017-18115-1

Chaban, B., Hughes, H.V., Beeby, M., 2015. The flagellum in bacterial pathogens: For motility and a whole lot more. Semin. Cell Dev. Biol. 46, 91–103. https://doi.org/10.1016/j.semcdb.2015.10.032

Chang, Y., Liu, J., 2019. Architecture and Assembly of Periplasmic Flagellum. Microbiol. Spectr. 7. https://doi.org/10.1128/microbiolspec.psib-0030-2019

Chang, Y., Zhang, K., Carroll, B.L., Zhao, X., Charon, N.W., Norris, S.J., Motaleb, M.A., Chunhao Li, J.L., 2020. Molecular mechanism for rotational switching of the bacterial flagellar motor. Nat Struct Mol Biol. 27, 1041–1047. https://doi.org/10.1038/s41594-020-0497-2.Molecular

Chatterjee, S., Debenedetti, P.G., Stillinger, F.H., Lynden-Bell, R.M., 2008. A computational investigation of thermodynamics, structure, dynamics and solvation behavior in modified water models. J. Chem. Phys. 128, 124511–124519. https://doi.org/10.1063/1.2841127

Cheatham, T.E., Case, D.A., 2013. Twenty-five years of nucleic acid simulations. Biopolymers 99, 969–977. https://doi.org/10.1002/bip.22331

Chevance, F.F. V, Hughes, K.T., 2008. Coordinating assembly of a bacterial

macromolecular machine. Nat. Rev. Microbiol. 6, 455–465. https://doi.org/10.1038/nrmicro1887

Chintoan-Uta, C., Cassady-Cain, R.L., Stevens, M.P., 2016. Evaluation of flagellum-related proteins FliD and FspA as subunit vaccines against Campylobacter jejuni colonisation in chickens. Vaccine 34, 1739–1743. https://doi.org/10.1016/j.vaccine.2016.02.052

Cohen, E.J., Nakane, D., Kabata, Y., Hendrixson, D.R., Nishizaka, T., 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. https://doi.org/10.1371/journal.ppat.1008620

Comeau, S.R., Gatchell, D.W., Vajda, S., Camacho, C.J., 2004. ClusPro: A fully automated algorithm for protein-protein docking. Nucleic Acids Res. 32, 96–99. https://doi.org/10.1093/nar/gkh354

Constantinescu, C.S., Arsenescu, R.I., Arenescu, V., n.d. Neuro-Immuno-Gastroenterology, 2016th ed. Springer International Publishing.

Correa, P., Houghton, J.M., 2007. Carcinogenesis of Helicobacter pylori. Gastroenterology 133, 659–672. https://doi.org/10.1053/j.gastro.2007.06.026

Cover, T.L., Blaser, M.J., 2009. Helicobacter pylori in Health and Disease. Gastroenterology 136, 1863–1873. https://doi.org/10.1053/j.gastro.2009.01.073

Dalhus, B., Saarinen, M., Sauer, U.H., Eklund, P., Johansson, K., Karlsson, A., Ramaswamy, S., Bjørk, A., Synstad, B., Naterstad, K., Sirevåg, R., Eklund, H., 2002. Structural basis for thermophilic protein stability: Structures of thermophilic and mesophilic malate dehydrogenases. J. Mol. Biol. 318, 707–721. https://doi.org/10.1016/S0022-2836(02)00050-5

Darden, T., York, D., Pedersen, L., 1993. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J. Chem. Phys. 98, 10089–10092. https://doi.org/10.1063/1.464397

Darnton, N.C., Turner, L., Rojevsky, S., Berg, H.C., 2007. On torque and tumbling in swimming Escherichia coli. J. Bacteriol. 189, 1756–1764. https://doi.org/10.1128/JB.01501-06

Deckert, G., Warren, P. V., Gaasterland, T., Young, W.G., Lenox, A.L., Graham, D.E., Overbeek, R., Snead, M.A., Keller, M., Aujay, M., Huber, R., Feldman, R.A., Short, J.M., Olsen, G.J., Swanson, R. V., 1998. The complete genome of the hyperthermophilic bacterium Aquifex aeolicus. Nature 392, 353–358. https://doi.org/10.1038/32831

DeLano, W.L., 2002. The PyMOL Molecular Graphics System.

Dey, S., Ritchie, D.W., Levy, E.D., 2018. PDB-wide identification of biological assemblies from conserved quaternary structure geometry. Nat. Methods 15, 67–72. https://doi.org/10.1038/nmeth.4510

Djennad, A., Lo Iacono, G., Sarran, C., Lane, C., Elson, R., Höser, C., Lake, I.R., Colón-González, F.J., Kovats, S., Semenza, J.C., Bailey, T.C., Kessel, A., Fleming, L.E., Nichols, G.L., 2019. Seasonality and the effects of weather on Campylobacter infections. BMC Infect. Dis. 19, 1–10. https://doi.org/10.1186/s12879-019-3840-7

Dodd, T., 2017. Simulation-based methods for model building and refinement in cryo-electron microscopy. Physiol. Behav. 176, 139–148. https://doi.org/10.1021/acs.jcim.0c00087.Simulation-based

Dokholyan, N., 2012. Computational Modeling of Biological Systems, 1st ed, Biological and Medical Physics, Biomedical Engineering. Springer New York, NY, New York. https://doi.org/https://doi.org/10.1007/978-1-4614-2146-7

Dugar, G., Svensson, S.L., Bischler, T., Wäldchen, S., Reinhardt, R., Sauer, M., Sharma, C.M., 2016. The CsrA-FliW network controls polar localization of the dual-function flagellin mRNA in Campylobacter jejuni. Nat. Commun. 7, 1–18. https://doi.org/10.1038/ncomms11667

ECDC, 2019. European Centre for Disease Prevention and Control. Campylobacteriosis. In: ECDC. Annual epidemiological report for 2017.

Edgar, R.C., 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. https://doi.org/10.1093/nar/gkh340

EFSA, 2021. The European Union One Health 2019 Zoonoses Report, EFSA Journal.

https://doi.org/10.2903/j.efsa.2021.6406

Eisenbach, M., 1996. Micro Review Control of bacterial chemotaxis. Mol. Microbiol. 20, 903–910. https://doi.org/https://doi.org/10.1111/j.1365-2958.1996.tb02531.x

Eisenbach, M., Wolf, A., Welch, M., Caplan, S.R., Lapidus, I.R., Macnab, R.M., Aloni, H., Asher, O., 1990. Pausing, switching and speed fluctuation of the bacterial flagellar motor and their relation to motility and chemotaxis. J. Mol. Biol. 211, 551–563. https://doi.org/10.1016/0022-2836(90)90265-N

Elgamoudi, B.A., 2016. The role of transducer-like proteins in Campylobacter jejuni Thesis submitted for the degree of Doctor of Philosophy At the University of Leicester By Department of Genetics University of Leicester. University of Leicester.

Elgamoudi, B.A., Ketley, J.M., 2018. Lighting up my life: a LOV-based fluorescent reporter for Campylobacter jejuni. Res. Microbiol. 169, 108–114. https://doi.org/10.1016/j.resmic.2017.10.003

Evans, L.D.B., Poulter, S., Terentjev, E.M., Hughes, C., Fraser, G.M., 2013. A chain mechanism for flagellum growth. Nature 504, 287–290. https://doi.org/10.1038/nature12682

Ewels, P., Magnusson, M., Lundin, S., Käller, M., 2016. MultiQC: Summarize analysis results for multiple tools and samples in a single report. Bioinformatics 32, 3047–3048. https://doi.org/10.1093/bioinformatics/btw354

Fiser, A., Sali, A., 2003. ModLoop: Automated modeling of loops in protein structures. Bioinformatics 19, 2500–2501. https://doi.org/10.1093/bioinformatics/btg362

Galaz-Montoya, J.G., Flanagan, J., Schmid, M.F., Ludtke, S.J., 2015. Single particle tomography in EMAN2. J. Struct. Biol. 190, 279–290. https://doi.org/10.1016/j.jsb.2015.04.016

Gaynor, E.C., Cawthraw, S., Manning, G., MacKichan, J.K., Falkow, S., Newell, D.G., 2004. The Genome-Sequenced Variant of Campylobacter jejuni NCTC 11168 and the Original Clonal Clinical Isolate Differ Markedly in Colonization, Gene Expression, and Virulence-Associated Phenotypes. J. Bacteriol. 186, 503–

517. https://doi.org/10.1128/JB.186.2.503-517.2004

Geis, G., Suerbaum, S., Forsthoff, B., Leying, H., Opferkuch, W., 1993. Ultrastructure and biochemical studies of the flagellar sheath of Helicobacter pylori. J. Med. Microbiol. 38, 371–377. https://doi.org/10.1099/00222615-38-5-371

Girón, J.A., 1995. Expression of flagella and motility by Shigella. Mol. Microbiol. 18, 63–75. https://doi.org/10.1111/j.1365-2958.1995.mmi_18010063.x

González-Pedrajo, B., Fraser, G.M., Minamino, T., Macnab, R.M., 2002. Molecular dissection of Salmonella FliH, a regulator of the ATPase FliI and the type III flagellar protein export pathway. Mol. Microbiol. 45, 967–982. https://doi.org/10.1046/j.1365-2958.2002.03047.x

Gonzalez, M.A., 2011. Force fields and molecular dynamics simulations. Collect. SFN 12, 169–200. https://doi.org/10.1051/sfn/201112009

Gordon, J.C., Myers, J.B., Folta, T., Shoja, V., Heath, L.S., Onufriev, A., 2005. H++: A server for estimating pKas and adding missing hydrogens to macromolecules. Nucleic Acids Res. 33, W368-371. https://doi.org/10.1093/nar/gki464

Green, E.R., Mecsas, J., 2016. Bacterial secretion systems: An overview. Virulence Mech. Bact. Pathog. 213–239. https://doi.org/10.1128/9781555819286.ch8

Green, J.C.D., Kahramanoglou, C., Rahman, A., Pender, A.M.C., Charbonnel, N., Fraser, G.M., 2009. Recruitment of the Earliest Component of the Bacterial Flagellum to the Old Cell Division Pole by a Membrane-Associated Signal Recognition Particle Family GTP-Binding Protein. J. Mol. Biol. 391, 679–690. https://doi.org/10.1016/j.jmb.2009.05.075

Grunenfelder, B., Gehrig, S., Jenal, U., 2003. Role of the Cytoplasmic C Terminus of the FliF Motor Protein in Flagellar Assembly and Rotation. J Bacteriol 185, 1624–1633. https://doi.org/10.1128/jb.185.5.1624-1633.2003

Guerry, P., 2007. Campylobacter flagella: not just for motility. Trends Microbiol. 15, 456–461. https://doi.org/10.1016/j.tim.2007.09.006

Guerry, P., Ewing, C.P., Schirm, M., Lorenzo, M., Kelly, J., Pattarini, D., Majam, G., Thibault, P., Logan, S., 2006. Changes in flagellin glycosylation affect Campylobacter autoagglutination and virulence. Mol. Microbiol. 60, 299–311.

https://doi.org/10.1111/j.1365-2958.2006.05100.x

Hara-Kudo, Y., Takatori, K., 2011. Contamination level and ingestion dose of foodborne pathogens associated with infections. Epidemiol. Infect. 139, 1505–1510. https://doi.org/10.1017/s095026881000292x

Health Protection Scotland, 2020. Annual summary of Campylobacter infections, 2019.

Henderson, L.D., Teige R. S. Matthews-Palmer, C.J.G., Ribardo, D.A., Beeby, M., Hendrixson, D.R., 2020. Diversification of Campylobacter jejuni Flagellar C-Ring Composition Impacts Its Structure and Function in Motility, Flagellar Assembly, and Cellular Processes. Mol. Biol. Physiol. 11, e02286–e02319. https://doi.org/https://doi.org/ 10.1128/mBio.02286-19

Hendriksen, J.J., Lee, H.J., Bradshaw, A.J., Namba, K., Chevance, F.F.V., Minamino, T., Hughes, K.T., 2021. Genetic Analysis of the Salmonella FliE Protein That Forms the Base of the Flagellar Axial Structure. MBio 12, e02392-21. https://doi.org/10.1128/mBio.02392-21

Hess, B., Bekker, H., Berendsen, H.J.C., Fraaije, J.G.E.M., 1997. LINCS: A Linear Constraint Solver for molecular simulations. J. Comput. Chem. 18, 1463–1472. https://doi.org/10.1002/(SICI)1096-987X(199709)18:12<1463::AID-JCC4>3.0.CO;2-H

Hoffman, C.L., Gonyar, L.A., Zacca, F., Sisti, F., Fernandez, J., Wong, T., Damron, F.H., Hewlett, E.L., 2019. Bordetella pertussis can be motile and express flagellum-like structures. MBio 10, e00787-19. https://doi.org/10.1128/mBio.00787-19

Hollingsworth, S.A., Dror, R.O., 2018. Molecular dynamics simulation for all. Neuron 99, 1129–1143. https://doi.org/10.1016/j.neuron.2018.08.011.

Hooft, R.W.W., Sander, C., Vriend, G., 1997. Objectively judging the quality of a protein structure from a ramachandran plot. Bioinformatics 13, 425–430. https://doi.org/10.1093/bioinformatics/13.4.425

Horii, T., Morita, M., Muramatsu, H., Muranaka, Y., Kanno, T., Maekawa, M., 2003. Effects of mupirocin at subinhibitory concentrations of flagella formation in

Pseudomonas aeruginosa and Proteus mirabilis. J. Antimicrob. Chemother. 51, 1175–1179. https://doi.org/10.1093/jac/dkg226

Horrocks, S.M., Anderson, R.C., Nisbet, D.J., Ricke, S.C., 2009. Incidence and ecology of Campylobacter jejuni and coli in animals. Anaerobe 15, 18–25. https://doi.org/10.1016/j.anaerobe.2008.09.001

Howitt, M.R., Lee, J.Y., Lertsethtakarn, P., Vogelmann, R., Joubert, L.M., Ottemann, K.M., Amieva, M.R., 2011. Chepep controls helicobacter pylori infection of the gastric glands and chemotaxis in the epsilonproteobacteria. MBio 2, e00098-11. https://doi.org/10.1128/mBio.00098-11

Huang, J., Mackerell, A.D., 2013. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. J. Comput. Chem. 34, 2135–2145. https://doi.org/10.1002/jcc.23354

Huang, J., Rauscher, S., Nawrocki, G., Ran, T., Feig, M., De Groot, B.L., Grubmüller, H., MacKerell, A.D., 2016. CHARMM36m: An improved force field for folded and intrinsically disordered proteins. Nat. Methods 14, 71–73. https://doi.org/10.1038/nmeth.4067

Hughes, K.T., 2017. Flagellum Length Control: How Long Is Long Enough? Curr. Biol. 27, R413–R415. https://doi.org/10.1016/j.cub.2017.04.008

Humphrey, S., Wigley, P., Williams, N., Chaloner, G., Davidson, N., Humphrey, T., Kipar, A., Kemmett, K., 2014. Campylobacter jejuni Is Not Merely a Commensal in Commercial Broiler Chickens and Affects Bird Welfare. MBio 5, 1–7. https://doi.org/10.1128/mbio.01364-14

Humphrey, W., Dalke, A., Schulten, K., 1996. Visual Molecular Dynamics. J. Mol. Graph. 14, 38–34. https://doi.org/https://doi.org/10.1016/0263-7855(96)00018-5

Ikeda, T., Oosawa, K., Hotani, H., 1996. Self-assembly of the filament capping protein, FliD, of bacterial flagella into an annular structure. J. Mol. Biol. 259, 679–686. https://doi.org/10.1006/jmbi.1996.0349

Imada, K., Minamino, T., Uchida, Y., Kinoshita, M., Namba, K., 2016. Insight into the flagella type III export revealed by the complex structure of the type III ATPase and its regulator. Proc. Natl. Acad. Sci. 113, 3633–3638.

https://doi.org/10.1073/pnas.1524025113

Inoue, T., Barker, C.S., Matsunami, H., Aizawa, S.I., 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. https://doi.org/10.1099/mic.0.000648

Jiang, H., Dong, H., Zhang, G., Yu, B., Chapman, L.R., Fields, M.W., 2006. Microbial diversity in water and sediment of Lake Chaka, an athalassohaline lake in northwestern China. Appl. Environ. Microbiol. 72, 3832–3845. https://doi.org/10.1128/AEM.02869-05

Johnson, S., Fong, Y.H., Deme, J.C., Furlong, E.J., Kuhlen, L., 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. https://doi.org/10.1038/s41564-020-0703-3

Johnson, S., Furlong, E.J., Deme, J.C., Nord, A.L., Caesar, J.J.E., Chevance, F.F.V., Berry, R.M., Hughes, K.T., Lea, S.M., 2021. Molecular structure of the intact bacterial flagellar basal body. Nat. Microbiol. 6, 712–721. https://doi.org/10.1038/s41564-021-00895-y

Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W., Klein, M.L., 1983. Comparison of simple potential functions for simulating liquid water. J. Chem. Phys. 79, 926–935. https://doi.org/10.1063/1.445869

Josenhans, C., Suerbaum, S., 2002. The role of motility as a virulence factor in bacteria. Int. J. Med. Microbiol. 291, 605–614. https://doi.org/10.1078/1438-4221-00173

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S.A.A., Ballard, A.J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., Back, T., Petersen, S., Reiman, D., Clancy, E., Zielinski, M., Steinegger, M., Pacholska, M., Berghammer, T., Bodenstein, S., Silver, D., Vinyals, O., Senior, A.W., Kavukcuoglu, K., Kohli, P., Hassabis, D., 2021. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589. https://doi.org/10.1038/s41586-021-03819-2

Kaakoush, N.O., Castaño-Rodríguez, N., Mitchell, H.M., Man, S.M., 2015. Global epidemiology of campylobacter infection. Clin. Microbiol. Rev. 28, 687–720. https://doi.org/10.1128/CMR.00006-15

Kaur, H., Kalia, M., Singh, V., Modgil, V., Mohan, B., Taneja, N., 2021. In silico identification and characterization of promising drug targets in highly virulent uropathogenic Escherichia coli strain CFT073 by protein-protein interaction network analysis. Informatics Med. Unlocked 25, 100704. https://doi.org/10.1016/j.imu.2021.100704

Kawamoto, A., Miyata, T., Makino, F., Kinoshita, M., Imada, K., Kato, T., Namba, K., 2021. Native structure of flagellar MS ring is formed by 34 subunits with 23-fold and 11-fold subsymmetries. Nat. Commun. 12, 4223. https://doi.org/10.1038/s41467-021-24507-9

Kazmierczak, M.J., Wiedmann, M., Boor, K.J., 2005. Alternative Sigma Factors and Their Roles in Bacterial Virulence. Society 69, 527–543. https://doi.org/10.1128/MMBR.69.4.527

Khan, S., Guo, T.W., Misra, S., 2018. A coevolution-guided model for the rotor of the bacterial flagellar motor. Sci. Rep. 8, 11754. https://doi.org/10.1038/s41598-018-30293-0

Kibbe, W.A., 2007. OligoCalc: An online oligonucleotide properties calculator. Nucleic Acids Res. 35, 43–46. https://doi.org/10.1093/nar/gkm234

Kihara, D., 2010. Protein Function Prediction in Bacteria, 1611th ed. Springer Protocols.

Kihara, M., Minamino, T., Yamaguchi, S., Macnab, R.M., 2001. Intergenic suppression between the flagellar MS ring protein FliF of Salmonella and FlhA, a membrane component of its export apparatus. J. Bacteriol. 183, 1655–1662. https://doi.org/10.1128/JB.183.5.1655-1662.2001

Kim, D.E., Chivian, D., Baker, D., 2004. Protein structure prediction and analysis using the Robetta server. Nucleic Acids Res. 32, 526–531. https://doi.org/10.1093/nar/gkh468

Kim, E.A., Panushka, J., Meyer, T., Carlisle, R., Baker, S., Ide, N., Lynch, M., Crane,

B.R., Blair, D.F., 2017. Architecture of the Flagellar Switch Complex of Escherichia coli: Conformational Plasticity of FliG and Implications for Adaptive Remodeling. J. Mol. Biol. 429, 1305–1320. https://doi.org/10.1016/j.jmb.2017.02.014

Kobayashi, H., Saito, H., Kakegawa, T., 2000. Bacterial strategies to inhabit acidic environments. J. Gen. Appl. Microbiol. 46, 235–243. https://doi.org/https://doi.org/10.2323/jgam.46.235

Koboldt, D.C., Chen, K., Wylie, T., Larson, D.E., McLellan, M.D., Mardis, E.R., Weinstock, G.M., Wilson, R.K., Ding, L., 2009. VarScan: Variant detection in massively parallel sequencing of individual and pooled samples. Bioinformatics 25, 2283–2285. https://doi.org/10.1093/bioinformatics/btp373

Kojima, S., Blair, D.F., 2001. Conformational change in the stator of the bacterial flagellar Motor. Biochemistry 40, 13041–13050. https://doi.org/10.1021/bi011263o

Kojima, S., Kajino, H., Hirano, K., Inoue, Y., Terashima, H., Homma, M., 2021. Role of the N- A nd C-Terminal Regions of FliF, the MS Ring Component in the Vibrio Flagellar Basal Body. J. Bacteriol. 203, 1–9. https://doi.org/10.1128/JB.00009-21

Komatsu, H., Hayashi, F., Sasa, M., Shikata, K., Yamaguchi, S., Namba, K., Oosawa, K., 2016. Genetic analysis of revertants isolated from the rod-fragile fliF mutant of Salmonella. Biophys Physicobiol 13, 13–25. https://doi.org/10.2142/biophysico.13.0_13

Konkel, M.E., Klena, J.D., Rivera-amill, V., Monteville, M.R., Biswas, D., Raphael, B., Mickelson, J., 2004. Secretion of Virulence Proteins from Campylobacter jejuni Is Dependent on a Functional Flagellar Export Apparatus. J. Bacteriol. 186, 3296–3303. https://doi.org/10.1128/JB.186.11.3296

Krah, A., Takada, S., 2016. On the ATP binding site of the ε subunit from bacterial F-type ATP synthases. Biochim. Biophys. Acta - Bioenerg. 1857, 332–340. https://doi.org/10.1016/j.bbabio.2016.01.007

Kremer, J.R., Mastronarde, D.N., McIntosh, J.R., 1996. Computer visualization of

three-dimensional image data using IMOD. J. Struct. Biol. 116, 71–76. https://doi.org/10.1006/jsbi.1996.0013

Krulwich, T.A., Sachs, G., Padan, E., 2011. Molecular aspects of bacterial pH sensing and homeostasis. Nat. Rev. Microbiol. 9, 330–343. https://doi.org/10.1038/nrmicro2549

Krzyzek, P., Gościniak, G., 2018. Morphology of Helicobacter pylori as a result of peptidoglycan and cytoskeleton rearrangements. Prz. Gastroenterol. 13, 182–195. https://doi.org/10.5114/pg.2018.78284

Kubori, T., Yamaguchi, S., Aizawa, S.I., 1997. Assembly of the switch complex onto the MS ring complex of Salmonella typhimurium does not require any other flagellar proteins. J. Bacteriol. 179, 813–817. https://doi.org/10.1128/jb.179.3.813-817.1997

Kuhlen, L., Abrusci, P., Johnson, S., Gault, J., Deme, J., Caesar, J., Dietsche, T., Mebrhatu, M.T., Ganief, T., Macek, B., Wagner, S., Robinson, C. V., Lea, S.M., 2018. Structure of the core of the type III secretion system export apparatus. Nat. Struct. Mol. Biol. 25, 1–8. https://doi.org/10.1038/s41594-018-0086-9

Kukol, A., 2008. Molecular Modeling of Proteins. Humana Press, Totowa, NJ. https://doi.org/10.1007/978-1-59745-177-2

Lake, I.R., Colón-González, F.J., Takkinen, J., Rossi, M., Sudre, B., Gomes Dias, J., Tavoschi, L., Joshi, A., Semenza, J.C., Nichols, G., 2019. Exploring campylobacter seasonality across europe using the european surveillance system (TESSy), 2008 to 2016. Eurosurveillance 24, 1–12. https://doi.org/10.2807/1560-7917.ES.2019.24.13.180028

Lam, K.-H., Ling, T.K.-W., Chan, S.-O., Au, S.W.-N., Lam, Y.-W., Ip, W.-S., 2012. Multiple Conformations of the FliG C-Terminal Domain Provide Insight into Flagellar Motor Switching. Structure 20, 315–325. https://doi.org/10.1016/j.str.2011.11.020

Lam, K.H., Lam, W.W.L., Wong, J.Y.K., Chan, L.C., Kotaka, M., Ling, T.K.W., Jin, D.Y., Ottemann, K.M., Au, S.W.N., 2013. Structural basis of FliG-FliM interaction in Helicobacter pylori. Mol. Microbiol. 88, 798–812.

https://doi.org/10.1111/mmi.12222

Lane, M.C., O'Toole, P.W., Moore, S.A., 2006. Molecular basis of the interaction between the flagellar export proteins FliI and FliH from Helicobacter pylori. J. Biol. Chem. 281, 508–517. https://doi.org/10.1074/jbc.M507238200

Langmead, B., Salzberg, S.L., 2012. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. https://doi.org/10.1038/nmeth.1923

Lee, L.K., Ginsburg, M.A., Crovace, C., Donohoe, M., Stock, D., 2010. Structure of the torque ring of the flagellar motor and the molecular basis for rotational switching. Nature 466, 996–1000. https://doi.org/10.1038/nature09300

Lele, P.P., Hosu, B.G., Berg, H.C., 2013. Dynamics of mechanosensing in the bacterial flagellar motor. Proc. Natl. Acad. Sci. U. S. A. 110, 11839–11844. https://doi.org/10.1073/pnas.1305885110

Letunic, I., Doerks, T., Bork, P., 2015. SMART: Recent updates, new developments and status in 2015. Nucleic Acids Res. 43, D257–D260. https://doi.org/10.1093/nar/gku949

Levenson, R., Zhou, H., Dahlquist, F.W., 2012. Structural insights into the interaction between the bacterial flagellar motor proteins FliF and FliG. Biochemistry 51, 5052–5060. https://doi.org/10.1021/bi3004582

Li, H., Sourjik, V., 2011. Assembly and stability of flagellar motor in Escherichia coli. Mol. Microbiol. 80, 886–899. https://doi.org/10.1111/j.1365-2958.2011.07557.x

Li, J., Gulbronson, C.J., Bogacz, M., Hendrixson, D.R., Thompson, S.A., 2018. Fliw controls growth-phase expression of campylobacter jejuni flagellar and non-flagellar proteins via the post-transcriptional regulator CsrA. Microbiol. (United Kingdom) 164, 1308–1319. https://doi.org/10.1099/mic.0.000704

Lloyd, S.A., Tang, H., Wang, X., Billings, S., Blair, D.F., 1996. Torque generation in the flagellar motor of Escherichia coli: Evidence of a direct role for FliG but not for FliM or FliN. J. Bacteriol. 178, 223–231. https://doi.org/10.1128/jb.178.1.223-231.1996

Lowder, B.J., Duyvesteyn, M.D., Blair, D.F., 2005. FliG subunit arrangement in the flagellar rotor probed by targeted cross-linking. J. Bacteriol. 187, 5640–5647.

https://doi.org/10.1128/JB.187.16.5640-5647.2005

Lowenthal, A.C., Hill, M., Sycuro, L.K., Mehmood, K., Salama, N.R., Ottemann, K.M., 2009. Functional analysis of the Helicobacter pylori flagellar switch proteins. J. Bacteriol. 191, 7147–7156. https://doi.org/10.1128/JB.00749-09

Lučić, V., Rigort, A., Baumeister, W., 2013. Cryo-electron tomography: The challenge of doing structural biology in situ. J. Cell Biol. 202, 407–419. https://doi.org/10.1083/jcb.201304193

Luo, J., Liu, Z., Guo, Y., Li, M., 2015. A structural dissection of large protein-protein crystal packing contacts. Sci. Rep. 5, 14214. https://doi.org/10.1038/srep14214

Lynch, M.J., Levenson, R., Kim, E.A., Sircar, R., Blair, D.F., Dahlquist, F.W., Crane, B.R., 2017. Co-Folding of a FliF-FliG Split Domain Forms the Basis of the MS:C Ring Interface within the Bacterial Flagellar Motor. Structure 25, 317–328. https://doi.org/10.1016/j.str.2016.12.006

MacKerell, A.D., Banavali, N., Foloppe, N., 2000. Development and current status of the CHARMM force field for nucleic acids. Biopolymers 56, 257–265. https://doi.org/10.1002/1097-0282(2000)56:4<257::AID-BIP10029>3.0.CO;2-W

Macnab, R.M., 2003. How Bacteria Assemble Flagella. Annu. Rev. Microbiol. 57, 77–100. https://doi.org/10.1146/annurev.micro.57.030502.090832

Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E., Simmerling, C., 2015. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. J. Chem. Theory Comput. 11, 3696–3713. https://doi.org/10.1021/acs.jctc.5b00255

Marrink, S.J., Risselada, H.J., Yefimov, S., Tieleman, D.P., De Vries, A.H., 2007. The MARTINI force field: Coarse grained model for biomolecular simulations. J. Phys. Chem. B 111, 7812–7824. https://doi.org/10.1021/jp071097f

Martínez, O.F., Cardoso, M.H., Ribeiro, S.M., Franco, O.L., 2019. Recent advances in anti-virulence therapeutic strategies with a focus on dismantling bacterial membrane microdomains, toxin neutralization, quorum-sensing interference and biofilm inhibition. Front. Cell. Infect. Microbiol. 9, 1–24. https://doi.org/10.3389/fcimb.2019.00074

McMurry, J.L., Murphy, J.W., González-Pedrajo, B., 2006. The FliN-FliH interaction mediates localization of flagellar export ATPase FliI to the C ring complex. Biochemistry 45, 11790–11798. https://doi.org/10.1021/bi0605890

McQuiston, J.R., Parrenas, R., Ortiz-Rivera, M., Gheesling, L., Brenner, F., Fields, P.I., 2004. Sequencing and Comparative Analysis of Flagellin Genes fliC, fljB, and flpA from Salmonella. J. Clin. Microbiol. 42, 1923–1932. https://doi.org/10.1128/JCM.42.5.1923-1932.2004

McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y.M., Buso, N., Cowley, A.P., Lopez, R., 2013. Analysis Tool Web Services from the EMBL-EBI. Nucleic Acids Res. 41, W597–W600. https://doi.org/10.1093/nar/gkt376

Medina, J.S., Prosmiti, R., Villarreal, P., Delgado-Barrio, G., Winter, G., González, B., Alemán, J. V., Collado, C., 2011. Molecular dynamics simulations of rigid and flexible water models: Temperature dependence of viscosity. Chem. Phys. 388, 9–18. https://doi.org/10.1016/j.chemphys.2011.07.001

Minamino, T., 2014. Protein export through the bacterial flagellar type III export pathway. Biochim. Biophys. Acta - Mol. Cell Res. 1843, 1642–1648. https://doi.org/10.1016/j.bbamcr.2013.09.005

Minamino, T., Imada, K., Namba, K., 2008. Molecular motors of the bacterial flagella. Curr. Opin. Struct. Biol. 18, 693–701. https://doi.org/10.1016/j.sbi.2008.09.006

Minamino, T., Kinoshita, M., Namba, K., 2019. Directional Switching Mechanism of the Bacterial Flagellar Motor. Comput. Struct. Biotechnol. J. 17, 1075–1081. https://doi.org/10.1016/j.csbj.2019.07.020

Minamino, T., Macnab, R.M., 2000. FliH, a soluble component of the type III flagellar export apparatus of Salmonella, forms a complex with FliI and inhibits its ATPase activity. Mol. Microbiol. 37, 1494–1503. https://doi.org/10.1046/j.1365-2958.2000.02106.x

Minamino, T., Nakamura, S., Imada, K., Namba, K., Morimoto, Y. V., Kinoshita, M., 2011. Structural Insight into the Rotational Switching Mechanism of the Bacterial Flagellar Motor. PLoS Biol. 9, e1000616. https://doi.org/10.1371/journal.pbio.1000616

Minamino, T., Yoshimura, S.D.J., Morimoto, Y. V., González-Pedrajo, B., Kami-Ike, N., Namba, K., 2009. Roles of the extreme N-terminal region of FliH for efficient localization of the FliH-FliI complex to the bacterial flagellar type III export apparatus. Mol. Microbiol. 74, 1471–1483. https://doi.org/10.1111/j.1365-2958.2009.06946.x

Mirel, D.B., Chamberlin, M.J., 1989. The Bacillus subtilis flagellin gene (hag) is transcribed by the σ28 form of RNA polymerase. J. Bacteriol. 171, 3095–3101. https://doi.org/10.1128/jb.171.6.3095-3101.1989

Mohammad, N., Karsabet, M.T., Amani, J., Ardjmand, A., Zadeh, M.R., Gholi, M.K., Saffari, M., Ghasemi, A., 2016. Design of a Chimeric Protein Containing Antigenic Fragments of ; A Bioinformatic Approach. Open Microbiol. J. 10, 97–112. https://doi.org/10.2174/1874285801610010097

Monticelli, L., Kandasamy, S.K., Periole, X., Larson, R.G., Tieleman, D.P., Marrink, S.J., 2008. The MARTINI coarse-grained force field: Extension to proteins. J. Chem. Theory Comput. 4, 819–834. https://doi.org/10.1021/ct700324x

Morgan, D.G., Owen, C., Melanson, L.A., Derosier, D.J., 1995. Structure Of Bacterial Flagellar Filaments At 11 Angstrom Resolution - Packing Of The Alpha-Helices. J. Mol. Biol. 249, 88–110. https://doi.org/https://doi.org/10.1006/jmbi.1995.0282

Morimoto, Y. V., Kami-ike, N., Minamino, T., Che, Y.-S., Hiraoka, K.D., Ito, M., Bai, F., Namba, K., 2014. Assembly and stoichiometry of FliF and FlhA in Salmonella flagellar basal body. Mol. Microbiol. 91, 1214–1226. https://doi.org/10.1111/mmi.12529

Mukherjee, S., Bhattacharyya, D., 2013. Influence of divalent magnesium ion on DNA: Molecular dynamics simulation studies. J. Biomol. Struct. Dyn. 31, 896–912. https://doi.org/10.1080/07391102.2012.713780

Muller, A., Beeby, M., McDowall, A.W., Chow, J., Jensen, G.J., Clemons Jr., W.M., 2014. Ultrastructure and complex polar architecture of the human pathogen Campylobacter jejuni. Microbiologyopen 3, 702–710. https://doi.org/10.1002/mbo3.200

Murray, T.S., Kazmierczak, B.I., 2006. FlhF Is required for swimming and swarming

in Pseudomonas aeruginosa. J. Bacteriol. 188, 6995–7004. https://doi.org/10.1128/JB.00790-06

Niehus, E., Gressmann, H., Ye, F., Schlapbach, R., Dehio, M., Dehio, C., Stack, A., Meyer, T.F., Suerbaum, S., Josenhans, C., 2004. Genome-wide analysis of transcriptional hierarchy and feedback regulation in the flagellar system of Helicobacter pylori. Mol. Microbiol. 52, 947–961. https://doi.org/10.1111/j.1365-2958.2004.04006.x

Nishikino, T., Hijikata, A., Miyanoiri, Y., Onoue, Y., Kojima, S., Shirai, T., Homma, M., 2018. Rotational direction of flagellar motor from the conformation of FliG middle domain in marine Vibrio. Sci. Rep. 8, 17793. https://doi.org/10.1038/s41598-018-35902-6

Noreen, F., Gina, S., Dennis, T., David, D., 1994. Isolation, Characterization and Structure of Bacterial Flagellar Motors Containing the Switch Complex. J. Mol. Biol. 235, 1261–1270. https://doi.org/https://doi.org/10.1006/jmbi.1994.1079

O'Rourke, J., Bode, G., 2001. Morphology and Ultrastructure, in: Mobley, H.L.T., Mendz, G.L., Hazell, S.L. (Eds.), Helicobacter Pylori: Physiology and Genetics. ASM Press, Washington, D.C., pp. 53–67. https://doi.org/10.1128/9781555818005.ch6

Ogawa, R., Homma, M., Kojima, S., Abe-Yoshizumi, R., Kishi, T., 2014. Interaction of the C-Terminal Tail of FliF with FliG from the Na + -Driven Flagellar Motor of Vibrio alginolyticus. J. Bacteriol. 197, 63–72. https://doi.org/10.1128/jb.02271-14

Parkhill, J., Wren, B.W., Mungall, K., Ketley, J.M., Churcher, C., Basham, D., Chillingworth, T., Davies, R.M., Feltwell, T., Holroyd, S., Jagels, K., Karlyshev, A. V, Moule, S., Pallen, M.J., Penn, C.W., Quail, M.A., Rajandream, M.A., Rutherford, K.M., van Vliet, A.H., Whitehead, S., Barrell, B.G., 2000. The genome sequence of the food-borne pathogen Campylobacter jejuni reveals hypervariable sequences. Nature 403, 665–8. https://doi.org/10.1038/35001088

Parkin, D.M., Bray, F., Ferlay, J., Pisani, P., 2005. Global Cancer Statistics, 2002. CA. Cancer J. Clin. 55, 74–108. https://doi.org/10.3322/canjclin.55.2.74

Parrinello, M., Rahman, A., 1981. Polymorphic transitions in single crystals: A new molecular dynamics method. J. Appl. Phys. 52, 7182–7190. https://doi.org/10.1063/1.328693

Pascoe, B., Williams, L.K., Calland, J.K., Meric, G., Hitchings, M.D., Dyer, M., Ryder, J., Shaw, S., Lopes, B.S., Chintoan-Uta, C., Allan, E., Vidal, A., Fearnley, C., Everest, P., Pachebat, J.A., Cogan, T.A., Stevens, M.P., Humphrey, T.J., Wilkinson, T.S., Cody, A.J., Colles, F.M., Jolley, K.A., Maiden, M.C.J., Strachan, N., Pearson, B.M., Linton, D., Wren, B.W., Parkhill, J., Kelly, D.J., van Vliet, A.H.M., Forbes, K.J., Sheppard, S.K., 2019. Domestication of campylobacter jejuni NCTC 11168. Microb. Genomics 5, e000279. https://doi.org/10.1099/mgen.0.000279

Patodia, S., 2014. Molecular Dynamics Simulation of Proteins: A Brief Overview. J. Phys. Chem. Biophys. 4, 1000166. https://doi.org/10.4172/2161-0398.1000166

Paul, K., Brunstetter, D., Titen, S., Blair, D.F., 2011a. A molecular mechanism of direction switching in the flagellar motor of Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. 108, 17171–17176. https://doi.org/10.1073/pnas.1110111108

Paul, K., Gonzalez-Bonet, G., Bilwes, A.M., Crane, B.R., Blair, D., 2011b. Architecture of the flagellar rotor. EMBO J. 30, 2962–2971. https://doi.org/10.1038/emboj.2011.188

Payão, S.L.M., Rasmussen, L.T., 2016. Helicobacter pylori and its reservoirs: A correlation with the gastric infection. World J. Gastrointest. Pharmacol. Ther. 7, 126–132. https://doi.org/10.4292/wjgpt.v7.i1.126

Peek, R.M., Crabtree, J.E., 2006. Helicobacter infection and gastric neoplasia. J. Pathol. 208, 233–248. https://doi.org/10.1002/path.1868

Perilla, J.R., Goh, B.C., Cassidy, C.K., Liu, B., Bernardi, R.C., Rudack, T., Yu, H., Wu, Z., Schulten, K., 2015. Molecular dynamics simulations of large macromolecular complexes. Curr. Opin. Struct. Biol. 31, 64–74. https://doi.org/10.1016/j.sbi.2015.03.007

Periole, X., Marrink, S.-J., 2013. The Martini Coarse-Grained Force Field, in: Monticelli, L., Salonen, E. (Eds.), Biomolecular Simulations: Methods and

Protocols. Humana Press, Totowa, NJ, pp. 533–563. https://doi.org/https://doi.org/10.1007/978-1-62703-017-5_20

Perry, S., Sanchez, M.D.L.L., Yang, S., Haggerty, T.D., Hurst, P., Perez-Perez, G., Parsonnet, J., 2006. Gastroenteritis and transmission of Helicobacter pylori infection in households. Emerg. Infect. Dis. 12, 1701–1708. https://doi.org/10.3201/eid1211.060086

Petrov, D., Margreitter, C., Grandits, M., Oostenbrink, C., Zagrovic, B., 2013. A Systematic Framework for Molecular Dynamics Simulations of Protein Post-Translational Modifications. PLoS Comput. Biol. 9, e1003154. https://doi.org/10.1371/journal.pcbi.1003154

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E., 2004. UCSF Chimera - A visualization system for exploratory research and analysis. J. Comput. Chem. 25, 1605–1612. https://doi.org/10.1002/jcc.20084

Platts-Mills, J.A., Kosek, M., 2014. Update on the burden of Campylobacter in developing countries. Curr Opin Infect Dis 27, 444–450. https://doi.org/10.1002/cncr.27633.Percutaneous

Public Health England, 2018. Campylobacer data 2007 to 2016 : National laboratory data for residents of England and Wales.

Qi, Y., Lee, J., Singharoy, A., McGreevy, R., Schulten, K., Im, W., 2017. CHARMM-GUI MDFF/xMDFF Utilizer for Molecular Dynamics Flexible Fitting Simulations in Various Environments. J. Phys. Chem. B 121, 3718–3723. https://doi.org/10.1021/acs.jpcb.6b10568

Qin, Z., Lin, W. ting, Zhu, S., Franco, A.T., Liu, J., 2017. Imaging the motility and chemotaxis machineries in Helicobacter pylori by cryo-electron tomography. J. Bacteriol. 199, e00695-16. https://doi.org/10.1128/JB.00695-16

Ramos Sasselli, I., Ulijn, R. V., Tuttle, T., 2016. CHARMM force field parameterization protocol for self-assembling peptide amphiphiles: The Fmoc moiety. Phys. Chem. Chem. Phys. 18, 4659–4667. https://doi.org/10.1039/c5cp06770g

Razvi, A., Scholtz, J.M., 2006. Lessons in stability from thermophilic proteins. Protein Sci. 15, 1569–1578. https://doi.org/10.1110/ps.062130306

Reid, S.D., Selander, R.K., Whittam, T.S., 1999. Sequence diversity of flagellin (fliC) alleles in pathogenic Escherichia coli. J. Bacteriol. 181, 153–160. https://doi.org/10.1128/jb.181.1.153-160.1999

Ren, F., Lei, T., Song, Z., Yu, T., Li, Q., Huang, J., 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. https://doi.org/10.1016/j.micres.2017.12.006

Renault, T.T., Abraham, A.O., Bergmiller, T., Paradis, G., Rainville, S., Charpentier, E., Guet, C.C., Tu, Y., Namba, K., Keener, J.P., Minamino, T., Erhardt, M., 2017. Bacterial flagella grow through an injection-diffusion mechanism. Elife 6, 1–22. https://doi.org/10.7554/eLife.23136

Reuter, M., van Vliet, A.H.M., 2013. Signal Balancing by the CetABC and CetZ Chemoreceptors Controls Energy Taxis in Campylobacter jejuni. PLoS One 8, 1–10. https://doi.org/10.1371/journal.pone.0054390

Rohl, C.A., Strauss, C.E.M., Misura, K.M.S., Baker, D., 2004. Protein Structure Prediction Using Rosetta. Methods Enzymol. 383, 66–93. https://doi.org/10.1016/S0076-6879(04)83004-0

Rohou, A., Grigorieff, N., 2015. CTFFIND4: Fast and accurate defocus estimation from electron micrographs. J. Struct. Biol. 192, 216–221. https://doi.org/10.1016/j.jsb.2015.08.008

Rosenberg, E., 2014. The Prokaryotes: Deltaproteobacteria and Epsilonbacteria, 4th ed. Springer Berlin, Heidelberg. https://doi.org/https://doi.org/10.1007/978-3-642-39044-9

Rütschlin, S., Böttcher, T., 2020. Inhibitors of Bacterial Swarming Behavior. Chem. - A Eur. J. 26, 964–979. https://doi.org/10.1002/chem.201901961

Sakai, T., Miyata, T., Terahara, N., Mori, K., Inoue, Y., Morimoto, Y. V., Kato, T., Namba, K., Minamino, T., 2019. Novel Insights into Conformational Rearrangements of the Bacterial Flagellar Switch Complex. MBio 10, 1–14.

https://doi.org/10.1128/mBio.00079-19

Sali, A; Blundell, T.L., 1993. Comparative protein modelling by satisfaction of spatial restraints. J. Mol. Biol. 234, 779–815. https://doi.org/https://doi.org/10.1006/jmbi.1993.1626

Salomon-Ferrer, R., Case, D.A., Walker, R.C., 2013. An overview of the Amber biomolecular simulation package. Wiley Interdiscip. Rev. Comput. Mol. Sci. 3, 198–210. https://doi.org/10.1002/wcms.1121

Sampriti, M., Kearns, D.B., 2016. The structure and regulation of flagella in Bacillus subtilis. Annu Rev Genet. 48, 319–340. https://doi.org/10.1146/annurev-genet-120213-092406.The

Santiveri, M., Roa-Eguiara, A., Kühne, C., Wadhwa, N., Hu, H., Berg, H.C., Erhardt, M., Taylor, N.M.I., 2020. Structure and Function of Stator Units of the Bacterial Flagellar Motor. Cell 183, 244-257.e16. https://doi.org/10.1016/j.cell.2020.08.016

Sardis, M.F., Economou, A., 2010. SecA: A tale of two protomers: MicroReview. Mol. Microbiol. 76, 1070–1081. https://doi.org/10.1111/j.1365-2958.2010.07176.x

Sarkar, M.K., Paul, K., Blair, D., 2010. Chemotaxis signaling protein CheY binds to the rotor protein FliN to control the direction of flagellar rotation in Escherichia coli. Proc. Natl. Acad. Sci. U. S. A. 107, 9370–9375. https://doi.org/10.1073/pnas.1000935107

Scallan, E., Hoekstra, R.M., Angulo, F.J., Tauxe, R. V., Widdowson, M.A., Roy, S.L., Jones, J.L., Griffin, P.M., 2011. Foodborne illness acquired in the United States-Major pathogens. Emerg. Infect. Dis. 17, 7–15. https://doi.org/10.3201/eid1701.P11101

Schuhmacher, J.S., Thormann, K.M., Bange, G., 2015. How bacteria maintain location and number of flagella? FEMS Microbiol. Rev. 39, 812–822. https://doi.org/10.1093/femsre/fuv034

Schwede, T., Kopp, J., Guex, N., Peitsch, M.C., 2003. SWISS-MODEL: An automated protein homology-modeling server. Nucleic Acids Res. 31, 3381–

3385. https://doi.org/10.1093/nar/gkg520

Sharma, P., Vakil, N., 2003. Review article: Helicobacter pylori and reflux disease. Aliment. Pharmacol. Ther. Suppl. 17, 297–305. https://doi.org/10.1111/j.1365-2036.2005.02607.x

Shaw, F.L., Mulholland, F., Le Gall, G., Porcelli, I., Hart, D.J., Pearson Bruce, M.M., van Vliet, A.H.M., 2012. Selenium-dependent biogenesis of formate dehydrogenase in Campylobacter jejuni is controlled by the fdhTU accessory genes. J. Bacteriol. 194, 3814–3823. https://doi.org/10.1128/JB.06586-11

Shelley, J.C., Shelley, M.Y., Reeder, R.C., Bandyopadhyay, S., Moore, P.B., Klein, M.L., 2001. Simulations of phospholipids using a coarse grain model. J. Phys. Chem. B 105, 9785–9792. https://doi.org/10.1021/jp011637n

Shen, M., Sali, A., 2006. Statistical potential for assessment and prediction of protein structures. Protein Sci. 15, 2507–2524. https://doi.org/10.1110/ps.062416606

Sircar, R., Borbat, P.P., Lynch, M.J., Bhatnagar, J., Beyersdorf, M.S., Halkides, C.J., Freed, J.H., Crane, B.R., 2015. Assembly states of FliM and FliG within the flagellar switch complex. J. Mol. Biol. 427, 867–886. https://doi.org/10.1016/j.jmb.2014.12.009

Slonczewski, J.L., Fujisawa, M., Dopson, M., Krulwich, T.A., 2009. Cytoplasmic pH Measurement and Homeostasis in Bacteria and Archaea, Advances in Microbial Physiology. https://doi.org/10.1016/S0065-2911(09)05501-5

Søndergaard, C.R., Garrett, A.E., Carstensen, T., Pollastri, G., Nielsen, J.E., 2009. Structural artifacts in protein-ligand X-ray structures: Implications for the development of docking scoring functions. J. Med. Chem. 52, 5673–5684. https://doi.org/10.1021/jm8016464

Song, W.S., Cho, S.Y., Hong, H.J., Park, S.C., Yoon, S. il, 2017. Self-Oligomerizing Structure of the Flagellar Cap Protein FliD and Its Implication in Filament Assembly. J. Mol. Biol. 429, 847–857. https://doi.org/10.1016/j.jmb.2017.02.001

Spohn, G., Scarlato, V., 1999. Motility of Helicobacter priori is coordinately regulated by the transcriptional activator FlgR, an NtrC homolog. J. Bacteriol. 181, 593–599. https://doi.org/10.1128/jb.181.2.593-599.1999

Steinbach, P.J., Brooks, B.R., 1994. New spherical-cutoff methods for long-range forces in macromolecular simulation. J. Comput. Chem. 15, 667–683. https://doi.org/10.1002/jcc.540150702

Stoakes, E.A., 2017. Campylobacter jejuni : Deciphering the role of FlhF in flagellar biogenesis. University of Warwick.

Stock, D., Namba, K., Lee, L.K., 2012. Nanorotors and self-assembling macromolecular machines: The torque ring of the bacterial flagellar motor. Curr. Opin. Biotechnol. 23, 545–554. https://doi.org/10.1016/j.copbio.2012.01.008

Suerbaum, S., Josenhans, C., Labigne, A., 1993. Cloning and genetic characterization of the Helicobacter pylori and Helicobacter mustelae flaB flagellin genes and construction of H. pylori flaA- and flaB-negative mutants by electroporation-mediated allelic exchange. J. Bacteriol. 175, 3278–3288. https://doi.org/10.1128/jb.175.11.3278-3288.1993

Suzuki, H., Marshall, B.J., Hibi, T., 2006. Overview: Helicobacter pylori and extragastric disease. Int. J. Hematol. 84, 291–300. https://doi.org/10.1532/IJH97.06180

Szymanski, C.M., Logan, S.M., Linton, D., Wren, B.W., 2003. Campylobacter - A tale of two protein glycosylation systems. Trends Microbiol. 11, 233–238. https://doi.org/10.1016/S0966-842X(03)00079-9

Takada, S., 2012. Coarse-grained molecular simulations of large biomolecules. Curr. Opin. Struct. Biol. 22, 130–137. https://doi.org/10.1016/j.sbi.2012.01.010

Takekawa, N., Kawamoto, A., Sakuma, M., Kato, T., Minamino, T., Namba, K., Homma, M., Imada, K., Kojima, S., Kinoshita, M., 2021. Two distinct conformations in 34 flif subunits generate three different symmetries within the flagellar ms-ring. MBio 12, 1–13. https://doi.org/10.1128/mBio.03199-20

Takekawa, N., Nishiyama, M., Kaneseki, T., Kanai, T., Atomi, H., Kojima, S., Homma, M., 2015. Sodium-driven energy conversion for flagellar rotation of the earliest divergent hyperthermophilic bacterium. Sci. Rep. 5, 1–9. https://doi.org/10.1038/srep12711

Terashima, H., Kojima, S., Homma, M., 2008. Flagellar Motility in Bacteria. Structure

and Function of Flagellar Motor, in: International Review of Cell and Molecular Biology. Elsevier Inc., pp. 39–85. https://doi.org/10.1016/S1937-6448(08)01402-0

Thomas, D., Morgan, D.G., DeRosier, D.J., 2001. Structures of bacterial flagellar motors from two FliF-FliG gene fusion mutants. J. Bacteriol. 183, 6404–6412. https://doi.org/10.1128/JB.183.21.6404-6412.2001

Thomas, D.R., Francis, N.R., Xu, C., DeRosier, D.J., 2006. The three-dimensional structure of the flagellar rotor from a clockwise-locked mutant of Salmonella enterica serovar typhimurium. J. Bacteriol. 188, 7039–7048. https://doi.org/10.1128/JB.00552-06

Thomas, D.R., Morgan, D.G., DeRosier, D.J., 1999. Rotational symmetry of the C ring and a mechanism for the flagellar rotary motor. Proc. Natl. Acad. Sci. U. S. A. 96, 10134–10139. https://doi.org/10.1073/pnas.96.18.10134

Thomas, M.T., Shepherd, M., Poole, R.K., Van Vliet, A.H.M., Kelly, D.J., Pearson, B.M., 2011. Two respiratory enzyme systems in Campylobacter jejuni NCTC 11168 contribute to growth on l-lactate. Environ. Microbiol. 13, 48–61. https://doi.org/10.1111/j.1462-2920.2010.02307.x

Thompson, A.P., Aktulga, H.M., Berger, R., Bolintineanu, D.S., Brown, W.M., Crozier, P.S., in 't Veld, P.J., Kohlmeyer, A., Moore, S.G., Nguyen, T.D., Shan, R., Stevens, M.J., Tranchida, J., Trott, C., Plimpton, S.J., 2022. LAMMPS - a flexible simulation tool for particle-based materials modeling at the atomic, meso, and continuum scales. Comput. Phys. Commun. 271, 108171. https://doi.org/10.1016/j.cpc.2021.108171

Togashi, F., Yamaguchi, S., Kihara, M., Aizawa, S.I., Macnab, R.M., 1997. An extreme clockwise switch bias mutation in fliG of Salmonella typhimurium and its suppression by slow-motile mutations in motA and motB. J. Bacteriol. 179, 2994–3003. https://doi.org/10.1128/jb.179.9.2994-3003.1997

Trabuco, L.G., Villa, E., Schreiner, E., Harrison, C.B., Schulten, K., 2009. Molecular dynamics flexible fitting: A practical guide to combine cryo-electron microscopy and X-ray crystallography. Methods 49, 174–180. https://doi.org/10.1016/j.ymeth.2009.04.005

Tsang, J., Hirano, T., Hoover, T.R., McMurry, J.L., 2015. Helicobacter pylori FlhA binds the sensor kinase and flagellar gene regulatory protein FlgS with high affinity. J. Bacteriol. 197, 1886–1892. https://doi.org/10.1128/JB.02610-14

Tsang, J., Hoover, T.R., 2015. Basal body structures differentially affect transcription of RpoNand FliA-dependent flagellar genes in Helicobacter pylori. J. Bacteriol. 197, 1921–1930. https://doi.org/10.1128/JB.02533-14

Tupiņa, D., Krah, A., Marzinek, J.K., Zuzic, L., Moverley, A.A., Constantinidou, C., Bond, P.J., 2022. Bridging the N-terminal and middle domains in FliG of the flagellar rotor. Curr. Res. Struct. Biol. 4, 59–67. https://doi.org/10.1016/j.crstbi.2022.02.002

Ueno, T., Oosawa, K., Aizawa, S.I., 1992. M ring, S ring and proximal rod of the flagellar basal body of Salmonella typhimurium are composed of subunits of a single protein, FliF. J. Mol. Biol. 227, 672–677. https://doi.org/10.1016/0022-2836(92)90216-7

van Gunsteren, W.F., Bakowies, D., Baron, R., Chandrasekhar, I., Christen, M., Daura, X., Gee, P., Geerke, D.P., Glättli, A., Hünenberger, P.H., Kastenholz, M.A., Oostenbrink, C., Schenk, M., Trzesniak, D., Van Der Vegt, N.F.A., Yu, H.B., 2006. Biomolecular modeling: Goals, problems, perspectives. Angew. Chemie - Int. Ed. 45, 4064–4092. https://doi.org/10.1002/anie.200502655

van Gunsteren, W.F., Daura, X., Hansen, N., Mark, A.E., Oostenbrink, C., Riniker, S., Smith, L.J., 2018. Validation of Molecular Simulation: An Overview of Issues. Angew. Chemie - Int. Ed. 57, 884–902. https://doi.org/10.1002/anie.201702945

Vartanian, A.S., Pazs, A., Fortgang, E.A., Abramsons, J., Dahlquist, F.W., 2012. Structure of flagellar motor proteins in complex allows for insights into motor structure and switching. J. Biol. Chem. 287, 35779–35783. https://doi.org/10.1074/jbc.C112.378380

Vouga, M., Greub, G., 2016. Emerging bacterial pathogens: The past and beyond. Clin. Microbiol. Infect. 22, 12–21. https://doi.org/10.1016/j.cmi.2015.10.010

Wadhwa, N., Tu, Y., Berg, H.C., 2021. Mechanosensitive remodeling of the bacterial flagellar motor is independent of direction of rotation. Proc. Natl. Acad. Sci. 118,

e2024608118. https://doi.org/10.1073/pnas.2024608118

Wassenaar, T.M., Bleumink-Pluym, N.M.C., Newell, D.G., Nuijten, P.J.M., Van der Zeijst, B.A.M., 1994. Differential flagellin expression in a flaA flaB+ mutant of Campylobacter jejuni. Infect. Immun. 62, 3901–3906. https://doi.org/10.1128/iai.62.9.3901-3906.1994

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., De Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T., 2018. SWISS-MODEL: Homology modelling of protein structures and complexes. Nucleic Acids Res. 46, W296–W303. https://doi.org/10.1093/nar/gky427

Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M., Barton, G.J., 2009. Jalview Version 2-A multiple sequence alignment editor and analysis workbench. Bioinformatics 25, 1189–1191. https://doi.org/10.1093/bioinformatics/btp033

Webb, B., Sali, A., 2016. Comparative protein structure modeling using MODELLER. Curr. Protoc. Bioinforma. 15, 5.6.1-5.6.30. https://doi.org/10.1002/cpbi.3

Welch, M., Oosawa, K., Aizawa, S.I., Eisenbach, M., 1993. Phosphorylation-dependent binding of a signal molecule to the flagellar switch of bacteria. Proc. Natl. Acad. Sci. U. S. A. 90, 8787–8791. https://doi.org/10.1073/pnas.90.19.8787

West, B., Brown, F.L.H., Schmid, F., 2009. Membrane-protein interactions in a generic coarse-grained model for lipid bilayers. Biophys. J. 96, 101–115. https://doi.org/10.1529/biophysj.108.138677

Wösten, M.M.S.M., van de Lest, C.H.A., van Dijk, L., van Putten, J.P.M., 2017. Function and regulation of the C4-dicarboxylate transporters in campylobacter Jejuni. Front. Microbiol. 8, 1–13. https://doi.org/10.3389/fmicb.2017.00174

Wösten, M.M.S.M., van Mourik, A., van Putten, J.P.M., 2008. Regulation of Genes in Campylobacter jejuni, in: Nachamkin, I., Szymanski, C.M., M.J.Blaser (Eds.), Campylobacter. ASM Press, pp. 611–624. https://doi.org/10.1128/9781555815554.ch34

Wösten, M.M.S.M., Wagenaar, J.A., van Putten, J.P.M., 2004. The FlgS/FlgR Two-component Signal Transduction System Regulates the fla Regulon in

Campylobacter jejuni. J. Biol. Chem. 279, 16214–16222. https://doi.org/10.1074/jbc.M400357200

Wroblewski, L.E., Peek, R.M., Wilson, K.T., 2010. Helicobacter pylori and gastric cancer: Factors that modulate disease risk. Clin. Microbiol. Rev. 23, 713–739. https://doi.org/10.1128/CMR.00011-10

Xue, C., Lam, K.H., Zhang, H., Sun, K., Lee, S.H., Chen, X., Au, S.W.N., 2018. Crystal structure of the FliF-FliG complex from Helicobacter pylori yields insight into the assembly of the motor MS-C ring in the bacterial flagellum. J Biol Chem 293, 2066–2078. https://doi.org/10.1074/jbc.M117.797936

Young, H.S., Dang, H., Lai, Y., DeRosier, D.J., Khan, S., 2003. Variable symmetry in Salmonella typhimurium flagellar motors. Biophys. J. 84, 571–577. https://doi.org/10.1016/S0006-3495(03)74877-2

Zhang, C., Raugei, S., Eisenberg, B., Carloni, P., 2010. Molecular dynamics in physiological solutions: Force fields, alkali metal ions, and ionic strength. J. Chem. Theory Comput. 6, 2167–2175. https://doi.org/10.1021/ct9006579

Zhou, J., Lloyd, S.A., Blair, D.F., 1998. Electrostatic interactions between rotor and stator in the bacterial flagellar motor. Proc. Natl. Acad. Sci. U. S. A. 95, 6436–6441. https://doi.org/10.1073/pnas.95.11.6436

Zhou, X., Roujeinikova, A., 2021. The Structure, Composition, and Role of Periplasmic Stator Scaffolds in Polar Bacterial Flagellar Motors. Front. Microbiol. 12, 639490. https://doi.org/10.3389/fmicb.2021.639490

Zwe, Y.H., Yuk, H.G., 2017. Food quality and safety in Singapore: Microbiology aspects. Food Qual. Saf. 1, 101–105. https://doi.org/10.1093/fqsafe/fyx016

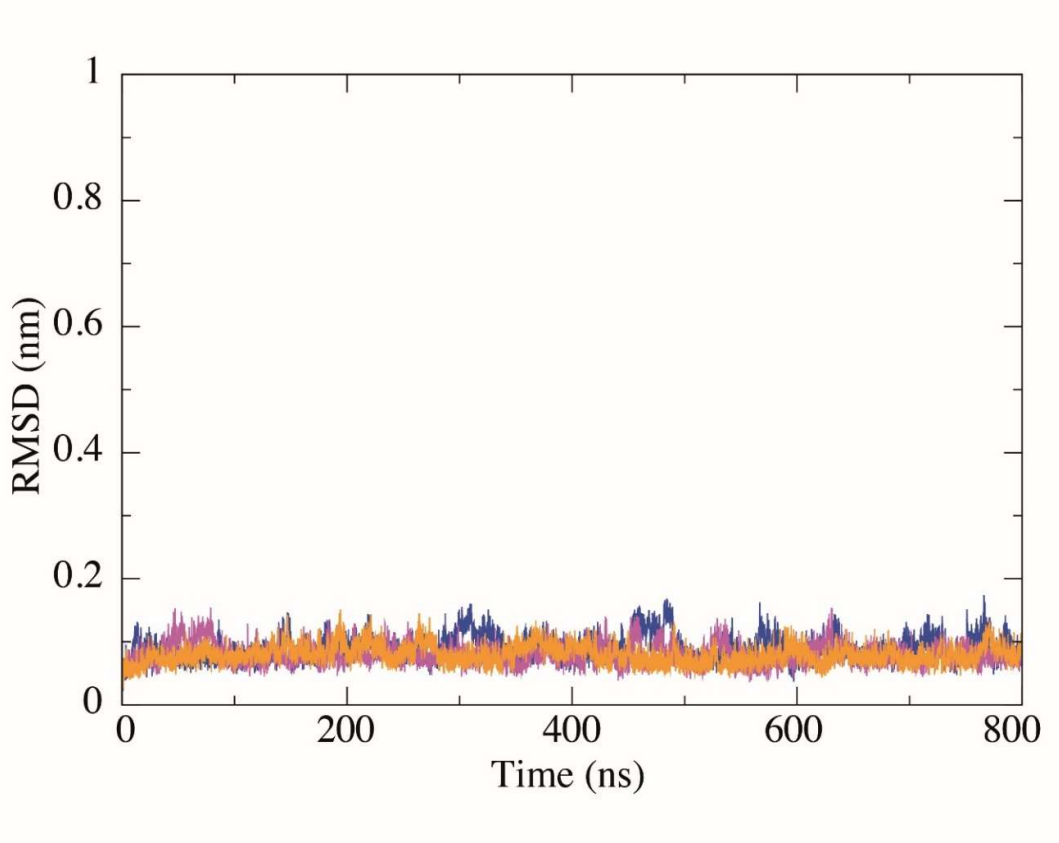



**Figure A. Core stability of 5WUJ complex in solution.** Backbone RMSD after fitting backbone of protein core (helix 2 to 5) against core (helix 2 to 5) in FliG-N/FliF-C system. Different colours represent three independent runs, system used CHARMM36m forcefield.

APPENDIX B

Rawtlt values of WT tilt series:

50.99
47.99
44.99
41.99
38.99
35.99
32.99
29.99
26.99
23.99
20.99
17.99
14.99
11.99
8.99
5.99
2.99
-0.03
-2.99
-5.99
-8.99
-11.99
-14.99
-17.99
-20.99
-23.99
-26.99
-29.99
-32.99
-35.99
-38.99
-41.99
-44.99
-47.99
-50.99

APPENDIX C

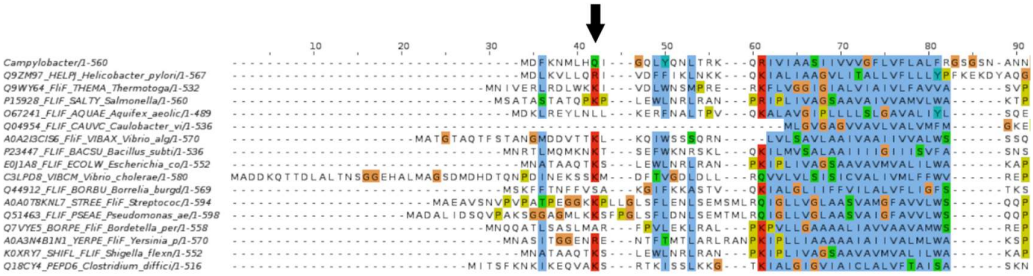



**Figure C. FliF sequence alignment with various FliF sequences.** Species that are not typically flagellated are also included unlike the alignment presented in the main text.

APPENDIX D

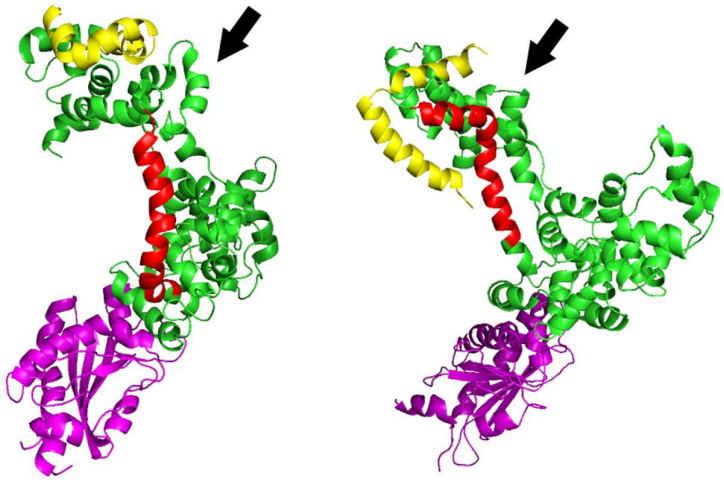


***Figure D1*. MDFF outcomes of simulating *H. pylori* complexes with C2 conformation of C-terminus.** FliF-C/FliG/FliM-M complex after MDFF in CW (left) and CCW (right) densities. Arrows point to the extensive contacts between FliG-C and FliG-N that served as basis for rejecting these models in final selection.

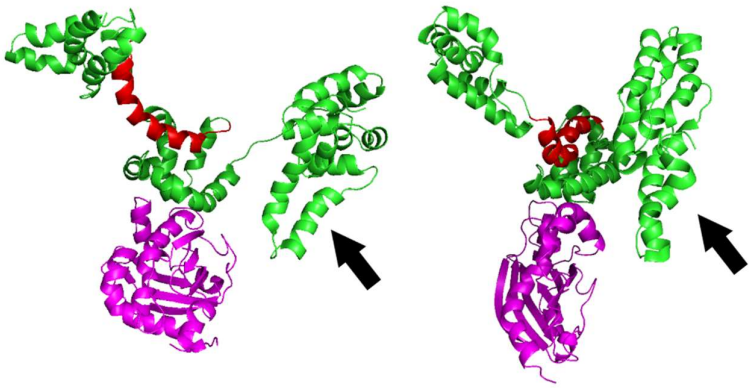


***Figure D2*. MDFF outcomes of simulating *A. aeolicus* complexes.** FliF-C/FliG/FliM-M complex after MDFF in CCW (left) and CW (right) densities. Arrow points to severe distortion and flattening of FliG-C (more apparent in trimer setting) that served as basis for rejecting these models in final selection.

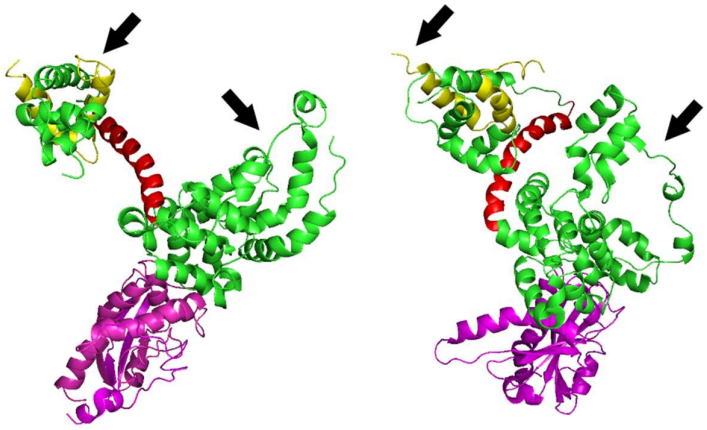


***Figure D3.*** **MDFF outcomes of simulating *H. pylori* complexes with 5TDY$_{eln}$ based N-terminus and C1 conformation of C-terminus.** FliF-C/FliG/FliM-M complex after MDFF in CCW (left) and CW (right) densities. Arrows point to unravelling of FliF-C, FliG-N and FliG-C that served as basis for rejecting these models in final selection.

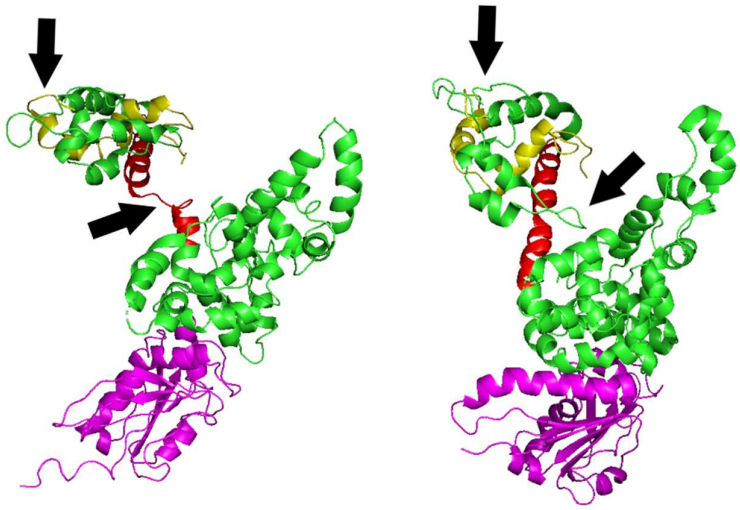


***Figure D4.*** **MDFF outcomes of simulating *H. pylori* complexes with 5TDY$_{eln}$ based N-terminus and C2 conformation of C-terminus.** FliF-C/FliG/FliM-M complex after MDFF in CCW (left) and CW (right) densities. Arrows point to unravelling of FliF-C, FliG-N and FliG-C, NM helix, and contacts between unravelled FliG-N loop and NM-helix that served as basis for rejecting these models in final selection.

APPENDIX E

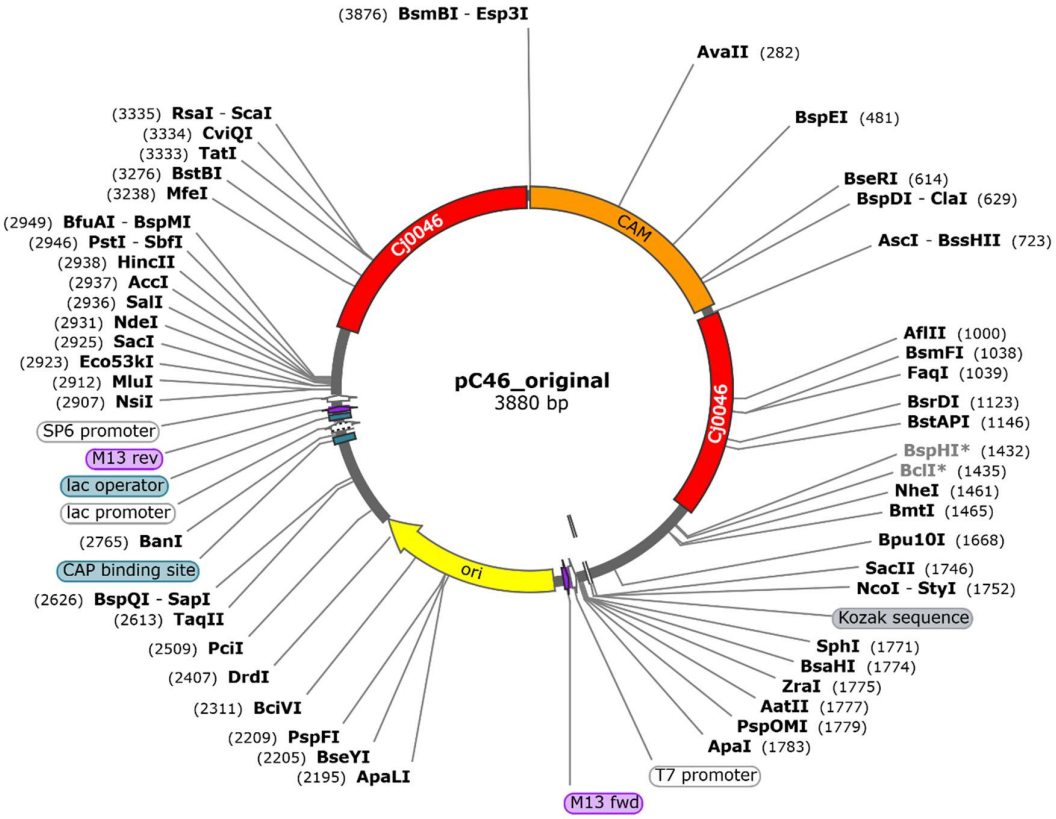



**Figure E. Plasmid map of pC46 plasmid.** This plasmid was used as basis for pC46-fdxA constructed by Emily Stoakes (Table 2.4) subsequently used to construct FliF containing variants for complementation (Table 2.4, Figure 5.2). Figure created in SnapGene Viewer (www.snapgene.com).