DNA-Templated Assembly of the Bacterial Flagellar Motor's Cytoplasmic Ring



A thesis submitted for the degree of Doctor of Philosophy

Joel Spratt

Brasenose College

University of Oxford

Trinity Term 2018

ABSTRACT

The bacterial flagellar motor is one of the most complex protein machines found in nature and how it self-assembles and produces force are very much open questions. In this thesis, I study the constituent parts of the motor *in vitro*, with the focus being primarily on FliG, a protein which helps to form the motor's cytoplasmic ring. The experimental approach employs a DNA scaffold to direct the *in vitro* formation of the FliG ring, with the aim of determining both the number of proteins in the complex and the means by which they associate to form a circular structure. One leading theory states that the FliG proteins interact through a process of domain-swap oligomerization and the primary goal of this work is to search for evidence of this process with the aid of *in vitro* DNA templates.

To attach the FliG protein to the underlying DNA structures, I produced FliG-DNA conjugates through two different means. First, I bound both tris- and pentakis-NTA-modified DNA strands to the Histidine-tag of FliG and showed that these conjugates could be arranged on simple, linear DNA templates. In order for the proteins to be arranged with the spacing observed in the cytoplasmic ring however, a different conjugation strategy was required in which a maleimide-modified DNA strand was reacted with a single-cysteine FliG mutant. After issues relating to Histidine-tag-mediated FliG dimerization were resolved, these conjugates were able to be organized on both linear DNA strands and origami tiles with the desired spacing between the proteins. No interactions between either template-bound or free-floating FliG proteins could be observed however, meaning a new strategy was needed.

One model of cytoplasmic ring assembly predicts that binding of FliG to the C-terminal domain of the membrane-bound FliF protein induces a conformational change in both partners which then triggers interactions between neighbouring FliG proteins. To test this, I designed a peptide corresponding to the $FliF_C$ domain of *Salmonella* and *E. coli* FliF and incorporated it into my system. Several biophysical assays indicated that the $FliF_C$ peptide stably bound to purified wild-type *Salmonella* FliG and that binding to the peptide increased the overall stability of the protein. This constitutes the first confirmation that the cytoplasmic domain of FliF is sufficient for FliG binding in enteric bacteria. Following this discovery, a maleimide-modified DNA strand was then conjugated to a single-cysteine version of the peptide and the peptide-DNA conjugates organized with various stoichiometries and spacing on DNA templates to create an array of 'binding platforms' for the FliG protein. FliG bound to the organized peptides in a one-to-one fashion but interactions between the organized FliG proteins were still not detected.

In a final effort to stimulate FliG-FliG binding, a radially-symmetric DNA nanostructure was designed with dimensions and a geometry which matched those found in the cytoplasmic rings of bacteria. Gels and AFM microscopy proved that the 12 designed strands assembled correctly into the desired structure and that $FliF_C$-DNA conjugates with FliG proteins bound could be organized on the circular arrangement of binding sites around the structure's outer rim. Even in such an arrangement, interactions between neighbouring proteins could not be observed, and though higher concentrations of $FliF_C$ and FliG were tested, issues with purification prevented accurate characterization of these complexes.

Thus, though the domain-swap model for FliG oligomerization was therefore neither confirmed or denied, the use of a DNA scaffold to organize protein molecules and study their interactions holds promise as a technique to investigate the proteins of other biological systems and will no doubt be employed to great effect in the coming years.

Now it is written,

I dedicate this thesis

To you, the reader.

ACKNOWLEDGEMENTS

This thesis would not have been possible without the help and support given to me by a great many people over the years both in Oxford and back in Canada. As they say, 'it takes a village to raise a thesis' and I am extremely grateful for each and every contribution, big or small.

First, I would like to thank my supervisors, Professors Andrew Turberfield and Richard Berry for the guidance and insight they have provided during my time in their labs. The independence and freedom that I was given throughout my work was very much appreciated, as were the suggestions and direction that came my way at the crucial points. The unique blend of enthusiasm and scientific rigour that they brought to the table underwrites all the work that I present and I have them to thank most for getting me started on such an interesting and challenging project.

For providing such an amazing environment in which to work and do science, I have all my wonderful labmates to thank. Putting up with my haikus, my constant pleas for an outdoor lunch, and my tendency to stack various pieces of lab equipment on top of one other I am sure took a great deal of patience, but thanks for indulging me. To those around when I began: Flo, Ibon, Le, Natalie, Wenjing, Céline, Mike, and Robert, and to those with me at the end: Sam, Jon, Antonio, Robbie, Juan, Emma, Katie, Behnam, Seham, Robert, Helen, and Jasmine, thanks so much for making my time in the lab such a good one.

Of my labmates, I would like to thank Sam Tusk in particular for his generosity and patience in showing me the ropes when I arrived and for always taking the time to talk through problems and ideas about the project with me. This thesis benefitted immensely from all his contributions. Likewise, I am also particularly grateful to Jon Bath. He has been a constant source of sound judgement and good advice over these past four years and is a truly one-of-a-kind scientist and person. Thanks for all you have done. In addition, before crossing the pond, I had the good fortune of undertaking work in the lab of Professor Gary Brouhard who encouraged me to pursue a doctorate and gave me my first insights into the world of scientific research. I will always be grateful of the direction and inspiration he provided.

Many others helped to make this thesis possible. I have benefitted a great deal from the assistance given by our collaborators on the project, in particular the lab of Lawrence Lee in Sydney who provided protocols, constructs, and many milligrams of knowledge and expertise. The Clarendon Fund provided the financial means which allowed me to undertake this degree while the HFSP funded most of the science I present. To both of these organizations I am much obliged. Thanks must also go to the Synthetic Biology CDT for the training and support they gave me, and to Brasenose College and the Department of Physics for their assistance on many levels.

Outside of science, I have always had tremendous backing from my friends and family. Thanks to each and every one of you for your unwavering support. In particular, thanks to the McCanns for providing such a lovely lake-side setting for my writing of section 3.2 and to my housemates at 438 Marston Road for the welcome distractions they provided as I put together this thesis.

Finally, the biggest thanks of all must go to my parents who have been there with motivation and encouragement every step of the way. They have given me so much over the years but, in the context of this thesis, I am perhaps most grateful for the love of learning and sense of curiosity they have passed on.

'Learn something new every day', they say. Thanks for that mom and dad.

TABLE OF CONTENTS

CHAPTER ONE – INTRODUCTION 1

1.1 **Proteins Are the Preferred Functional Molecule of Life** **1**
Proteins Exhibit a Hierarchy of Organization 3
Proteins Use a Small Number of Techniques to Carry Out Their Function 6

1.2 **Protein Complexes Are Needed for the Cell's Most Difficult Tasks** **10**
Protein Complexes Provide Added Benefits When Compared to Monomers 10
Protein Complexes Self-Assemble into Dynamic, Often Symmetric, Structures 14

1.3 **The Bacterial Flagellar Motor is a Remarkable Molecular Machine** **19**
Bacteria Can Integrate Signals from Their Environment and Swim in a Directed Manner 19
The Bacterial Flagellar Motor is Made of Hundreds of Individual Components 23

1.4 **The Structure and Assembly Pathway of the Cytoplasmic Ring Are Still Unknowns** **29**
FliG is a Central Hub of the Bacterial Flagellar Motor 30
A Domain-Swap Polymerization Model Outlines an Assembly Pathway for FliG 38

1.5 **The Tools of DNA Nanotechnology Provide a Way to Test the FliG Domain-Swap Hypothesis** **48**
Nearly Four Decades of Work Has Produced Major Advances in DNA Nanotechnology 49
DNA Nanostructures Can Position Proteins with Nanometre Precision 52

1.6 **The Primary Goal Is to Test the Domain-Swap Hypothesis Using an *in vitro* DNA Template** **57**
An *in vitro* DNA Template Enables Many Experiments Impossible to Do Inside a Living Cell 58
The Work Presented in This Thesis Builds Sequentially from Smaller to Larger Structures 63

CHAPTER TWO –THE FIRST HURDLE 68

2.1 **NTA Modification of a DNA Oligonucleotide** **69**
NTA Groups Form Coordination Complexes with Nickel Ions and Histidine Residues 69
Problems with the NTA-DNA Led to a Search for Alternative Conjugation Strategies 72

2.2 **Expression and Purification of the FliG Protein** **73**
FliG Constructs Are Expressed in *E. coli* and Purified Using His-Tags 74

2.3 **DNA-Templated Protein Conjugation** **77**
DTPC Uses a Guide Strand to Position a Reactive Strand at a Specific Conjugation Site 78
The Reactive Strand Was Modified with DSS, a Homobifunctional Cross-Linker 81
Initial Conjugation Attempts Showed a Stable Protein-DNA Species on Gels 83
The Stable Assemblies Seen on Gels Were in Fact Complexes of NTA-DNA and FliG 86

2.4 **Templating Experiments Using NTA-DNA** **88**
Five NTA Groups on a DNA Strand Bind Stronger to a 10x His-Tag Than Three NTA Groups 90
Three FliG Molecules Were Bound to a Linear DNA Template Via the NTA-His Linkage 92
Conjugates Using 5-NTAs Could Not Be Organized with the Spacing Found in the C-Ring 94

CHAPTER THREE – A SIMPLER STRATEGY 97

3.1 **Maleimide Modification of a DNA Oligonucleotide and Its Conjugation to Cysteine-Modified FliG 98**
A Maleimide Group Was Introduced to the 5′ End of a DNA Strand Using Sulfo-SMCC 100
Conjugation of Maleimide-DNA to Cysteine-FliG Required HPLC Purification of the DNA 101

3.2 **Histidine-Tag-Mediated FliG Dimerization 104**
Initial Results Spurred Investigation into the Possibility of Domain-Swapped FliG Dimers 106
Characterization Showed the Dimers Were Broken with Heat and Not Domain-Swapped 110
Dimerization of FliG Proteins Was Mediated by Their Histidine Tags 115

3.3 **Experiments Using Permanent FliG-DNA Conjugates on a Five-Site Linear Template 119**
FliG-DNA Conjugates Can Be Placed on a Template at the Correct Physiological Spacing 121
Native Polyacrylamide Gels Showed No Evidence of FliG-FliG Interactions on Templates 123

3.4 **Organizing FliG on an Origami Tile 127**
Issues in the Tile Strands Were Resolved, Generating a Suitable Template for the FliG Ring 129
Streptavidin but Not FliG Was Visualized Bound at the Correct Sites on the Origami Tile 133

3.5 **Disulfide Cross-Linking Experiments with Double-Cysteine FliG 138**
Cross-Linking Produced No Conclusive Evidence of DNA-Template-Induced Oligomerization 141

CHAPTER FOUR – A NEW LAYER 146

4.1 **Design and Preparation of the FliF$_C$ Peptide 147**
Both Wild-Type and Single-Cysteine Versions of the FliF$_C$ Peptide Were Designed 148
Initial Attempts at Observing FliF$_C$-FliG Binding Were Unsuccessful Due to Solubility Issues 151

4.2 **Biophysical Characterization of FliG and FliF$_C$ 153**
CD Spectroscopy Showed That the Degree of α-Helicity Varies Between FliG Constructs 154
A Tryptophan Fluorescence Assay Demonstrated One-to-One Binding of FliG to FliF$_C$ 159
A Thermal Shift Assay Indicated That the Stability of FliG Increases upon FliF$_C$ Binding 162

4.3 **Conjugation of a DNA Oligonucleotide to a Cysteine-Modified FliF$_C$ Peptide 167**
HPLC Purification Isolated the Conjugate from Excess Maleimide-DNA and Cysteine-FliF$_C$ 168

4.4 **Organization of FliFC-DNA Conjugates on Five-Site Linear DNA Templates 172**
One to Five FliF$_C$ Peptides Were Arranged on DNA Templates in Various Configurations 173

4.5 **Binding of the FliG Protein to DNA-Templated FliF$_C$ Peptides 176**
Only Wild-Type FliG and Not Mutant Constructs Bound to the FliF$_C$ Peptide 177
Native PAGE Demonstrated One-to-One Binding Between FliG and DNA-Templated FliF$_C$ 183

4.6 **Expression and Purification of the FliM-FliN Complex 187**
No Interactions Could Be Detected Between the Purified FliM-FliN Complex and FliG 189

CHAPTER FIVE –RING BUILDING 197

5.1 **Design and Characterization of the Hexagon DNA Nanostructure** **200**
The Hexagon DNA Nanostructure is 35nm in Diameter and Presents 30 FliG Binding Sites 201
Gels and AFM Showed That the Hexagon Nanostructure Assembled as Designed 203

5.2 **Assembly of $FliF_C$ and FliG on the Hexagon DNA Nanostructure** **208**
An Excess of $FliF_C$-DNA and FliG is Needed to Fully Fill the Hexagon's Binding Sites 208
Several Techniques Were Used in an Attempt to Remove Unbound $FliF_C$-DNA and FliG 212

CHAPTER SIX – CONCLUSIONS 218

6.1 **A Brief Summary** **218**

6.2 **The Key Lessons** **223**

6.3 **Future Directions** **228**

SUPPLEMENTARY INFORMATION 233

Materials **233**

Methods **236**

References **241**

CHAPTER ONE

INTRODUCTION

To put into context the results that I present and conclusions that I make in this thesis it is necessary to first take several steps back for a broader overview of the landscape in which this work sits. Step by step I will narrow in on the system which I have studied with the hope of providing both a useful background and helpful introduction to the problem at hand.

1.1 Proteins Are the Preferred Functional Molecule of Life

We live in a protein world. This is to say that of all possible biological macromolecules, evolution has decided on amino acids strung together by peptide bonds as the building material for life. And not only are they merely the raw materials, the bricks and mortar. Proteins are the builders as well. Zooming down to the scale of a cell, one can see that living things are built *of* and built *by* proteins (Fig. 1.1).

The functional dominance of this single class of molecule has taken time to establish itself. RNA is thought to be the original functional biological macromolecule, with most believing that the early stages in the evolution of life were those of an "RNA world"[3]. The functions afforded by a palette of 20 amino acids however go beyond what is possible with RNA and though vestiges of the RNA world remain, it is in proteins that today we find the endpoint of biological information transfer. Carried from generation to generation by DNA and moved through the cell as RNA, the blueprints for life end up in the hands of proteins, and it is proteins which turn these instructions into action.

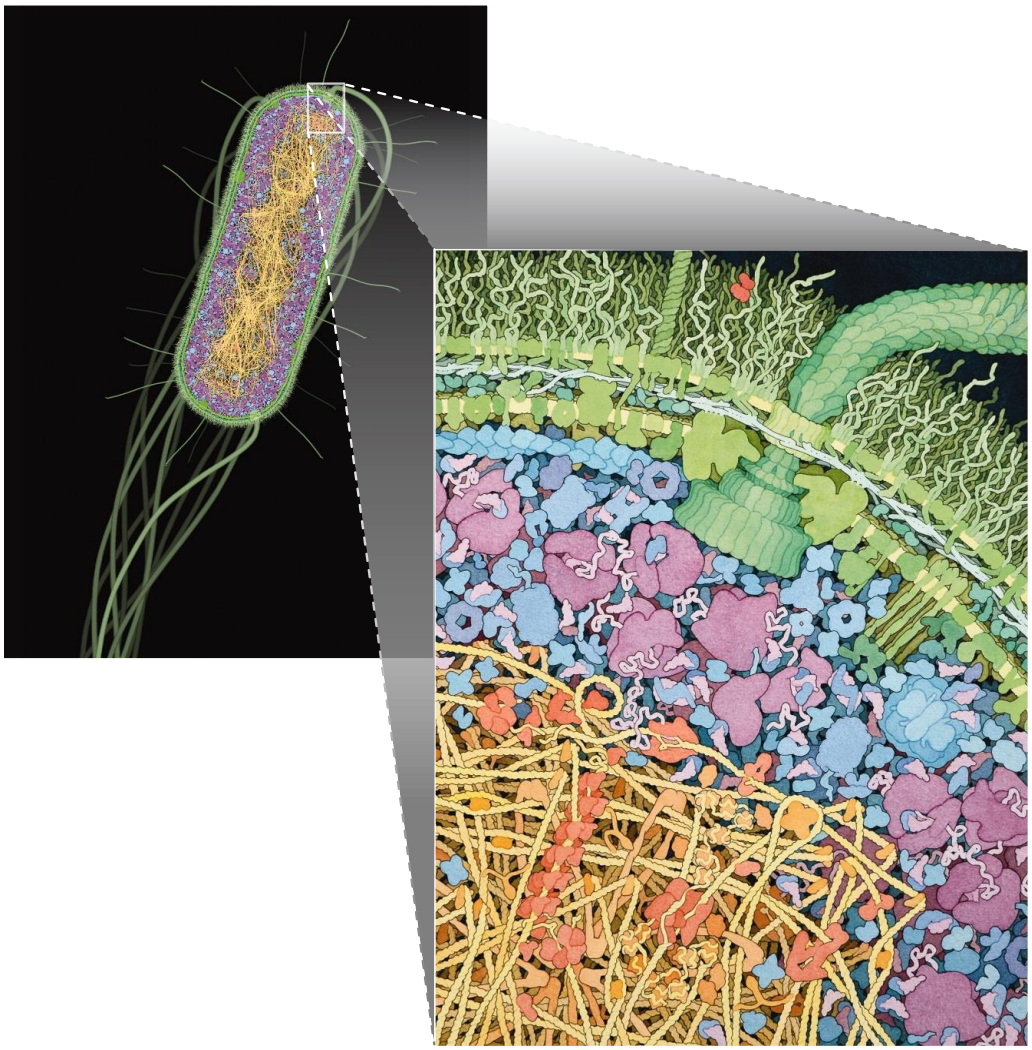


Figure 1.1 | Living things are made of and made by proteins. An *E. coli* bacterium is thought to contain around 3 million protein molecules making up approximately 55% of the dry weight of the cell[1]. Taking information from DNA and RNA (yellow), ribosomes (purple), translate this into an amino acid sequence which folds into a protein. These are the functional molecules of the cell and can reside both in the cytoplasm (blue) and embedded in membranes (green). Illustrations from (2) .

They push and they pull, they are gates, they are walls, they build and they destroy. The information encoded in these strings of amino acids gives rise, (in the words of one describing life at a slightly larger length scale), to “endless forms most beautiful and most wonderful”[4]. It is estimated that around six million different types of protein exist in humans[5] with millions upon millions of different proteins expressed across the spectrum of living organisms, each occupying a special niche in the workings of a cell. Making sense of these molecular ecosystems has been the life’s work of many scientists, trying to discern not only the structure and behaviour of each protein but how these individual components interact and work with one another. In this thesis, I examine a tiny piece of a tiny puzzle in a tiny organism, one which seems impossibly small when viewed in the light of such abundance and such complexity, but one which nonetheless has its own compelling story, millennia in the making.

Proteins Exhibit a Hierarchy of Organization

For proteins, “form follows function”[6], and all the tasks that they must carry out depend crucially on the subtle organization of their amino acids in three-dimensional space. Their structures can be viewed as having four sequential levels of organization, building hierarchically towards a final form.

The first level of organization is the simple sequence of amino acids which emerges following the process of translation, termed the **primary structure** of the protein (Fig. 1.2A). The identity of the amino acid at each position in the chain is dictated by the underlying sequence of the mRNA, with sequential peptide bond formation catalyzed by the ribosome. Somewhat arbitrarily, a short chain of around 30 or fewer amino acids is called a **peptide** whereas longer chains are referred to as **polypeptides**. A **protein** is a polypeptide with a well-defined three-dimensional structure[9].

This linear polymer gains its next level of organization through the formation of hydrogen bonds between its subunits, creating several well-defined geometries known as **secondary structures**. The α-helix and the β-sheet are the two most commonly occurring secondary structures, making up around 60 percent of the average polypeptide chain[7] (Fig. 1.2B). The rest of the chain consists

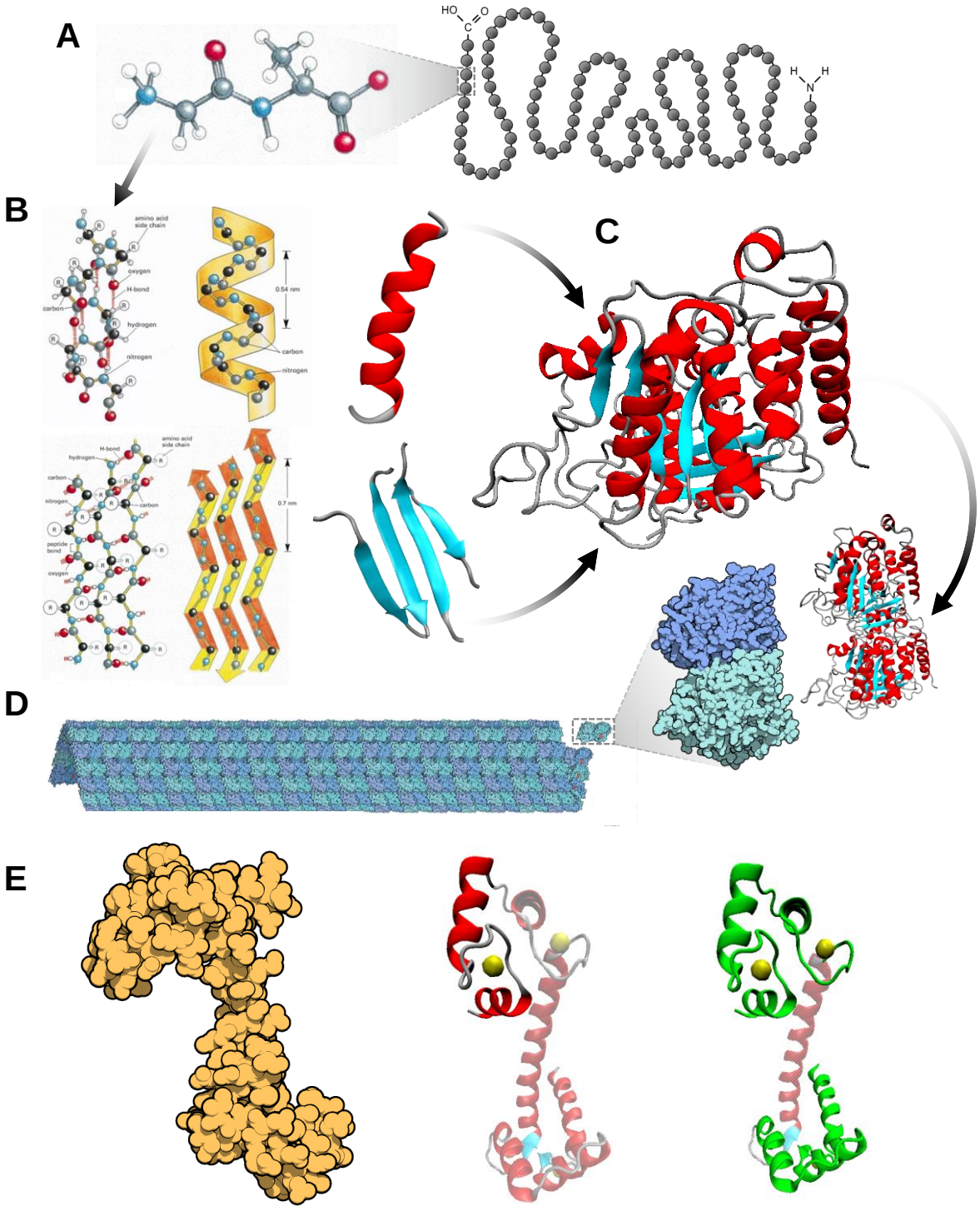


Figure 1.2 | Proteins are structured hierarchically. A) Linear strings of amino acids connected by peptide bonds constitute the primary structure of a protein. **B)** Hydrogen bonding creates secondary structures such as the commonly seen α-helix and β-sheet. **C)** Tertiary structure, such as that shown here for α-tubulin, represents the overall three-dimensional spatial organization of a single polypeptide chain. **D)** Multiple polypeptide chains can come together to form a quaternary structure, such as the microtubule shown, composed of many α/β-tubulin dimers. **E)** The protein calmodulin contains four helix-loop-helix motifs (one is highlighted) capable of binding a calcium ion. The protein as a whole is composed of two globular domains (highlighted in green) connected by an α-helix . Illustrations adapted from (7) and (8).

mainly of unstructured, random coil, lending a degree of flexibility to the overall architecture. These secondary structures are a common currency among proteins from across the tree of life and are thought to have taken root very early on in their evolutionary history[10].

Tertiary structure constitutes the overall conformation of the polypeptide chain and is the highest level of organization for a single protein (Fig. 1.2C). The interactions which make up a tertiary structure are in general weaker and more changeable than those of secondary structures, and are more diverse in their nature. As opposed to the hydrogen bonds of α-helices and β-sheets, the tertiary structure of a protein is held together by a combination of hydrophobic interactions, electrostatic interactions, and hydrogen bonds as well as the occasional disulfide bridge covalently linking two nearby cysteine residues[9]. The mouldable nature of the tertiary structure of a protein is often key to its function as changes to the conformation of a protein are one of the main means by which they carry out their tasks.

Organization beyond the tertiary level involves the incorporation of multiple polypeptide chains into a single **quaternary structure** (Fig. 1.2D). As quaternary structures are a key aspect of this thesis, I will discuss these in more depth in section 1.2.

Within this hierarchy of organization, it is worth pointing out two definitions that will be useful in what is to follow. The first is a **motif**. This is a unique three-dimensional organization of secondary structures found within the overall architecture of a protein which usually performs a single structural or functional role[7] (Fig. 1.2E). Motifs having nearly identical structure and performing the same function can be seen scattered around a wide range of different proteins, conserved in evolution due to their particular usefulness. Motifs can for example be finger-like projection used to bind to DNA, or a helix-loop-helix structure capable of binding a small ion.

The second term to define is a **domain**. This belongs to a higher level of organization than a motif, being composed of both secondary and tertiary structures which can either fold independently from the rest of the protein (a structural domain) or which constitutes the minimal unit needed for one particular function of the protein (a functional domain)[9] (Fig. 1.2E). Key to the definition of a

domain is the fact that a domain can both evolve and exist independently from the rest of the polypeptide chain. Domains are regularly shuffled around in the course of evolution and large proteins can often be identified as a series of domains strung together via flexible linkers. Proteins can also be divided into domains topologically based on where the domain is located within the cell. Most commonly this is the case with membrane-embedded proteins which can have for example extracellular, transmembrane, and cytoplasmic domains.

Proteins Use a Small Number of Techniques to Carry Out Their Function

As all proteins posses at their heart the same constituent amino acids and, as just described, the same hierarchical organization of these amino acids, it comes as no surprise that though there may be millions of different proteins, they exhibit only a small, shared suite of ways by which they carry out their functions. Though their actions and tasks are as diverse as the living things they make up, a few common “idioms” appear again and again when studying the workings of proteins.

The first and most universal fact concerning proteins is that they **bind**. It is through physical interaction with other species (termed *ligands*) that they carry out their tasks and most actions accomplished in a cell can be traced back to the direct contact of ligands with proteins. This binding is remarkably specific, with the majority of proteins having only a select few binding partners amongst the thousands possible within a cell[11]. To mediate these interactions, patches of residues on the surface of the protein, brought in close proximity by the complex folds of the structure, form a **binding site** which possesses a unique three-dimensional geometry able to accommodate a specific ligand. The ligand (commonly another protein with a complementary geometry) fits like a hand in a glove into the binding site, with the binding stabilized via the combined actions of multiple non-covalent interactions across the interface[9] (Fig. 1.3A). The nature of each binding site is unique, its charge, hydrophobicity, geometry, and other factors determined by the amino acid side-chains that make it up, with the specific composition and spatial arrangement of these side-chains leading to the specificity of binding that proteins require to function.

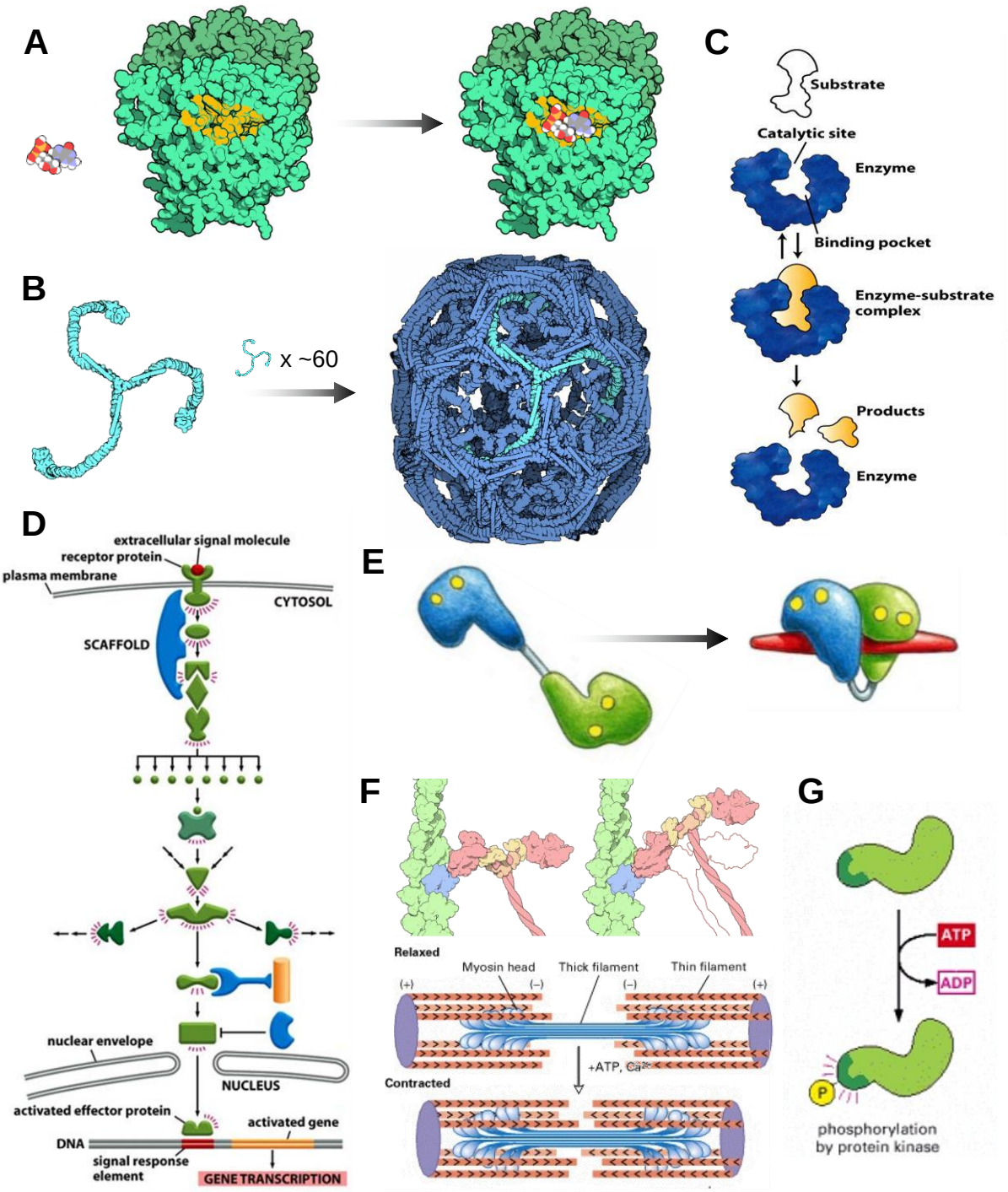


Figure 1.3 | Recurring themes emerge when studying protein function. A) Proteins work by binding ligands. Here the small molecule GTP binds to a small, specific site on the surface of α-tubulin. **B)** Some proteins bind to build structures like the clathrin cage shown here, made of around 60 clathrin proteins. **C)** Other proteins are enzymes, catalysing chemical reactions involving the ligands which they bind. **D)** Signalling proteins relay information from one part of the cell to the next through specific interactions with proteins in a pathway or cascade. **E)** An input often triggers a conformational change in a protein. Here the protein calmodulin takes a more compact form upon binding a target protein. **F)** Motor proteins transform chemical into kinetic energy to generate large scale motion within the cell. Here a conformational change in myosin called the 'power stroke' leads eventually to the contraction of a muscle cell. **G)** Proteins are the target of many different 'post-translational modifications', with different chemical groups being conjugated to their amino acids. The phosphorylation of the protein shown here is a commonly used method in signal transduction pathways. Illustrations adapted from (7) (8) (9).

Binding in and of itself is often the end goal of a protein whose job it is to create a larger structure within the cell, as large assemblages are formed by the sequential addition of subunits through repeated binding events (as described in section 1.2) (Fig. 1.3B). Aside from this class of structurally-oriented proteins however, there are others for which binding is only the first step in their task. Enzymes are a class of proteins which modify the ligand to which they bind, using the chemistry and geometry provided in the binding site to catalyze a chemical reaction involving their substrate (Fig. 1.3C). For another class of proteins, involved in signalling and regulation inside the cell, binding of a ligand constitutes the delivery of a message which they must act upon (Fig. 1.3D). Typically, these proteins act by binding another protein or molecule downstream and often form binding cascades with sequential binding events, eventually leading to the catalysis of a reaction or the assembly of a structure.

Among the assemblers and catalyzers and messengers, another common “idiom” appearing time and again in the workings of proteins is that of the **conformational change** (Fig. 1.3E). Proteins, by and large, are shape-shifters, having several different conformations that they can take, each state possessing different properties[12]. Triggered by an input, such as the binding of another protein, the secondary structures and tertiary layout of a protein can change dramatically, with side chains emerging or being sequestered away, helices and sheets disappearing or forming, and whole domains rearranging and reorienting.

Often this conformational change produces a new form of the protein capable of undertaking a task its previous form could not, for example through the creation of a new binding site. In the most extreme cases however, a conformational change of the protein can be used to induce large-scale movement within the cell. The class of molecules capable of this, called **motor proteins**, use their unique geometries to amplify the small-scale motions associated with ligand binding into motions capable of transporting vesicles, separating chromosomes, or contracting a muscle cell for example (Fig. 1.3F). As this thesis deals heavily with this class of protein, more will be discussed about them later.

Though shape-changing is nearly universal amongst proteins, on a more general level, modification is a key theme. After a protein has emerged from the ribosome, the majority undergo a series of **post-translational modifications**, wherein their residues are bound either covalently or non-covalently by various small molecules (Fig. 1.3G). Though a palette of 20 amino acids can accomplish many things, these small molecule attachments extend the capabilities of proteins beyond what their residues can offer and are important in regulation, catalysis, structural assembly, and all other ways in which proteins function. Lipid attachments make them suited for membrane insertion, glycosylation prepares them for extracellular work, binding of a metal ion forms special active sites, while phosphorylation often allows for signal transduction. The list goes on but the concept is the same throughout. Using the remarkable functionalities of proteins as a starting point, nature has pushed further to enhance or extend their functions with non-amino-acid attachments.

Though changes in conformation and addition of chemical groups are important, the simple presence or absence of a protein has the biggest effect on the final outcome of a task within the cell and regulation of protein number is one of a cell's key responsibilities. The levels of both the expression and degradation of each protein are constantly adjusted, tuning the amount of each type of protein present based on its needs. Additionally, proteins are shuttled to different sub-cellular locations either to be sequestered away or to engage in a certain spatially-constrained function. Thus, at any moment and at any location within the cell, the proteins present are under strict spatial and temporal control and their actions highly regulated.

Taken together, a protein's function is determined by its levels at a certain location within a cell, by how its amino acids have been modified, by its current conformational state, and most importantly, to what it binds. This binding in turn is dictated by the nature of the protein's binding sites which in turn are governed by the protein's overall structure. Thus, a protein's actions are inextricably linked to its three-dimensional geometry, its function to its form.

1.2 Protein Complexes Are Needed for the Cell's Most Difficult Tasks

The most common ligand for a protein to bind to is in fact another protein. When surveying the contents of a cell it is rarely the case that a protein is found acting alone and monomeric, and when this is the case its solitude is most often driven by the need for quick diffusion or another such evolutionary pressure[13]. The default state, observed much more frequently, is one in which proteins act together in oligomeric protein complexes. Though the methods by which they are characterized may lend a degree of bias to the classification, in the selection of cell types that have been surveyed, over half of the proteins expressed have been found to carry out their function only when in a complex with other protein molecules[14]. These multiprotein complexes (the quaternary structures briefly touched on in the previous section) carry out all the central processes of a cell, from replication to protein synthesis to locomotion, and without them living things would grind to a halt.

In this section I will provide an overview of the benefits afforded by forming these complexes and touch on a few key characteristics that they exhibit. Throughout I will bring forward examples of impressive multimeric protein structures and highlight the equally impressive subcellular feats they accomplish through the concerted action of their constituent proteins. I will also briefly introduce a class of complexes known as molecular machines, one of which the work of this thesis centres around.

Protein Complexes Provide Added Benefits When Compared to Monomers

Why have cells decided to build protein complexes? Why add another level of organization and complication to a macromolecule which already exhibits such a wide array of functionalities?

First, bringing together different subunits (i.e. the constituent proteins) into a complex allows for the coordination of each of their distinct functions. Having them all in one place means that their actions can be linked together and regulated more efficiently, allowing for more fine-grained control of the process or task which they are carrying out. This aspect of protein complexes is

seen most clearly in the spatial organization of a group of enzymes which catalyze a series of steps in a reaction pathway (Fig. 1.4A). Placing the enzymes side-by-side means that the product emerging from one enzyme does not have to diffuse far before becoming the reactant of the next enzyme in the chain. Linking the steps in such a way greatly speeds up the overall rate of the process, in some cases producing more than a 20-fold increase over the rate shown by enzymes working separately[16]. In addition, these complexes also often contain regulatory modules which can provide quick feedback to the pathway, either accelerating or slowing it down based on the needs of the cell[17].

These benefits seem obvious but a question then arises. Why could the same not be done with one long polypeptide chain, a string of different functional modules held together by covalent peptide bonds? One reason for tacking together modules with non-covalent protein-protein interactions lies in the limits of the cell's protein production machinery. Though massive proteins can be and are made (titin for example is 30 000 amino acids long[18]) it becomes more and more difficult for the transcription and translation machinery of the cell to produce error-free, fully-functioning proteins as the length of a polypeptide increases. In addition, the chances that the protein will fold incorrectly also increase as more residues are added[19]. So, though there may be a benefit for larger proteins with a wider range of functions, evolution seems to have settled on an optimal size of around 30 to 50kDa on average[18], providing both the range of abilities a cell needs from its proteins and the assurance that it will have properly-folded, fully-functioning products at its disposal.

Genetically, smaller proteins are much more economical as well, as the regions which code for them in the genome take up less space. Viruses provide a good illustration of this, where the length of the genome is tightly constrained and every bit must be used efficiently. In making the capsids which enclose their genetic material, viruses use only a small handful of proteins, transcribed and translated many times, with the subunits then binding one another in highly symmetric, ornate geometries, to create these large, complicated structures often hundreds of nanometres in diameter[20] (Fig. 1.4B). Although the constraints on genome length are not nearly

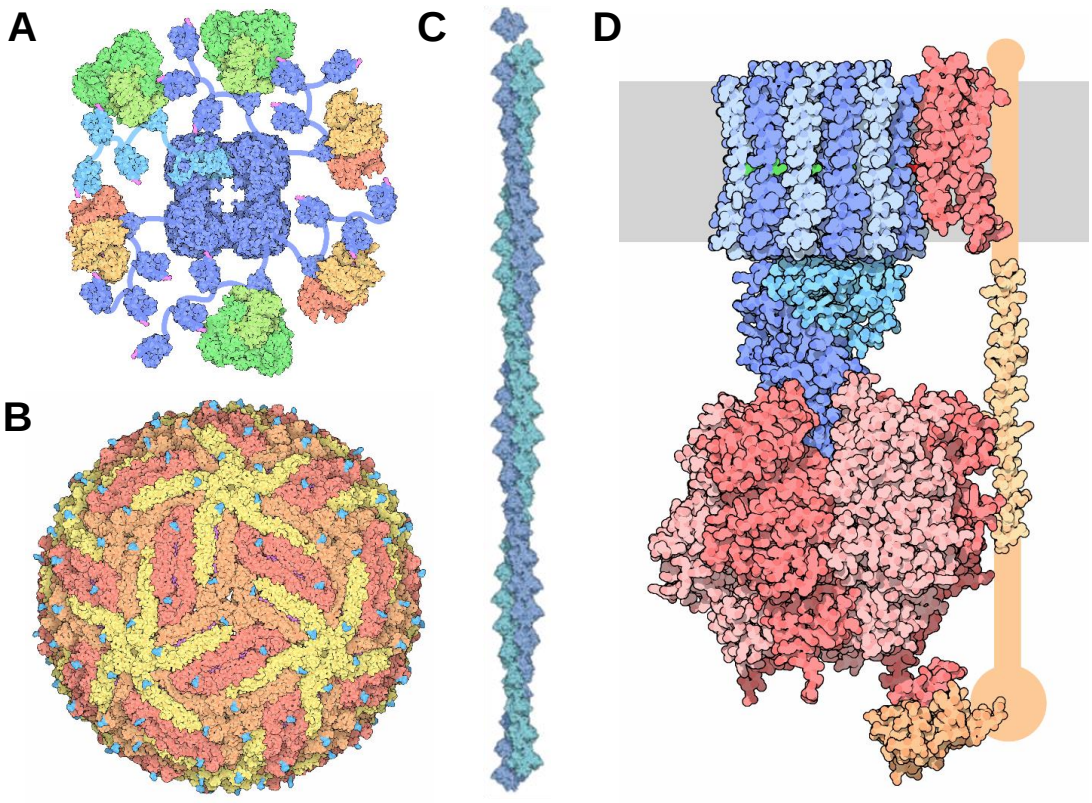


Figure 1.4 | Why build protein complexes? A) Tethering different functional modules together can speed up a multi-step process. Here the pyruvate dehydrogenase complex brings together several enzymes involved in the same reaction pathway (orange and green) through their binding to flexible protein tails emanating from the central complex (blue). **B)** The large zika virus capsid is made 180 copies of two different proteins. Using small proteins copied many times saves space on the constrained viral genome. **C)** Structures like the long actin filament shown here would be very hard to assemble from a single polypeptide chain. Oligomerization allows new length scales to be reached by protein structures in the cell. **D)** The F_1F_O ATP synthase is a molecular machine built of dozens of different proteins, all acting together to couple the flow of protons across a membrane with the generation of ATP molecules. Structures like this demonstrate the complex, coordinated actions that multi-protein machines are capable of. Illustrations from (15).

as tight in higher organisms, it would be quite difficult to imagine a single polypeptide chain being produced and then folding into such a structure without an error.

More generally however, building with breakable non-covalent bonds is much more useful in a cell which must constantly change and adapt to cues and stimuli. Being able to mix and match modules, to discard and replace faulty pieces, and dynamically respond to external inputs are all much easier to accomplish when subunits are not covalently bound to one another. In addition, due to the wide range of possible interactions between protein side-chains (ex. electrostatic, hydrophobic, hydrogen bonds) the strength with which each subunit is bound to the complex can be tuned, creating a range of possible stabilities.

Looking at the overall stability of a large multimeric structure reveals another advantage of using large protein complexes to accomplish a task. As the size of a structure increases, the ratio of its surface area to volume decreases, meaning proportionally fewer residues are exposed to potentially harmful external agents. The structure is thereby less likely to be degraded and exhibits a greater degree of stability than could be achieved with small monomeric proteins[21]. As most of the key processes a cell must undertake crucially depend on large protein complexes, this increased stability is quite important.

Finally, the fact that cells use protein complexes for the majority of their most important tasks ties into one of the main benefits for building them. In many cases, cells require more than what a monomeric protein, or even a collection of monomeric proteins, can offer either owing to the sheer size of the structure required or a need for a sophisticated, coordinated action. Simply due to their dimensions, large storage containers, pores through membranes, and long cytoskeletal filaments (Fig. 1.4C) could not be built from single proteins, while complexes like the ribosome, polymerases, and ATP synthase (Fig. 1.4D) require proteins assembled in such a way that they become more than the sum of their parts. This final class of complexes, the so-called **molecular machines**, are some of nature's most remarkable inventions capable of undertaking its most

difficult jobs. More will be discussed on these in relation to the next topic and the topics of section 1.3.

Protein Complexes Self-Assemble into Dynamic, Often Symmetric, Structures

As I have described, cells benefit in many ways from the construction of large multimeric protein complexes. These structures are as diverse as the proteins that make them up or the organisms that they are part of yet despite this diversity, at a fundamental level, they all share several important traits.

The first, and perhaps most remarkable property of protein complexes is that they **self-assemble**. Without external guidance, the individual components come together to fit, piece by piece, into a larger, more complicated structure. This is something rarely seen on a macroscopic level. To quote a commonly used analogy, it would be like shaking up a box of bricks and having a house emerge fully-formed. In a cell, the shaking and mixing is accomplished by the thermal motions of diffusing molecules, which in such a crowded environment (see Fig. 1.1) means they are constantly bumping into one another[21]. In most cases, these collisions produce transient non-specific interactions which break almost as soon as they are formed, but on the rare occasion that the correct protein finds its correct binding site, a more stable bond is created, leading to a longer-lived structure. These small-scale local interactions of proteins add up, building bit by bit through intermediate stages, until a large complex has assembled, an ordered structure arising from what was disorder a short time before (Fig. 1.5A).

The instructions for building these complexes are written in the amino acid sequences of their constituent components, passed down from DNA sequences through transcription and translation. Folding of the protein into its correct tertiary form is in most cases the first step of these instructions to be put into action, yet beyond this, for many protein complexes, the pathway by which they assemble is still not known. (In fact, for some complexes, the first step is not even the folding of its constituent monomers, as in a variety of cases the monomer has been found to change from a disordered string of amino acids to a fully-folded protein only upon binding to the larger structure[23]).

For certain complexes, a single assembly pathway may always be taken, while for others it may be possible to take multiple routes and arrive at the same final finished product. In general, it appears that the strongest, most stable interactions tend to be the first to form yet this is far from the rule and the order of assembly on the whole seems to be dictated by a more complicated mix of factors than stability alone[24]. One common theme among many assembly pathways is that the proteins exhibit different forms in their monomeric, pre-assembly state and when they are bound to the structure[25]. As described in section 1.1, a conformational change upon binding is a common occurrence, and can act to lock the protein in place or set it up for the next binding event in the pathway.

So, while stability may dictate which assemblies form first, and it appears to be common for the monomeric components to change conformation upon binding, it is hard to draw up general rules for protein complex assembly pathways. Overall there seem to be more questions than answers on this topic which makes it an area of keen scientific interest. The assembly pathway of one particular complex is a focal point of this thesis and the outstanding questions concerning the assembly of the system of study will be discussed further in the sections to come.

Besides the fact that they can self-assemble, another common feature of protein complexes is that by and large they are **dynamic**. Though some structures, like membrane-spanning pores and ion channels, appear to be static and stable for as long they exist[26], these are the exceptions rather than the rule and in most cases changes in the number or nature of the subunits result in multiple states that the structure can take. In fact, even those which appear to be static are often constantly being renewed and rejuvenated as subunits turn over, with older proteins being swapped in for new ones[27].

Molecular machines (touched on previously) are the poster children for dynamic protein assemblies. These complexes undergo large coordinated motions of different subunits, generating the large-scale movement a cell requires for a specific task. Examples include the ribosome translating mRNA into proteins, ATP synthase (see Fig. 1.4D) coupling the flow of protons to the

generation of the cell's main energy-storage molecule, the proteasome chopping up the cell's unwanted proteins, and larger assemblies such as the mitotic spindle in eukaryotes (Fig. 1.5B) faithfully segregating one copy of each chromosome into each daughter cell. They all require an energy input (often in the form of ATP or GTP), and use it to generate mechanical force and motion. The keys to this motion are the intrinsic flexibility of their subunits and their ability to coordinate and propagate conformational changes from one region of the structure to another. They are truly some of the high-points in evolutionary design and their complexity on such a small scale has led to long and painstaking scientific inquiry into their workings.

In general, seeing as all living things are in a constant state of flux, it may come as no surprise that protein complexes tend to be changeable and dynamic. The remarkable **symmetry** that is a common trait of many of these complexes however comes as a bit more of a surprise. The evolutionary arguments for the perfectly symmetric rings of a needle complex (Fig. 1.5C) or the stunning icosahedral cages of a virus capsid (see Fig. 1.4C) are not immediately obvious yet symmetry appears again and again when dealing with protein assemblies.

There are several possible explanations for this. The first is that a symmetric complex tends to be a stable complex[28]. Although there are examples of stable, asymmetric protein assemblies, it appears that one of the easiest ways to generate strong, specific protein-protein interactions is through complementary mutations across an interface between two monomers of the same species. Should a binding-site-strengthening mutation take place, its positive effects are doubled since both monomers of the "homodimer" pair will posses it and thereby create two such bonds symmetrically placed at the interface[29]. This greatly speeds up the pace at which a strong and stable interaction can be evolved and helps to explain the over-representation of symmetric homodimers seen across the spectrum of protein species. To create larger structures, this same process can be repeated multiple times to build dimers of dimers and so forth (complexes with so-called dihedral symmetry[21]).

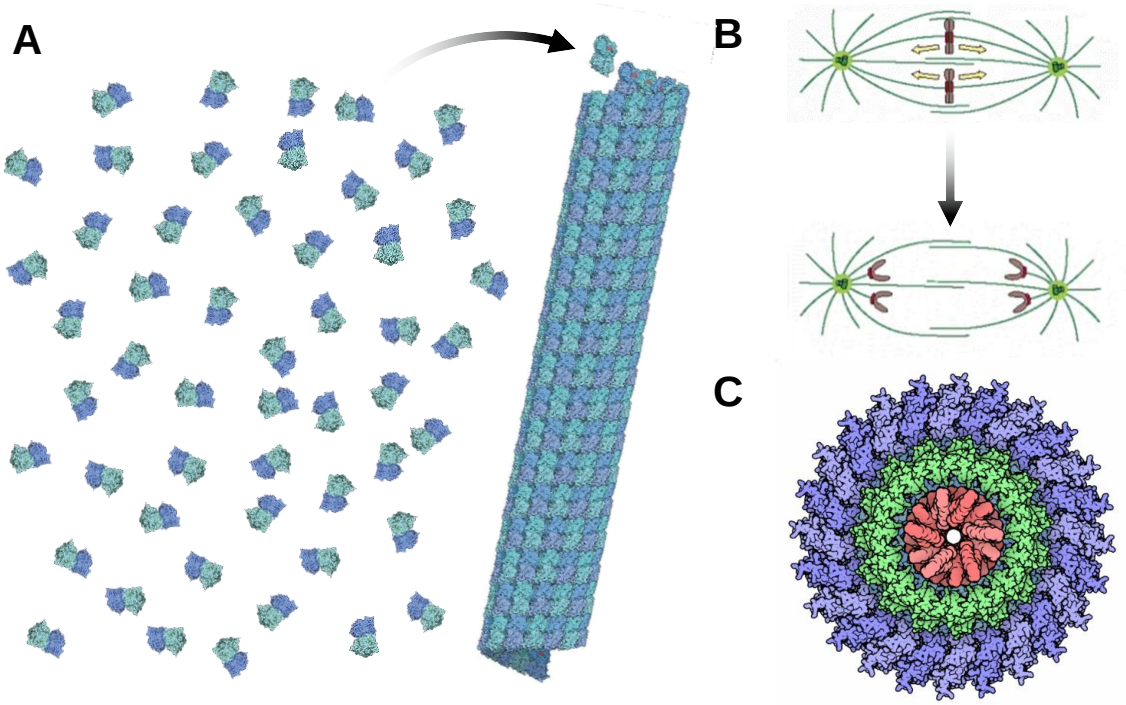


Figure 1.5 | Protein complexes share several important characteristics. A) Protein complexes self-assemble into their final form. Long cytoskeletal filaments called microtubules are made up of hundreds of individual α/β-tubulin dimers which polymerize into tubes microns in length. **B)** Many protein complexes are dynamic. The mitotic spindle is made of many microtubules and motor proteins working together to segregate sister chromatids. Dynamic shrinking of the microtubules drives segregation. **C)** Many protein complexes display a high degree of symmetry. The needle tip complex of *Salmonella's* type III secretion system, used to inject pathogenic proteins into its host, is cyclically symmetric. Illustrations from (7) (15).

For complexes displaying cyclic symmetry like ATP synthase (Fig. 1.4D) and the needle complex (Fig. 1.5C), a different explanation may be at work. In these cases, when the cell has had a pressing need for a structure with a precise, finite dimension (for example the diameter of a ring encircling the DNA double helix[30]), cyclical symmetric assembly provides the simplest way of forming oligomers with a finite number of subunits and hence size[21]. Closing these structures cyclically into a ring and thereby tying up any loose ends, also aids in their stability and prevents the unbounded growth of protein oligomers (as seen in the amyloid plaques of Alzheimer's disease[31]).

So, taking these cyclic assemblies together with the cell's propensity for homodimer pairings, symmetry appears to be a trick evolution uses to build stable protein complexes of a specific finite size. It is not that these structures are symmetric for symmetry's sake, rather it is the secondary benefits of stability and control that symmetric assembly provides which explains why a survey of structures in a cell finds so many with an impressive degree of symmetry.

Thus, in sum, proteins are remarkable in their ability to self-assemble into dynamic, often symmetric, complexes. The assembly of such complexes provides a way for the cell to combine and organize multiple functional units, to coordinate processes, and to quickly modulate and regulate its behaviour. Multimeric structures have the added benefits of being in general more stable than small monomers and of being more manageable for the cell's protein manufacturing machinery to produce than a single long string of amino acids. Finally, protein complexes extend the possibilities of functions beyond that which monomers provide and, as exemplified by the cell's molecular machines, are capable of an astounding range of movements and actions. The next section brings to light one of the pinnacles of molecular machinery, a remarkable protein complex which demonstrates what it is possible to achieve given a palette of 20 amino acids and billions of years of evolution.

1.3 The Bacterial Flagellar Motor is a Remarkable Molecular Machine

Even the smallest of organisms have need for protein complexes capable of integrated, coordinated actions. Bacteria, some of the oldest and simplest living things on Earth, have evolved one of the most complex protein structures found in nature in response to a need for directed motion. Called the **bacterial flagellar motor** (or the BFM), this structure is a remarkable feat of molecular engineering. Made up of over two dozen different proteins in various stoichiometries, its parts list includes hundreds of individual components[32], all working together to propel a bacterium through its liquid environment. It is a textbook example of a self-assembling, dynamic molecular machine and perhaps the most well-studied protein complex in nature. That being said, the more that it is studied, the more questions emerge as to both its structure and its functions, with each level of detail providing new and exciting discoveries as well as new puzzling mysteries.

In this section I will set out to provide an overview of the bacterial flagellar motor, the system of study which this thesis revolves around. I will first illustrate the movement it makes possible by describing the run-and-tumble motility of *Escherichia coli*, and will then review the overall architecture of the protein complex. Along the way I will go through the current state of knowledge concerning two key aspects of the motor: its ability to generate torque, and its ability to switch the direction of its rotation. This naturally ties into ideas at present of how the motor is thought to self-assemble, though this process is still relatively poorly understood. My goal is not to give an exhaustive description of every aspect of this protein complex, but instead to highlight and provide background for the key features of the motor that will be focussed on in the work that is to follow. It is a complicated machine and I work with only four of its proteins so it is worthwhile concentrating primarily on these four and highlighting their actions amidst the workings of the whole complex.

Bacteria Can Integrate Signals from Their Environment and Swim in a Directed Manner

Far back in the history of the evolution of life, organisms developed locomotion: the ability to actively move about in their environment. In contrast to organisms with a sessile lifestyle, creatures which can move are not dependent on the state of their local surroundings for survival but can

instead travel to places with more favourable conditions. Bacteria, one of the very earliest branches of the tree of life, were quick to adopt motile lifestyles, using directed, self-propelled movement to avoid danger and access areas with beneficial resources. Surveying the world of bacteria today, one finds that many species are capable of some type of active, directed motion, and that their ability to move is a key factor in their successful colonization of most environments on Earth[33].

There are a large number of different ways that bacteria move around, from inching along surfaces using extensions called pili, to coordinated strokes of oar-like appendages, but the most common method for generating motion is through the propeller-like rotation of **flagella**[34]. These are long, thin, helical filaments around 10μM in length and 20nm in diameter which extend from the body of the bacteria out into the extracellular environment[35]. Each flagellum is connected at its base via a flexible hook to a single bacterial flagellar motor (BFM), a complicated structure which is embedded in the membranes of the bacteria[36] (Fig. 1.6A). The rotation of the BFM spins its corresponding flagella, and the interaction of this spinning flagella with its liquid environment generates a propulsive force which either pushes or pulls the bacteria. The rotation is exceptionally fast, commonly seen having frequencies on the order of 100Hz, with the highest observed being 1700Hz by a species of *Vibrio*[39]. This propels the bacteria along at speeds around 30μm/s which, if translated via body lengths per second to human length scales, would put it roughly on par with a sailfish (the fastest fish in the ocean) in terms of swimming speed[40].

Though many bacteria have flagella, different species use them in different ways to move in a directed fashion. While some have only one flagellum, others have several clustered at one pole, while others still have many spread all over the cell body[41]. In addition, certain species have motors which spin only in a single direction[42] while others have motors which can spin in both clockwise (CW) and counterclockwise (CCW) directions[43]. This diversity means there are many strategies taken to both move and change direction. Species with single-direction motors can change their course by either stopping or slowing down their motors to let Brownian motion reorient the cell[44]. Others, with a single, two-directional polar flagellum, use the two directions of rotation to either

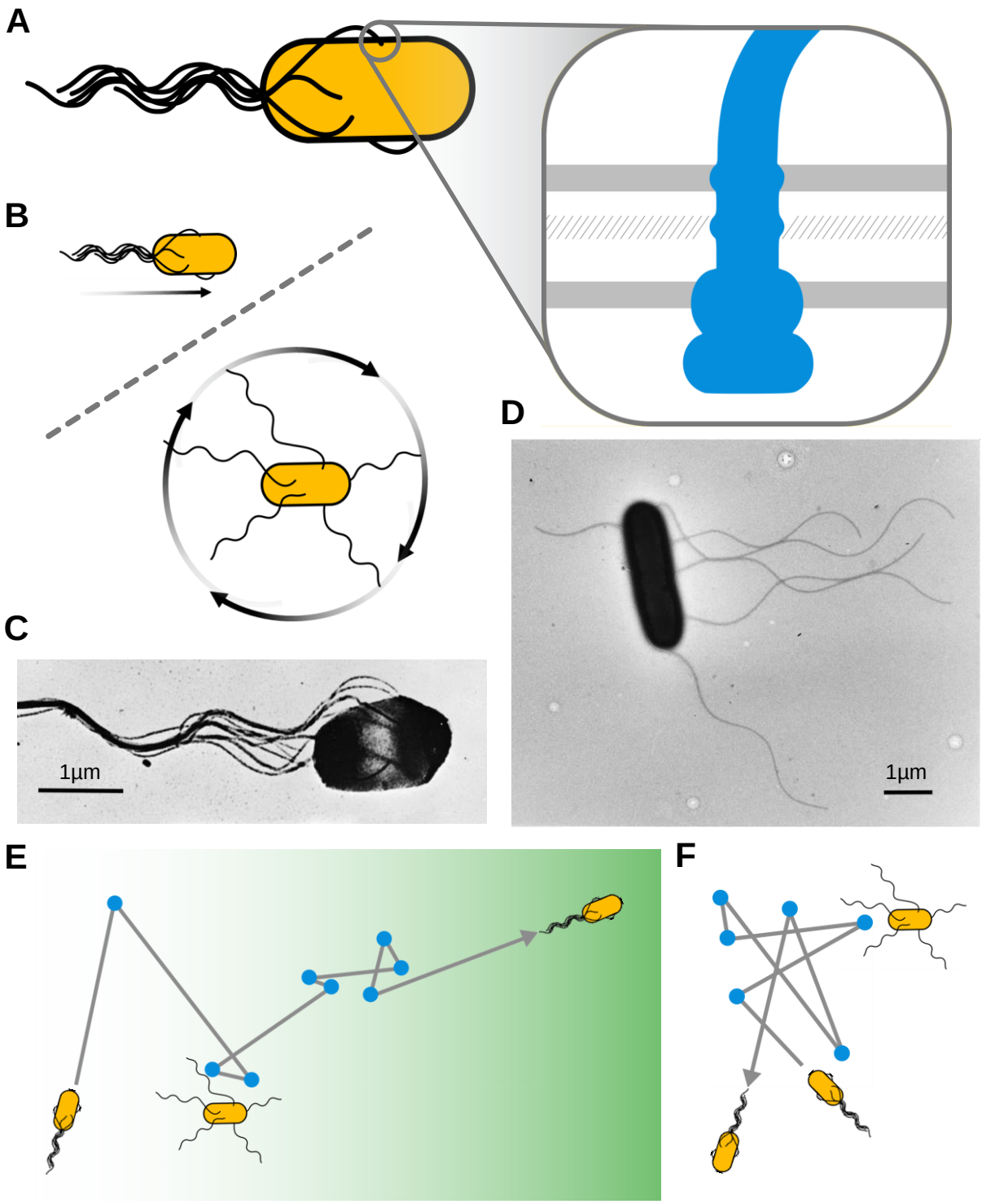


Figure 1.6 | *E. coli* and *Salmonella* bacteria exhibit directed, 'run-and-tumble' motion driven by the bacterial flagellar motor. A) Many bacteria are propelled through their liquid environment by the rotation of long, helical flagella. At the base of each flagellum, embedded in the membranes, is a bacterial flagellar motor (blue), the molecular machine which drives this rotation. **B)** When the flagella of a bacteria are all rotating CCW, they form a bundle, and the cell moves smoothly forward (a 'run'). When one flagellum switches direction, the bundle is disrupted, and the cell 'tumbles'. Electron micrographs of *Salmonella* bacteria show flagella in a run[37] **(C)** or tumble **(D)** state[38]. **E)** Bacteria achieve directed motion towards an attractant (green) by favouring 'runs' when moving up the concentration gradient and favouring 'tumbles' if moving in the wrong direction. **F)** Without an external gradient the bacterium moves randomly.

push or pull the cell, with reorientation the result of a small, random 'flick' of the flagellum during the CCW to CW transition[45]. The bacteria of the most heavily-studied class, which includes such species as *Escherichia coli (E. coli)* and *Salmonella enterica (S. enterica)*, possess a large number of flagella which can spin in both directions, and they move in a directed fashion by using a **'run-and-tumble'** mechanism[46]. As the proteins which I study come from this class of bacteria, I will briefly give an overview of this method of motility.

When an *E. coli* is swimming during the 'run' phase of its motion, its multiple flagella bundle together, all rotating in the same direction (CCW, looking from outside to inside the cell)[47] (Fig. 1.6B, C, E). The cell travels in roughly a straight trajectory and moves quickly through its environment. If, however, one flagellum in the bundle has the direction of its rotation reversed by a stochastic switching of its motor from a CCW to a CW state, this flagellum will be expelled from the bundle with the disruption causing a 'tumble', that is, a stopping and random reorientation of the cell[48] (Fig. 1.6B, D, F). Should there be no external input given to the bacteria, this running and tumbling, the bundling and unbundling of flagella, will carry on in a random fashion and produce no net movement.

The cell is capable of directed motion however and will travel up or down gradients of chemicals in its external environment, moving either towards favourable areas (such as a food source) or away from unfavourable ones (such as an area of toxicity). To do this it uses receptors on its surface to detect substances around it and when one is detected it engages the chemotactic signalling pathway (possibly the most thoroughly studied signalling pathway in science)[49]. The endpoint of this pathway is a protein called **CheY** whose phosphorylation state determines whether or not it will interact with the motor[50]. When left unphosphorylated, it does not interact with the motor and the smooth CCW rotation of the bundle leads to the 'run' state. If phosphorylated, it binds and biases the motor towards a CW state (more on this later), leading to an unbundling and to 'tumbling' reorientation[51]. Thus, if the cell is moving along a gradient of more and more food or less and less toxin, its CheY is left unphosphorylated and 'runs' are favoured. If it is heading towards unfavourable regions, CheY phosphorylation increases and 'tumbles' are favoured,

turning it around until it finds the right direction to ‘run’ in. This system is finely tuned and can adapt and respond to chemical concentrations over five orders of magnitude[52]. In addition, the chemotaxis machinery integrates and coordinates many different signals and chemical concentrations at once, deciding based on these inputs whether or not CheY should be phosphorylated and a ‘run’ be turned into a ‘tumble’.

To summarize, directed motion, seen in many forms across a spectrum of species, is one of the most complex actions a bacterium undertakes. The majority have settled on a solution which uses long, filamentous flagella spinning like propellers to move themselves around and at the heart of this motion is the bacterial flagellar motor, one of the most complex molecular machines in all of nature. It spins at over 100Hz and can change the direction of its rotation in response to external inputs like the protein CheY, a key component in the ‘run and tumble’ motion of *E. coli* and *Salmonella* bacteria. Like with any protein or protein complex however, the form of the BFM dictates its function, and in order to gain insight into both how torque is generated and how the direction of rotation can be switched, it is worth taking time to go through the overall architecture of this machine.

The Bacterial Flagellar Motor is Made of Hundreds of Individual Components

As just illustrated, bacteria are remarkably adept at getting from one place to another using their flagella, and the bacterial flagellar motor is the driving force behind this motility. But what sort of machine is the BFM and how does it work? What sort of structure can spin at over 100Hz in both CW and CCW directions?

Figure 1.7 illustrates the BFM of an *E. coli* or *Salmonella* bacterium. Several important features can be seen at first glance. First, the complex is embedded in the membranes and in the peptidoglycan layer of the bacterium with its components extending from the cytoplasm out into the extracellular medium (Fig. 1.7A, B). This membrane-spanning architecture is key to how the BFM functions in that this layout allows it to use the concentration gradient of ions across the inner membrane (established by ion pumps such as ATP synthase, see Fig. 1.4D) as an energy source

to power its rotation. The second observation is the great deal of cyclic symmetry present, with the motor showing many rings of proteins in the cryo-electron microscopy (cryo-EM) reconstruction (Fig. 1.7C). This is not surprising for a rotary motor and connections between components around these rings are needed for the complex to rotate and switch direction as a single unit. Finally, the scale shows that the BFM is quite a large complex, around 50nm in diameter and even larger in the membrane-spanning direction (Fig. 1.7C). Weighing in at over 11MDa[43], the BFM demonstrates many of the benefits for building large complexes as highlighted in section 1.2. By bringing so many proteins together it can combine and coordinate each of their individual functions, making it capable of movements not possible with a single protein alone.

So, the BFM is a large, membrane-spanning complex with rings of proteins. To appreciate how it functions however, one must delve deeper into the workings of each of its components. At a finer-grained level, the BFM can be divided into distinct functional or structural regions and I will now touch on each of these, building a working knowledge of the motor in preparation for the work to be described in the later chapters.

At the core of the BFM is a ring embedded in the inner membrane composed of a single protein called **FliF** (Fig. 1.7D). This so-called **MS ring**, is what the rest of the motor is built around and the current leading thought is that it is the first structure to assemble[56]. Unlike many of the other parts of the BFM, it is thought to be relatively static, a solid base anchoring the complex in the membrane.

Around this ring, embedded both in the inner membrane and in the peptidoglycan layer are the **stator complexes** (Fig. 1.7D), named as such due to the fact that as the rest of the motor turns, these structures remain stationary, fixed to the cell wall. Around a dozen of these complexes, each with four copies of the protein **MotA** and two of **MotB**, are scattered around the MS ring and form channels spanning the inner membrane[56.57]. It is through these channels that protons flow from the intermembrane space into the cytoplasm, travelling down a concentration gradient established elsewhere by active proton pumps[58]. This flow of protons is the energy source for the rotation of the motor and the movement of protons down the channel in the stator is thought to drive cycles

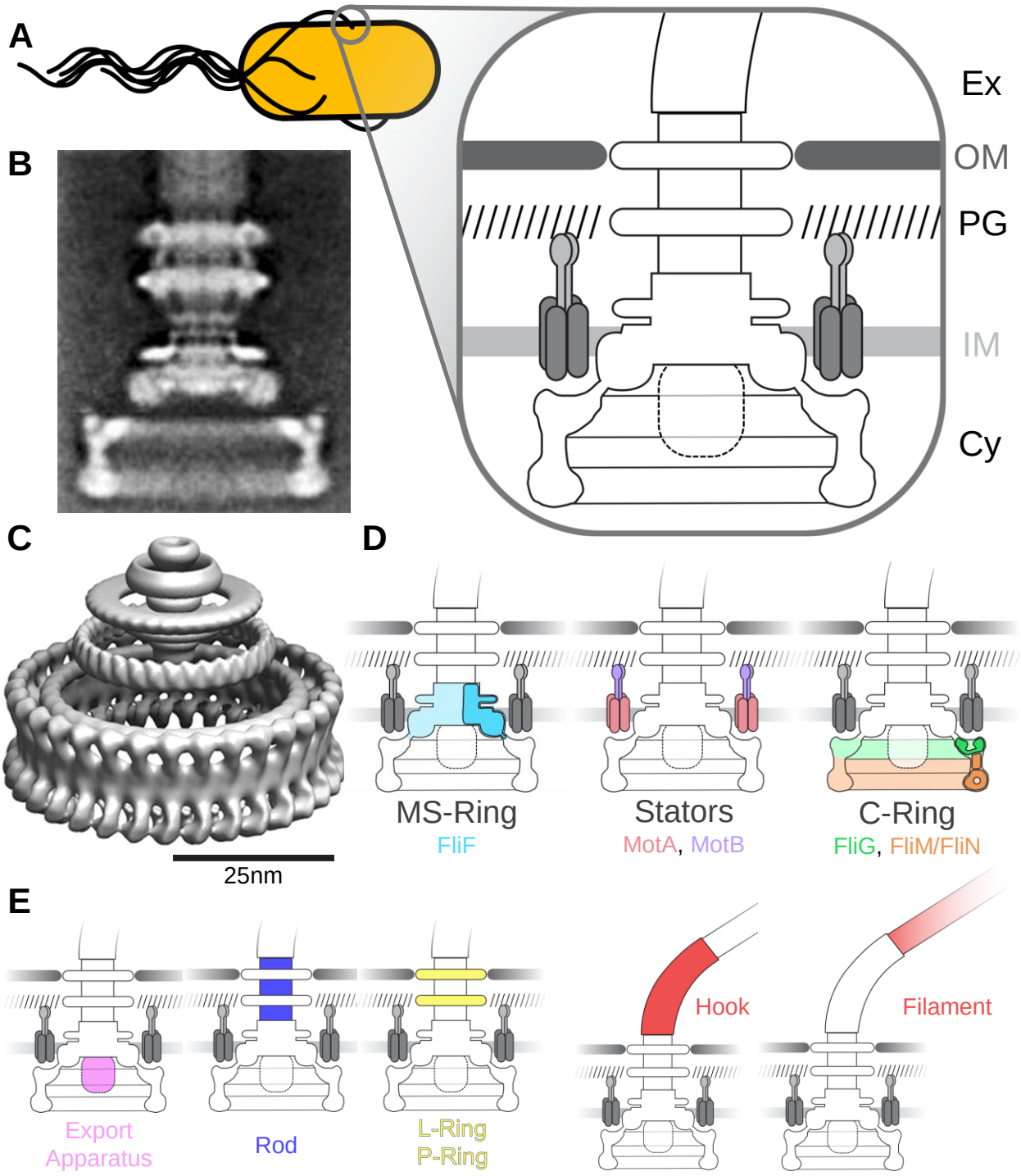


Figure 1.7 | The BFM can be divided into several structural and functional units. A) The BFM stretches from the cytoplasm (Cy) to the external environment (Ex), passing through the inner membrane (IM), the peptidoglycan layer (PG) and the outer membrane (OM). Electron micrographs[53] **(B)** and cryo-EM reconstructions[54] **(C)** of the *Salmonella* BFM show the large size and cyclical symmetry of the complex. **D)** The MS-ring is embedded in the inner membrane and is composed solely of FliF. Stators use the flow of H^+ ions as an energy source as they push on and thereby rotate the C-ring below. The conformation of FliG, FliM, and FliN proteins in the C-ring determines if the rotation is CW or CCW. **E)** The export apparatus selectively delivers components from the cytoplasm through its central pore during motor assembly. First, the rod grows outwards through the peptidoglycan layer and outer membrane. The L- and P-rings form around it to act as bushings. Next, a flexible hook assembles, coupling the rotation of the motor to that of the filament. The filament is composed of several helical protofilaments and can extend over 10µm into the external environment. Its rotation propels the swimming bacterium.

of conformational changes in the protein MotA, akin to the power strokes of the protein myosin[59] (see Fig. 1.3F).

A ring lying below the stators, in the cytoplasm, is repeatedly pushed on by MotA's cyclic 'power strokes' and torque is generated, causing the ring to spin (Fig. 1.8A). Composed of the C-terminal region of FliF (**FliF$_C$**), as well as the proteins **FliG**, **FliM**, and **FliN**, this underlying **cytoplasmic ring** (also known as the C-ring or torque ring, Fig. 1.7D) is between 30 and 50 nm in diameter and through its connection (via the interaction between FliF and FliG) to the MS ring, its rotation drives the rotation of the rest of the motor. The interaction at the torque-generation interface between the C-ring and the stator, of all the interactions that make up this molecular machine, is perhaps the most important one for the overall function of the BFM. Without this interaction, the motor cannot rotate and the bacterium cannot swim[63]. In addition, subtle changes in the conformation of the components at this interface determine which direction the rotation will take place, meaning the same cyclic power strokes of MotA can drive both a CW or CCW spin[63].

This remarkable ability of the motor to switch the direction of its rotation is driven by the remarkable ability of the C-ring as a whole to adopt two different conformations. Switching between the two happens stochastically, does so incredibly quickly (1-100ms[61]), and is finely tuned to concentrations of phosphorylated-CheY in the cell, with greater CheY-P levels biasing the ring towards a CW (tumbling) conformation[51]. In addition, switching appears to be an all-or-none response, meaning that the same conformation (either CW-favouring or CCW-favouring) is found in each subunit of the ring[64]. This quick, coordinated transition has been explained by a 'conformational spread' model in which a switch from one conformation to another in a single subunit propagates cooperatively (Hill coefficient ~20) around the subunits of the ring[61] (Fig 1.8B). How exactly this propagation takes place at a molecular level is still yet to be understood and this mystery forms one of the key motivations for the work that is to follow.

This core of the motor, the MS-ring, the C-ring, and the stators, can be thought of as the engine of the BFM, driving the rotation and enabling the switching of direction, yet on its own it does very little for the cell. To swim, bacteria need to build a flagellum, the part which actually propels the

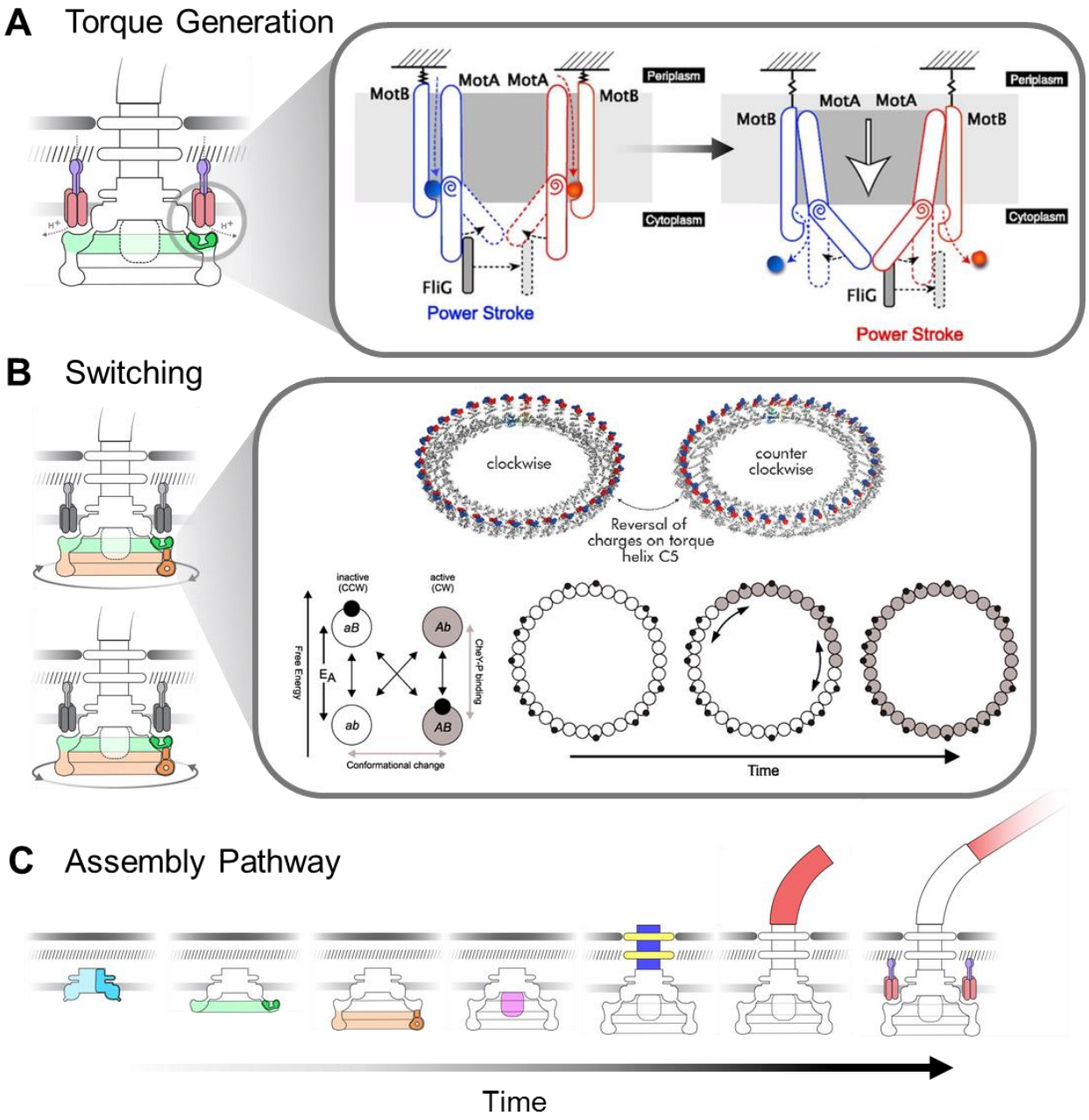


Figure 1.8 | The BFM self-assembles into an ion-driven motor capable of rotating in both CW and CCW directions. A) In the current view, torque is generated by a MotA 'power stroke' in which the transit of a proton through the stator channel is coupled to a conformational change in the MotA protein. This change pushes on a nearby FliG protein, thereby spinning the C-ring, and hence the rest of the motor. Illustration from (59). **B)** The motor can spin in either a CW or CCW direction depending on the conformation of its C-ring. The two different conformations are thought to place a charged region of FliG in opposite orientations (top[60]), meaning the same MotA power stroke can generate rotation in two different directions. All subunits of the ring take on the same configuration with a rapid transition between the two, thought to occur via a 'conformational spread' (bottom[61]). The binding of the CheY-P protein biases the conformation of a subunit towards the CW state, and this state cooperatively spreads around the ring. **C)** How the BFM self-assembles is still not known but the leading theory postulates a sequential addition of components as follows. First, the MS-ring forms in the inner membrane from FliF proteins, followed soon after by FliG, and FliM/FliN, the components of the C-ring. Perhaps at the same time the export apparatus forms and begins delivering components out through its central channel first for the rod (around which the L- and P-rings form) and then for the hook. The apparatus also pumps out repressor proteins, thereby releasing the inhibition of the final assembly stage in which the filament forms and stators assemble and begin to rotate the complex.

cell through its liquid environment. They do this in quite an unusual way. Either concurrently or shortly after the cytoplasmic ring is assembled, an **export apparatus** (Fig 1.7E) is built in the cytoplasm at the base of the BFM[65] (Fig. 1.8C). This complex has both membrane-spanning and cytoplasmic components which connect up with the rest of the motor and form a selective channel through the inner membrane[66]. The apparatus then controls the sequential assembly of the flagellum by selectively delivering each component part through its channel in a step-by-step manner.

First to travel out through the channel are the components for the **rod** (Fig 1.7E). These form a hollow tube, growing outwards, hydrolyzing and pushing through the peptidoglycan and then the outer membrane, finishing when the extracellular environment has been reached[67] (Fig. 1.8C). Surrounding this rod at the regions where it passes through the peptidoglycan layer and outer membrane are the **P- and L-rings** respectively (Fig. 1.7E). These act as bushings, separating the membrane and cell wall from the quickly rotating shaft[68].

Once the rod is complete, the next proteins which are exported form the flexible **hook** (Fig. 1.7E). This section connects the rotation of the motor to the rotation of the filament and its flexibility allows for the bundling of the many flagella of an *E. coli* or *Salmonella* bacteria[69]. Once the hook has grown to a certain length (thought to be specified by the molecular ruler FliK[70]), the **filament** (Fig. 1.7E) begins to grow through the export of the protein **FliC** (also known as flagellin)[71] (Fig 1.8C). Over 20 000 copies of FliC oligomerize into a filament often over 10μm long, composed of helical protofilaments which give it its spiralling nature[72]. A switch in direction of the rotation of the BFM is thought to change the handedness of this helix which leads to the expulsion of a filament from a bundle and the ‘tumbling’ of the bacteria[73].

In sum then, not including the many copies of FliC, the parts list for the BFM runs well into the hundreds of components. To a great degree, how these parts fit together and interact with one another has been slowly but surely worked out over a period stretching back to the 1950s, with greater and greater detail being obtained with each investigation. Current work is trying to untangle several outstanding puzzles concerning the function of the BFM at a molecular level of detail. The

coarse-grained nature of how the motor spins, switches, and assembles has roughly been determined (see Fig. 1.8), yet the field is only now reaching the point where the interactions between each of the protein subunits are being resolved, amino acid by amino acid.

The BFM may, in fact, be the best studied protein complex in nature, yet with each new breakthrough in understanding comes more questions. How many of each protein subunit are found in the motor? How do they interact with one another? Moving beyond this simple, static picture, questions about the dynamics of the motor arise. By what pathway do the components self-assemble? Is the self-assembly process strictly sequential or are there parallel, overlapping steps? Once built, are the components of the motor held unchangeably in place or can the subunits turn over dynamically? These questions and others concerning the structure and dynamics of the complex drive most of the current research in the field and form the basis for the work to follow in this thesis.

In the chapters to come, I work exclusively on the cytoplasmic ring of the BFM, and, more specifically, with four proteins: FliF, FliG, FliM, and FliN. In the following section I will take a closer look at these four components with a focus on FliG, the protein most central to my work. I will then present a leading theory to explain the composition and assembly of the cytoplasmic ring before proceeding later in this introduction with the techniques used to test and investigate this theory. In light of the many components and remarkable abilities of the BFM just presented in this section, the cytoplasmic ring appears to be a small piece of the overall puzzle, yet it is perhaps the most key component in the motor, and, as the work to come illustrates, harbours many unanswered questions.

1.4 The Structure and Assembly Pathway of the Cytoplasmic Ring Are Still Unknown

Narrowing in from protein complexes to the BFM, I will now narrow once more to focus on the structure which forms the focal point for this thesis, the cytoplasmic ring. As just described, the

cytoplasmic or C-ring is one of the first components of the BFM to assemble and it is necessary for both torque generation and for the switching of the motor's rotation from a CW to CCW direction. It consists of multiple copies of four different proteins: FliG, FliM, FliN, and the C-terminal domain of FliF, each of which play different roles. Studies relating to the C-ring proteins have been published as far back as the 1960s[74], yet it was only in the early 1990s that a purification technique was developed which allowed intact C-rings to be visualized as part of the BFM[75]. Since that time, electron microscopy as well as continued genetic and biochemical studies have shed more and more light on the structures of both the C-ring as a whole and of its constituent proteins.

Though it appears to vary from species to species[76], the C-rings studied so far have diameters between ~30 and ~50nm[77] and in cryo-EM reconstructions show rotational symmetry corresponding to between roughly 20 and 40 subunits[78,79]. As can been seen in figures 1.7 and 1.8, the C-ring sits fully in the cytoplasm just below the MS-ring with the stator units positioned just above it in the inner membrane. The upper surface of the C-ring by the membrane is composed of the protein FliG which forms one half of the torque-generating interface and the rotation of the ring by the stators at this interface leads to the rotation of the rest of the motor. The proteins FliM and FliN are found farther from the membrane and form the bulk of the C-ring density found in the cryo-EM reconstructions[75]. FliG is the key protein in the work to follow however so in describing the current picture of the structure of the C-ring I will use it as the focal point.

FliG is a Central Hub of the Bacterial Flagellar Motor

The history of investigations into the structure and function of FliG begins at the same time as the first genetic studies of the flagellar motor. These consisted mainly of mutational screens conducted in the 1960s which picked out areas in the genome needed for swimming, switching, and flagellar assembly. The gene that would eventually be called FliG was one of the first to be identified as necessary for these phenotypes[74]. In the early 1980s, directed mutations of the FliG gene were used to confirm its necessity in the processes of switching, flagellar rotation, and assembly[80] and work with other nearby genes placed it in a proposed 'switch complex'[81]. Only once the purification

method which preserved intact C-rings was introduced in the early 1990s was this 'switch complex' linked to the cytoplasmic ring and the location of FliG within the motor began to take shape[75].

One of the most important steps in the story of FliG however came just before the first images of cytoplasmic rings. In 1992, a *Salmonella* mutant was accidentally created in which the genes for FliF and for FliG were fused, created a single FliF-FliG polypeptide chain[82]. Surprisingly, this produced functional motors and motile cells, thereby indicating that the two proteins were very close by in the motor and most likely had the same stoichiometry. This insight was quickly incorporated into the new picture of the BFM informed by the first C-ring-inclusive EM images of the motor, placing FliG correctly somewhere in close proximity to the cytoplasmic side of the MS-ring[83].

With its location within the motor for the most part defined, the structure of the FliG protein itself became a primary goal. In 1999, over 30 years after the first delineation of the FliG gene, the first crystal structure of part of the protein was solved, in this case roughly the C-terminal third of FliG from *Thermatoga maritima* (*T. maritima*)[84] (Fig. 1.9A). Building on this, in 2002, roughly two-thirds of the protein were crystallized[85] (Fig. 1.9B), and then finally, in 2010, the crystal structure of the entire *Aquifex aeolicus* (*A. aeolicus*) FliG protein was solved[60] (Fig. 1.9C). This was a major milestone, and represents the culmination of nearly half a century of work, starting from the hazy outline of a gene and ending with the precise three-dimensional location, atom by atom, of every residue in FliG. Using this structure as the model, it is beneficial for what is to come to now take an in-depth look, domain by domain, at FliG, highlighting its binding partners along the way.

A quick look at the structures in Figure 1.9 shows that FliG is entirely α-helical with flexible linkers connecting the helices. The residues are organized into roughly three globular domains joined together via two long α-helices. The three domains, (amino-terminal ($FliG_N$), middle ($FliG_M$), and carboxy-terminal ($FliG_C$)), each contribute a different function to FliG, and by being linked together, these functions can be integrated and coordinated. Thus FliG acts as a sort of hub in the motor, taking the torque-generating power strokes of the stators, the switching signals from the

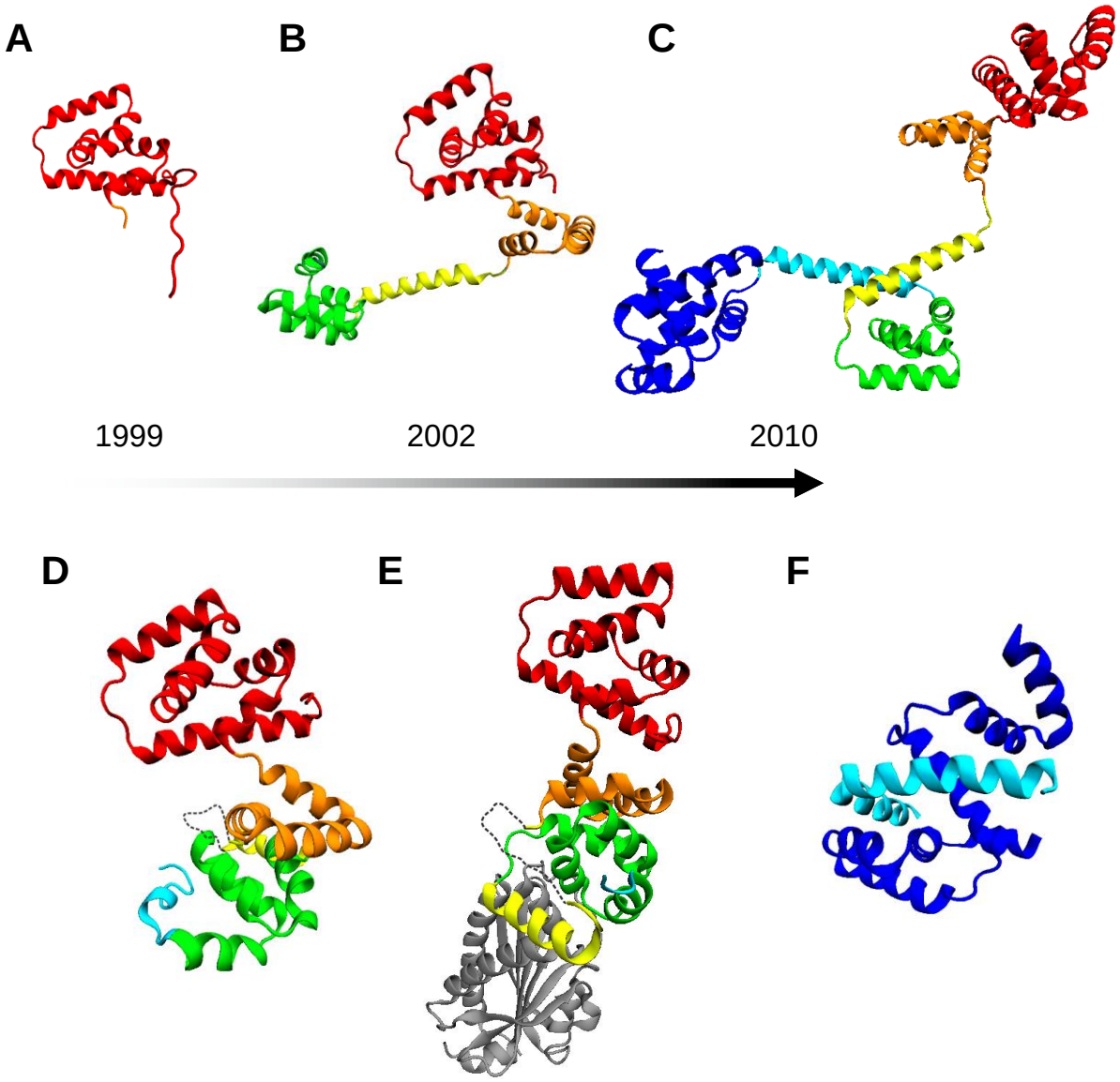


Figure 1.9 | The last two decades have seen nearly a dozen crystal structures of FliG published. A) The first crystal structure of FliG was published in 1999 and was of only the C-terminal amino acids[84]. **B)** In 2002 a crystal structure of both the middle and C-terminal domains of *T. maritima* FliG was published[85]. **C)** The first full-length structure of FliG was published in 2010 using proteins from *A. aeolicus*[60]. Structures published since have shown an alternate, unstructured confirmation for the region between the two ARM motifs (shown in **(D)** and **(E)** as a dashed line) which leads to a more compact conformation of the protein[86]. Several recent structures show FliG in complex with its binding partners. **E)** The M- and C-terminal domains of *T. maritima* FliG were crystallized bound to FliM (grey)[87]. **F)** The N-terminal domain of *T. maritma* FliG was crystallized bound to the cytoplasmic domain of FliF ($FliF_C$ in cyan)[88]. The colours of FliG's domains and motifs correspond to those shown in Figure 1.10.

chemotaxis pathway, and the underlying rotational symmetry of the MS ring and bringing them together to orchestrate the assembly of a rotating, switchable, cytoplasmic ring.

Starting from the N-terminus then, the first globular domain, called **FliG$_N$**, consists of around 90 amino acids arranged into five α-helices (blue in Fig. 1.10A), the first three of which are responsible for the main function of this domain: binding FliF[89]. This binding is a very important interaction in the motor and one of the first in the assembly pathway (see Fig. 1.8), acting to anchor the C-ring to the MS-ring and thereby coupling the rotation of the C-ring by the stators to the rotation of the rest of the motor. As such, the interaction between FliG and FliF is quite strong. Studies *in vitro* with purified FliG$_N$ and peptides corresponding to various lengths of the C-terminal region of FliF have demonstrated interactions with a K_d in the low nanomolar range (~40nM)[90] and as this interaction is repeated between 20 and 40 times around the ring of proteins, it appears that, once formed, the connection between the C- and MS-rings does not easily come undone.

These studies also determined that of the entire 60kDa FliF protein, only ~40 cytoplasmic amino acids at the C-terminus are responsible for binding to FliG$_N$[90]. Termed the cytoplasmic or **FliF$_C$** domain (Fig. 1.10B), these amino acids are also the only part of FliF to have a crystal structure and, even then, have only been crystallized in a complex with FliG$_N$ [88,91] (see Fig. 1.9F). This is due in part to the fact that FliF, like many other large membrane proteins, has proved very hard to work with *in vitro*, and that its FliF$_C$ domain had a great deal of flexibility, dangling like a fishing hook into the cytoplasm, until it binds to FliG[88].

The binding of FliG$_N$ to FliF$_C$ has gathered interest in the last decade as crystal structures of the two domains in complex have been published[88,91] (see Fig. 1.9F), and it is important to note it as the interaction is key to several of the results to be discussed in this thesis. In both *T. maritima* and *Heliobacter pylori* (*H. pylori*), peptides corresponding to the ~40 C-terminal amino acids of FliF have been synthesized and shown to bind to and crystallize with FliG$_N$, with this binding causing conformational changes in both partners[88]. In fact, it has been proposed that, upon binding, FliF$_C$ and FliG$_N$ fold around one another into a single functional domain similar in tertiary

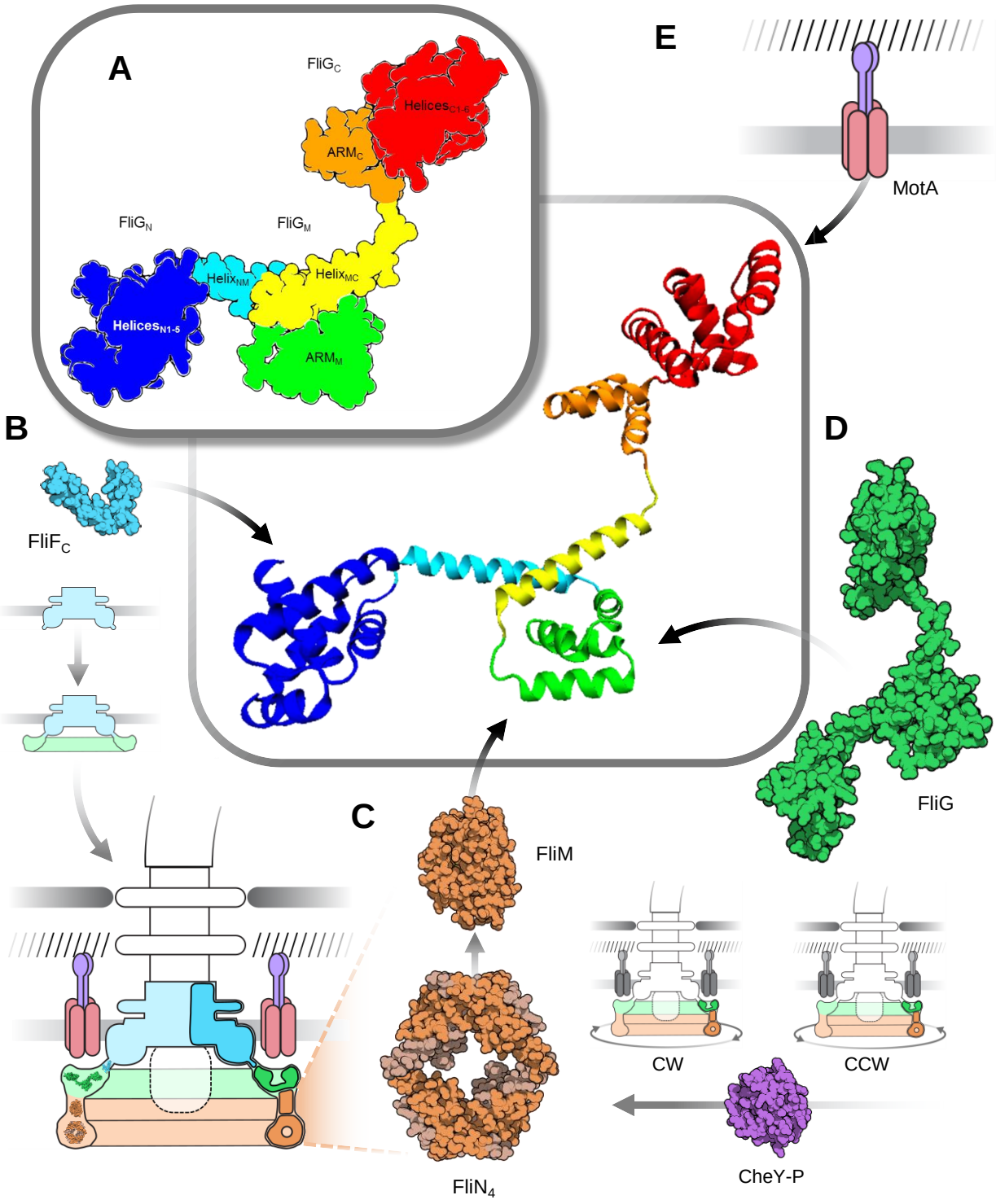


Figure 1.10 | FliG is a multi-domain protein with at least four binding partners: a hub of the BFM. A) The structure of FliG can be seen as three globular domains (N, M, C) connected by two long alpha-helices. An Armadillo motif sits within each of the M and C domains[60]. **B)** The cytoplasmic domain of FliF binds strongly to FliG$_N$ during the assembly of the BFM[89]. **C)** Phosphorylated CheY induces switching by binding to FliN which binds as a tetramer to FliM which binds to FliG$_M$. A conformational change propagates through the C-ring upon CheY-P binding.[98] **D)** FliG$_M$ is also involved in FliG-FliG interactions.[101] **E)** Torque-generation depends on the interaction of the 'torque helix' of FliG$_C$ with the stator protein MotA. The orientation of the torque helix determines the direction of the motor's rotation.[60]

structure to the middle and C-terminal domains of FliG[88]. From an evolutionary point of view, it makes sense that the same tertiary structure is repeated three times in close proximity as duplication of the domain followed by subtle mutations or its division between two polypeptides chains would be a simple explanation for such similar folds repeated not only within a single protein but over a two-protein complex.

Evolutionary history aside, once formed, the interaction is the strongest of any found in the C-ring and the most resistant to changes in temperature and pH[90,92]. This strength comes mainly from a great deal of hydrophobic interaction between the proteins, as $FliF_C$ forms into two α-helices with hydrophobic regions along their length which fit snugly into a large hydrophobic groove in the $FliG_N$ domain. Some electrostatic interactions around the hydrophobic patches shore up the binding and contribute to the low observed K_d[90]. In addition, some recent models indicate that binding of $FliF_C$ to $FliG_N$ induces a further conformational change in $FliG_N$, exposing hydrophobic regions which allow for interactions between neighbouring $FliG_N$ domains[91]. This is definitely an interesting proposition as it implies the coupling of one phase of the motor's assembly to the next and feeds into the overall discussion concerning interaction between neighbouring FliG molecules, a key topic in this thesis.

Moving beyond the N-terminal domain of FliG, towards the C-terminus, the next region encountered is a long α-helix called **$Helix_{NM}$** (cyan in Fig. 1.10A) which bridges the N- and M-domains of the protein. This helix is not involved in binding of other components and for the most part acts as a structural component, spacing out the N- from the M-domain[60]. The middle domain however, called **$FliG_M$** (green in Fig. 1.10A), is involved in a great deal of protein-protein binding. It is composed of four α-helices joined by flexible linkers, spiralling around in what has been termed an Armadillo motif (here **ARM_M**) due to its similarity to other spiralling, helical repeats of α-helices seen in eukaryotic proteins such as β-catenin[94] (see Fig. 1.12A). How this domain sits within the structure and interacts with other neighbouring FliG proteins is controversial and will be discussed shortly but it is known with great confidence to be the key binding site for the protein **FliM**[87] (Fig. 1.10C).

Four residues (EHPQ) on $FliG_M$ have been shown across many species to bind a four residue motif (GGXG) on FliM[94] with a K_d in the low micromolar range[95]. The position of the binding site on $FliG_M$ places FliM on the opposite side of FliG to FliF, meaning the FliM protein sits entirely in the cytoplasm. Its main functions lie both in sustaining the structure of the C-ring through interactions with its neighbouring FliMs and in coordinating the switching of the ring from CCW to CW conformations upon the binding of CheY-P[50]. A domain at its N-terminus, connected to the rest of the protein by a long, flexible tether is thought to bind CheY-P and sequester it, enabling its association with another region on the nearby protein **FliN**[96] (Fig. 1.10C). Along with FliM, FliN makes up the majority of the mass of the C-ring and it has been shown that four copies of FliN first pre-assemble into a homotetramer with dihedral symmetry and make a $FliM_1$-$FliN_4$ complex prior to joining the C-ring[97]. The binding of CheY-P to the FliN tetramer is thought to induce a conformational change which propagates up through FliM and finally via $FliG_M$ to the critical FliG residues at the torque-generating interface, although the molecular details of the different conformations are yet to be worked out definitively[98]. One final interesting aspect of FliM and FliN is that recent work seems to indicate a great deal of turnover of the $FliM_1$-$FliN_4$ complexes in the ring[99,100]. This is to say that even after the BFM has been assembled and while it is rotating, it appears that $FliM_1$-$FliN_4$ units come and go from the structure, varying their overall stoichiometry and even the size of the C-ring. It is puzzling to think how this is possible but should not come as a surprise considering the dynamic nature of the majority of protein complexes found in nature.

Carrying on then towards the C-terminus of FliG, after the middle domain comes the second of the two long domain-bridging α-helices, this one called **$Helix_{MC}$** (yellow in Fig. 1.10A). In contrast to $Helix_{NM}$ the conformation of this helix is controversial. As can be seen in Figure 1.9, some crystal structures show this stretch of amino acids to be one long α-helix (Fig. 1.9B, C), whereas others show most of it as an unstructured, flexible linker (Fig. 1.9D, E). Debate currently centres on which of the two conformations is physiological and which is an artifact of the crystallization procedure. Many believe that the conformation with most of the helix unstructured is what is found *in vivo* as all the crystal structures showing FliG bound to FliM display the helix in this conformation[87].

However, one recently-introduced model says that both conformations are in fact physiological and that FliG can transition between the two[101]. Testing of this model forms the backbone of the work in this thesis and it will be discussed more thoroughly in the sections to come.

Rounding out the structure of FliG is its C-terminal domain, **$FliG_C$** (orange and red in Fig. 1.10A). This is better thought of as two subdomains which can flexibly rotate around one another. The first, more N-terminal of these is another spiralling, Armadillo motif termed **ARM_C** and is composed of three α-helices (orange in Fig. 1.10A), while the six most C-terminal α-helices carefully position the key amino acids of the torque-generation interface (red in Fig. 1.10A). ARM_C will be dealt with in relation to ARM_M, $Helix_{MC}$, and a model concerning how they may fit into the overall C-ring structure in the following section but it can be thought of as mostly involved with FliG-FliG interactions[101] (Fig. 1.10D).

The primary responsibility of the C-terminal helices on the other hand is interaction with the MotA protein of the stator complexes (Fig. 1.10E). The fifth of these six α-helices (called the **torque-helix**) sits directly below the cytoplasmic residues of MotA and forms one half of the torque-generating interface[60]. Its positioning means that the power stroke of MotA pushes directly on the torque-helix and generates force on the protein which is translated into rotational motion by the circular nature of the C-ring (see Fig 1.8A). In addition, rotation has been found to require charged residues on the torque-helix which implies an electrostatic component to the interaction along with the presumed steric one[102]. To enable the switching of the rotation direction, current thinking dictates that binding of CheY-P to FliN generates a wave of conformational changes propagating through the C-ring and manifesting itself in the rotation and reorientation of the charged residues of the torque-helix in relation to MotA[60]. This flips the order of charges and in theory could reverse the direction of rotation without the power stroke of MotA changing[59]. The main stumbling block in confirming or denying this picture of torque-generation and switching is the lack of a detailed, three-dimensional structure of MotA[103], and this remains a key goal in the field.So, to bring this picture together, with its three connected globular domains, composed entirely of a series of linked α-helices, FliG acts as a major hub in the BFM, coordinating its most important actions. Its N-

terminal domain sets the shape and size of the C-ring through its binding to the underlying FliF ring template, while its middle domain mediates switching through contact with the CheY-sensing $FliM_1$-$FliN_4$ complex. Finally, its C-terminal domain forms an essential part of the torque-generation interface and is actively pushed by MotA of the stator complex, resulting in the rotation of the BFM. Thus, the self-assembly, switching, and rotation of the motor require FliG. Being such a central component, some of the biggest remaining questions concerning this protein complex have FliG at their heart and one of these key questions concerns the architecture of the C-ring itself and how the FliG proteins within the ring interact. In the next section I will present a leading theory which proposes an arrangement for the FliG ring and dictates the specifics of FliG-FliG interactions. Testing this theory forms the bulk of the work in this thesis so in addition to stating its claims, I will also provide some background on the general mechanism from which it has been developed.

A Domain-Swap Polymerization Model Outlines an Assembly Pathway for FliG

After over 50 years of work with FliG, its binding partners are fairly well established and the likes of FliF, FliM, and MotA have been introduced in the previous section. Importantly however, biochemical evidence indicates that FliG also interacts with itself[104-106]. In the cytoplasmic ring, with FliG molecules side-by-side, determining the nature of this FliG-FliG interaction is important in determining the overall architecture of the C-ring. Do its components form a continuous ring or discrete protein patches? Can its subunits easily be pulled apart or are they tightly interlocked? If a conformational change is necessary for switching, how does this propagate between proteins and around the ring?

Prior to 2010, many tentative theories had been published describing the inter- and intra-subunit architecture of the C-ring yet for the most part they fell short of describing one key aspect or another of the overall structure. Along with the publication of the full-length *A. aeolicus* FliG sequence came a follow-up study in 2016 with a fleshed-out model for how this structure would fit into both a CW and CCW-spinning ring and how it could transition between the two forms[101]. It

was well-developed and took into account all of FliG's domains and residues, something that had yet to be done due to the lack of a full-length structure.

Although it describes the actions and locations of each part of the protein, the major claims of the model centre on the ARM_M, ARM_C, and $Helix_{MC}$ regions of the sequence (see Fig. 1.10A), stating that these regions in particular are essential for FliG-FliG interactions. Its main breakthrough and most contentious point is that these ARM motifs mediate inter-protein interaction through a process of **domain-swap oligomerization**, a bold claim that caused much discussion in the field of BFM study[104,106].

To give some background, the process of domain-swapping was generalized as a concept in the early 1990s[107] to explain certain patterns seen in several oligomeric crystal structures of proteins dating back to the 1960s[108]. In this process, a 'mobile structural element' from one polypeptide chain (a domain for example) is exchanged with an identical structural element found in another, separate polypeptide chain of an identical sequence. These elements are generally attached to the rest of the protein by a flexible linker which allows them to leave their monomeric context and stretch over to sit in the context of the partner protein's amino acids in the same location as it would in its monomeric form (Fig. 1.11A). From its ubiquity across many oligomeric structures, it appears that domain-swapping is a common evolutionary trick for generating an inter-subunit interaction. Rather than slowly mutating single residues one-by-one until a binding site is developed, incorporating a flexible linker to allow regions of the protein to open up and explore around for other 'opened' proteins means that if they find one another, there is a stable, ready-made binding site available to generate a long-lasting interaction.

The most common domain-swapped structures are domain-swapped dimers in which the same domain on each protein in a pair is mutually swapped, creating a closed structure with two domain-swapped interactions[107] (Fig. 1.11B). The same process however can create closed oligomers of higher order, with the majority forming symmetric rings with subunits bound in a head-to-tail fashion[109] (Fig. 1.11C). In the most extreme case of domain-swap oligomerization, the subunits do

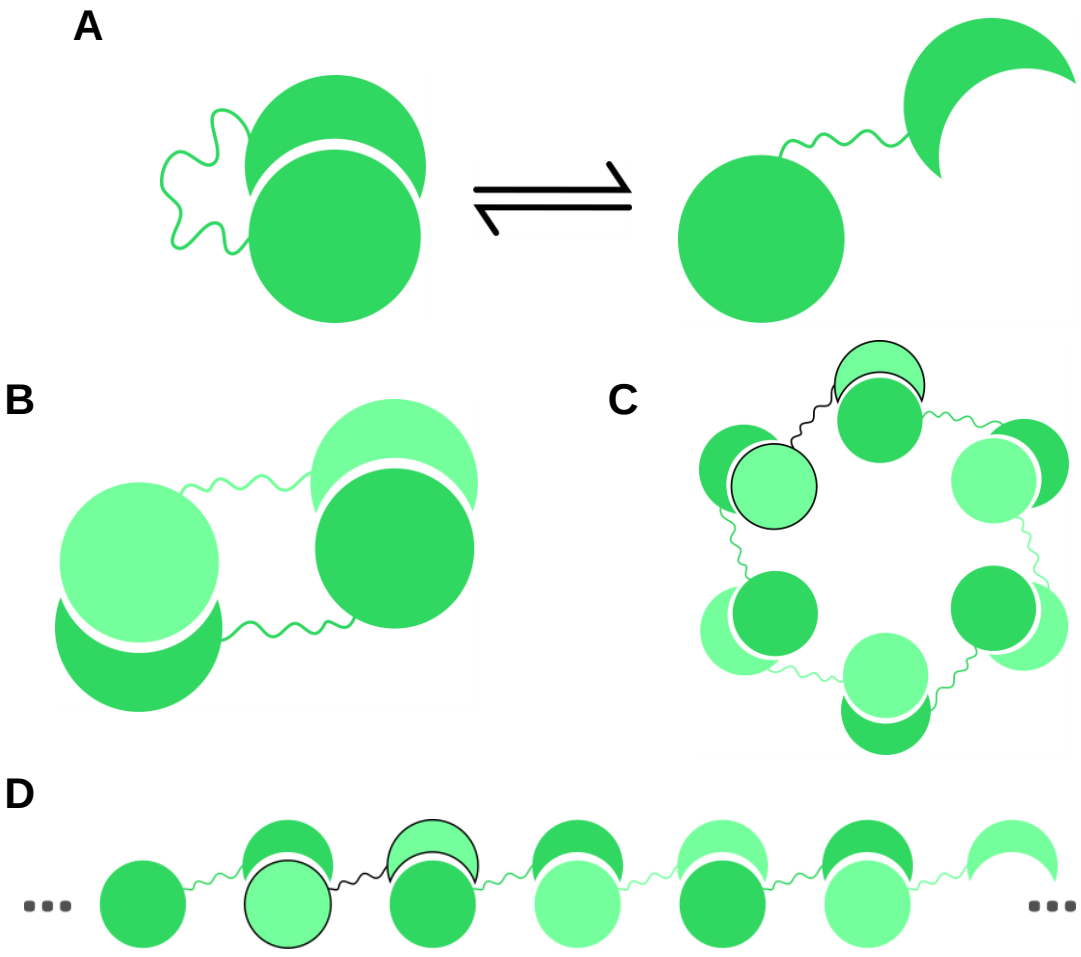


Figure 1.11 | Domain-swap is a common process by which proteins oligomerize. A) Proteins capable of domain-swapping contain a 'mobile structural element' (here the crescent) connected to the rest of the structure by a flexible linker. The protein transitions from its compact form (left) to extended form (right) when the mobile element breaks from the rest of the protein[107]. **B)** The most common domain-swapped oligomer is a domain-swapped dimer wherein two identical proteins mutually swap domains. **C)** Larger oligomers can also form, creating closed rings with subunits connected in a head-to-tail fashion. **D)** Large, open-ended polymers can form through domain-swap as is the case in the amyloid fibrils of Alzheimer's disease[31].

not close up into a ring and instead leave free, open ends available to build from, which can lead to the formation of 'runaway polymers', continually adding subunits onto the open ends to make larger and larger aggregates (Fig. 1.11D). Several types of amyloid fibrils have been shown to polymerize in such a fashion, creating the plaques seen in both Alzheimer's and Creutzfeldt-Jakob disease[31].

Taking this knowledge of domain-swapping as a starting point, the model proposed in 2016 set FliG's two ARM motifs, ARM_M and ARM_C, as the 'mobile structural elements', hypothesizing that FliG-FliG interactions had at their core **ARM_M-ARM_C interactions**. The ARM motifs are especially strong candidates for mediating FliG oligomerization as, akin to the ARM motifs seen in eukaryotic proteins (Fig. 1.12A), they both possess a very large hydrophobic surface on one of their faces (Fig. 1.12B). These hydrophobic regions do not like to be exposed to an aqueous environment and in all the crystal structures of FliG to date they are hidden away from solution, with the hydrophobic patch of ARM_M binding and sequestering the commensurate patch of ARM_C[60, 85-87].

Although in every crystal structure an ARM_C is found with an ARM_M, surveying the results shows that there are in fact two distinct forms present. In the first, the ARM motifs of a single polypeptide chain are bound to one another intramolecularly. Crucially, in this class, $Helix_{MC}$ is found to be partially unstructured, with the majority of the amino acids which connect the two ARM motifs taking the form of a flexible loop[87,111] (see Fig. 1.9D, E and Fig. 1.12C left). In the second class of structures found, the two ARM motifs are instead separated by a fully rigid and α-helical $Helix_{MC}$, making them unable to contact one another in an intramolecular interaction[60,85] (see Fig. 1.9B, C and Fig. 1.12C right). Instead, the ARM_C motif of one polypeptide chain stretches over to interact with the ARM_M motif of a neighbouring chain, creating an intermolecular bond (Fig. 1.12D). In such structures, the ARM motifs can be considered to be domain-swapped and the crystals show head-to-tail domain-swapped oligomers of FliG molecules bound together via ARM_M-ARM_C interactions[60,85].

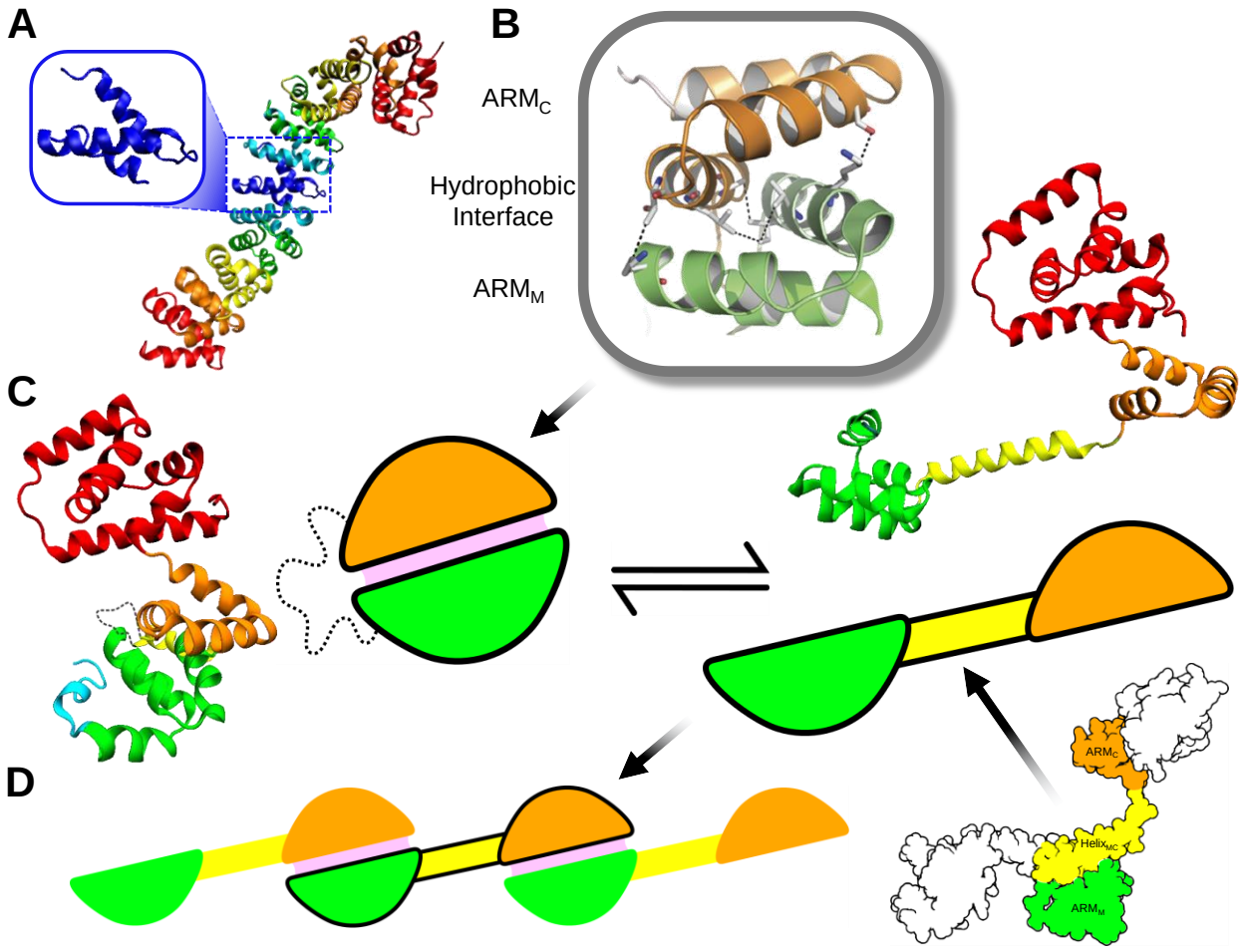


Figure 1.12 | One model proposes that hydrophobic interactions between Armadillo motifs are responsible for FliG-FliG bonds. A) Eukaryotic proteins such as β-catenin contain repeats of spiralling helices known as Armadillo (ARM) motifs[110]. **B)** FliG possesses two structures similar to eukaryotic ARM motifs, one in its C-domain (ARM_C) and one in its M-domain (ARM_M). Crystal structures show these two motifs held together by a hydrophobic interaction[101]. **C)** Two different conformations are present for FliG in published crystal structures. The first shows the ARM motifs stacked in a compact, intramolecular fashion[111]. The second shows the structure extended with a rigid $Helix_{MC}$ and ARM motifs interacting in an intermolecular fashion (as in **(D)**)[60]. A domain-swap model proposes that the two structures can interconvert[101].

With two observed forms, which one is correct? Which of the two is physiologically relevant and present in the C-ring and which is a crystallographic artefact? The model introduced in 2016 addresses these questions by saying that, in fact, both forms are physiologically relevant and that FliG can transition between the two[101]. This is perhaps its most substantial claim and the one that, before 2016 had no experimental verification. The study however provided small angle X-ray scattering (SAXS) and *in vivo* cross-linking data to back up their claims of a transition from intra- to intermolecular ARM-ARM interactions and it has proved quite popular due to its well-developed explanation of how FliG might assemble in the C-ring.

The model (see Fig. 1.13) states that at first, FliG resides in the cytoplasm as a monomer with its ARM motifs hidden away in an intramolecular interaction. This prevents any unwanted FliG polymerization which could lead to potentially cytotoxic protein aggregates. During the assembly of a BFM however, a FliF ring forms in the inner membrane with its $FliF_C$ domain dangling into the cytoplasm. These dangling 'fishing hooks' bind to FliG via its N-terminal domain and sequester the monomers one beside the next in a ring whose size and shape is templated by that of the FliF ring. Once sequestered in such close proximity, random thermal fluctuations would destabilize intramolecular ARM-ARM links and at some point would lead to two 'opened' FliG monomers side-by-side, allowing for the formation of an intermolecular ARM-ARM bond. These intermolecular ARM-ARM bonds would propagate around the ring, with each FliG finally ending up in an open, 'extended' form and the subunits all strongly linked together via the hydrophobic ARM-ARM interactions between neighbours.

The strengths of this model are that it gives a very descriptive pathway for FliG assembly into the C-ring (something most previous models did not), and that the interconnectedness of its subunits feeds nicely into the experimental observations of quick, coordinated switching of the rotation direction of the ring. It is also quite bold however as, should it prove true, this would be the first system where the extent of domain-swapped polymerization is controlled by an underlying template (in this case the FliF ring). In all other cases the extent of polymerization is controlled

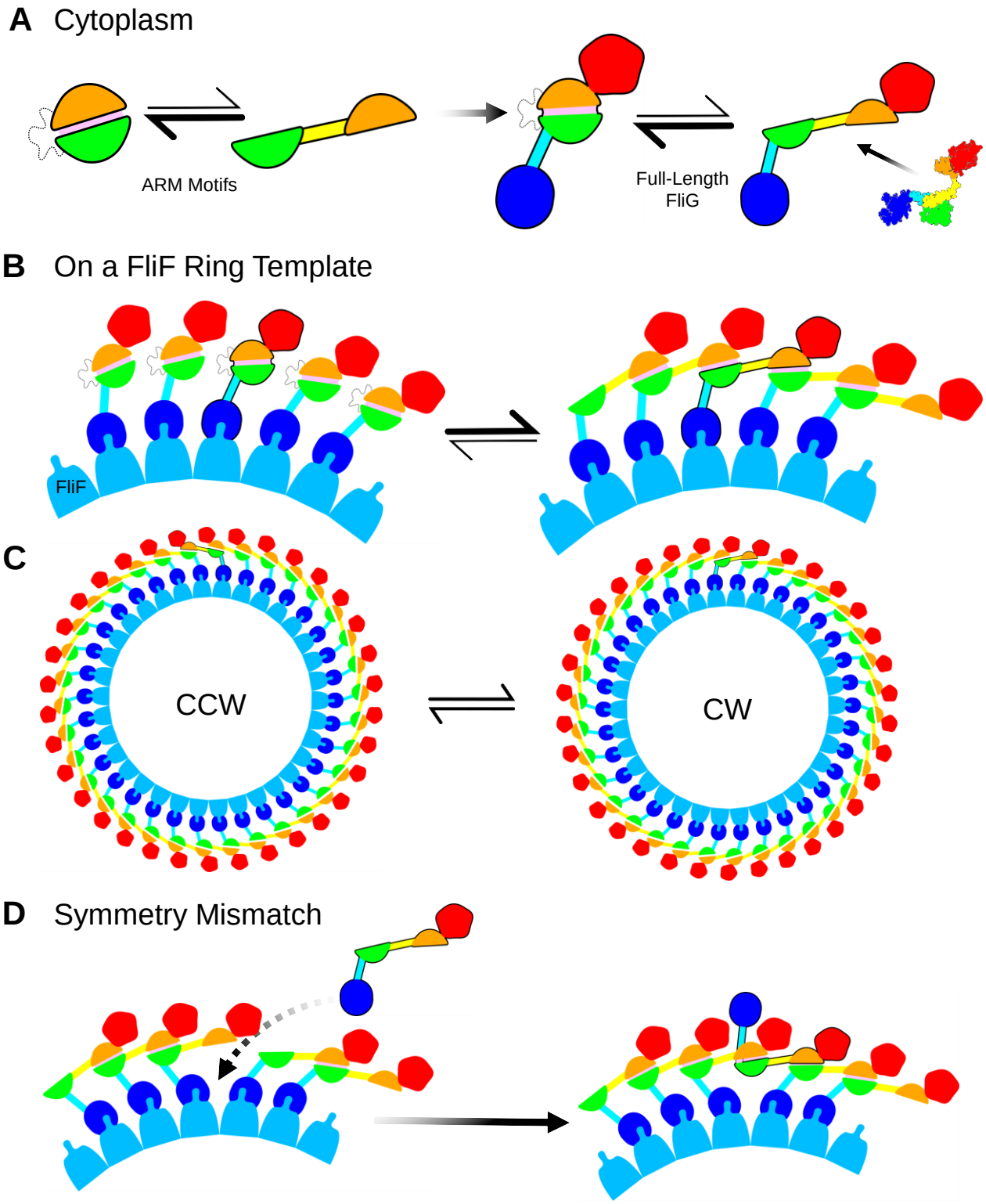


Figure 1.13 | In the domain-swap model of FliG polymerization, intermolecular interactions are triggered upon binding to a template. **A)** Compact and extended forms of FliG interconvert but in the cytoplasm the compact form is favoured due to its stable intramolecular ARM_C-ARM_M interaction. **B)** When FliG binds to the FliF ring, the close proximity of its neighbours mean stable intermolecular ARM_C-ARM_M interactions can form. These are favoured over intramolecular interactions due to the lower free energy of a fully α-helical $Helix_{MC}$. **C)** FliF controls the extent of FliG polymerization by constraining it to a finite ring. Conformational changes in the subunits dictate if the ring spins CCW or CW. **D)** A discrepancy in the observed symmetries of the FliF and FliG rings has been explained by the incorporation of FliG molecules through their binding to neighbouring FliGs and not to the underlying FliF ring.

intrinsically by the geometry of the subunits, or, as in the case of amyloids, it is not controlled at all and leads to runaway polymerization[31].

On a side note, when the model was published it also gave an explanation for one of the other big questions in the field. Cryo-EM reconstructions of the motor's MS- and C-rings had shown a symmetry mismatch between the two components, with the MS-ring appearing to be made of 26 subunits and the C-ring to be made of 34[79]. The ability of the FliG proteins to open up once bound to FliF and create a free, unbound ARM motif meant that it was possible for a cytoplasmic FliG to come and bind to this ARM motif and thereby enter the ring, without even contacting the underlying FliF, thus increasing the overall FliG count beyond that of FliF (Fig. 1.13D). Some recent work has cast doubt on the stoichiometries given for the MS- and C-rings in these previous studies however, hinting that the stoichiometries of the two rings are in fact the same (at 34 subunits each) and that the previous mismatch can be explained by errors due to the averaging procedures in the reconstructions *(unpublished, K. Namba)*.

Though it is a well-formulated model, a detailed examination of the data presented in the 2016 study shows that, though highly probable, definitive proof that this template-induced domain-swap oligomerization of FliG does in fact take place in the bacterium is still lacking. It presented three main lines of evidence supporting this claim: evolutionary covariance analysis of the ARM residues, SAXS data for FliG in buffers of various ionic strengths, and *in vivo* cross-linking of double-cysteine FliG mutants, and though they all lend some degree of support to the theory, an in-depth look shows that there is still work to be done before the domain-swap model can be universally accepted.

The evolutionary covariance analysis of the residues in over 2000 FliG protein sequences showed that the hydrophobic ARM-ARM interaction was in fact conserved and therefore important[101], yet it is not possible to say whether the ARM-ARM interactions this method picks out are of an intramolecular or intermolecular nature. The SAXS data on the other hand showed that the FliG molecule can in fact take two different forms, as placing the protein in buffers with varying salt

concentrations produced two distinct shapes: one with a radius of gyration (R_g) equal to that expected for a compact monomer, the other with an R_g closer to that expected for an extended FliG molecule[101]. Although this convincingly shows that the protein can transition between two different conformations, with the simplest explanation being that these are the compact and extended structures predicted by the model, it does not prove that this transition is physiological and it is just as likely that one of the two forms observed is never found in the bacterium.

The final experiments, in which double-cysteine FliG mutants were cross-linked and their oligomeric state compared in either the cytoplasmic or membrane fraction of the cell, looked to prove that this transition was in fact physiological and that the template provided by the motor triggers this change. Though monomers were seen in the cytoplasm and oligomers in the membrane fraction as the model predicts[101], serious objections have been raised concerning the methods used[104,106]. These experiments did not take into account the fact that the cytoplasmic fraction was greatly diluted after cell lysis, meaning that oligomers that might have been present in the cytoplasm could have fallen apart before the cross-linking. The differences in cross-linking shown by the paper could therefore be due to differences in the concentration of the FliG proteins and not as a result of template-induced oligomerization.

Thus, though the covariance analysis shows that the ARM-ARM interaction is important and the SAXS data shows there are two possible forms that FliG can transition between, there is still no strong evidence that both forms are physiological or that transitioning from one form to the next is triggered by incorporation of FliG into the assembling BFM. In order to definitively prove the domain-swap model, one would have to convincingly show a transition from the compact, monomeric conformation of FliG to the extended, oligomeric form upon binding to an underlying template.

Taken in sum then, the two ARM motifs appear to be strong candidates for mediating FliG-FliG interactions. The variable nature of the amino acids connecting ARM_M to ARM_C, with some crystal structures showing a flexible chain and others showing a rigid α-helix, has fed into a model where

the two forms interchange. It dictates that the compact form with flexible linker predominates in the cytoplasm and the extended form with rigid helix predominates in the ring. Following a common oligomerization technique observed in many other protein complexes, the model proposes that intermolecular FliG interactions are the product of a domain-swap polymerization between the ARM motifs. Its boldest claim however is that this polymerization is held in check due to the underlying presence of a circular template, namely the FliF ring. **The main focus of this thesis is to test the validity of this domain-swapped model for FliG oligomerization** with the work to follow attempting to experimentally verify its claims.

But how can this model be tested? How can its predictions be either confirmed or denied? When asking these questions, almost immediately technical considerations begin to get in the way. Obtaining atomic-resolution images of a working C-ring goes beyond what the current cryo-EM techniques are capable of and at the moment these can at most provide only vague outlines of where the proteins sit. More indirect methods are required which infer positions of subunits and secondary structures based on other readouts. Ideally these methods would be applied to a working, rotating BFM but trying to measure anything accurately in a living cell, even one as simple as a bacterium, is historically quite a challenge. An inviting alternative is to instead investigate the protein components *in vitro*, taking them outside of their complicated cellular environment and studying them piece-by-piece in isolation. Some work has been done in the past with purified BFM components[92] but a major stumbling block has been the isolation of its membrane-embedded components, most prominently the protein FliF. Due to its large size and membrane-spanning nature, the FliF protein is very challenging to work with *in vitro* and, apart from its small, cytoplasmic $FliF_C$ domain, no crystal structures for the protein exist. Yet the model's biggest claim revolves around the actions of a FliF template on the polymerization of FliG. A simplified *in vitro* system provides many benefits but without the FliF template how can the model truly be tested?

In the next section, I will describe the primary tool employed in this thesis to move past these technical hurdles. It is a deviation from standard biochemical methods for studying proteins so it pays to first describe it more generally before then moving on to how it can be implemented for

use in the study of the C-ring. Most crucially, this tool provides a way to test the domain-swap model's central claim and I will describe in depth the technique by which one can generate nanoscale templates for FliG to assemble on, made not from FliF but from DNA.

1.5 The Tools of DNA Nanotechnology Provide a Way to Test the FliG Domain-Swap Hypothesis

The domain-swap model for FliG oligomerization depends crucially on the proximity of one FliG molecule to another. Floating freely in the cytoplasm, FliG-FliG interactions are rare and a compact monomer is favoured (Fig. 1.13A). Being closely packed once tethered to the FliF template however leads to the extended state being favoured and to the formation of FliG oligomers (Fig. 1.13B). A key requirement for testing this model then is to be able to test the affect of FliG molecules' proximity on their oligomeric state. And there is the added possibility that perhaps the *in vivo* FliF template does more than simply increase the local FliG concentration. Perhaps the spacing and orientation that it gives to its bound FliG molecules are also crucial factors in the oligomerization process. To effectively test this model, there needs to be a way to place FliG molecules in specific locations with nanometre precision and then monitor if (and ideally how) they interact. In addition, any method used has to be compatible with conditions approximating an *in vivo* setting for FliG, working well in a salty aqueous environment.

Although there are several biocompatible ways to arrange proteins[112,113], given the precision needed to approximate the FliF template at the nanometre scale, there is one material which stands out from the others when looking to construct a FliF-mimicking template, and which provides by far the best means for testing the domain-swap hypothesis. As the information storage molecule of living things, DNA (deoxyribonucleic acid) is robust, stable, is (almost by definition) biocompatible, and beyond this has several key factors which make it easy to work with and ideal for the system at hand. Since the discovery of its structure in 1953[114], knowledge and protocols for working with DNA have rapidly accumulated and now a whole suite of tools and techniques exists surrounding the molecule. It can be cut with restriction enzymes, pasted together with

ligases, amplified through the polymerase chain reaction (PCR), and synthesized base-by-base quickly and cheaply by automated machines. And although it may not have the same in-built functionalities that amino acids give to proteins, DNA is easily-modified chemically, thereby extending its capabilities beyond simple information storage.

The most important point about DNA in the context of DNA nanotechnology however is its ability to predictably and reliably form a double-stranded duplex between two strands possessing specific, complementary nucleotide sequences (Fig. 1.14A). Taking the Watson-Crick base pairing of an A (adenosine) with a T (thymine), and a G (guanine) with a C (cytosine) as a foundation and coupling that knowledge with a few simple design rules, DNA sequences can be generated which bind exclusively to one another even in a sea of other strands with different sequences. Going beyond a straightforward duplex, applying these same design rules can produce more complex, self-assembling nanoscale DNA shapes and geometries held together simply by the hydrogen bonds between nucleotides.

This simplicity and design-ability forms the basis for all of DNA nanotechnology and in this section I will briefly highlight some of the key milestones and innovations in the field as it has developed over the course of nearly four decades. Following this I will narrow in on the use of DNA as a means to template and organize macromolecules such as proteins before tying this into the FliG system at hand. The field is now far too broad to give a comprehensive overview so I will focus mainly on aspects or developments which are pertinent for the work to come and which relate in some way to my main focus, the DNA-templated assembly of the bacterial flagellar motor's cytoplasmic ring.

Nearly Four Decades of Work Has Produced Major Advances in DNA Nanotechnology

The discovery of DNA's double-helical structure in the 1950s was a true breakthrough but it was not until the 1980s that the molecule began to be used not in its biological, genetic capacity, but for its properties as a building material. The first structures to be built with DNA were based off of

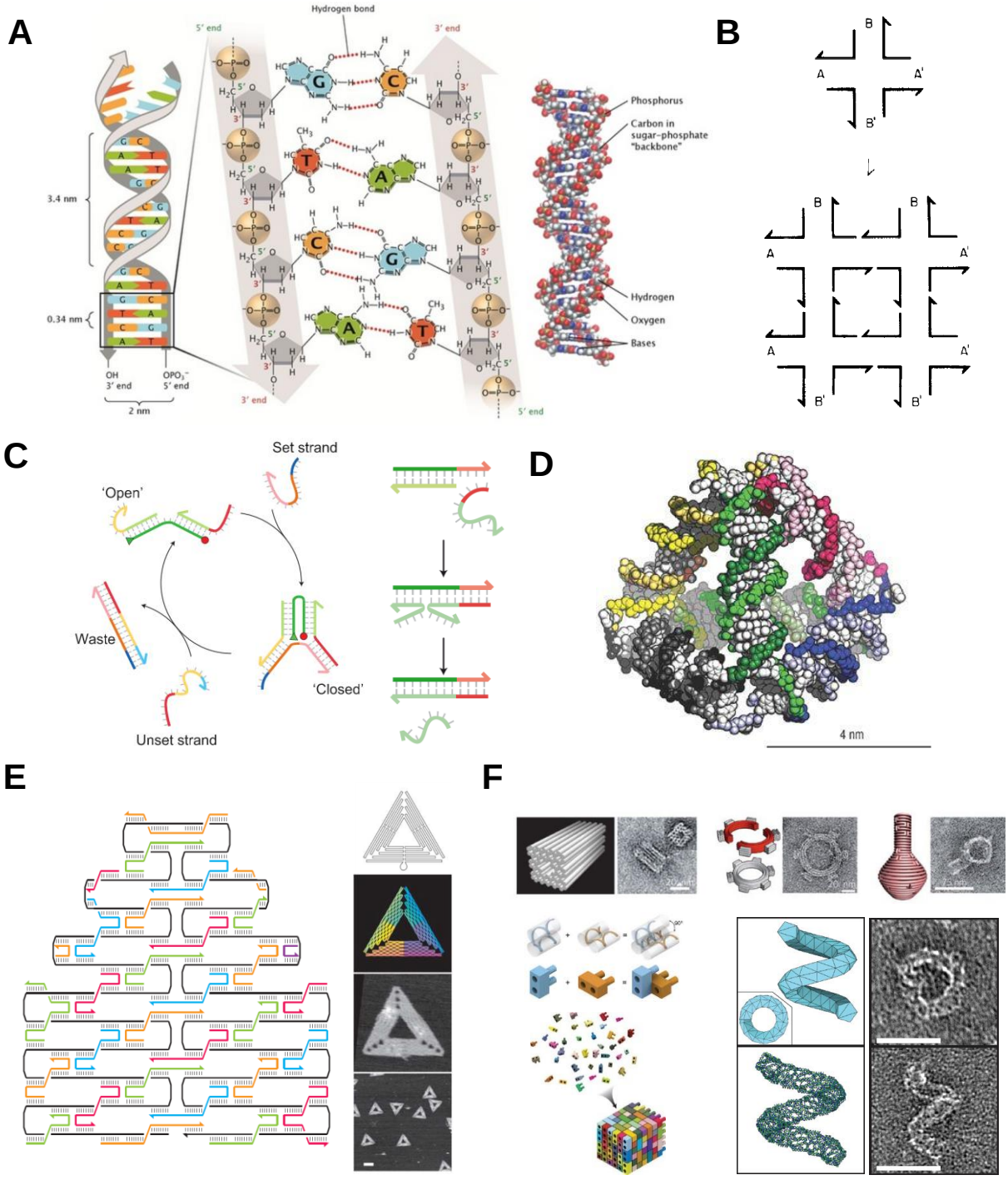


Figure 1.14 | Over the course of nearly 40 years, DNA nanotechnology has seen some major innovations. A) The DNA double helix completes a full turn in ~3.4nm or 10.5 bases[115]. **B)** The first DNA nanostructures proposed were based off of four-strand junctions which could bind via 'sticky ends' into regular arrays[116]. **C)** The first dynamic DNA nanostructure, the 'molecular tweezers', could be opened and closed using toehold-mediated strand displacement (right)[117]. **D)** A small tetrahedron was one of the first 3-D objects to fold in high yields[118]. **E)** In DNA origami, a long scaffold strand (black) is folded into a shape by short staple strand (coloured)[119]. **F)** Many new design concepts have been introduced in recent years. 3-D and curved origami (top row) appeared in 2009 [120,121]. Structures made out of DNA bricks (bottom left)[122] and based off of triangle meshes (bottom right) have also appeared within the last decade[123].

a motif found in nature called the Holliday-Junction and consisted of the hybridization of four synthetic oligonucleotides into a cross-like shape[116]. The crosses were left with unpaired single-stranded 'sticky ends' on their arms which allowed for their association into tiled arrays (Fig. 1.14B). These simple DNA arrays paved the way for all that came next with the 1980s and 1990s producing a wide range of two-dimensional DNA lattices and short-oligonucleotide-based structures.

From this static, two-dimensional foundation the field soon broadened and explored the various capacities of DNA as a building material. Moving away from the static DNA arrays, dynamic structures were also developed, capable of transitioning between multiple states. The first, a system called 'molecular tweezers'[117] appeared in the year 2000 and took advantage of the now ubiquitous process of 'toehold-mediated strand displacement' wherein a short unpaired region of one strand is used by another 'invading' strand to peel it away from its duplex partner (Fig. 1.14C). Other dynamic, multi-state systems soon followed, with many researchers developing complicated pathways capable of responding to a specific molecular input, often incorporating logic gates and basic computing functions[124] while others pursued the design of motor-protein-mimicking 'DNA walkers' capable of directed motion along a DNA track[125]. Another branch of the field looked to move beyond two-dimensions to construct three-dimensional structures out of short, self-assembling oligonucleotides. The first design to accomplish this successfully at high yields was a small DNA tetrahedron developed in 2004[118] (Fig. 1.14D) and several others came soon after.

One of the most important developments however came in 2006, with the introduction of a new design paradigm for DNA structures called DNA origami[119]. Rather than building structures out of small oligonucleotides, DNA origami takes a long 'scaffold strand' (often a circular viral plasmid) and folds it into a desired shape through the use of small 'staple strands' which bind and bring together different sections of the scaffold (Fig. 1.14E). This concept allowed for the high-yield production of self-assembling structures with dimensions of around 100nm exhibiting an amazing range of shapes and geometries (Fig. 1.14E). Though these first shapes were built in only two-dimensions, this new paradigm sparked a large wave of projects and designs, and only three years

later, in 2009, three-dimensional and curved origamis were being produced[120,121] (Fig. 1.14F). Alongside these structures, computational design tools were developed which made it even easier to construct arbitrary shapes and geometries through the simple folding of a scaffold strand[126] and this ease of use has led to the adoption of DNA origami structures as a tool in many fields quite distant from DNA nanotechnology.

In the last decade, new ways of building with DNA have been introduced, with some structures being constructed from 'DNA bricks'[122], and others taking an open, wireframe architecture[123] (Fig. 1.14F). Alongside these developments in design, larger, more complex, and more dynamic assemblies have been produced at an accelerating pace. One recent structure, the largest finite one to date, is a dodecahedron incorporating over 50 000 nucleotides which reaches over 400nm in diameter and over a gigadalton in weight[127]. Another is a DNA pore which can poke holes in lipid membranes and form a channel between two regions[128]. Other designs form cages which encapsulate and deliver drug molecules to specific targets, opening up to release their cargo[129].

So, in less than 40 years, DNA has been moved from its natural context as the simple double-helical molecule of heredity to the realms of nanotechnology as the building block for tubes and tiles, channels and cages. Among the many useful properties of DNA that have been exploited by this wide range of structures, in this project, for constructing a FliG template, the main utility of DNA nanostructures is the fact that they can be functionalized at very specific locations with specific molecules. Many previous studies have taken advantage of this fact to attach various different kinds of molecules to DNA structures for a wide range of purposes and it is worth highlighting a few of these now before defining what sort of template would be needed to study FliG.

DNA Nanostructures Can Position Proteins with Nanometre Precision

DNA in and of itself is relatively inert. It does not catalyze a reaction or possess a wide range of chemical properties as proteins do and this limits the possible applications that a DNA nanostructure might have. It is however, very easily chemically-modified with functional molecules

and a great deal of knowledge exists on how to attach reactive groups and other moieties at various points along the phosphate backbone and amongst the bases. This has led to the incorporation of various molecules in DNA-based designs to create nanostructures with a wide range of functions.

Some of the first uses of functionalized DNA occurred as far back as the 1990s, when thiol-modified strands were bound to gold nanoparticles, with the hybridization of the strands bringing the particles together into an aggregate with new optical properties[130]. Since that time, gold nanoparticles have been incorporated into many DNA nanostructures through the same thiol-gold interaction, often to take advantage of the particles' optical or electronic properties[131,132]. Other non-biological components such as carbon nanotubes[133], organic dyes[134], and various small molecules[135] have also been incorporated into DNA structures and a whole subfield has developed with the aim of templating the construction of nanowires for use in nanoelectronics[136] (Fig. 1.15A).

In the founding papers of the field however one of the key motivations for generating regular DNA lattices was to template the assembly of a regular array of proteins for use in crystallographic studies[116] and in the four decades that have followed, there has been a continuous and concerted effort to produce protein-DNA nanostructures both for that purpose and others. Some of the most successful studies have used proteins which naturally bind to specific DNA sequences and have used nanostructures as tools to study these DNA-binding proteins. These studies have given insights into how certain proteins bend the DNA duplex[145], or methylate certain sequences[146], and in general have paved the way to demonstrate the utility of DNA nanostructures in understanding protein function.

It is only relatively recently however that researchers have started to use DNA nanostructures to look at the interactions between individual proteins by taking advantage of the ability of these structures to precisely position molecules. In 2005 zinc finger motifs were engineered into two halves of a split GFP (green fluorescent protein) molecule and the binding of the zinc fingers to nearby regions on a DNA strand brought the two halves of GFP together, leading to its

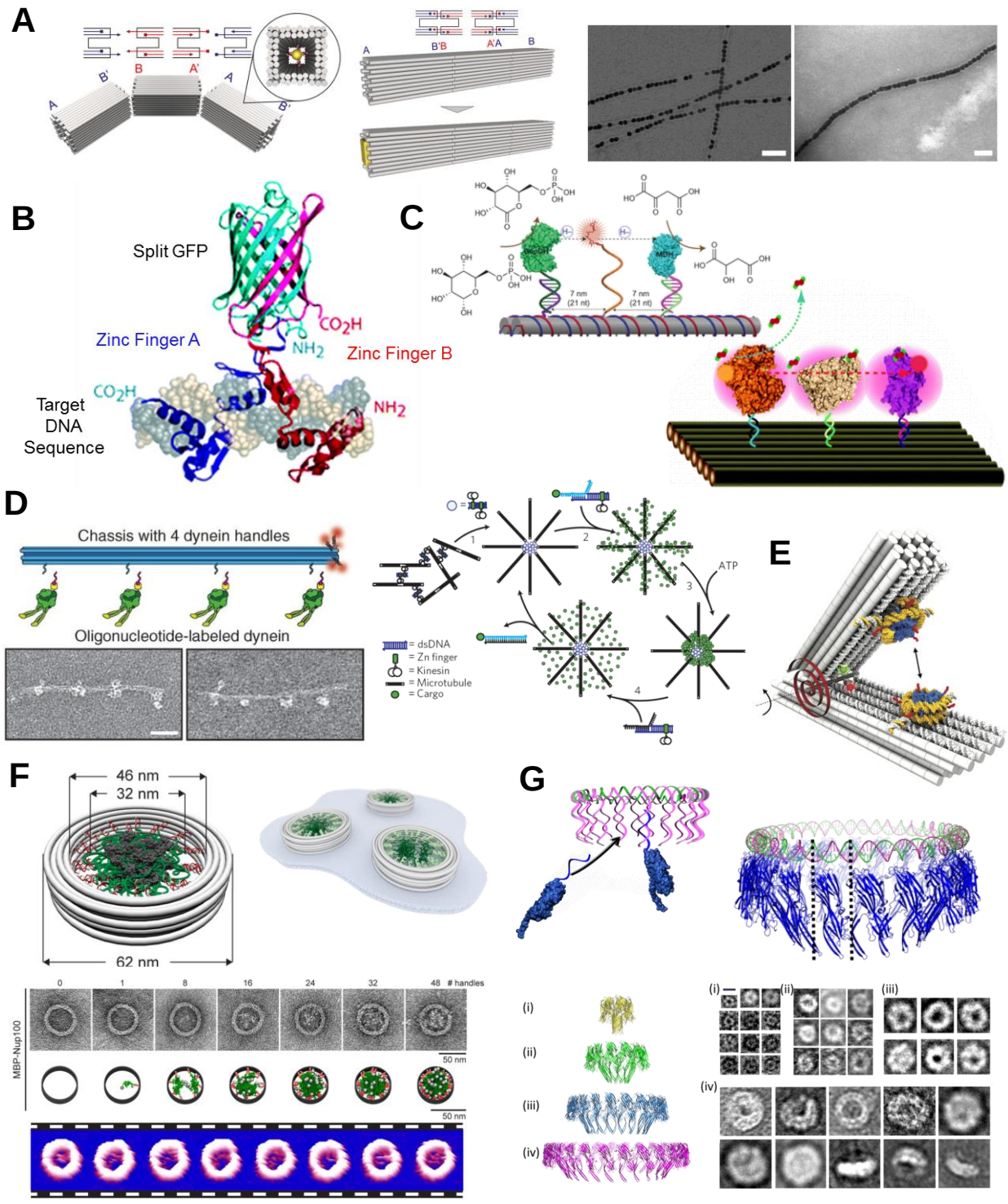


Figure 1.15 | DNA nanostructures have been used to precisely position and thereby study a wide range of molecules including proteins. **A)** Gold nanoparticles are anchored within a DNA origami 'mold' which confines the direction of the particles' growth into long 'nanowires'[136]. **B)** Two halves of a split GFP molecule are brought together by the binding of engineered Zn finger domains to a DNA template[137]. **C)** Enzyme cascades have been organized on DNA scaffolds, thereby increasing the efficiency of the path from reactant to product[138,139]. **D)** Dynein and kinesin molecules have been arranged along a DNA origami rod to investigate their combined actions on microtubules (left)[140]. Kinesins linked together via Zn-finger-to-DNA interactions can form radial microtubule arrays (right)[141]. **E)** A DNA origami 'force spectrometer' was used to investigate the interaction between nucleosomes[142]. **F)** Various nucleoporins were arranged inside a DNA origami cylinder to investigate how they occluded the inner channel[143]. **G)** α-Hemolysin proteins were arranged on circular DNA templates with varying numbers of binding sites. Rings with various diameters were seen using EM[144].

fluorescence[137] (Fig. 1.15B). More advanced work followed, with several studies placing enzymes involved in a single reaction cascade together side-by-side on a DNA scaffold and demonstrating an increase in final product formation due to their co-localization[138,139] (Fig. 1.15C). Other studies have arranged the motor proteins kinesin and dynein along DNA constructs to study the effects of their combined efforts. The first of these pitted kinesin against dynein bound to a single scaffold and monitored the net movement of the microtubules to which they bound[140] while the second connected kinesins together via DNA in such a way that the actions of the DNA-bound kinesins created a radial microtubule network[141] (Fig. 1.15D).

Apart from the simple recombination of split GFP, in the studies just listed, proteins interact only indirectly, either by passing substrate between one another or, in the case of the motor proteins, through their action on the same microtubule. Placing separate protein molecules on a template with nanometre precision and physically inducing the direct interaction between the bound components is at the cutting edge of current work in the field and only a few studies have been published so far showing this templated interaction. One such study in 2016 looked at the interactions between histone octamers by placing one nucleosome on either arm of a DNA 'nanocaliper'. Interactions between the histones brought the two arms together, as monitored by EM imaging, and by varying the placement of the histones along the arms, an idea of the strength of binding was obtained[142] (Fig. 1.15E). Another study, this in 2018, assembled different types of unstructured nucleoporin molecules on the inside of DNA origami cylinders and monitored how well the central channel of the cylinder was occluded with different numbers, densities, and types of nucleoporins using high-speed AFM (atomic force microscopy)[143] (Fig. 1.15F). A final study from late 2017, used a simple circular DNA template to assemble α-hemolysin proteins into rings[144] (Fig. 1.15G). The proteins naturally form heptamer barrels but various templates were able to accommodate larger numbers of proteins appearing as bigger rings in EM images and letting more current through across a membrane.

In all the studies just listed however, the proteins in question naturally interact with one another even without a template: histones bind to one another in chromatin[147], nucleoporins form dense

aggregate meshes *in vitro*[148], and α-hemolysin forms solid heptameric barrels[149]. The DNA structures used are tools to study the protein-protein interactions rather than the assembly process of the final structure. The C-ring of the BFM is more complex. At cytoplasmic concentrations, its subunits FliG and FliM are monomeric[101, 104], exhibiting no long-lasting protein-protein interactions. To function as a rotating oligomeric ring, it seems that there needs to be a template to induce interactions. As opposed to all the studies which have come before, in this system a template is not an optional addition but rather a functional necessity.

The work to follow then, is attempting something not done before. **This project is the first to use a DNA scaffold to study the assembly pathway of a large protein complex.** Though it is ambitious, the system has the benefit of nearly 70 years of accumulated knowledge on the BFM and nearly 40 years of work in the field of DNA nanotechnology. From the first lattices, through the first dynamic and three-dimensional objects, on through the introduction of DNA origami to the wide range of varied DNA structures and devices seen today, this work represents a new test for this versatile tool: to investigate one of the most complex, dynamic molecular machines found in nature. A great many studies have brought together DNA with inorganic components like gold nanoparticles or carbon nanotubes and others have incorporated a wide range of proteins, but no system investigated so far quite approaches the complexity of the BFM or even of the C-ring.

In the final section of this introduction, I will bring the previous sections together to give a focussed look at the project's goals and its plan for achieving those goals. As mentioned previously, testing the domain-swap polymerization model for FliG is the project's primary aim and in this next section I will outline several testable predictions which the model presents. The next section also functions to lay out how the work in the following chapters is organized, providing a guide to the experiments and results to come.

1.6 The Primary Goal is to Test the Domain-Swap Hypothesis Using an *in vitro* DNA Template

So, looking back now, I have described the hierarchy of protein structure from primary sequence to quaternary assemblies, and I have introduced the common suite of techniques that proteins use to perform their functions such as binding to various partners and changing their conformation (1.1). Following on from this I then described a few of the benefits for building the large multi-protein complexes that are so common inside a cell such as the ability to coordinate various proteins' functions and the increase in overall stability of complexes as compared to monomers. In this same section I also highlighted three key traits of protein complexes: their ability to self-assemble, their dynamic nature, and their high degree of symmetry (1.2). This set the stage for an introduction to the bacterial flagellar motor, a self-assembling, dynamic, and highly symmetric protein complex, as I highlighted both the 'run-and-tumble' motion that it makes possible and some of the nearly two dozen proteins which make up its overall architecture (1.3). Narrowing in, I then focussed on the cytoplasmic ring of the motor and its main components: $FliF_C$, FliM, FliN, and FliG, with FliG being a central hub of the motor and the key focus of this project. To conclude this section, I outlined the main hypothesis to be tested in this thesis, the domain-swap polymerization model for FliG assembly (1.4). Finally, in the section preceding this one, I gave some background on the main technique to be employed in testing the domain-swap hypothesis, first by presenting a historical overview of the development of the field of DNA nanotechnology and then by describing how DNA nanostructures have been used to organize components with nanometer precision (1.5).

Before proceeding to the results of this thesis, it is worthwhile now to clarify the experimental approach in more detail and how a DNA template can be used to look for domain-swapping in FliG by highlighting some of the theory's testable predictions. I will touch on both the benefits and drawbacks of templating purified proteins *in vitro* and will also outline what my work has aimed to achieve in the context of the larger questions still present around the BFM. To lead into the material in the coming chapters I will finish by giving a general outline of what I have accomplished and

how that will be presented in the remaining sections. The project is ambitious and not without its technical difficulties but my work is quite easily organized, sequentially building step-wise from smaller to larger structures.

An *in vitro* DNA Template Enables Many Experiments Impossible to Do Inside a Living Cell

Taken in a wider context, the BFM is an important system to study due to its involvement in many of today's pressing concerns. It is the target of many antibiotics and necessary for the pathogenicity of many bacterial species causing diseases ranging from typhoid to Lyme disease[151]. With nearly 70 years of research behind it, it is also perhaps the best studied of all large protein complexes and presents a model system for the study of self-assembling, dynamic molecular machines. As well, its provides a source of inspiration and design principles for the burgeoning field of nanotechnology, with many in that field hoping to create artificial 'nanoswimmers' to deliver drugs and other materials from one place to the next[152].

In the scope of this project however, the grand aim would be to build a functional, switching C-ring *in vitro* with all of its component parts in order to gain a complete understanding of how it assembles, rotates, and switches direction. This could fit into the subsequent construction of an *in vitro* rotating motor but the need for an energized membrane and hundreds of individual proteins makes this a very long-term aim to be realized only decades in the future. As mentioned previously, the more manageable, yet still ambitious, goal of this work is to study the domain-swap polymerization model for FliG, a small, yet important part of a large and complicated puzzle.

As was also mentioned previously, the way in which this model will be tested is by investigating the C-ring components *in vitro* using a DNA-template to replace the FliF-ring template found in the motor (Fig. 1.16A). From cryo-EM reconstructions and other work, it is known that FliG molecules are spaced somewhere between 3 and 6nm away from one another in the C-ring[54,77], a perfect length-scale for a DNA template as these in general present uniquely addressable sites with nanometre spacing. Thus, the dimensions of the C-ring and DNA nanostructures match. But even

if it is technically feasible, why do this *in vitro* with artificial components? What are the advantages of this approach over working with functioning BFMs in living cells?

Firstly, the environment within a living cell is incredibly complex (see Fig. 1.1) and the data obtained from *in vivo* measurements is in general much noisier than that obtained *in vitro*. When trying to do something as precise as count proteins or look at possibly weak protein-protein interactions, working with purified components outside of the cell is much more likely to produce accurate results and there are a much wider range of techniques available *in vitro*. In examining proteins via fluorescent imaging for example, *in vivo* work requires the genetic fusion of an often bulky fluorescent protein to the protein of interest, something that often disturbs the protein's function. Working *in vitro* however means that purified proteins can be conjugated to robust organic dyes which are much smaller and more stable than fluorescent proteins.

An *in vitro* system also lends itself to a bottom-up approach, meaning that all the components added to the system are determined and controlled by the experimenter with concentrations and stoichiometries very precisely known. The environment that these components work in can also be more carefully controlled than can be done within a cell, with the ionic strength, pH, hydrophobicity, and many other parameters of the solution easily tuned. This level of control in the components and their environment makes it easier to link cause and effect and directly implicate one particular factor, isolated from the complicated, tangled web of interactions found within a cell.

Thus an *in vitro* system is conducive to a simpler, bottom-up approach. But why use a DNA template? Why not build an *in vitro* system incorporating the FliF protein in a more natural MS-ring template for FliG? As touched on before, the main obstacles to taking this approach are technical considerations concerning the FliF protein. Being large and membrane-embedded, it is quite difficult to work with and represents a particular challenge during isolation and purification for *in vitro* work. And though a purified MS-ring would be the best stand-in for the natural template (with a caveat that the purification procedure could alter it in some ways), the fine-grained control

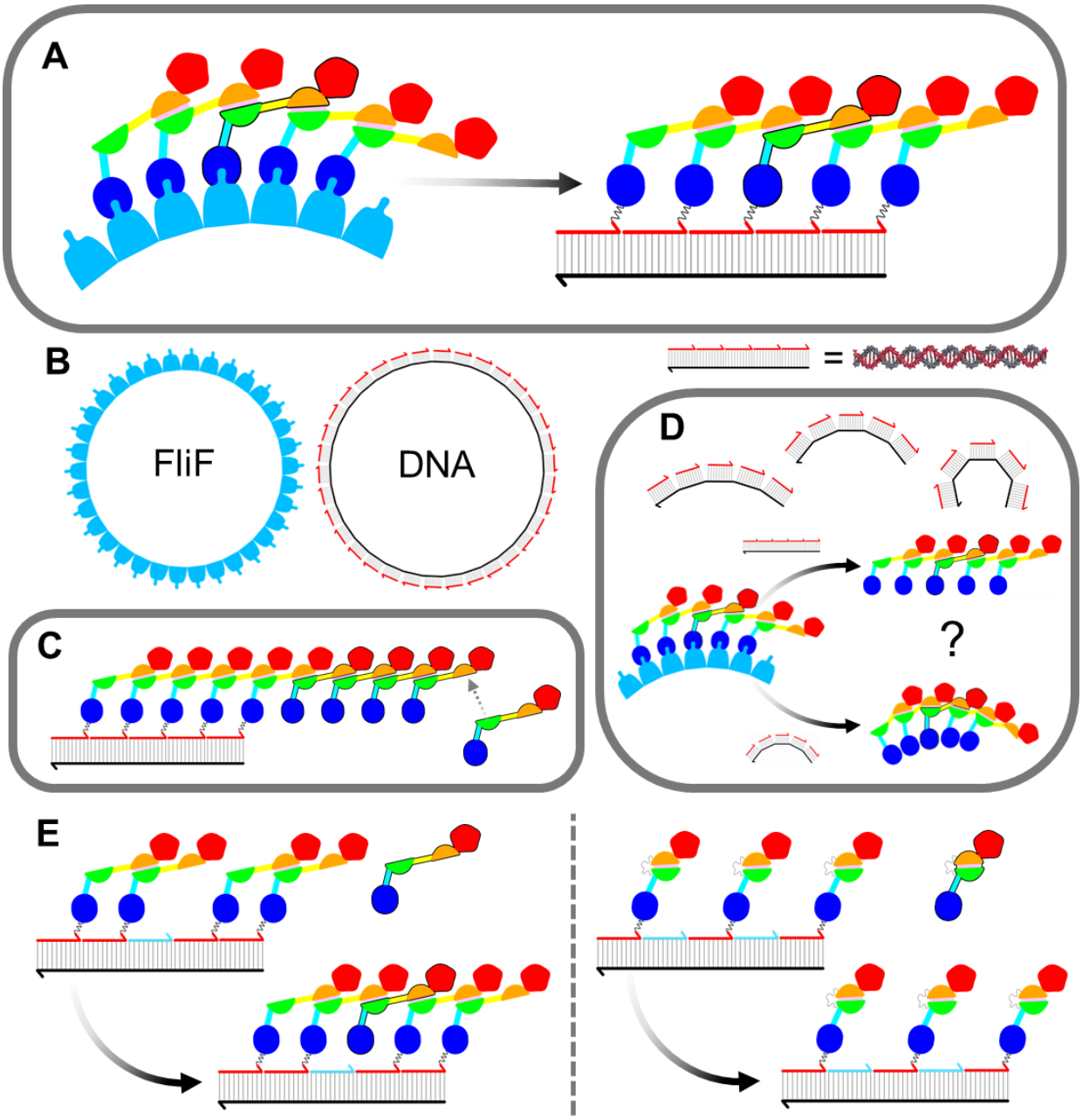


Figure 1.16 | A DNA template for FliG molecules can be used to investigate the domain-swap model. A) Replacing the FliF proteins underlying the C-ring with a DNA scaffold is feasible due to the fact that DNA nanostructures can be constructed which position FliG molecules with nanometre precision. **B)** Both linear and circular structures can be created, exhibiting varying degrees of similarity to the FliF ring. Importantly, circular templates are closed, leaving no room for any exposed ARM motifs. **C)** Linear templates on the other hand are open at both ends, and exposed ARM motifs could stimulate runaway polymerization, as predicted by the domain-swap model. **D)** DNA templates can be constructed with a range of curvatures. These can be used to explore if there is any built-in curvature to the FliG protein or if different templates can produce different FliG shapes. **E)** Another prediction of the domain-swap hypothesis is gap-filling. According to the model, gaps in the FliG oligomer could be filled by free-floating FliG interacting with exposed ARM motifs. Only certain patterns of gaps, with extended FliGs on either side, would be likely to be filled however. Gaps of various lengths and spacing can be introduced in a DNA template to test this prediction.

and ability to manipulate and quickly change a DNA template opens up more room for exploration and more ways to test the domain-swap hypothesis, a few of which I will now list.

Going through possibilities for experiments that can be performed using a DNA template, one main test is the comparison of a linear with a circular template. A circular template like that in the BFM would allow a FliG ring to close up on itself and thereby hide away any exposed ARM motifs (Fig. 1.16B). A linear template by contrast would by its nature leave exposed ARM motifs at either end. The domain-swap model predicts that these exposed ARM motifs would recruit more free FliG protein, building a larger and larger assembly in what has been termed a **'runaway polymerization'** (Fig. 1.16C). Finding a large runaway FliG polymer would be quite easy with standard *in vitro* techniques and hence this presents a major testable prediction of the model.

Another manipulation that would be easily achieved with a DNA template is to alter its curvature away from that found in the motor to see if the ring-forming nature is 'built in' to the FliG molecules or imprinted onto them by the underlying MS-ring (Fig. 1.16D). This would also tie into the question of FliG stoichiometry as an in-built curvature to the FliG molecule would indicate a tendency towards a constant diameter ring and hence a strictly defined number of subunits. If FliG was instead to take the shape and curvature of its underlying template, it would indicate that the stoichiometry is more malleable and dictated by the geometry of the MS-ring to which FliG binds.

The ability to manipulate the DNA sequences which connect the BFM components to the DNA structure and thereby tune the strength of the interaction between FliG and its template is another advantage of this approach. Whereas the binding of FliG to FliF has a single, defined strength (K_d = 40nM)[90], the length and sequence of the DNA adaptors can be designed to produce FliG-template interactions ranging from transient and fleeting, to extremely stable. Tuning the strength of this binding would be one way to investigate the kinetics of the FliG assembly process and to determine the relative contributions of FliG-FliG interactions and FliF-FliG interactions during the building of the ring. Tying into this, using techniques such as toehold-mediated strand

displacement, the entire DNA template could be removed after FliG assembly to investigate the stability of an isolated FliG oligomer and its disassembly kinetics.

One final testable prediction of the domain-swap model is the proposed ability of a FliG molecule to slot into empty spaces in a templated polymerizing oligomer, a process termed **'gap-filling'** (Fig. 1.16E). Should a region on the template exist without FliG that is flanked by two exposed ARM motifs (perhaps due to the domain-swap oligomerization of the flanking FliG molecules), the model predicts that a free FliG molecule can enter this region, bind the exposed ARM motifs and thereby fill the gap. Designing templates which place FliG proteins at various distances from one another, thereby incorporating gaps, is very straightforward with DNA and the incorporation of a FliG molecule into these designed gaps can be easily read out using various *in vitro* techniques.

But along with all the benefits of an *in vitro* system and the wide range of experiments that can be done with DNA templates, there are also several drawbacks. The most obvious question that purified components and an artificial template brings up is whether or not the structures observed are physiological. A DNA template may lack certain key characteristics of the MS ring, such as its precise geometry or flexibility, or the simplified system could be missing key BFM proteins which are present and necessary for C-ring-formation in *E. coli* or *Salmonella*. In addition, it is possible that the *in vitro* environment may not have the correct salts, pH, small molecules, or overall viscosity for natural cytoplasmic FliG behaviour and this may skew the observed results. Another consideration is that, with the cytoplasmic ring, it is particularly difficult to prove the physiological relevance of any assembly that is formed until it gets to a point where it can recapitulate a physiological function (such as switching for example). Due to the complexity of the ring there is quite a lot of building to be done before anything resembling a C-ring takes shape which means there is a considerably long, uncertain phase when progressing from the construction of small, tenuous assemblies to the formation of a fully-formed ring which can be compared to those found *in vivo*.

Despite these drawbacks, the control and ease of manipulation that an *in vitro* DNA template offers when testing the domain-swap model of FliG polymerization, along with the wealth of interesting experiments that this system enables, makes the prospect hard to pass up. Undoubtedly, on a broader scale, as the field of BFM study moves forward, an interplay of *in vivo* and *in vitro* work will be necessary to progress past the current incomplete picture of BFM assembly and function, but in this thesis all experiments that are presented have been done *in vitro* with purified components from *E. coli* and *Salmonella* bacteria.

With this in mind, in the next section I will briefly outline the work to be presented in the following chapters. The main experiments that I have just highlighted, such as testing for runaway polymerization and gap-filling, require a great deal of work prior to their realization and in my presentation of the results these key experiments come only after a lengthy period devoted to the preparation of the individual components. In addition, the amount of space in this thesis that each of the topics to be discussed takes up is not proportional to the amount of time spent working on them so I will provide a brief chronological summary of the works presented for context as to how they fit into the full timeline of the project.

The Work Presented in This Thesis Builds Sequentially from Smaller to Larger Structures

The four chapters following this introduction present the results obtained over approximately three years of work stretching from May 2015 to May 2018 (Fig. 1.17). For the most part, the chapters flow chronologically from one to the next and throughout there is a constant focus on the main goal: using a DNA template to study the domain-swap model of FliG oligomerization. This fits into the wider aim of understanding the assembly of the C-ring as a whole and as the work progresses, more of the C-ring's components are incorporated alongside FliG. My results however, fit into a broader collaboration between four laboratories, which began work on the project before I started, so it is worthwhile to note here their contributions and to give a picture of the state of affairs when I began my work.

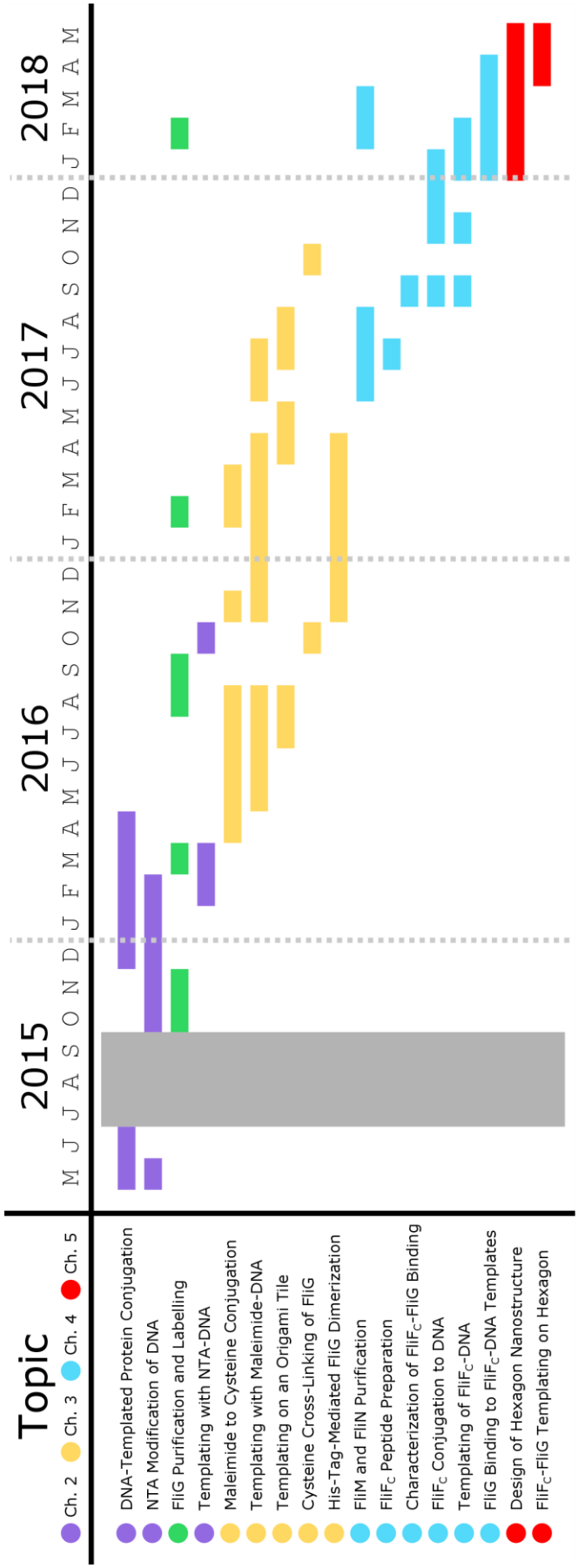


Figure 1.17 | A chronology of the work presented in this thesis. The results of chapters 2 to 5 were obtained over the course of three years, stretching from May 2015 to May 2018. The major topics focussed on in each month are shown colour-coded by chapter. A period of three months in 2015 was spent working on an unrelated short project in another lab (grey rectangle). The purification and modification of FliG took place at several points throughout the work (green) and appears in multiple chapters.

The framework and initial ideas for this project were developed jointly by Andrew Turberfield, Richard Berry, Lawrence Lee, and Keiichi Namba who began a collaboration in 2013 with the overarching, long-term goal of using DNA nanostructures to artificially synthesize the BFM. The work was to be divided evenly between the four labs with the preliminary phase involving the construction and characterization of DNA structures in the Turberfield lab along with the production of FliG constructs and mutants in the Lee lab. Following this, the Lee lab would use biophysical techniques such as surface plasmon resonance (SPR) while the Berry lab would use single-molecule fluorescence microscopy to investigate the binding strength, kinetics, and stoichiometries of FliG proteins on the DNA templates. The Namba lab would then produce cryo-EM images of the DNA-protein assemblies to obtain more fine-grained structural detail of the constructs.

With this framework in place, the work began a little less than two years before I joined the Berry and Turberfield labs. Many techniques and tools had been developed even before that time which allowed for a relatively rapid preliminary phase during which various components were obtained and developed. In the Turberfield lab, several DNA nanostructures were planned and a few were built including the short linear templates and the origami tile which appear later in this thesis. In addition, the Lee lab designed and constructed a large range of FliG mutants from both *E. coli* and *Salmonella* and were successful in developing a FliG purification procedure, the same procedure I used for purifying my FliG proteins. Alongside this, a graduate student in the Berry and Turberfield labs, Sam Tusk, worked extensively on developing a fluorescence microscopy set-up to enable the visualization of dye-labelled FliG proteins. After establishing a protocol for binding FliG-DNA assemblies to a microscope slide, he was able to image single proteins on a surface and had begun to create software to allow for the automated counting of FliG molecules bound to DNA templates. Also, back in 2009, the Turberfield lab had developed a procedure to modify DNA with a specific chemical group, thereby allowing it to bind non-covalently to certain residues on a protein (as is discussed in the next chapter)[153]. When I started on the project this was the main

means for joining FliG to the DNA nanostructures and a great deal of work was going into optimization of the procedure for the chemical-modification of the DNA.

So, taken together, I was starting with something far from a blank slate when I arrived. All the previous work provided an excellent platform for me to build from and the advice shared and support given between all the collaborators ensured a great deal of forward momentum for the project. As the chapters unfold, I will make sure to credit the work of the collaborators when it arises and be clear to specify any results presented which are not my own. Especially in the next chapter, whose material comes mostly from the earliest period of my work, I partnered very closely with Sam Tusk and I will highlight his contributions as well when they come up.

Giving a brief overview of what is to come, I have divided the results into four separate but connected chapters, divided generally by topic and arranged roughly chronologically (Fig. 1.17). The chapters build on one another both in terms of the content they introduce, and physically in the components they bring together, with larger and more complicated assemblies being produced as the work progresses. On a very basic level, the main focus of Chapter 2 is the investigation of how to graft together a DNA nanostructure and FliG, a protein which naturally does not interact with DNA. Chapter 3 begins with one solution to that problem and then proceeds to focus on how to arrange FliG on various different templates. The broad focus of Chapter 4 is the incorporation of other C-ring proteins apart from FliG into the system, such as $FliF_C$, FliM, and FliN, and the study of how they interact with one another. The final chapter of results, Chapter 5, highlights the latest experiments done and describes the design and assembly of a new DNA nanostructure for templating the assembly of C-ring proteins.

Thus, from the BFM to DNA nanotechnology through to the domain-swap model of FliG polymerization, the appropriate material has been introduced in preparation for the results to follow. In these chapters, aside from the main results presented, there will be some space devoted to introducing smaller, more technical matters necessary for the understanding and interpretation of the work, and there will be, in addition, some words devoted to the discussion of the results as

they are brought forward. One final discussion will come in the last chapter of this thesis aimed at bringing together the different aspects presented in the preceding chapters, and concentrating on how they relate to the central focus of the project: the DNA-templated assembly of the bacterial flagellar motor's cytoplasmic ring.

Before this though comes Chapter 2, where the assembly process begins.

CHAPTER TWO

THE FIRST HURDLE

Conjugation of NTA-DNA to His-Tagged FliG

To implement a system in which a DNA scaffold is used to template proteins, the most immediate consideration is how to join the DNA with the proteins. There must be some sort of connection between the two layers, some way of transferring the organization of the DNA nanostructure on to the proteins, and in most cases this connection has to be carefully designed into the system. In the simplest case, when using proteins which naturally bind to DNA, incorporating a specific exposed sequence into the DNA structure is enough to generate a binding site which can bring the DNA-binding protein to a defined location within the assembly. Most proteins however, do not naturally bind DNA and require a bit more work before they can be coupled to a nanostructure.

FliG is in this majority of proteins which do not normally interact with DNA and in this chapter I discuss several of the techniques that I used to try to conjugate a DNA strand to the FliG molecule. With a DNA oligonucleotide tag conjugated to it, FliG would be able to bind to complementary sequences designed into an underlying, templating nanostructure and thereby could be positioned with nanometre precision (see Fig. 1.16). Conjugation however proved to be a major hurdle to overcome in the course of my work and the story of joining FliG to DNA carries on past the end of this chapter, forming perhaps the longest narrative in this thesis. In addition, aside from conjugation strategies, one section of this chapter is devoted to the topic of FliG expression and purification. Purified FliG is central to most of the experiments to follow so it is worthwhile describing the process here in the early stages of the presentation of the results.

2.1 NTA Modification of a DNA Oligonucleotide

As mentioned, FliG does not naturally interact with DNA so the protein or the DNA or both must be in some way modified in order for an interaction to take place. Choosing how to modify the protein or DNA is quite difficult as there are many different modifications and protocols that have been used to conjugate DNA strands to proteins and, due to the complexity of the molecules, determining the most appropriate strategy from the outset is close to impossible. Every protein is different and finding which method works best is often a matter of educated guessing followed by trial and error.

With no previous attempts published looking to join FliG to DNA (or to any molecule for that matter), the first strategy implemented was chosen more for practical than for scientific reasons. Being developed and commonly used in the Turberfield lab, the materials and expertise were close at hand to apply a method based on coordinate covalent interactions between a short string of residues engineered into a protein and a string of chemical groups incorporated into a DNA strand[153]. This strategy was developed back in 2009 and most of the work done in applying it to the FliG system was accomplished by Sam Tusk and others before I entered the project, but it is useful to describe the strategy and highlight my contributions to its development as it returns in several other sections of this thesis.

NTA Groups Form Coordination Complexes with Nickel Ions and Histidine Residues

At its core, the link between protein and DNA in this conjugation strategy is made up of coordinate covalent bonds. A central nickel (Ni^{2+}) ion bridges the two molecules by providing six coordination sites for six coordinate bonds, four of which are taken up by an **NTA** (nitrilotriacetic acid) moiety attached to the DNA while the remaining two are occupied by one imidazole group each from two histidine residues on the protein (Fig. 2.1A). In the published protocol, the overall interaction is strengthened through the incorporation of three of these NTA-Ni^{2+}-histidine linkages, thereby

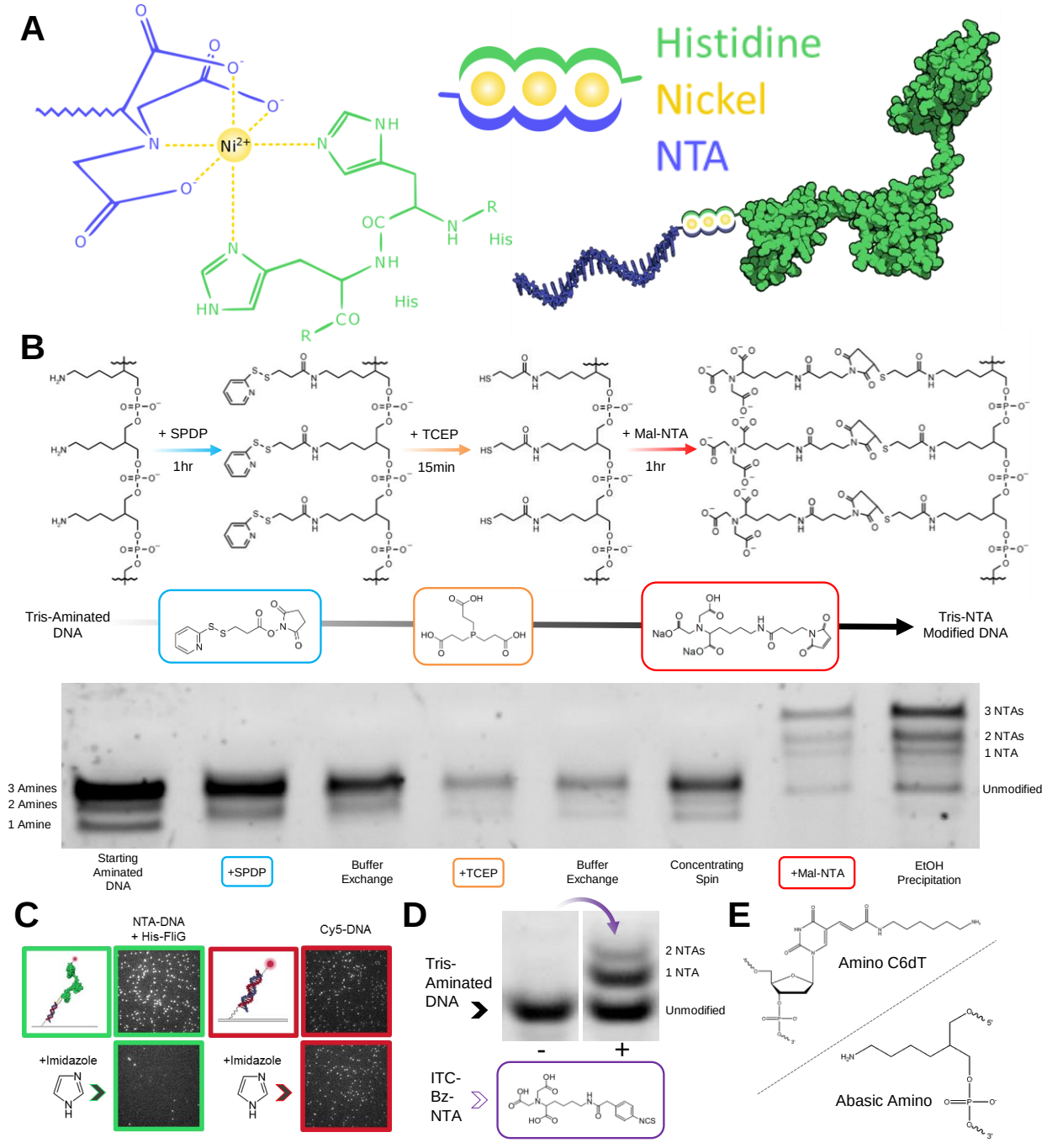


Figure 2.1 | NTA-modified DNA binds to the His-tag of FliG. A) A Ni^{2+} ion forms four coordinate covalent bonds with an NTA group added to the DNA strand and two with the His-tag of FliG (one per histidine). (Adapted from (155)). The connection is strengthened through the incorporation of three NTA groups per DNA strand. **B)** The three-step synthesis protocol for tris-NTA-modified DNA begins with a tris-aminated strand to which is added SPDP, TCEP, and finally maleimide-NTA (Adapted from (153)). Washing and occasionally concentrating of the strand occurs between each step. A gel shift can only be seen upon incorporation of the NTA group. Note the heterogeneous nature of the starting aminated DNA. **C)** To verify that the NTA-modified DNA bound to His-tagged FliG, fluorescently-labelled FliG was washed over NTA-modified strands bound to a complementary DNA strand anchored to a microscope slide. TIRF imaging showed bright spots corresponding to FliG binding. To confirm binding was specifically via an NTA-His interaction, imidazole was washed over the slide. Fluorescent spots disappeared on slides with FliG (green) but remained on slides bearing a fluorescent Cy5-DNA in place of the NTA-DNA (red). **D)** Variations on the NTA-modification protocol showed no increase in yields. A one-step reaction using isothiocyanobenzyl-NTA resulted in poor labelling of a tris-aminated strand. **E)** Experiments with strands incorporating amines on a modified thymine rather than an abasic site also showed no increase in yields.

involving six histidine residues in total. Six histidines are not naturally found in close proximity in most proteins but in modern-day molecular biology so-called **His-tags**, composed of a string of six to ten histidine residues, are ubiquitous additions to many protein sequences for use during affinity purification[154]. In fact, the NTA-Ni^{2+}-histidine interaction found between protein and DNA in this conjugation strategy is the same that takes place on standard affinity columns used to purify proteins.

The Lee lab, which designed the FliG constructs used throughout this thesis, incorporated a His-tag at the N-terminus of FliG, to be used both in purifying the protein and in this conjugation strategy for binding to the NTA groups incorporated into the DNA. The N-terminal domain is where the $FliF_C$ domain of the natural FliF template binds to FliG so positioning the interaction with the DNA at that end of the protein was thought to be the most physiologically relevant.

Though adding a His-tag to a protein is a fairly standard procedure, the NTA-modification of DNA is not and had to be developed step-by-step in the Turberfield lab for the 2009 publication. The protocol involves three main reactions to get from the tris-aminated DNA starting material to a final tris-NTA-modified DNA strand (Fig. 2.1B). First, the heterobifunctional cross-linker SPDP (*N*-succinimidyl 3-(2-pyridyldithio)propionate) is added to the tris-aminated DNA strand and its NHS (*N*-hydroxysuccinimide) esters react with the amines on the strand. Next, the reducing agent TCEP (tris(2-carboxyethyl)phosphine) is added which acts to break the disulfide bonds incorporated by the SPDP, thereby revealing free thiol groups. These thiols can then be attacked in the final step of the procedure by maleimide-NTA, leaving the strand with three NTA groups projecting off its phosphate backbone.

One important factor about this linkage is that it can be broken and remade by the simple addition or removal of standard chemicals. Adding in a chelating agent such as EDTA (ethylenediaminetetraacetic acid) strips the nickel ion from the complex thereby breaking the interaction, while the addition of free imidazole leads to competition with the imidazole groups of the His-tag for binding sites on the nickel ion, with the protein being displaced at high enough free imidazole concentrations. Removing these chelating agents or competitors through buffer

exchange or dialysis allows the complexes to reform. Taken in the context of the many other published methods for conjugation, this reversibility of the interaction is perhaps its most unique and useful trait and makes it stand out from the other, predominantly permanent, conjugation techniques.

Problems with the NTA-DNA Led to a Search for Alternative Conjugation Strategies

So, due to the in-house expertise surrounding the technique and the added benefit of its reversibility, the NTA-His linkage was the first to be implemented for the FliG system, some time before I arrived in the lab. Sam Tusk had succeeded in producing tris-NTA DNA strands and had confirmed their binding to His-tagged FliG using his TIRF microscopy set-up (Fig. 2.1C) but several problems remained. When trying to attach FliG molecules to a template with a specified number of binding sites using this NTA-His linkage, Sam repeatedly observed fewer FliG molecules bound than there were binding sites. He also was having trouble purifying the NTA-modified strands as had been done in the 2009 publication using gels and HPLC (high-performance liquid chromatography), with unmodified, mono-, and bis-NTA strands also present in the final solution. The predominant hypothesis when I entered the lab was that these two problems were connected in that the improperly modified DNA strands left over after the purification attempt were binding to the template and occupying sites which a tris-NTA strand was supposed to bind to. FliG could not bind to these improperly-modified strands and hence the observed stoichiometries on the templates were lower than they should be.

One of my first tasks on entering the lab was to try several variations on the standard NTA-modification protocol in an effort to produce higher modification yields. I attempted a single-step modification of the aminated strands using isothiocyanobenzyl-NTA (Fig. 2.1D), tested different variations of the starting tris-aminated strand having the amines either on a modified thymine base or at an abasic site (Fig. 2.1E), and varied the concentrations and different washing steps in the protocol yet nothing gave appreciably better yields than seen in the standard, published procedure. It appeared that the biggest factor determining the yields was the nature of the starting

aminated strand, as often the material which we received from the oligonucleotide synthesis company was in fact a heterogeneous mix of unmodified, mono-, bis-, and tris-aminated species.

Thus, with Sam still encountering difficulties in purifying the tris-NTA strand, and my attempts at variations of the protocol resulting in no increase in yield, we began to think about other possible conjugation strategies which could more reliably attach FliG to the DNA template. There was some concern as well surrounding the strength of the linkage (the K_d was measured previously as 14nM[156]) and there were thoughts that the connection could be perhaps broken by a moving AFM tip, by fluid flowing over a microscope slide in a TIRF set-up, or as the complexes passed through the porous meshwork of a gel. This led us to begin looking into more permanent conjugation techniques which could irreversibly link the protein to its DNA tag.

In section 2.3 I describe one such conjugation technique applied to FliG with the hopes of obtaining a site-specific, permanent linkage, but first, in the next section, I will briefly go through the expression and purification protocol for the FliG protein, as purified FliG is used in almost all of the subsequent sections and a quick aside is useful in clarifying some points which come up later.

2.2 Expression and Purification of the FliG Protein

As can be seen from the Gantt chart in Fig. 1.17, I spent a good deal of time throughout the course of my work expressing, purifying, or labelling FliG. Thankfully for me however, very little of that time was spent developing protocols or testing conditions, as the Lee lab had done that work previously for other projects and provided me with a reliable set of procedures. The Lee lab also provided all of the FliG constructs that I worked with throughout my time on the project, meaning that rather than going through the often time-consuming process of engineering and testing different mutants or variants of the protein, I could simply ask for one of the constructs already in their library and receive a plasmid on a piece of filter paper some time later in the mail. This was a very important factor in the progress that I made as it freed up a great deal of time to work through various conjugation procedures.

In the Lee lab's library of constructs there were many variants of FliG from several different species, either full-length or containing only specific domains, with various amino acid substitutions, and incorporating various purification tags. For my work however, I only used FliG from two species, *E. coli* and *Salmonella*, worked only with mutants incorporating cysteine residues, and purified the proteins solely through the use of His-tags incorporated at the termini (Fig. 2.2A). I will give a brief overview here of the protocol for expression and purification but more detailed descriptions can be found in the Supplementary Information.

FliG Constructs Are Expressed in *E. coli* and Purified Using His-Tags

The benefits of working in an *in vitro* system, as described previously, include the precise control of which components are included at what concentrations inside the tube. Intrinsic to this approach then is the need to work with pure, well-characterized components isolated from any contaminants. Thankfully, the protocol for expressing and purifying FliG is both robust and reliable, using standard affinity chromatography approaches to provide very pure isolated proteins (Fig. 2.2B).

Starting from a plasmid sent from the library of constructs in the Lee lab, the first step was to amplify it to levels at which it could be transformed into a protein-expressing strain of bacteria later on. To do this, an *E. coli* strain developed for copying plasmids was used and standard kits allowed for the plasmids produced to be isolated from the bacteria. These isolated plasmids were then transformed using heat shock into another strain of *E. coli* developed for protein production and these bacteria were grown up in the presence of an antibiotic. Along with the gene for FliG, the plasmid also incorporated a standard antibiotic resistant gene and an inducible promoter meaning that it was straightforward to both check if the plasmid was transformed properly and to produce the FliG protein at a particular point in the bacterial growth curve.

With the plasmid transformed into the appropriate *E. coli* strain, large volumes of liquid bacterial culture were grown up and, at the appropriate point, a chemical inducer was added to start FliG protein production. The concentration of the FliG protein then built up in the cytoplasm of the

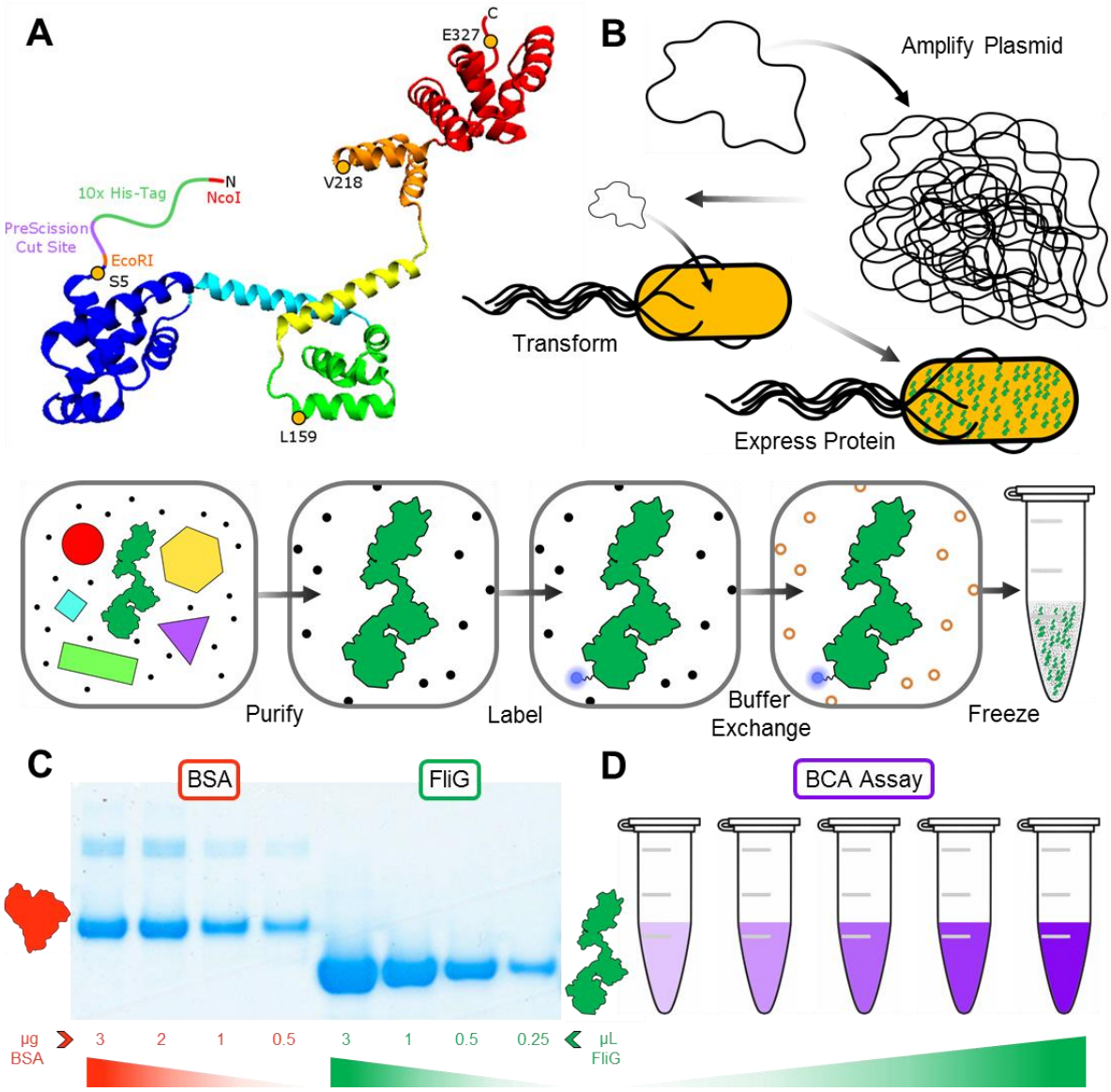


Figure 2.2 | FliG constructs are expressed, purified, and characterized before use in the *in vitro* system. A) FliG constructs used in this work come from either *E. coli* or *Salmonella*. Most incorporated an N-terminal 10x His-tag which could be cleaved from the protein using the PreScission protease. Four sites (near the N-terminus, on the two ARM domains, and near the C-terminus) which were mutated to cysteine residues in different constructs are noted. **B)** To obtain FliG for *in vitro* work, plasmids were first amplified and then transformed into *E. coli* cells. Expression of the protein was induced and following overnight production the cells were lysed in a French press. The cytoplasmic components of the lysate were passed through an affinity column which pulled FliG out of the mix via its His-tag. In some cases the protein was then labelled with a dye and buffer exchanged before being frozen in aliquots for storage at -80°C. **C)** A rough idea of the concentration and purity of the protein was obtained by running it on an SDS-PAGE gel and staining with Coomassie dye. A range of BSA standards of known concentrations were run for comparison. **D)** For a more accurate determination of the protein concentration, a colourimetric BCA assay was used with absorbance values mapped to a standard BSA curve.

bacteria until a specified time when the cells were lysed using a French Press. Following this lysis, the purification steps began as the lysate was spun at high speeds in a centrifuge to remove cellular debris and membranes before the clarified lysate supernatant containing the cytoplasmic contents of the bacteria was poured over an affinity chromatography column prepared with a nickel-charged IMAC (immobilized metal affinity chromatography) resin.

The constructs of FliG that I used all contained ten histidine residues at either their N- or C-terminus, forming a His-tag which binds to the nickel-charged resin using the same chemistry as found in the NTA-Ni^{2+}-His coordination complexes described in the last section. As the lysate passes over the beads, the proteins bind to the immobilized nickel ions and are thereby pulled out of the mix of other cytoplasmic proteins which do not contain stretches of residues which can bind tightly to the nickel ion. Thus, discarding the flow-throughs after a few washes to remove any weakly bound species, I was left with purified FliG protein bound to the resin.

With the protein thus purified and immobilized, I had the choice of manipulating it while still bound or eluting it first before later steps. In most cases where dyes were used to fluorescently label FliG, these reactions were done while FliG was still bound to the resin as its immobilization provided a simple means for washing away free dye, but in most other cases the protein was eluted first with a high concentration of imidazole, bringing it off the resin and into solution, before any modifications or manipulations took place. Most commonly, the final steps in the process were the use of a gravity column to buffer exchange the FliG into a more standard working buffer without such high imidazole concentrations, and the addition of glycerol as a cryo-protectant before aliquots were frozen in liquid nitrogen for storage at -80°C.

Characterizing the protein by gels (Fig. 2.2C) and other means following this procedure in general showed isolated FliG at high levels of purity with concentrations usually in the range of 10μM to 50μM (~0.4mg/mL to 2mg/mL) as determined by a standard BCA (bicinchoninic acid) assay (Fig. 2.2D). This meant that the FliG aliquots were ready to be used straight out of the freezer without any other modifications and could be incorporated into conjugation reactions and templating experiments with a high degree of certainty as to their final concentrations.

Thus, out of all the protocols undertaken in the course of this work, FliG expression and purification proved to be one of the most reliable, time and again producing pure, isolated FliG in relatively high quantities. But though the purification of FliG was consistent, its conjugation to a DNA strand was more challenging. As the previous section demonstrated and the ones to follow will show, a great deal more time and effort was needed before a FliG-DNA conjugate could be as reliably produced.

2.3 DNA-Templated Protein Conjugation

So, though there was plenty of purified FliG on hand, obtaining a pure NTA-modified strand to conjugate to it was proving a challenge. Unmodified and partially-modified NTA-strands were getting in the way of experiments looking to template FliG on a DNA scaffold and it became apparent that a different conjugation strategy was needed. To avoid worries that the reversible nature of the NTA-His link brought up, it was determined that a permanent covalent linkage between the DNA and the protein was the best way forward. Unlike the NTA-His linkage, a permanent covalent link would not fall apart in different buffers, could not be broken by a pushing AFM tip or fluid flow, and in general would simplify the system. Rather than worrying about what ratios and concentrations of nickel, NTA-DNA, and FliG to add to a solution, one could simply add a purified FliG-DNA conjugate ready to bind to a DNA scaffold.

Thus the benefits of a permanent, covalent linkage were clear and while Sam continued work on NTA-DNA purification, I set out to look into a new method. When researching various published techniques for protein conjugation however, one key question kept surfacing: how can a DNA strand be conjugated to a protein in a site-specific manner? Of all the residues on the FliG protein, how can a particular amino acid in a particular location be targeted with a single DNA strand? In the NTA-His method, the placement of the histidine-tag at the N-terminus naturally positioned the DNA at that end of the protein, the region which, *in vivo*, binds to the underlying FliF scaffold. Targeting the N-terminal domain with a permanent covalent attachment would require a different

approach however, using reactions and bonds of a different nature than those found in the NTA-Ni^{2+}-His complex.

Most methods for conjugating a molecule to a protein in a site-specific manner make use of unique amino acids or chains of amino acids introduced into the sequence of the protein through various genetic engineering techniques. In some cases, where a protein by chance has a rare amino acid, such as a cysteine, in the right location for the conjugation, no engineering is required and the procedure is quite straightforward. The FliG protein though contains no cysteines throughout its entire sequence and, in its wild type form, exhibits no particularly good targets at its N-terminal domain for a conjugation. Thus, for a site-specific linkage, it seemed like there was a need for engineering to be done at the N-terminus in order to introduce a conjugation site.

Surveying the literature revealed the incorporation of various amino acids and protein tags into the sequence of the species to be conjugated to, ranging from the standard cysteines[157], to unnatural amino acids for click chemistry[158], to larger constructs like SNAP- [159] and Halo-tags[160] and the SpyCatcher peptide[161]. There was a great deal of documentation around engineering these components into a protein of interest and many different systems had used them before, but a new technique called DNA-templated protein conjugation (DTPC), published in 2014 by the lab of Kurt Gothelf, caught my eye as they promised permanent, site-specific conjugation of a DNA-strand to a protein using only a histidine-tag[162], something which my FliG constructs already had. This meant that I could work with what was already in the freezer, without any further genetic engineering or creation of new constructs.

DTPC Uses a Guide Strand to Position a Reactive Strand at a Specific Conjugation Site

The basic steps of the approach are as follows (Fig. 2.3). First, as described in section 2.1, an NTA-modified DNA strand called the '**guide strand**' is bound to the His-tag of the protein through the addition of nickel (Fig. 2.3A). Next, a strand complementary to the guide strand is added, forming a duplex with most of the guide's nucleotides, but leaving a toehold for later use (Fig.

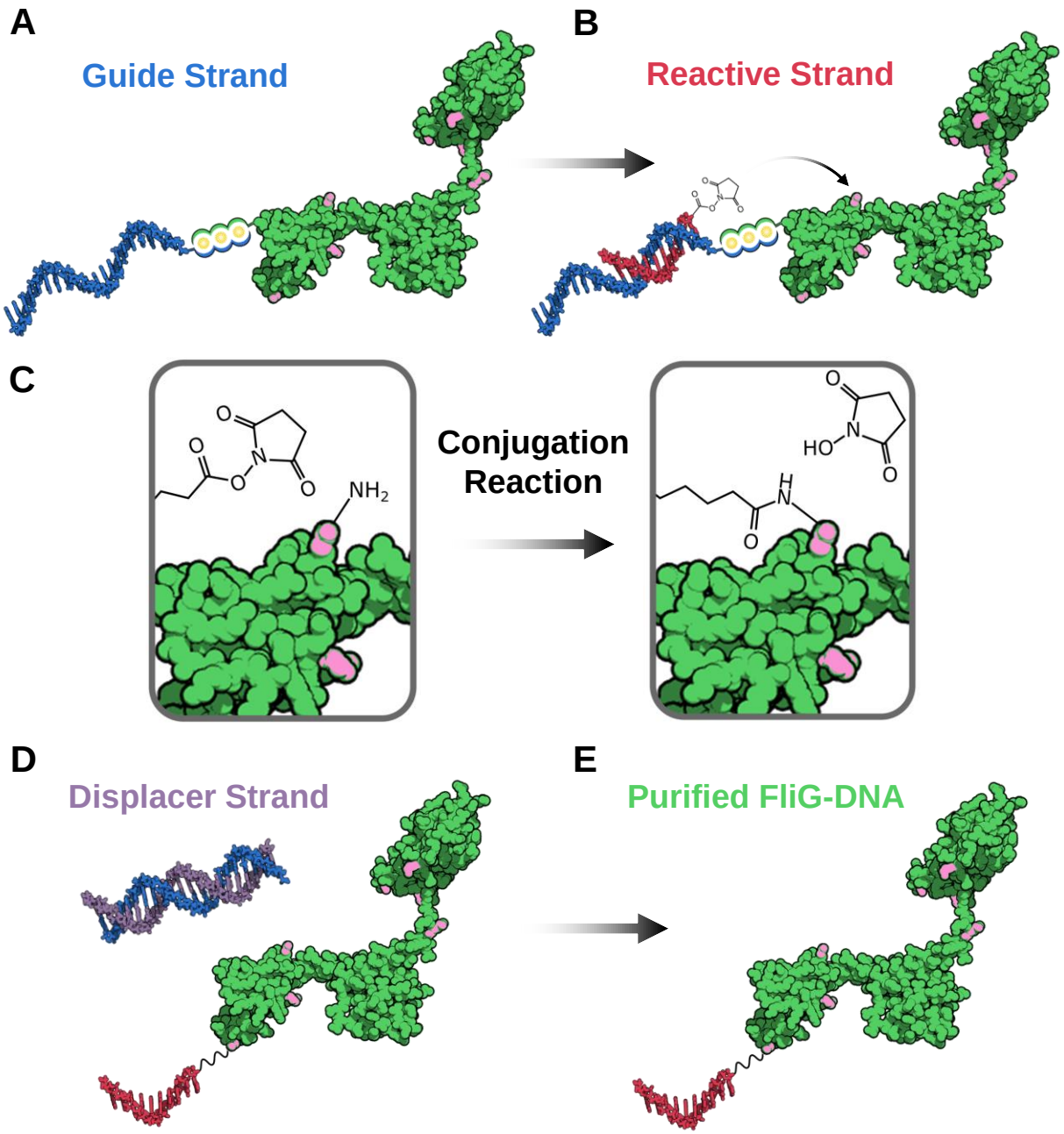


Figure 2.3 | DNA-templated protein conjugation uses a guide strand to direct the conjugation reaction to lysines at a specific location on the protein. A) First, an NTA-modified 'guide strand' binds to the protein via its His-tag. **B)** Next, a 'reactive strand' is added which binds complementarily to the guide strand, thereby positioning its reactive NHS ester moiety close to particular lysines (pink) on the protein. **C)** The conjugation reaction takes place between the NHS ester of the reactive strand and the primary amine of a nearby lysine. **D)** Both imidazole and a displacer strand are added to remove the guide strand from the reactive complex, leaving only the reactive strand bound to the protein via a covalent link. **E)** Purification removes the free DNA strands to leave purified FliG-DNA, with the DNA tag positioned on the N-terminal domain.

2.3B). Introduced into one end of this complementary strand is a moiety called an NHS ester which can react with primary amines to form a stable amide bond. This strand is therefore termed the '**reactive strand**'. Accessible primary amines are found on proteins at the N-terminus and at the end of lysine residues and, with lysine being a relatively common amino acid, this means there are multiple sites at which the DNA-to-protein reaction can take place. The innovation of this approach is the incorporation of the guide strand to direct the reaction towards a specific site by spatially positioning the reactive strand close to a particular primary amino acid. Bringing the NHS ester of the reactive strand in close proximity to the primary amines of a specific lysine for example favours the linkage of the DNA to that particular site, thus leading to a site-specific conjugation.

Following the guided NHS-ester-to-amine reaction (Fig. 2.3C), the guide strand is then removed through the addition of both EDTA or imidazole to break the NTA-Ni^{2+}-His linkage and a 'displacer strand', complementary to the full length of the guide, added to peel away the strand through toehold-mediated strand displacement (Fig. 2.3D). This then leaves the protein with a single DNA strand covalently bound at a specified location and this protein-DNA conjugate can be purified away from the other DNA species through techniques such as PAGE (polyacrylamide gel electrophoresis) gel extraction (Fig. 2.3E).

Thus the site for the conjugation is determined by the positioning of the reactive strand which is determined by the positioning of the guide strand which in turn is determined by the positioning of the His-tag on the protein. Therefore, with an N-terminal His-tag on FliG, a conjugation reaction could in theory be targeted specifically to N-terminal lysines. Looking at the structure of FliG showed a handful of lysines in close proximity to the N-terminal His-tag so it was determined that while Sam continued with the NTA-DNA linkage, I would apply this DTPC protocol to the FliG protein, with the hopes of obtaining a protein-DNA conjugate with the DNA strand attached permanently to the N-terminal domain of FliG.

The Reactive Strand Was Modified with DSS, a Homobifunctional Cross-Linker

To begin implementing the DTPC protocol with the FliG protein, I first had to obtain the correctly-modified DNA strands. For the sake of simplicity, I chose as the guide strand the 20-nucleotide-long NTA-modified strand that we were already using in conjugations. For the reactive strand I chose a 10-nucleotide-long oligonucleotide with the N10 sequence designed by Sam to lack secondary structure and to have a high melting temperature with its target once bound. This sequence was complementary to the first 10 nucleotides of the guide strand, thereby leaving a lengthy 10-nucleotide toehold for the displacer strand to bind to.

With the guide strand already in the freezer and the displacer simply ordered from a DNA synthesis company, that left the reactive strand modification as the sole task before a conjugation reaction could be attempted. Working from the original protocol[162], the strand was purchased with an amine on its 3' end and reacted with the homobifunctional cross-linker DSS (disuccinimidyl suberate), a molecule consisting of two NHS esters separated by a short chain of six carbons (Fig. 2.44A). Due to its symmetric nature, the cross-linker in theory could react with two separate aminated DNA strands to form an unreactive and useless bridged dimer. To try to prevent this, the stoichiometry of the cross-linker relative to the DNA strand was kept very high and any unreacted cross-linker was removed from the solution following the reaction using a size exclusion spin column, with the protocol incorporating an ethanol precipitation as the final step.

Unfortunately, even with these precautionary measures, gels showed a pronounced band forming after the DSS reaction running to approximately where a bridged dimer might be expected to run to (Fig. 2.4B). Due to the small, uncharged nature of the DSS cross-linker, no gel-shift was expected following the reaction so the presence of this upper band confirmed that in all likelihood this unwanted bridging was occurring. Nevertheless, I wanted to know if there was any of the correct, single-stranded species in solution which might have a chance at reacting with FliG so I looked to find an appropriate way of determining whether the desired species was present.

Due to the ease of working with PAGE gels, I chose to work on a simple gel-shift assay to test for the presence of the correctly-modified reactive strand (Fig. 2.4C). Testing the addition of several

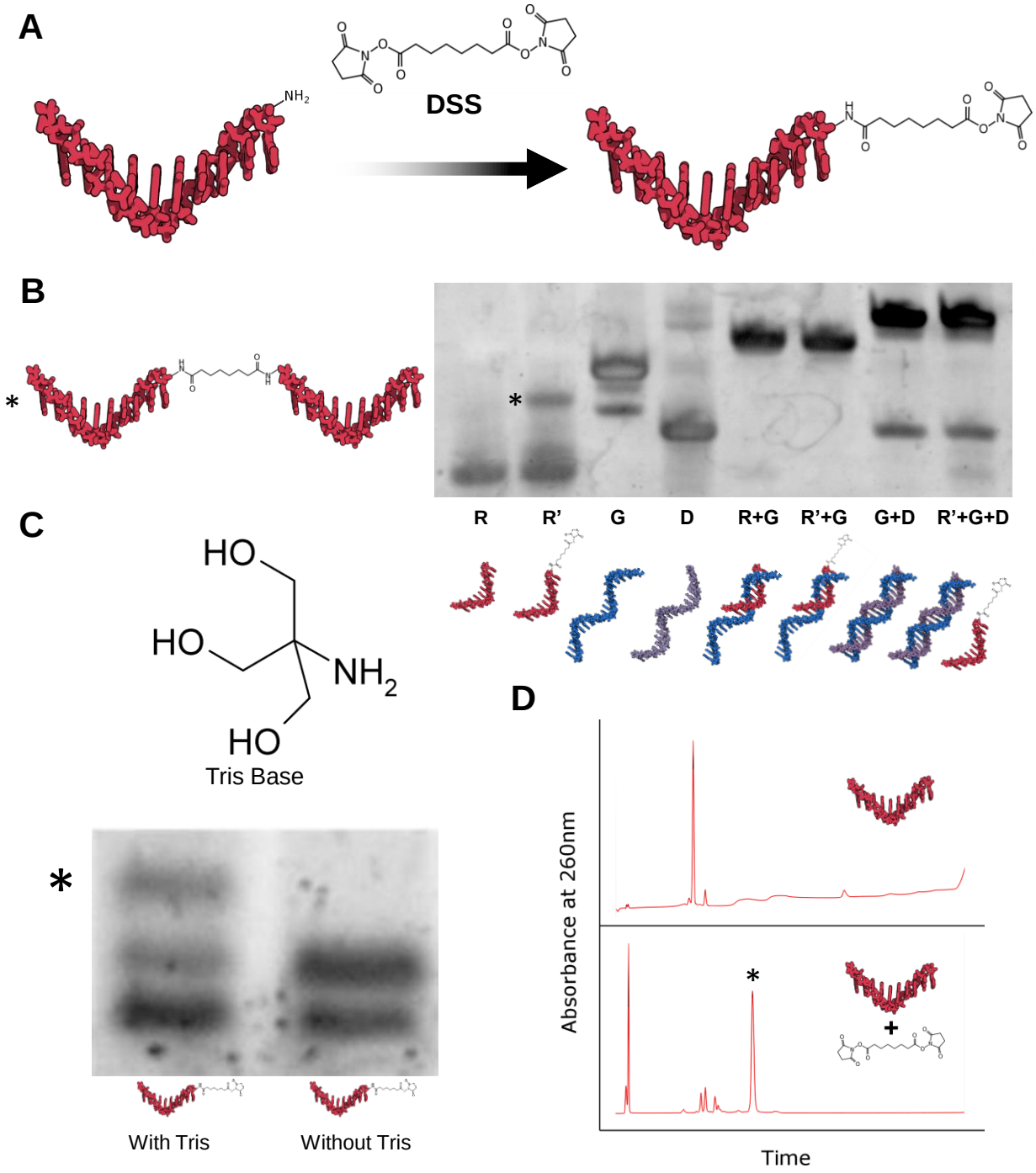


Figure 2.4 | The aminated reactive strand is modified by the cross-linker DSS and binds to the guide strand. A) The homobifunctional cross-linker DSS introduces an NHS ester at the 3' end of the reactive strand. **B)** Side-products of the reaction include an unreactive 'bridged dimer' which can be seen on gels (*). Gels also show that the reactive strand successfully binds to and then is displaced from the guide strand. **C)** A gel-shift assay was developed to identify the presence of unbridged, reactive strands wherein the DNA was mixed with Tris base. A higher band formed (*) following the reaction of the strand with the amine on the Tris. **D)** RP-HPLC showed a clear difference between the elution time of the major peak in runs with the unmodified strand (top) and reacted with DSS (bottom). (The first peak in the bottom trace is due to signal from excess DSS cross-linker.) The major DSS-modified peak (*), once purified, showed no reactivity to Tris however meaning it may correspond to the bridged dimer species.

different primary-amine-containing compounds to the reaction mixture showed that there was a reliable band shift observed upon the addition of several of the compounds, with Tris base showing the clearest and largest shift. A considerable fraction of the total DNA (comparable to the fraction in the bridged dimer band) was shifted upon addition of Tris, indicating that there were in fact DNA strands with reactive NHS esters present in solution and in theory available for reaction to FliG.

This was promising so I decided to try to purify this modified species away from the unmodified strands and bridged dimers using reversed-phase HPLC (separating on the basis on hydrophobicity). Scanning for DNA using absorbance at 260nm during the separation showed only one dominant, shifted peak after the modification reaction (Fig. 2.4D) and when purified, this peak did not react with Tris base and hence was not the species I was looking for. After several attempts at other purification strategies, I decided to proceed with the unpurified reaction mix to see if I could obtain a protein-DNA conjugate. My logic was that it might be easier to purify the protein-DNA conjugate than the DNA strands alone due to greater differences in size of the respective species in the mix of protein and DNA. Before thinking about purification however my goal was to see at the very least if I could obtain a protein-DNA conjugate band on a gel which appeared only in the presence of the guide strand, the test that the developers of the DTPC technique used to indicate a site-specific conjugation reaction.

Initial Conjugation Attempts Showed a Stable Protein-DNA Species on Gels

So with the individual components to a greater or lesser extent ready for the reaction, I implemented the DTPC technique with a fluorophore-labelled and histidine-tagged FliG protein. I followed the protocol that had been used in the publication, merely substituting in my FliG protein for the His-tagged GFP (green fluorescent protein) that they had conjugated a DNA strand to in their work[162]. Combining the guide strand, reactive strand, nickel and FliG protein at sub-micromolar concentrations so as to avoid a non-templated reaction, I let the components interact overnight before adding the displacing strand and EDTA the following morning in order to remove the guide strand from the complex. As was done in the publication, I assayed whether or not the conjugation had been successful using polyacrylamide gels, looking for bands with overlapping

DNA and FliG signals (through SYBR Gold staining and the conjugated Alexa Fluor 647 dye respectively).

In addition, I controlled for the possibility of a non-templated (and hence non-site-specific) reaction by including a trial in which I omitted the guide strand. A protein-DNA band in this control lane would indicate that a conjugation reaction had occurred even without the assistance of the guide and would imply that any protein-DNA bands seen on the gel could be the product of an interaction between a freely floating reactive strand and the protein. With this control in place and the displacer and EDTA added as described in the protocol, I was confident that I could identify any protein-DNA bands seen on the gel as either a templated or non-templated conjugate. In addition, I was told that an NTA-His linkage could not survive a run through a gel, meaning that any overlapping DNA and protein bands would indicate a permanently linked species.

Thus, when I observed the results of the first conjugation it appeared that the reaction had been successful. Around 40% of the total AF647 signal from the protein could be seen in a band with overlapping DNA signal. Moreover, in the control lane without the guide strand, no protein-DNA band was observed (Fig. 2.5A). Based on the factors just stated, I was convinced that this protein-DNA species was the product of a covalent conjugation between the NHS ester of the reactive strand and the amine of one of FliG's N-terminal lysines. Therefore, I decided to proceed first by attempting to increase the conjugation yields from the observed level of 40% and then to try to gel-purify this conjugate so that it could be used in templating experiments.

In the course of these attempts to improve the yield, I varied the ionic strength, pH, and buffering salts, added small amounts of surfactant, and tested ranges of concentrations for each of the components being added to the mix. In one of these tests, I performed the conjugation reaction as described, added the displacer strand, and then added varying amounts of EDTA or imidazole to a set of aliquots from the reaction mixture. Running the aliquots on a gel, I was surprised to see that above a certain concentration of imidazole, the protein-DNA band that I had repeatedly observed disappeared, leaving only separate, non-overlapping protein and DNA bands (Fig. 2.5B).

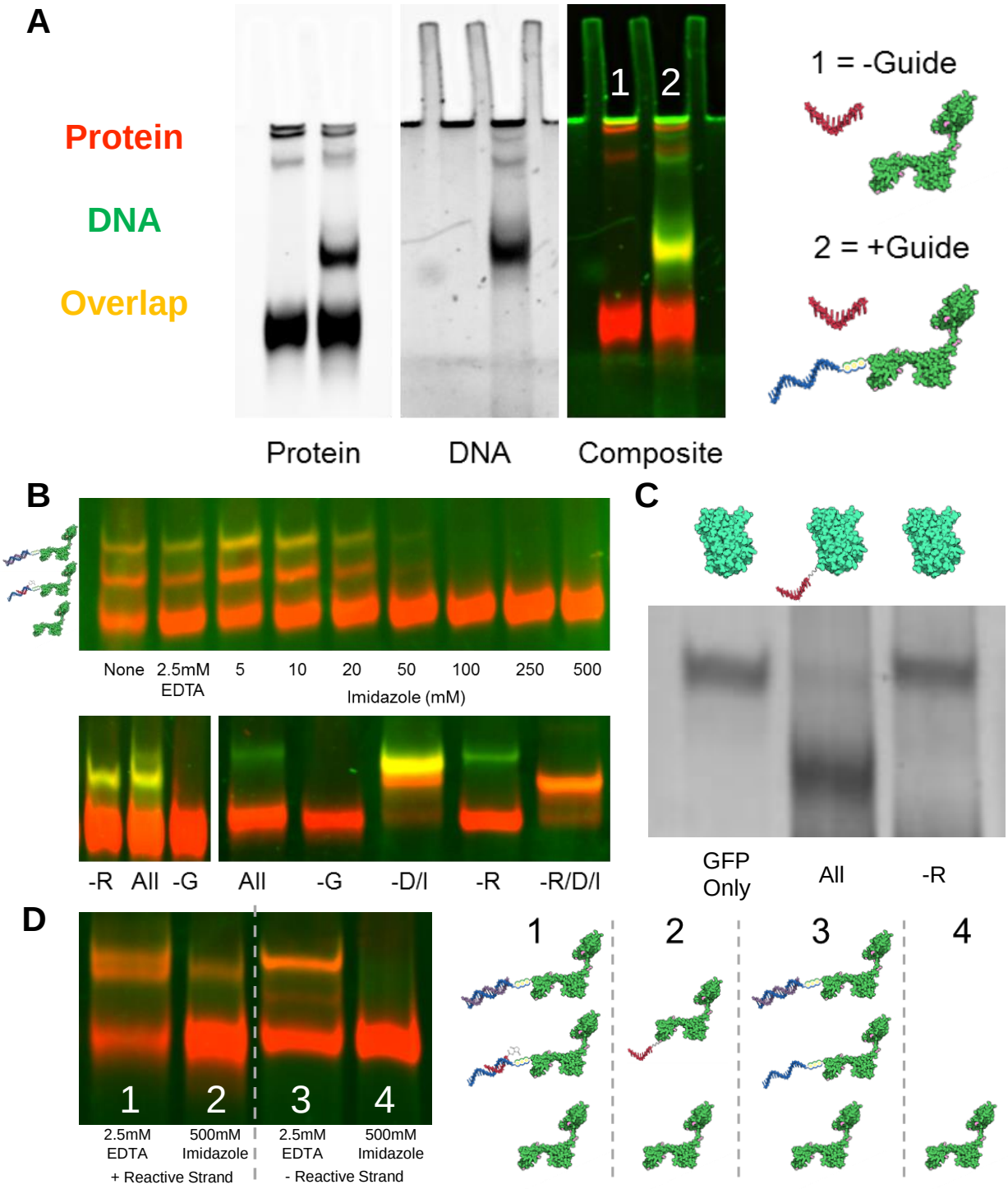


Figure 2.5 | Trials of the DTPC protocol for FliG produced conjugates in low yields but showed that NTA-His interactions could be seen on gels. A) Initial trials showed a prominent protein-DNA band that did not appear when the guide strand was omitted. **B)** Experiments with strand displacement showed that the protein-DNA species disappeared after a certain threshold imidazole concentration (top). Further trials (bottom) showed that the upper band disappeared when the guide strand was omitted (-G) and was still present even without the reactive strand (-R). Displacement used either 2.5mM EDTA (left gel) or 500mM imidazole (right gel). D = displacer, I = imidazole **C)** DTPC with GFP produced a conjugate between the protein and the reactive strand. **D)** After testing many variations of the protocol, FliG-reactive-strand conjugates were produced in low yield. The gel shows a conjugate band appearing in the presence of the reactive strand which is not removed with 500mM imidazole (2). Treating a mixture which omitted the reactive strand with 500mM imidazole broke all protein-DNA species (4). 2.5mM EDTA was not strong enough to displace strands (1), (3).

How could the addition of imidazole lead to the breaking up of the protein-DNA species? If, as I had previously assumed, this was a complex between the reactive strand and FliG joined by an amide bond, imidazole should have no effect. On the other hand, if the species observed was in fact a complex between FliG and the guide strand, held together by an NTA-Ni^{2+}-His linkage and somehow preserved even after passing through the gel, the effect of the imidazole would be much easier to explain. Imidazole competes with the His-tag for binding sites on the nickel ion (see section 2.1) and therefore at a certain concentration would break the guide-to-protein link. Thus, the simplest explanation seemed to be that the bands I had observed previously were in fact protein-guide-strand complexes and with this result in hand, I had to take a step back and re-evaluate the protein-DNA conjugates.

The Stable Assemblies Seen on Gels Were in Fact Complexes of NTA-DNA and FliG

My first test was a very simple set of controls in which I added to FliG all of the possible combinations of reactive, guide, and displacer strands as well as each of the strands on their own to see which strand was actually in the protein-DNA ‘conjugate’ band (Fig. 2.5B). The results were fairly conclusive. In all lanes without the guide strand the band in question disappeared whereas in all lanes which included the guide strand the band was present, with a definitively prominent signal even in lanes which omitted the reactive strand. Thus the species observed on the gels was in fact a guide-strand-FliG complex that had somehow survived the displacement process and a run through a gel and was not a reactive-strand-FliG conjugate.

The presence of a guide-strand-protein species on these gels led me to three conclusions. The first was that the protocol laid out in the publication for the removal of the guide strand from the protein via the addition of the displacer strand and 2.5mM of EDTA does not in fact break the guide-to-protein link, at least for FliG. As my imidazole titration (Fig. 2.5B) and later tests confirmed, it in fact takes over 200mM of imidazole to fully disrupt this linkage. The publication makes great use of denaturing PAGE gels in its read-outs of complex formation and it is possible that the SDS (sodium dodecyl sulfate), heat, or other disruptive factors involved in this technique

made up for the ineffective displacement by the EDTA and led to guide-protein dissociation. Tying in to this thought was the second conclusion that I drew: that on the native, non-denaturing gels that I was running, the guide strand was able to remain bound to the histidine-tag of the protein. To my knowledge **this is the first time that an intact NTA-Ni^{2+}-His complex has been observed using PAGE.** Having gels as a tool to visualize assemblies held together by NTA-Ni^{2+}-His links would open up possibilities for useful experiments in this project and, as is shown in section 2.4, I soon took advantage of this newfound technique. The final conclusion that I drew from this result, however, was that the conjugation reaction was not in fact working; no covalent amide bonds were being formed between the reactive strand and the lysines of the FliG protein. Thus my most immediate task was to determine what was going wrong and how I could get the conjugation reaction working.

My first test was to attempt the conjugation reaction on the His-tagged GFP protein used in the publication rather than on FliG. Using the same DNA strands and conditions as had been used for FliG and incorporating a fuller set of controls incorporating various subsets of the strands, I was in fact able to confirm a successful conjugation between the reactive strand and GFP, with a distinct conjugate band forming only in the presence of the reactive strand (Fig. 2.5C). This was reassuring but led me to think that the problems in fact stemmed from unique characteristics of the FliG protein such as the accessibility of its lysines or other parameters that I could not change.

Varying factors such as pH, salts, buffers, and concentrations as I had done previously, I was successful, after a considerable length of time, in obtaining a reactive-strand-FliG conjugate but at very low yields (Fig. 2.5D) which made purification infeasible. Attempts at improving the yield were not successful and the procedure proved to be quite unreliable, with conjugates appearing in some trials but not in others having seemingly the same components and conditions. Though one factor could not be isolated as the cause of this variability, the main culprit was most likely the unreliability of the reactive strand modification. As was discussed previously, the presence of unreactive ‘bridged’ dimers and the inability to successfully purify the correctly-modified species meant that a mixture of compounds was being added to the reaction which most likely had a very

low concentration of the strand bearing an NHS ester at one end. In addition, NHS esters are notoriously unstable in aqueous environments[163] so there was also a good chance that the reactive strand species being added to the mix had had their esters hydrolyzed.

Aside from these factors, after the trials with GFP, other questions had arisen concerning how appropriate this technique was in fact for modifying FliG. Direct comparison of the two proteins indicated that it was far easier to obtain a conjugate with GFP than with FliG leading to questions around the accessibility of FliG's lysine residues. In addition, GFP is quite a solid, barrel-like protein with few moving parts, whereas FliG contains multiple domains connected to each other by flexible linkers (see Fig. 2.2A) meaning they can explore a wide region of space. Thus it might be possible, even with the reactive strand anchored to the N-terminal His-tag, for lysines on other domains to stretch around and find themselves in close proximity to the strand's NHS ester, leading to a conjugation reaction product in which the DNA tag was linked to the M- or C-domain, the crucial regions in the domain-swap hypothesis.

Bringing these results together then, after many attempts I did manage to obtain a DNA-protein conjugate linked by a permanent amide bond, but low yields, purification difficulties, and doubts as to whether the linkage was in fact site-specific led me to reconsider whether the DTPC technique was the best way forward to produce a FliG-DNA conjugate for use in templating experiments. Along the way however I had made an important discovery: that NTA-Ni^{2+}-His assemblies could be visualized on polyacrylamide gels. This new tool opened an interesting avenue for exploration and, as I describe in the next section, I began to think about experiments in which I could use the guide strand to organize FliG on a template, observing and characterizing these complexes on gels.

2.4 FliG Templating Experiments Using NTA-DNA

Before the discovery that NTA-Ni^{2+}-His interactions could be seen on gels, the main technique that was being used to verify assembly was TIRF microscopy (see Fig. 2.1C). As described earlier,

dye-labelled FliG would be introduced to a system with NTA-DNA anchored to a microscope slide and the presence of fluorescent spots would indicate binding. This was a technically-challenging and labour-intensive technique which required a great deal of time to be spent in preparing the microscope slides and called for vanishingly small concentrations of the protein in the trials. In addition, information that could be gained using the TIRF set-up about larger FliG assemblies templated using NTA-DNA was at that point limited to the stoichiometry of the proteins in the complex, measured indirectly (and with a great deal of uncertainty) through the photo-bleaching of the protein's fluorophores[164].

Gels on the other hand provided a great many benefits. Not only were they simple and straightforward to run, they also took very little time and allowed different tests and trials to be conducted at a much higher throughput. In addition, protein and DNA staining techniques with gels meant that the FliG protein did not necessarily need a fluorophore attached in order for it to be detected and therefore a nearly-wild-type construct could be introduced into the system, one whose only non-native amino acids were the ten histidines of the His-tag at its N-terminus. Finally, though there are many advantages to the single-molecule approach of the TIRF microscope, the bulk measurements in gels meant that the overall spread of behaviours of the entire population of FliG molecules in the trial could be read out in a single scan rather than through the compilation and statistical analysis of thousands of spots on a slide.

Due to these benefits and others, from this point onwards gels became my main assay, the primary means by which I questioned and characterized the assembly of the protein-DNA complexes. My goal was to use the gels to build a picture of how FliG was being organized and behaving on DNA templates, by trying to determine the stoichiometry and spacing of the proteins, and, crucially, whether the proteins on these templates were interacting with one another. As I describe in this section, the NTA-DNA strands gave me the first glimpses of DNA-templated FliG assembly, with the complexes created forming the starting point for a narrative carried through the chapters to follow of a more and more fine-grained picture of more and more complex structures. First though,

in transferring the guide strand from its use in DTPC to its use in templating, a small but important modification was made which allowed for a stronger, more robust FliG-DNA conjugation.

Five NTA Groups on a DNA Strand Bind Stronger to a 10x His-Tag Than Three NTA Groups

The protein-DNA bands observed during the course of my trials with the DTPC technique proved that a tris-NTA-modified strand could stably bind to the His-tag of FliG and survive the passage through a native polyacrylamide gel. The His-tag in question however possessed 10 histidines, and as each NTA-Ni^{2+}-His complex incorporates two histidine residues, this means that in theory there are enough binding sites present for a total of five NTA groups. My goal was to create the strongest link possible between the DNA and protein, and it seemed logical that a linkage incorporating five NTAs would be stronger than that incorporating three NTAs, so I began work on producing a pentakis-NTA-modified DNA strand.

Sam Tusk rejoined me at this point and after some delay in acquiring a DNA strand modified with five amines, we were together able to successfully produce a **pentakis-NTA-DNA** strand in decently high yields using the same protocol as outlined in section 2.1 (Fig. 2.6A). While Sam then turned to TIRF microscopy, I turned to gels in order to characterize this pentakis-NTA-DNA and determine how well it bound to FliG in comparison to the tris-NTA-DNA strand. I applied a simple test in which I added increasing amounts of imidazole to both tris-NTA-DNA-FliG conjugates and pentakis-NTA-DNA-FliG conjugates to see at which point the complexes fell apart (Fig. 2.6B). The results showed very conclusively that the pentakis-NTA linkage was stronger than the tris-NTA linkage as it took only 50mM imidazole to break half of the tris-NTA connections as compared to 100mM imidazole for the pentakis-NTA links (Fig. 2.6C). This was reassuring as no one in the literature had worked with a pentakis-NTA-DNA structure in the past and there was a possibility that steric or entropic effects could have hampered the functionality of five NTA groups in a row[156].

Thus, with this modification in place, the guide strand from the DTPC procedure could now be put to use in templating experiments with a slightly greater assurance as to the stability of its

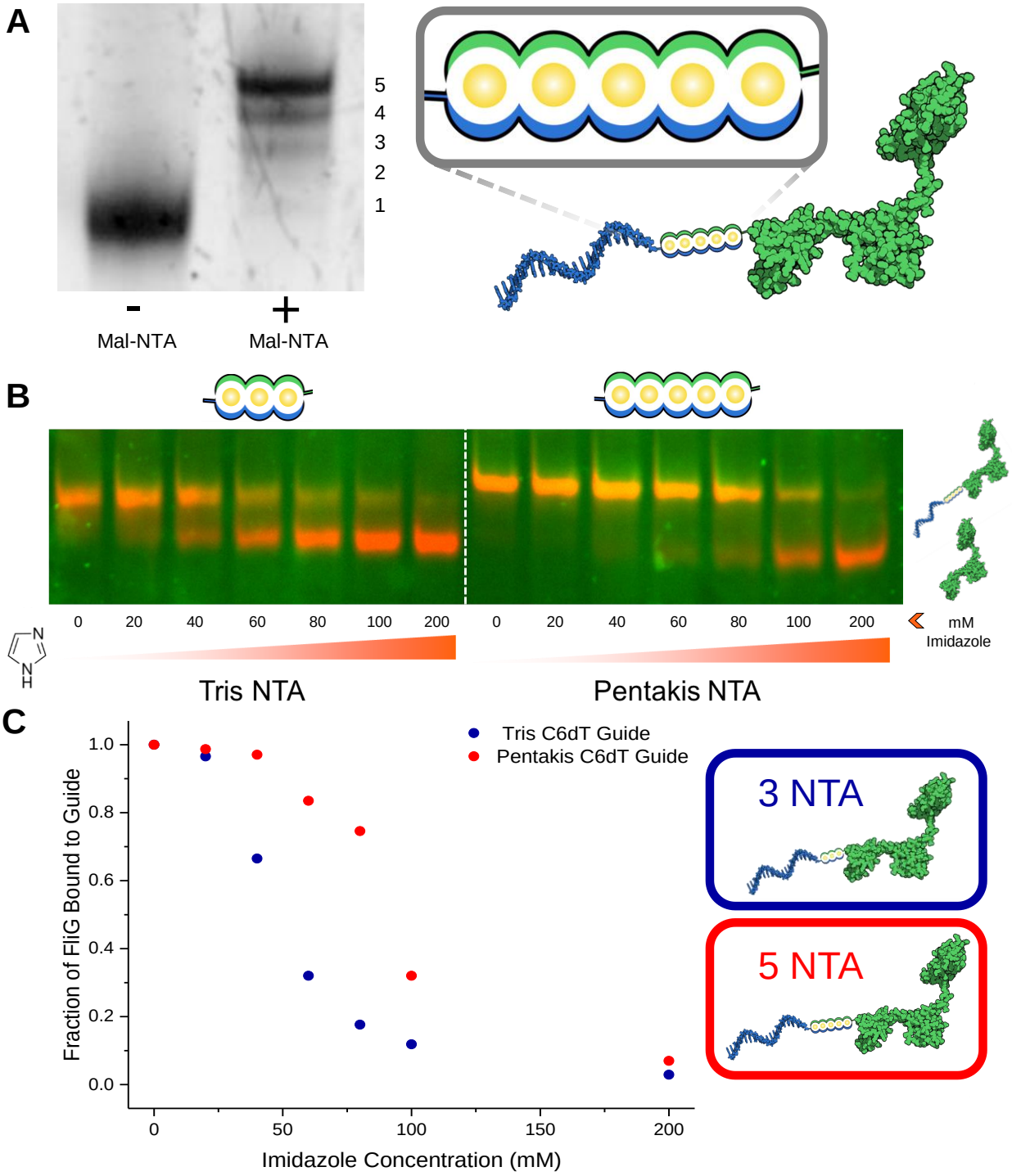


Figure 2.6 | Pentakis-NTA-modified DNA binds more strongly to FliG's His-tag than tris-NTA-modified DNA. A) The procedure that was used to modify a DNA strand with three NTA groups was modified and applied to a DNA strand with five amines at its 5' end. This resulted in the major species being a pentakis-NTA DNA strand (gel from Sam Tusk)[155]. **B)** Imidazole was added at higher and higher concentrations to FliG bound to either the tris-NTA DNA strand or the pentakis-NTA DNA strand. Dissociation was seen at higher concentrations for the pentakis-NTA strand than for the tris-NTA strand. **C)** Quantifying the results of the imidazole titration showed half of the complexes were dissociated by ~40mM imidazole for the tris-NTA strand and ~90mM for the pentakis-NTA strand.

connection to FliG. It was still not permanent like an amide bond would be but seemed to be as strong as was needed for arranging FliG on a DNA scaffold. The question now was what scaffold to template the conjugates on.

Three FliG Molecules Were Bound to a Linear DNA Template Via the NTA-His Linkage

While working to count FliG molecules using his TIRF microscopy set-up, Sam Tusk had previously designed a simple, linear, single-strand DNA template possessing three, 20-nucleotide-long binding sites accommodating three guide strands (Fig. 2.7A). Termed the **T3Guide** strand, it also possessed a 20-nucleotide extension at its 5' end for anchorage to a microscope slide, but for my purposes on gels this region was not needed and I focussed my attention on binding three FliG molecules to the three side-by-side binding sites at the 3' end of the strand.

In theory, the formation of a three-FliG complex templated by the T3Guide strand simply requires mixing the components together at the right temperature, ionic strength, and concentrations and allowing the self-assembling DNA duplexes to do the rest. Obtaining a clear picture of the species on gels however took some work. I tested various running buffers, times, voltages, gel percentages, ratios of acrylamide to bis-acrylamide, salt concentrations, and temperatures in order to obtain a gel-running protocol which could reliably give clear, separated bands of all the DNA, protein, and protein-DNA species in the mixture and allow me to identify the components in each of these bands.

Having this protocol worked out allowed me to clearly visualize the assemblies created by mixing the T3Guide strand, the guide strand, and His-tagged FliG at various concentrations. Titrating in more and more FliG or less and less template produced higher molecular weight species on the gel in a clear and distinct ladder of bands (Fig. 2.7B). Varying FliG concentrations however, I found that above a certain point, FliG dimers began to appear which seemed to impede their binding to the template. More will be discussed about this in Chapter 3 but for my purposes with the T3Guide

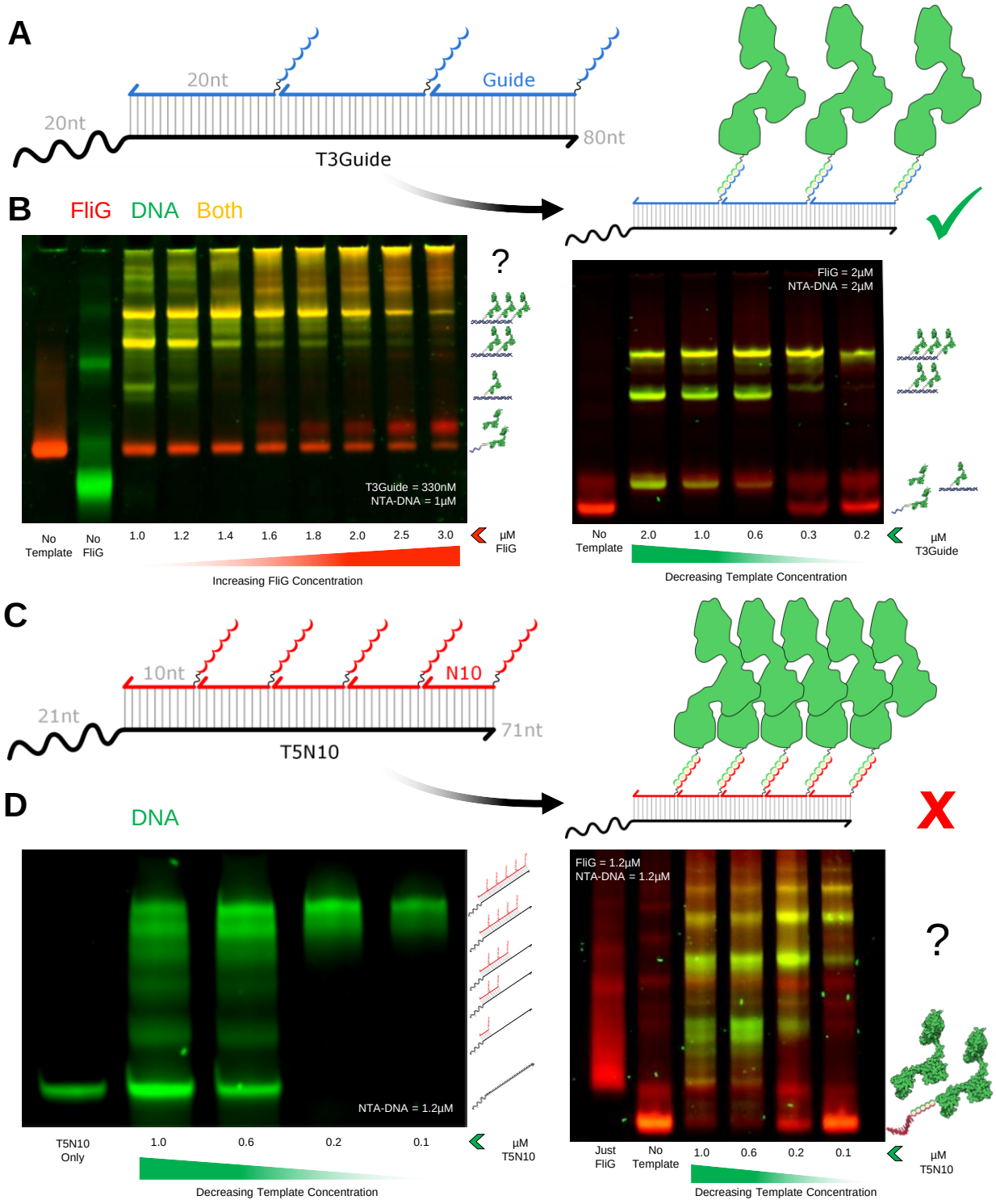


Figure 2.7 | NTA-modified DNA can be used to organize FliG on linear templates but not at the correct physiological spacing. A) The T3Guide strand is 80 nucleotides long, consisting of three binding sites for 20-nucleotide-long guide strands and a 20 nucleotide region used previously for binding to a microscope slide but unused in these experiments. **B)** Polyacrylamide gels allow for the visualization of FliG binding to the T3Guide template via the pentakis-NTA guide strand. Bands containing one, two, and three FliGs on the template could be seen along with higher weight bands which may have been caused by dimerization of the template or dimerization of the FliG protein. Using a lower concentration of the FliG protein resolved the issue of the higher weight bands (right gel, by Sam Tusk)[155]. **C)** The T5N10 template is 71 nucleotides long, consisting of five binding sites for the 10-nucleotide-long N10 strand and an unused 21 nucleotide region. **D)** Pentakis-NTA-modified N10 strands could be fully loaded on the T5N10 template (left) but the addition of FliG produced a range of species and incomplete filling of the template (right). Both gels by Sam Tusk[155].

strand, I chose to keep the FliG at 1µM and vary the template concentration, thereby avoiding the dimerization issue for the most part.

The ladder of bands visualized after these titrations allowed me to confirm that there were in fact three FliG molecules binding to the three sites on the template via their attached pentakis-NTA guide strands (Fig. 2.7B). This was the first time that we were able to determine the stoichiometry of FliG molecules on a DNA template using gels and confirmed the utility of the technique. Some problems remained however. Aside from the FliG dimerization issue, there were problems with the T3Guide template. A higher molecular weight species could be seen in lanes with the T3Guide strand alone implying some secondary structure or a tendency for the template to dimerize (Fig. 2.7B). In addition, the 20-nucleotide binding sites on the template placed FliG molecules farther apart than they would naturally be in the C-ring of the BFM. Thus, moving forward, it was determined that a new template should be designed which better captured the spacings found in the BFM.

Conjugates Using 5-NTAs Could Not Be Organized with the Spacing Found in the C-Ring

Based on the best current images of the BFM using cryo-EM and other techniques, the diameter of the C-ring ranges from around 30 to around 50nm while stoichiometry estimates for FliG range from around 26 at the low end to around 34 at the high end. Thus, each FliG molecule can be expected to be found somewhere between around 3 to 6nm from its neighbour. A single turn of the DNA double helix is completed after 10.5 nucleotides or 3.4nm so it made sense for the purposes of this project to design a template with 10-nucleotide-long binding sites. In this way the FliG molecules would be placed at roughly the same radial position around the duplex and be spaced at approximately the correct distance from one another.

Sam Tusk designed a single-stranded, linear template based on the 10-nucleotide N10 sequence he had used earlier in TIRF microscopy experiments. Accommodating five identical binding sites each 10-nucleotides long, this **T5N10** template could in theory line up five FliG molecules side-by-

side with the spacing found in the BFM (Fig. 2.7C). More will be discussed on this, and variations of this, template in the chapters to come but Sam was successful in showing that it could accommodate five 10-nucleotide-long DNA strands which had undergone the pentakis-NTA modification (Fig. 2.7D). When incorporating FliG into the system however, the same filling of the template could not be achieved. Even with excess FliG in the mixture, using the gel procedures that I had developed, Sam consistently found the templates under-filled with free, unoccupied binding sites present in the dominant species (Fig. 2.7D). Thus it appeared that with the components being used it was not possible to organize the FliG molecules with the correct, physiological spacing.

After much deliberation, it was concluded that our problems most likely stemmed from the bulkiness of the pentakis-NTA groups. Such a large modification at the end of the DNA strands could perhaps lead to steric hindrance and blocking of the binding sites on the template, making them inaccessible for incoming FliG-DNA conjugates. Thus it was decided that, moving forward, a different conjugation strategy should be employed which introduced a smaller, more compact linkage. Working on this new conjugation strategy became my focus over the coming months as Sam, having completed his degree, was now finished with the project.

Before moving on to my work on this new method for conjugation, which is the focus of Chapter 3, it is useful to briefly summarize the progress made in this chapter. Sam and I were able to successfully modify a DNA strand with three NTA groups yet had difficulty in purifying the correctly modified strand from the others (2.1). While the modification of the DNA was problematic, the expression and purification of the FliG protein followed a robust and reliable protocol which allowed for many different constructs to be isolated via the interaction of the protein's histidine-tag and an IMAC column (2.2). This same interaction was used in a new conjugation strategy which I employed called DNA-templated protein conjugation (2.3). The technique used an NTA-modified guide strand to position an NHS-ester-bearing reactive strand close to specific lysine residues on the protein, thereby encouraging a site-specific reaction. I was able to produce conjugates using this method but at too low yields to purify and with doubts as to the site-specificity of the reaction.

This technique was therefore abandoned as I chose to explore a discovery that I had made during the process: that NTA-Ni^{2+}-His interactions were stable and visible on a native polyacrylamide gel. This discovery proved quite useful as, using gels, I was able to visualize three FliG proteins organized on a three-site DNA template through connection to a pentakis-NTA-DNA strand (2.4). This same connection was used to try to template five FliG proteins at the physiological spacing seen in the BFM but the attempt was unsuccessful due perhaps to the steric bulk of the pentakis-NTA groups.

Thus I moved on to a new conjugation strategy, as the next chapter describes, looking to employ a technique with a more compact and permanent linkage between the DNA and protein. As I will show, this proved to be quite successful, and combined with the read-outs possible using PAGE, allowed for direct testing of the domain-swap hypothesis.

CHAPTER THREE

A SIMPLER STRATEGY

Conjugation of Maleimide-DNA to a Single-Cysteine FliG Mutant

My work with NTA-modified DNA and His-tagged FliG had produced some useful advances for the project. First, it had led to the development of gel-based techniques for looking at DNA-templated FliG assemblies. Experiments with gels were much faster and higher throughput than the experiments done previously with the TIRF microscope and produced clear evidence of FliG-DNA conjugates binding to short linear templates. These short linear templates were another useful development to come from the NTA-DNA work as they provided an ideal testing ground, to be used before moving to larger, more ring-like structures, for determining how well the conjugates bound to their respective binding sites and at which spacing of the binding sites the conjugates could fill the template.

As was discussed at the end of the previous chapter however, the NTA-DNA based techniques had run into a dead-end. With the bulky NTA-Ni^{2+}-His complexes joining the DNA to the protein, the conjugates were not able to fully fill a template designed to have binding sites with the appropriate, physiological spacing. The priority then became to search for a new conjugation technique, one with a more compact, permanent linkage. Ideally the method would also be a simpler one than the NTA-to-His conjugation with fewer dials to turn. Rather than worrying about concentrations of DNA, nickel, and protein, an ideal technique would allow for a single, purified protein-DNA conjugate of a known concentration to be added to the mix, alleviating worries about the effects of excess DNA or FliG in solution.

In this chapter, I describe one solution to this problem and demonstrate that conjugates produced through this new technique can be bound to and fully fill sites of templates with the correct, physiological spacing. With worries about conjugation then secondary, the chapter proceeds with my testing of a more ring-like template based on an origami tile and my investigation of FliG-FliG interactions through *in vitro* cross-linking and various gel-based techniques. The story of these interactions was not quite as simple as would have been hoped however and a great deal of work was needed before a system could be obtained which ensured any non-native interactions were removed.

The removal of these non-native interactions fits into a general trend seen throughout this project of bringing the *in vitro* system more and more in line with the *in vivo* complex and as this chapter proceeds I will highlight how the decisions made throughout were motivated by a drive for a closer and closer mimicry of the C-ring.

3.1 Maleimide Modification of a DNA Oligonucleotide and Its Conjugation to Cysteine-Modified FliG

To produce a covalent protein-DNA species in a site-specific manner, I chose one of the most popular and trusted reactions in the field of bioconjugation, one which results in a much more compact linkage than that seen with the pentakis-NTA-modified DNA strand. Using a **maleimide-modified DNA strand**, one can produce stable conjugates in an aqueous, biologically friendly buffer through the maleimide's reaction to the thiol group of a cysteine on the protein[157] (Fig. 3.1A). And as this reaction is one of the most tried and tested, being developed over several decades, there are many published protocols and commercially-available cross-linkers to assist in applying the process to one's system of choice.

In my case, I chose the sulfo-SMCC (sulfosuccinimidyl 4-(N-maleimidomethyl)cyclohexane-1-carboxylate) cross-linker (Fig. 3.1B) as the means by which I would introduce a maleimide group into the DNA strand to be conjugated to FliG. Its small size and high levels of solubility in aqueous buffers make it one of the most well-used cross-linkers for bioconjugation reactions and its

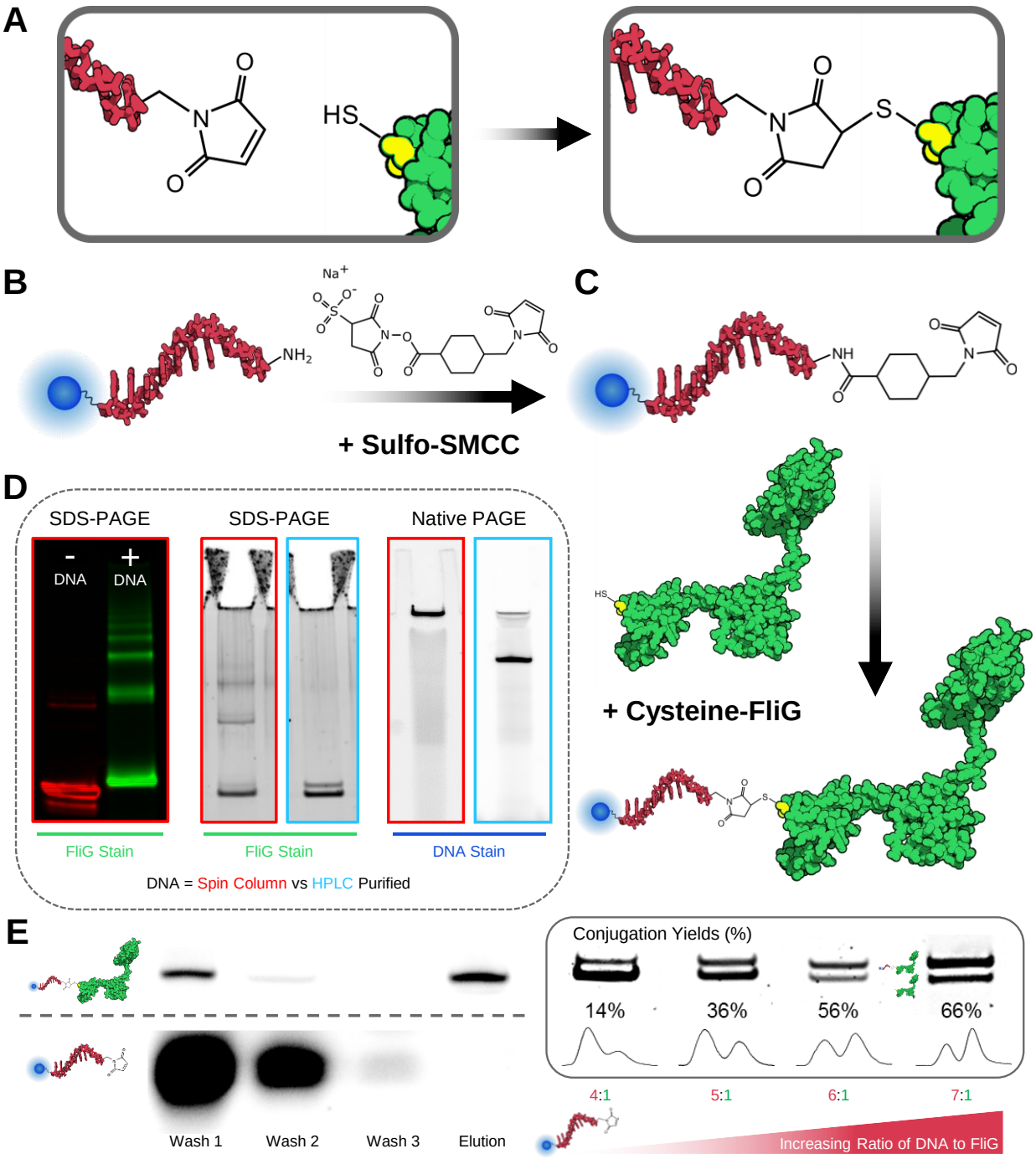


Figure 3.1 | Maleimide-modified DNA must be HPLC purified before conjugation to a single-cysteine FliG mutant. A) A maleimide group reacts with the thiol of a protein's cysteine residue to produce a stable thioether bond. **B)** The sulfo-SMCC cross-linker introduces a maleimide group at the 5' end of a fluorophore-labelled DNA strand via the reaction of its NHS ester with the amine of the DNA strand. **C)** The maleimide-modified DNA then reacts with the thiol of a single-cysteine FliG mutant to create a permanent FliG-DNA conjugate. **D)** Upon addition of spin-column-purified maleimide-DNA (+ in left gel), covalently-bound oligomers of FliG were observed which formed aggregates in native PAGE gels (red boxes). HPLC-purification removed excess sulfo-SMCC which led to the disappearance of the oligomers and aggregates (blue boxes). **E)** Purification of the conjugates was achieved by binding the His-tag of the protein to an IMAC resin and washing away the unbound excess maleimide-DNA. Conjugates were then eluted from the resin with a high concentration of imidazole. Adding an increasing amount of maleimide DNA led to better yields for the conjugation, with close to 70% of the protein being conjugated to DNA in the most successful trials.

heterobifunctional nature, having an NHS ester at one end and a maleimide group at the other, meant that I would not run into the problem of bridged DNA dimers as I had with the DSS cross-linker during the reactive strand modification in the DTPC protocol.

For the protein end of the reaction, I would use a construct recently characterized and added to the library of plasmids developed by the Lee lab in which a single-amino-acid mutation had placed a cysteine near the N-terminus of the FliG protein from *Salmonella* (named **S5C FliG**) (Fig. 3.1C). Luckily for my purposes, FliG in its wild type state does not possess any cysteine residues, meaning that the introduced cysteine would present the only thiol on the protein for reaction with the maleimide-DNA, guaranteeing a site-specific conjugation.

Thus, with a small and permanent link insured by the sulfo-SMCC cross-linker and the site-specificity dictated by the single cysteine of the new FliG construct, a plan was in place that would overcome the difficulties experienced with the NTA-DNA strands. As the first step in that plan I set about to produce a maleimide-modified DNA strand.

A Maleimide Group Was Introduced to the 5’ End of a DNA Strand Using Sulfo-SMCC

For simplicity, the strand I chose to use possessed the same 10-nucleotide N10 sequence as had been used previously for the reactive strand of the DTPC protocol and for the pentakis-NTA-modified strands binding to five-site templates. The 5’ end of this strand was modified with an amine for maleimide modification while at the 3’ end a fluorophore was attached for easier identification on gels and to allow for later use in TIRF microscopy experiments (Fig. 3.1B). This strand was dissolved in a buffered, aqueous solution with a slightly alkaline pH and the sulfo-SMCC cross-linker, solubilized in DMF (dimethylformamide), was added in great stoichiometric excess to ensure that as many of the aminated DNA strands as possible would react with the NHS ester of a sulfo-SMCC molecule. The reaction was left shaking at room temperature and, once completed, the mixture was passed through a size-exclusion spin column in an effort to remove the excess, unreacted cross-linker from the modified DNA strands. The solution was then

concentrated down to a pellet using a vacuum concentrator, ready for use in the conjugation reaction.

Though the DNA strand showed no change in migration on polyacrylamide gels before and after the addition of the cross-linker, this was not surprising due to the small size and low charge of the sulfo-SMCC molecule. Running a small aliquot through a reversed-phase HPLC column however showed a reassuring shift of the main peak towards the hydrophobic end of the trace after the addition of the cross-linker, indicating that in all likelihood the modification had been successful. The best way to truly test this however was simply to mix the strands with thiol-bearing molecules, so I chose to proceed directly with the conjugation to the single-cysteine FliG mutant.

Conjugation of Maleimide-DNA to Cysteine-FliG Required HPLC Purification of the DNA

The FliG mutant bearing the single cysteine near its N-terminus (S5C FliG) (Fig. 3.1C) was expressed and purified as described in section 2.2 and small aliquots were prepared for use in preliminary tests with the maleimide-modified DNA. The protein was stored in a buffer at pH 7.2, well within the range in which the maleimide group reacts specifically with cysteines[165], and at concentrations on the order of 10μM. To ensure as high yields as possible of the protein-DNA conjugate, the protein was added directly to a stoichiometric excess of the pelleted DNA strands, thereby keeping concentrations high. In addition, the molecule TCEP was added to ensure that the conjugation reaction was carried out in reducing conditions, thereby breaking any disulfide bridges that might have formed between the cysteines of two FliG molecules.

Following overnight incubation of the DNA with the protein to ensure the reaction had gone to completion, the mix was run on a native polyacrylamide gel to determine the extent of the conjugation. The extra size and negative charge given to the FliG protein upon attachment of the DNA strand would in theory produce a gel-shift as seen with the experiments in Chapter 2. In addition, because the DNA strand bore a fluorophore label, it would be very easy to determine

which bands contained the DNA and a large shift was to be expected between unconjugated DNA and DNA bound to the FliG protein.

Initial results however were not promising. On native gels it was seen that after addition of the DNA, the protein formed large aggregates in the wells, being too large to enter the gel (Fig 3.1D). My initial thought was that the conjugation of the DNA to the protein had led to denaturation of the FliG, perhaps due to instability of the FliG-DNA conjugates. An unfolded protein molecule would expose hydrophobic patches which would bind to other hydrophobic patches of other unfolded protein molecules leading to the formation of the aggregates seen on the gels. To test this thought I decided to run the reaction products on a denaturing SDS-PAGE gel. This gel would break apart any hydrophobic interactions in the aggregate, leaving only covalent connections and my expectation was that I would see a monomeric conjugate band.

Surprisingly however, the SDS-PAGE gels showed ladders of protein-DNA conjugate species (Fig. 3.1D). Large oligomers of covalently-attached FliG proteins with a range of stoichiometries had formed and these most likely corresponded to the aggregated species in the wells of the native PAGE gels. Thinking through what might have caused this laddering, I ruled out both the maleimide-DNA and the protein itself. The maleimide-DNA has only one reactive group and hence could not connect multiple FliG molecules, while the only covalent bond that a FliG protein could make would be a disulfide bridge via its cysteine. This at most would make a dimer, not a ladder of FliG molecules and had in theory been eliminated through the use of TCEP. Thus, most likely, there was another species in solution doing the cross-linking and thinking through the possibilities, the most probable culprit was excess, unreacted sulfo-SMCC cross-linker. The NHS ester at one end of the cross-linker could react with a lysine on one protein while the maleimide at the other end could react with the cysteine of a neighbouring protein. Oligomers could form through the head-to-tail linking of lysines to cysteines on multiple proteins via multiple sulfo-SMCC molecules.

In my protocol I had used size-exclusion spin columns to remove the excess sulfo-SMCC but perhaps this step had not been quite as successful as I had previously thought. To test this, I replaced the spin columns in my protocol with a run of the reaction mixture through the HPLC

system. After purifying the main peak seen in the HPLC trace, I concentrated the mixture to a pellet and then mixed the pellet with S5C FliG and TCEP. The results were quite dramatic. On native gels the aggregates disappeared leaving only a single monomeric FliG-DNA band while on denaturing gels the ladders were no longer present and a clear covalently-linked FliG-DNA conjugate band could be seen above the unconjugated FliG band (Fig. 3.1D). This made me reasonably confident that the original problem was in fact excess cross-linker in solution and it appeared that HPLC purification successfully removed the stray sulfo-SMCC that had previously reacted with the lysines and cysteines of the FliG molecules.

Confident that I now had a protocol established that could produce monomeric, covalent FliG-DNA conjugates, I then proceeded to develop a purification technique which could remove excess DNA from the solution. This was crucial if the conjugates were to be used in templating experiments as excess DNA would bind to and block binding sites on the templates which should be occupied by FliG-DNA conjugates. As opposed to previous conjugations using NTA-modified DNA, the His-tag of these conjugates was free and could be used for purification purposes. This was quite a useful benefit of the technique as I was able to use IMAC resin to bind the conjugates as I washed away the unbound, excess DNA (Fig. 3.1 E). Unfortunately, this technique did not allow me to separate conjugated from unconjugated FliG molecules (as both have His-tags) and attempts to purify away unconjugated FliG by size were also unsuccessful. This was not a serious problem as free, unconjugated FliG proteins would not block binding sites on the template and could be differentiated from FliG-DNA conjugates due to their lack of a fluorophore. In addition, work to optimize the conjugation reaction led to a procedure that conjugated a single DNA strand to around 70% of the total FliG molecules, meaning that unconjugated FliG was the minority species (Fig. 3.1E).

Thus, with procedures developed to produce and purify maleimide-DNA, to then conjugate that DNA to single-cysteine FliG, and finally to purify the conjugates away from excess DNA, I was ready to begin experiments in which I would bind these new conjugates to DNA templates. The coming sections highlight the progress made with these templating experiments and demonstrate

the successes that this technique enabled, allowing for fundamental questions to be probed relating to the FliG domain-swap hypothesis. Before these successes could be achieved and questions be investigated however there was one final obstacle to overcome, a puzzle which took nearly half a year to solve and which held the project at a standstill during that period.

3.2 Histidine-Tag-Mediated FliG Dimerization

With the conjugation procedure now working well, the next task was to start binding these conjugates to a DNA template as had been done with the NTA-DNA conjugates. Ideally with their compact maleimide-cysteine linkage, the new conjugates could be organized on a template with the appropriate physiological spacing as could not be done using the previous conjugation method. To test whether this was in fact the case, the first set of trials with the new conjugates was a simple repetition of the experiments that had been done with the old conjugates in which they were added to the five-site T5N10 linear DNA template (see Fig. 2.7), a template which incorporates 10-nucleotide-long binding sites placing the FliG proteins at roughly the correct physiological spacing.

Using the fluorophore on the DNA strand of the conjugate to monitor its position on native polyacrylamide gels, these initial tests clearly showed a gel-shift upon addition of the template (Fig. 3.2A). Unfortunately, these gels did not show a single gel-shifted band as might be expected for five conjugates bound to the template but instead exhibited a range of bands created by a wide array of different species whose identity could not easily be ascertained simply through analysis of the gel. Although trials with variants of the template, different salts, and stoichiometries of the components were conducted, it appeared that these extra bands were an effect of the conjugate itself and not the template or the environment in which it was placed. Looking at lanes in which the conjugate was run on its own without any template added showed two prominent bands rather than the one that would be expected for a single, monomeric species (Fig. 3.2A), and it seemed logical that this extra species might be the root cause of the confusing spread of bands that the templating gels were showing.

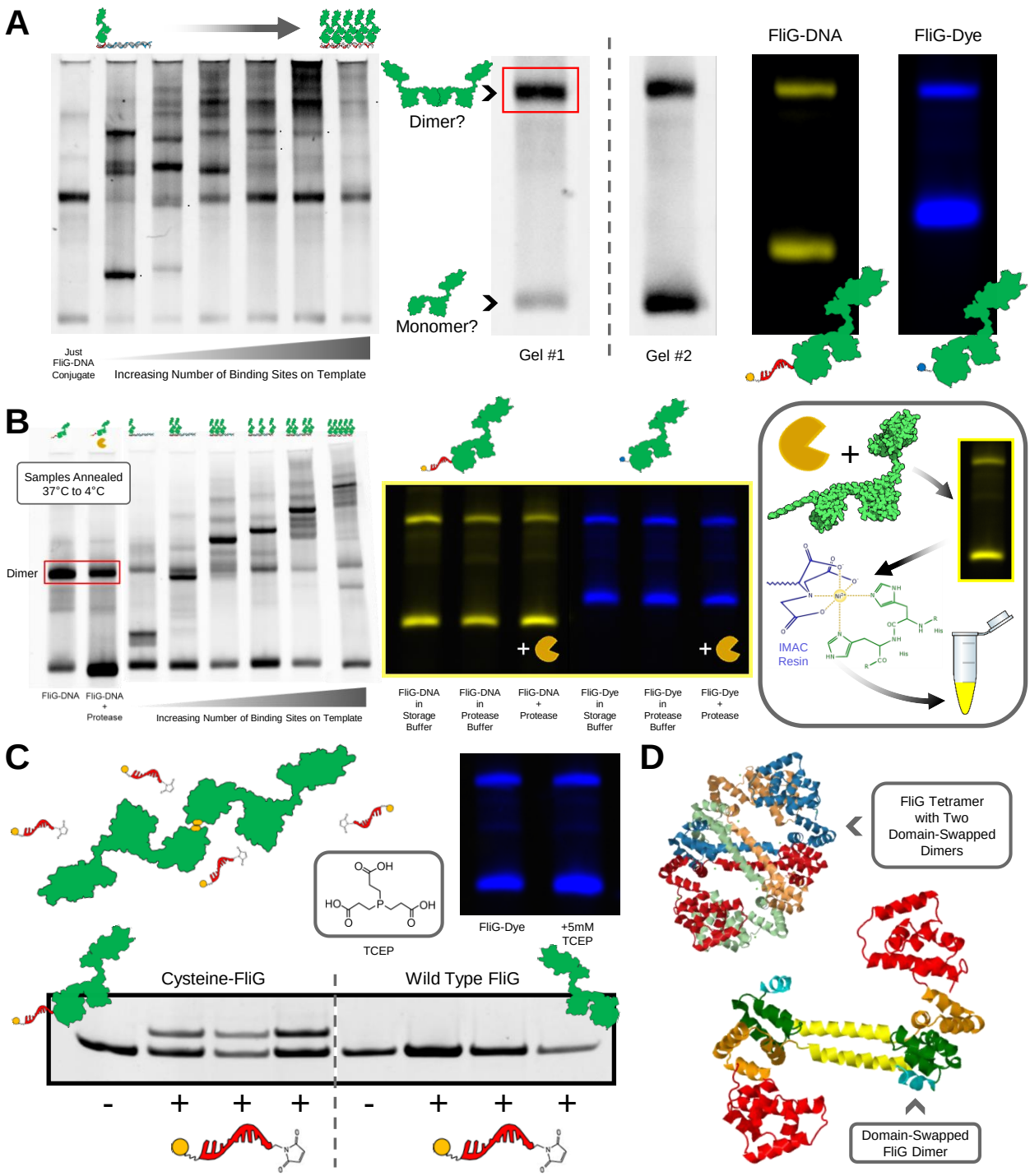


Figure 3.2 | Initial attempts at characterizing the FliG-FliG dimer species seemed to discount His-tags and cysteines as the cause of the interaction. A) The first experiments seeking to template the FliG-DNA conjugates constructed via the maleimide-cysteine reaction were unsuccessful, producing gels with indecipherable bands. The root cause of this seemed to be a dimeric species (red box) present in lanes containing only FliG-DNA conjugates. This species was seen to varying degrees on different gels but appeared in several different constructs including the DNA- (yellow) and dye-labelled (blue) FliG shown. **B)** One outlying gel showed a reduction in dimeric species and a decipherable ladder of bands when the conjugate was added to various templates. A His-tag-cleaving protease had been added in these mixes but attempts to repeat this result were unsuccessful. Conjugates treated with the protease and passed over an IMAC resin were shown to flow right through, seemingly implying that the His-tags had been cut. **C)** Cysteine-cysteine dimers were investigated but addition of TCEP did not break the species and non-specific binding of the maleimide-DNA to wild-type FliG was not observed. **D)** Domain-swapped dimers were found in one crystal structure[85], in which the proteins were held together by reciprocal, head-to-tail interactions of opposing ARM motifs.

Comparing the bands seen in the conjugate's lane with other standards of known weight and charge, it appeared that the lower band was the single, monomeric conjugate I was looking for while the upper band ran to approximately where **a dimer of FliG-DNA conjugates** would be expected to run to. Testing the conjugates on a denaturing SDS-PAGE gel showed only a single band meaning that the interaction that was producing this 'dimer band' must be a reversible, non-covalent one. The domain-swap polymerization model for FliG assembly says that without the aid of a template to increase the local concentration, FliG should be in a monomeric form. In addition, in a search through the literature I found no mention of FliG dimers or of any oligomeric state for that matter apart from monomers *in vitro* and the full rings *in vivo*. Thus, the species that was present on my gels was one that had not been previously observed or described and before I could obtain any useful results from templating experiments with these conjugates, I needed to understand why and how it was forming.

What interactions were holding the conjugates together? If it was dimerization, what was causing it? What would it take to break the dimers and return the conjugate to monomeric form? I set out to determine what was happening to the FliG-DNA conjugates I had created by investigating and characterizing this higher-molecular-weight species. Determining what in fact it was and how to remove it to allow the templating experiments to proceed took the better part of six months. In this section I describe the process by which I narrowed down the cause of this dimerization and how in the end I managed to produce monomeric FliG-DNA conjugates which could be used to test the domain-swap hypothesis.

Initial Results Spurred Investigation into the Possibility of Domain-Swapped FliG Dimers

Following on from my review of the literature which produced no evidence for FliG dimers in the protein's native, *in vivo* setting, I concluded that the dimer species I was observing must be the result of non-native interactions between components that had been added to the FliG construct I was working with. This meant there were three possibilities to test. Of the three non-native elements in my construct, the first to test, which to me seemed the most likely element to introduce

a non-native interaction, was the relatively large and highly charged polymer I had just conjugated to the protein in the form of its DNA tag. Aside from this, the protein also had a 10-residue-long His-tag added to its N-terminus, and a single cysteine residue introduced for the conjugation, secondary possibilities to be considered should the DNA tag be shown not to be involved in the dimerization. Thus, I set about testing these three components in turn, attempting to eliminate each individually and monitoring via gels the extent to which the higher-molecular-weight species was forming.

To begin, I looked to determine if the DNA strand which had been conjugated to the FliG protein was in some way responsible for the dimerization. Using native polyacrylamide gels, I ran samples of the same construct which had not had a DNA strand conjugated to it or had had a small organic dye conjugated at the same site. These unlabelled or dye-labelled constructs clearly showed the same dimer-sized band present as with the DNA-labelled construct (Fig. 3.2A). Thus, it was fairly definitive that the DNA-tag was not responsible for the interactions seen on the gels and what seemed originally to be the most compelling explanation was rejected.

Moving on, I next tested the possibility that the Histidine-tag was responsible for the dimerization. This seemed a rather unlikely option as I could find no account in publications working with other protein systems of a His-tag introducing spurious, dimerizing interactions. In fact, work with FliG to obtain several different crystal structures had left the His-tag attached to the protein with no obvious detrimental effect during the crystallization procedure[90]. The possibility was quite straightforward to test however as the construct I was working with incorporated an amino acid sequence between the His-tag and the rest of the protein which could be recognized and cleaved by a human rhinovirus (HRV) 3C protease, thus removing the His-tag from FliG (see Fig. 2.2).

Adding in the protease to the DNA-labelled FliG constructs initially showed promising but unreliable results. In one trial, I added the protease to both the conjugate on its own and to samples in which the conjugate was mixed with various linear templates, annealing the mixes from 37°C to 4°C with the hopes that his would aid in the templating process. (As will be later shown, this annealing was a key factor in the result obtained but this was not known to me at the time.) I was

surprised to see that the distribution of species in the lane containing only the conjugate had clearly shifted towards the monomer and that in the other lanes bearing templated conjugates I had obtained the clearest indication so far of the conjugates binding in the correct stoichiometries (Fig. 3.2B). This seemed to be a definitive result, pointing to the His-tags as the key factors responsible for the dimers and disrupting the templating, but it was a result which I could not repeat. Further trials adding in the protease to the conjugate showed no change in the distribution of dimer to monomer (Fig. 3.2B) and the templating gels again produced indecipherable bands.

Puzzled, I looked to confirm that the His-tags were in fact being cleaved by the protease, thinking that perhaps something was wrong with the batch of protease that I was using in these more recent trials. To do so I passed the protease-treated conjugates through an IMAC resin and monitored for the presence of the conjugates at the various binding, washing, and elution phases. If the His-tags had been cleaved, the conjugates should pass right through without binding, while if they had not, it would take elution with imidazole to remove them from the resin. As the conjugates were labelled with fluorophores on their DNA strands it was easy to visually track where they were during the process and it was immediately obvious that they were not binding to the resin (Fig. 3.2B). Gels confirmed that this was in fact the case and I thereby assumed that the His-tags had in fact been cleaved even in these dimer-forming samples. Thus, I decided to test the final non-native element in my construct: the single N-terminal cysteine, to see if it was in some way responsible for the dimerization.

A cysteine-cysteine dimerization seemed a more likely possibility than a His-tag-mediated one as I had often observed cysteine-cysteine dimers when running unconjugated, single-cysteine FliG on denaturing SDS-PAGE gels with the reducing agent omitted, thereby allowing disulfide bridges to form. With the FliG-DNA conjugates however, a dimer of this sort would indicate that the cysteines were free and accessible to form a disulfide bridge rather than being taken up in a thioether bond with the maleimide-DNA. The protein-DNA species seen on gels would therefore have to be the result of non-specific binding of the DNA to alternative sites on the surface of the protein, perhaps the amines of lysine residues[166]. I was slightly skeptical of this non-specific

labelling (and by extension cysteine-cysteine dimerization) as it appeared that FliG was only ever conjugated to a single DNA strand but it was simple to test and it seemed to be my last reasonable explanation for the dimerization so I started investigating the possibility.

Two experiments confirmed quite definitively that the dimers observed were not mediated by the cysteine residues of the construct. The first was the simple addition of the reducing agent TCEP. At concentrations which broke the disulfide bridges seen in unconjugated FliG on denaturing gels, the dimer band was still present in the FliG-DNA conjugates on native gels (Fig. 3.2C). The second experiment proved that the DNA labelling was in fact specifically via the cysteine residue. Adding maleimide-DNA to a FliG construct identical to the single-cysteine version but missing the cysteine showed no FliG-DNA conjugation while the same buffers and reaction conditions showed a clear conjugate band for the single-cysteine construct (Fig. 3.2C). Thus, its seemed very unlikely that the N-terminal cysteine residue was responsible for the dimerization and it was confirmed that my conjugation procedure had indeed produced site-specific conjugates.

Having now seemingly eliminated the three main non-native elements of the construct: the DNA, the His-tag, and the cysteine residue, it was left to conclude that the interaction seen on the gels was mediated by the native residues of the FliG protein. Though nothing of this sort had been reported for FliG *in vivo*, it was not impossible that FliG dimers had been missed in previous publications working with the complex environment of a living cell. I therefore began to search for possible ways in which the FliG protein could dimerize in its native state. Looking through various crystal structures of the protein, it came to my attention that crystals of a truncated form of *T. maritima* FliG in one publication were made up of dimerized FliG building blocks[85], and though these interactions were not claimed to be physiological, it was still an interesting possibility that they might be.

What stood out about this structure was that the motifs responsible for the dimerization were the same as those hypothesized to be involved in the domain-swap model for FliG polymerization. In fact, as Fig. 3.2D shows, the two FliG proteins in this structure were held together in a head-to-tail fashion by domain-swapping of the ARM motifs. Domain-swapped dimers such as this are

common occurrences in biology with many different proteins using this technique to bind to their partners (see section 1.4). It seemed a logical explanation then that the dimers I was observing on my gels were in fact a natural product of the interactions between ARM motifs and that perhaps this dimerization was physiologically-relevant. Moving forward I decided to further characterize the interaction by screening for its stability in a variety of conditions with the hope that, once understood, I could somehow fit these dimerized species into the model for FliG-ring assembly and work them into the templating experiments.

Characterization Showed the Dimers Were Broken with Heat and Not Domain-Swapped

In the domain-swap model for FliG polymerization, the cytoplasmic form of the protein is thought to be a closed, inert monomer in which its ARM motifs form an intramolecular interaction. This form hides away the hydrophobic patches of the ARM motifs and prevents any unwanted oligomerization until the protein is bound to its FliF template and in close proximity to its neighbours[101] (see section 1.4). The closed, cyclic architecture of a domain-swapped FliG dimer also hides away the ARM motifs of the proteins making it inert and preventing further oligomerization. Thus, it could be possible that instead of the closed monomers (or most likely in equilibrium with them), perhaps domain-swapped dimers constitute the inert, cytoplasmic form of the FliG protein. These dimers could bind to the FliF template and oligomerize into rings, thereby generally taking the role that in the current model the closed monomers do (Fig. 3.3A).

To determine if the species I was observing were in fact domain-swapped dimers and to thereby give support to the modified version of the domain-swap oligomerization model just described, I needed to characterize the dimer-forming interactions to better understand what they were composed of and how strong they were. I therefore conducted various tests and put the constructs through a range of conditions to narrow in on the nature of the interaction.

My first test was simply to look for dimers in different *Salmonella* FliG constructs as well as in FliG from *E. coli* to see if the interaction I was seeing was a peculiarity of the single-cysteine S5C FliG

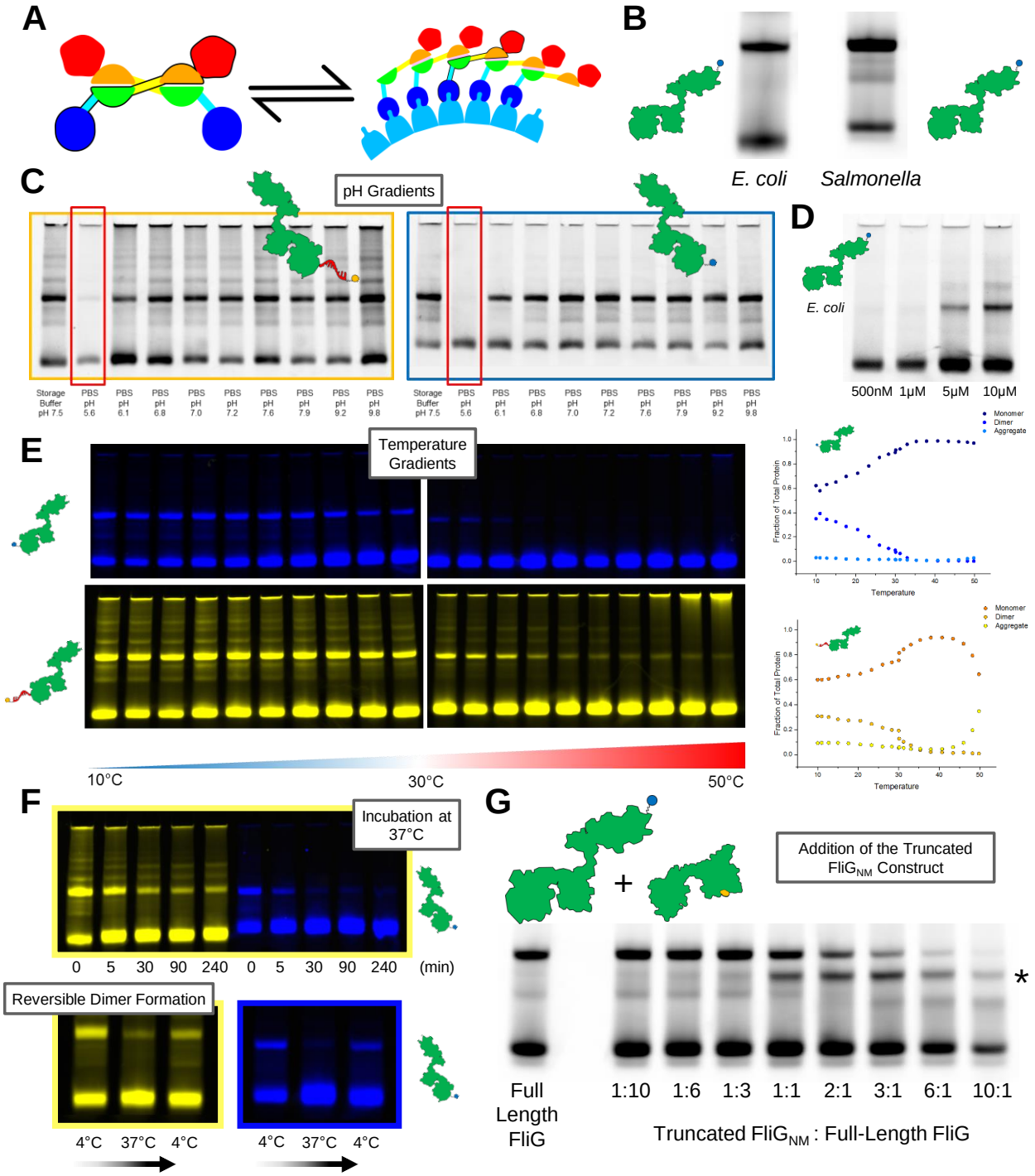


Figure 3.3 | Characterization of the FliG-FliG dimers showed their formation was sensitive to several different factors. A) A domain-swapped FliG dimer could in theory act as the inert, cytoplasmic form for the protein, burying the hydrophobic patches of the ARM motifs until binding to the FliF template. **B)** Dimer bands were seen across many different constructs including the C-terminally-dye-labelled proteins from *E. coli* and *Salmonella* shown here. **C)** Incubation of both DNA- and dye-labelled FliG in buffers at a range of pHs showed that dimers disappear below pH 6.0. **D)** Concentration gradients showed that dimers disappear at sub-micromolar concentrations. **E)** Temperature gradients from 10°C to 50°C showed that up to around 37°C, heat decreases the overall percentage of dimers in solution in favour of monomers. Beyond 40°C however, FliG begins to denature and forms aggregates in the wells. **F)** Incubation of conjugates at 37°C showed that by 30 minutes a new equilibrium is established (top). Returning the conjugates to 4°C leads to the reformation of dimers (bottom). **G)** Adding in FliG$_{NM}$, a truncated version of FliG missing the C-terminal domain, to a full-length FliG construct at a range of stoichiometries showed that with the truncated form in excess, a new dimer band forms (*) presumably containing one full-length and one truncated FliG molecule.

I was using. Screening a series of other constructs on native polyacrylamide gels, I found dimers present in all cases with prominent dimer bands seen in both wild type and single-cysteine versions, in *Salmonella* and *E. coli* FliG proteins (Fig. 3.3B). Thus, the dimerization was not confined to a single construct but common to all the FliG proteins tested, lending support to the thought that the interaction was a more general one, perhaps between opposing ARM motifs.

I next tested the effect of pH on the linkage by incubating both dye-labelled and DNA-labelled, single-cysteine FliG in phosphate-buffered solutions with pHs ranging from 9.8 down to 5.6. Interestingly, the dimer band disappeared below pH 6.0, leaving only the monomer (Fig. 3.3C). This was quite an unambiguous result but a puzzling one if these were to be domain-swapped dimers. The interactions holding together ARM motifs are for the most part hydrophobic and a change in pH in theory should not affect their binding. Thus, I began to doubt whether ARM-ARM interactions were in fact holding this species together.

I then looked into the effect of concentration on the distribution of monomers and dimers in the FliG solution. A serial dilution of the protein from 10μM down to 500nM showed that concentration did in fact have a profound influence on the extent of dimerization (Fig. 3.3D). A decrease in the concentration led to a decrease in the percentage of dimerized FliG, with dimers completely disappearing below 1μM for the construct used. Even if it was being formed through the native amino acids of the FliG protein, this result cast doubt on the physiological relevance of the dimeric species. The best estimate of FliG's cytoplasmic concentration within the bacterium is around 500nM[101], and in the titration I had done, 500nM FliG was completely monomeric. Thus, if the estimates for *in vivo* FliG concentration were correct, it seemed unlikely that a dimerized form of FliG would exist to any great extent in the cytoplasm. Even so, the dimers were having a significant impact in my *in vitro* system and still needed to be understood.

The next factor that I looked at in relation to the dimerization was temperature. As with any protein, FliG was expected to unfold and denature at a high enough temperature but it was my hope that I could find a temperature below this point at which the protein remained folded, but which provided enough thermal energy to break apart the FliG-FliG interaction. Finding such a temperature could

give me an idea of the strength of the interaction and allow me to quantify the binding with various thermodynamic parameters. I therefore placed aliquots of both dye-labelled and DNA-labelled single-cysteine FliG on a thermocycler with a gradient of temperatures ranging from 10°C to 50°C and incubated the aliquots at these temperatures over several hours. Running these aliquots on a gel after the incubation showed that the extent of dimerization decreased with increasing temperature and that by 37°C the dimers had mostly disappeared (Fig. 3.3E). It also showed, as expected, that at temperatures above ~40°C the FliG protein began to denature and form aggregates in the wells, but if I worked at 37°C I should be able to break apart the dimers without having to worry too much about denaturation.

Calculating the distribution of monomers to dimers from the gels and assuming a simple, one-step dimerization, the binding energy of the interaction at 37°C was determined to be $\Delta G^{37°C}$ = -20 kJ/mol. This value is similar to that for the binding energy calculated *in vivo* for the interaction between FliM and FliG[95] and hence seemed a reasonable value for interactions between FliG molecules. No accurate measure of the strength of FliG-FliG interactions had previously been calculated but most theories for the assembly of the C-ring postulated that the strength was roughly similar to that seen in the FliG-FliM and FliM-FliM bonds of the structure[101], implying that the interactions I was seeing were of roughly physiological strength.

With temperature as a newfound tool for manipulating the state of FliG, I then conducted several other experiments in order to determine to what extent I could use heating and cooling for later work. I again found that 37°C was the best temperature at which to work in order to break dimers without denaturing the protein and determined that beyond 30 minutes of incubation at this temperature, the equilibrium between monomers and dimers did not change (Fig. 3.3F). In addition, I tested whether or not the interactions which were broken upon bringing the protein to 37°C could be reformed if the mix was brought back to a cooler temperature. Incubating overnight first at 37°C and then for 2 hours at 4°C showed that the dimers did in fact reform in the cooler temperatures and that, as predicted, the interaction was reversible (Fig. 3.3F).

Taking the results of these experiments with temperature together, it appeared that though the strength of the interaction roughly matched the predicted strength for the *in vivo* FliG-FliG bond, the dimers were most likely not physiologically relevant as they broke apart at 37°C, the cytoplasmic temperature for most *E. coli* and *Salmonella* bacteria. This matched the previous assertions from the experiments with concentration and left me with the thought that all future experiments should be conducted at 37°C. Unfortunately, this high temperature interferes with the hybridization of FliG's DNA tag with the template, so for my current system this was not an option.

Aside from physiological relevance, in my final set of experiments to characterize this dimerization I looked to see if I could in any way confirm or deny that the interaction was mediated by the ARM motifs. To test this, I expressed and purified a truncated version of FliG called $FliG_{NM}$ which omitted its C-terminal domain and hence one of its two ARM motifs: ARM_C. The thought was that if the dimers were ARM-ARM domain-swapped, adding this construct in at high enough concentrations would allow the free ARM_M motif of $FliG_{NM}$ to bind to the ARM_C motif of one full-length FliG in the dimer, displacing the ARM_M motif of the other, thereby breaking the domain-swapped structure (Fig. 3.3F). There was also a thought that after this displacement, perhaps the structure formed containing one full-length FliG and one $FliG_{NM}$ with an exposed ARM_M motif would act as a nucleus for runaway FliG polymerization as described in section 1.4.

Titrating the truncated $FliG_{NM}$ into a solution with dimerized full-length FliG showed that with the $FliG_{NM}$ in excess, the dimers were in fact broken and a lower molecular-weight species formed which I took to be a complex of one full-length and one truncated FliG molecule (Fig. 3.3G). No runaway polymer aggregate was seen however meaning that if this complex was in fact forming, it was not acting as a nucleus for FliG polymerization. This set of experiments then proved somewhat inconclusive and did not definitively determine whether or not the interaction was in fact an ARM-ARM domain-swapped dimer as had been hoped.

Looking back at what I had done to characterize the dimerization, I found that several important discoveries had been made. The fact that dimers could be seen across several mutants and species of the FliG protein, as well as the fact that the interaction was sensitive to temperature

and concentration, and that its strength was on par with that proposed for a FliG-FliG interaction all indicated that the dimers could in fact be a relevant species *in vivo*. Looking quantitatively at the temperatures and concentrations at which the dimer formed however seemed to indicate that the species was not in fact physiologically-relevant, as at 500nM and 37°C, the conditions found in the cytoplasm of a bacterium, there were only monomers. The experiments with pH and $FliG_{NM}$, though inconclusive, also seemed to hint that the species was not being held together via the ARM motifs, leaving me again puzzled as to what could be going on.

With all this information consolidated, I considered overhauling my system of DNA templates so that I could work at 37°C, abandoning my search for what was causing the dimerization and simply avoiding it altogether by working at a higher temperature. Before doing so however, I revisited one outlying result that I had obtained back in my work with His-tags. Addition of a His-tag-cleaving protease had seemed to shift the distribution of FliG molecules towards a monomeric state and had produced my only clean, understandable templating gel (see Fig. 3.2B). This result was never repeated and, still curious as to how this could be explained, I looked to see if perhaps the recent data obtained relating to the dimerization could shed new light on this result and the dimerization problem in general.

Dimerization of FliG Proteins Was Mediated by Their Histidine-Tags

Looking back at the steps I had taken which had produced the outlying gel several months previously, I realized that one key factor in the process had not been given appropriate attention. I had added the His-tag-cleaving protease to the mixtures of conjugates and templates, did not bother removing the protease, and then annealed the solutions from 37°C down to 4°C in an attempt to improve formation of the correct DNA duplexes. In light of my recent work with temperature, I now realized that the period of time spent at 37°C during the annealing process would have broken apart the FliG-FliG dimers leaving the conjugates in their monomeric form. My experiments with temperature also suggested however that returning the solutions to 4°C would allow the dimers to reform (see Fig. 3.3F) but in the case of this particular gel, a clear shift towards the monomer species remained even at 4°C. To try to explain this I formulated a theory for the

dimerization which could be tested with the tools at hand, and, if correct, would allow me to obtain monomeric conjugates which could be predictably organized on DNA templates.

My thoughts went as follows. Rather than a cysteine-cysteine or domain-swapped ARM-ARM dimer, it was the His-tags that were causing the dimerization and that in my previous trials which had discounted the His-tag due to the presence of dimers after addition of the protease (see Fig. 3.2B), I had incorrectly assumed that the His-tag had been cleaved. I had seen that protease-treated dimers did not bind to an IMAC resin and hence had assumed they lacked His-tags but had not run the key control of passing untreated dimers through that same resin. Now it occurred to me that the constructs had run freely through the resin not because their His-tags had been cleaved but because the His-tags were hidden and sequestered away from the solution upon dimer formation, meaning that both the IMAC resin and the protease could not access them (Fig. 3.4A). Heating to 37°C however appeared to break the interaction and reveal the previously hidden His-tags, meaning that when I had annealed the mixes for the gel months ago, I had in fact made the His-tag cleavage site available for the protease which I had not bothered to remove from the solution. Thus, in this particular case, the His-tags had in fact been cleaved and the His-tag-cleaved conjugates could take a monomeric form, thereby explaining why the conjugates had bound in a predictable fashion to the template.

This seemed a promising explanation for the results I had seen and quite a simple one to test. To do so I simply heated the conjugates to 37°C to break apart the dimers, incubated them for some time with the His-tag-cleaving protease, brought the solution back down to 4°C, and monitored via gels for the presence or absence of dimers. The results were definitive. Whereas with untreated conjugates dimers reformed when cooled, heat-and-protease-treated conjugates remained monomeric when returned to 4°C (Fig. 3.4B). Refining the protocol further, I was able to cleave and make monomeric nearly all of the conjugates in solution and to then remove the protease from the mix via affinity chromatography using its GST (glutathione S-transferase) -tag (Fig 3.4B). Thus, it was concluded, after half a year of investigation, that the dimers were in fact His-tag-

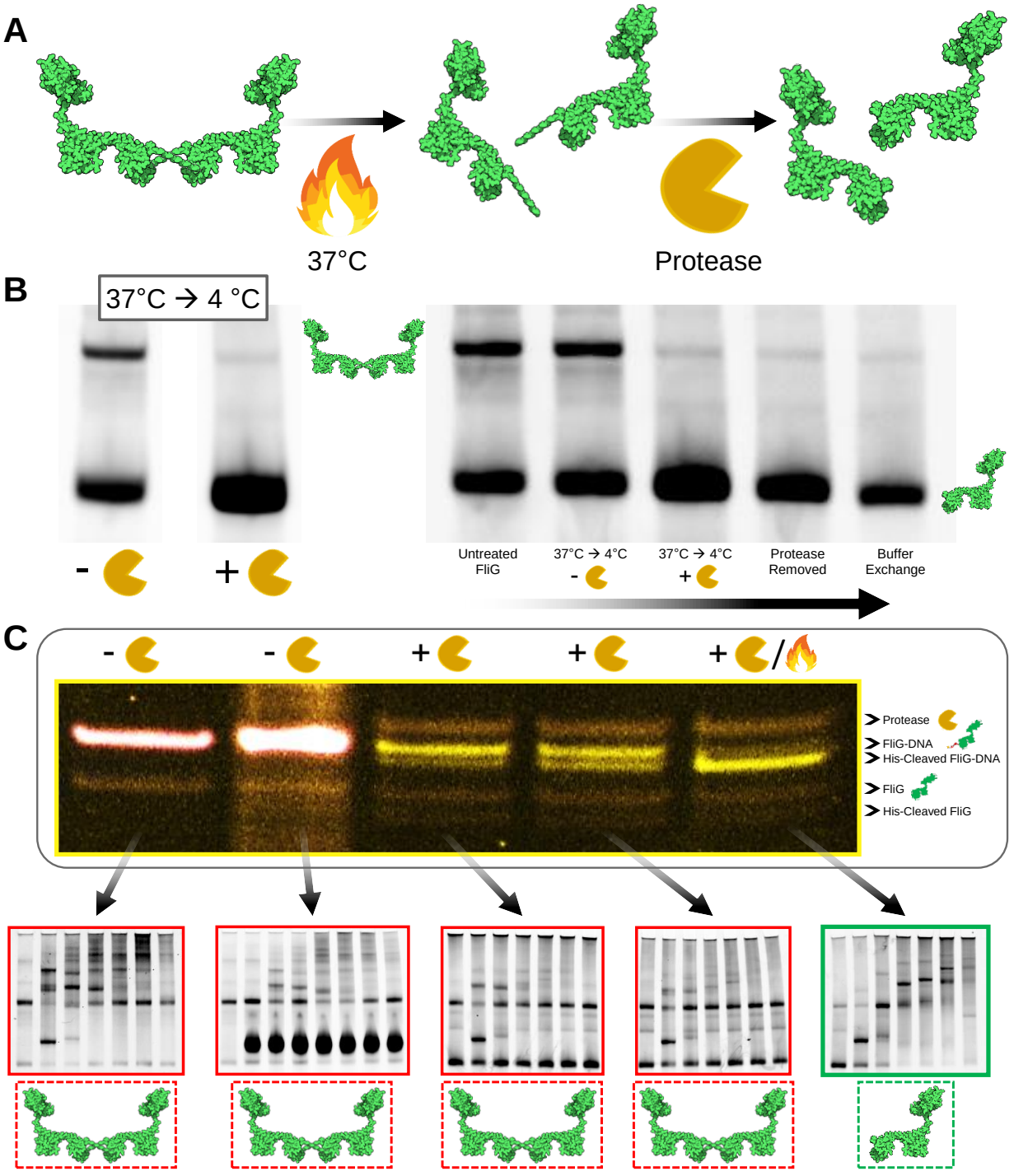


Figure 3.4 | FliG-FliG dimers can be removed by first incubating at 37°C and then treating with a His-tag-cleaving protease. A) His-tagged FliG constructs dimerize via their His-tags. These dimers can be removed by first incubating at 37°C for at least 30 minutes to break the interaction and expose the His-tags, and then by adding a protease to cleave at the site between FliG and the His-tag. **B)** Once the His-tag is cleaved, the dimers do not reform when the solution is returned to 4°C (left). The protease can be removed and the FliG constructs returned to their preferred storage buffer with minimal losses (right). **C)** The degree of His-tag cleavage can be monitored using denaturing SDS-PAGE. Distinct bands corresponding to species with or without their His-tag cleaved can be observed. Templating experiments were attempted with each of the five samples shown. The first two were not treated with protease and dimers dominated (the second included BSA protein). The third and fourth samples were treated with the protease but were not incubated at 37°C prior to treatment and though some cleavage is observed, dimers still are the predominate species. The final sample underwent the full treatment as outlined in (A). This removed the dimers, producing mostly monomeric FliG. This monomeric FliG then bound to templates in a predictable and decipherable manner.

mediated, and that the native, monomeric form of FliG could be regained simply through cleavage of the His-tags.

This conclusion fit well with my previous results as all the previous constructs that had shown FliG-FliG dimerization had been expressed with His-tags. In addition, the fact that the proteins monomerized below pH 6.0 during my pH screen was quite telling (see Fig. 3.3C). Histidine has a pK_a around pH 6.0 meaning that its imidazole moiety becomes protonated below that point[167], thereby changing the electrostatic properties of the His-tag as a whole, which, as it seems, breaks its interaction with its partner in the FliG-FliG dimer. Molecular particulars of the dimerization aside, what was important was the fact that now, at long last, I could obtain monomeric FliG-DNA conjugates at high yields for use in templating experiments and I was eager to move forward with these.

To conclude this section then, my initial attempts at templating experiments with the FliG-DNA conjugates I had created via the maleimide-cysteine reaction had ended in confusion. The main cause of these indecipherable gels appeared to be a dimer-sized species appearing in solutions containing the conjugates and these dimers were interfering with the templating. Initial attempts to remove this dimeric species had resulted in a single successful experiment which could not be repeated with trials then shifting focus towards the possibility of cysteine-cysteine and then ARM-ARM domain-swapped dimers. Further characterization showed that the dimers disappeared below pH 6.0, at concentrations below 1μM, and at temperatures above 37°C but could reform if returned to 4°C. Re-analysis of the earlier outlying success showed that temperature had played a key role and that the dimers were most likely His-tag-mediated. This was confirmed as treating the conjugates with heat and then a His-tag-cleaving protease produced monomeric FliG-DNA conducive to templating experiments (Fig. 3.4C).

With these monomeric conjugates now available, experiments with various DNA templates could now proceed and, as section 3.3 demonstrates, questions concerning the domain-swap polymerization model of FliG assembly could finally be investigated.

3.3 Experiments Using Permanent FliG-DNA Conjugates on a Five-Site Linear Template

After this long six-month sidetrack dealing with dimers, the focus once again returned to the organization of FliG-DNA conjugates on DNA templates. The goal at this stage of the project was to place the conjugates at the correct physiological spacing from one another to mimic the organization seen in the C-ring of the BFM, as this had not been accomplished with the NTA-DNA conjugates. The thought was that placing the conjugates at the correct spacing from one another and thereby increasing their local concentration might be enough to induce domain-swap polymerization and several different experiments using native polyacrylamide gels would allow me to determine whether or not these interactions were actually taking place.

As my template for these experiments, I used the same **T5N10** linear DNA strand that had been used in previous unsuccessful attempts to template NTA-DNA conjugates. This strand contained five binding sites, each with the 10-nucleotide-long **N10 sequence** and thus could accommodate five FliG-DNA conjugates (see section 2.4). To accurately assign bands present in gels to the appropriate FliG-DNA complexes however, I also incorporated several other variants of the T5N10 template into my system. These strands possessed from one to five binding sites for the FliG-DNA conjugate at various locations along the 50-nucleotide stretch, with the regions of the template which did not constitute a binding site being composed of domains bearing the 10-nucleotide-long **dN10 sequence** (Fig. 3.5A). These dN10 domains could be filled by a complementary 10-nucleotide-long strand or left single-stranded, thereby changing the overall rigidity of the template and, in some designs, introducing longer or shorter gaps between the FliG-DNA binding sites.

With this new toolbox of strands and with monomeric FliG-DNA conjugates free of their His-tags, I set out to investigate the effects of organizing FliG molecules on a template with spacings at which the molecules could interact with one another. In this section I will show that I was successful in organizing the conjugates as I had hoped, and this enabled me to probe for the presence of intermolecular FliG-FliG interactions. As I will also show however, immediate confirmation of the

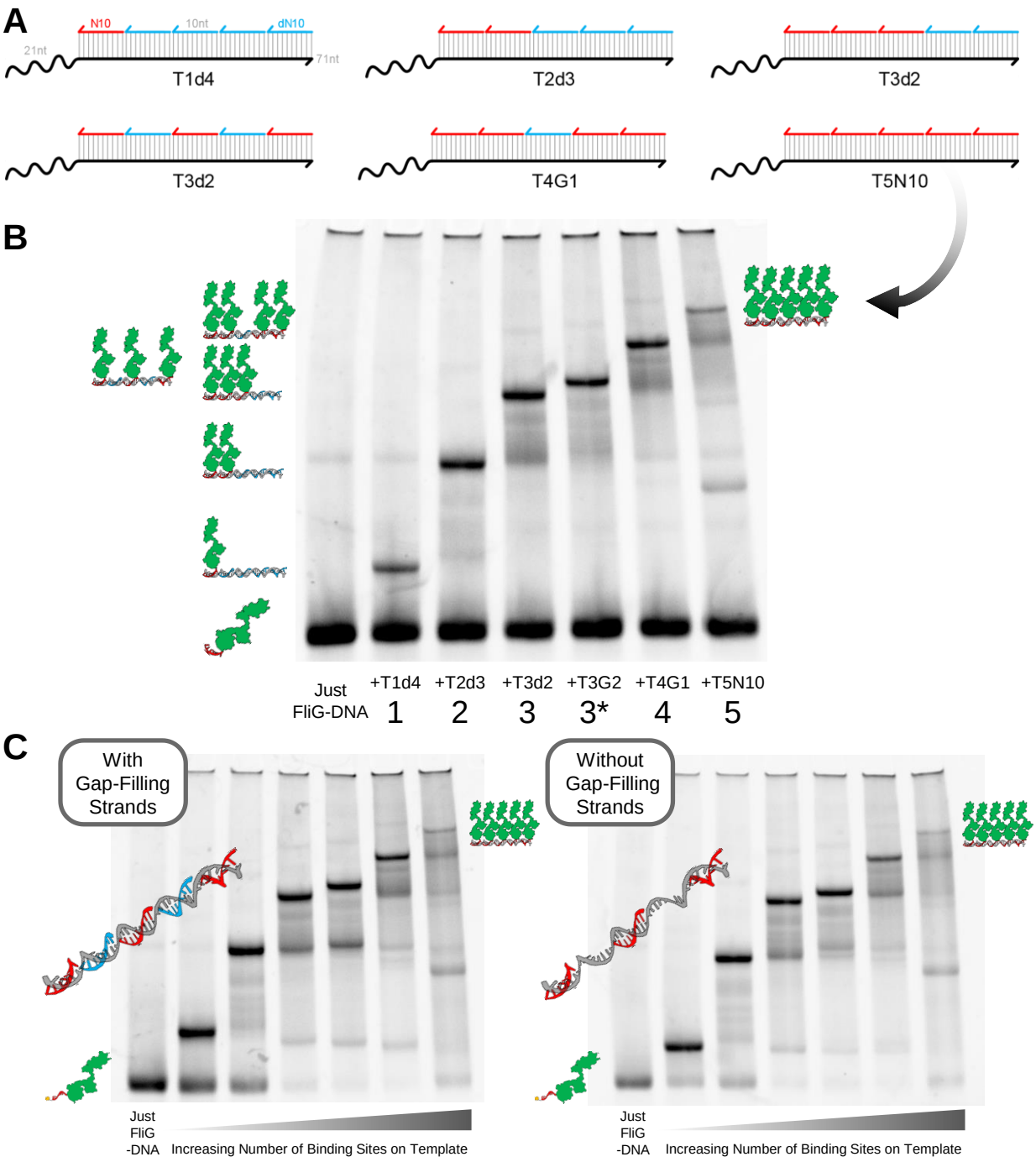


Figure 3.5 | Native polyacrylamide gels demonstrate that FliG-DNA conjugates can be organized with stoichiometries and configurations dictated by the sequence of the underlying DNA template. A) Variants of the five-site T5N10 template were used which incorporated from one to five FliG-DNA binding sites (red) at various spacings from one another. Spacing regions contained the 10-nucleotide dN10 domain (blue). **B)** Mixing excess FliG-DNA with the six different templates and running the mixes on a native polyacrylamide gel showed that the templates bound the correct number of FliG-DNA conjugates as dictated by the configuration of their binding sites. A stepwise series of bands corresponding to a range from one to five proteins bound to the template can be observed. In addition, two different configurations of a three-site template (T3d2 and T3G2) produced bands running to different lengths (3 and 3*), indicating that the gels could distinguish complexes to some degree based on the spacing of the templated molecules. **C)** Adding or omitting the 10-nucleotide gap-filling strands during the templating procedure produced little change in the banding patterns of the gels. Note the lack of high-molecular-weight species above the dominant band in each lane. This indicates that both gap-filling and runaway polymerization by FliG molecules is not occurring.

domain-swap hypothesis was not achieved and as my work beyond this section demonstrates, experiments with these simple five-site templates constituted only the very beginning of a long search for FliG-FliG interactions.

FliG-DNA Conjugates Can Be Placed on a Template at the Correct Physiological Spacing

Although initial attempts at adding maleimide-cysteine-linked FliG-DNA conjugates to DNA templates had produced definite band-shifts (see Fig. 3.2), it was not until the issue of His-tag-mediated dimerization was resolved that clear and distinguishable species could be picked out on gels. Rather than a spread of different bands observed upon addition of the template, with dimerization eliminated, a single dominant species could be seen in each lane, with a different species observed depending on which template was added (Fig. 3.5B).

Significantly, the size of the band-shift seen on these native polyacrylamide gels corresponded very nicely with the number of binding sites the added template possessed. Addition of a template with a single FliG-DNA binding site resulted in a band shifted only a small distance upwards from the FliG-DNA band whereas addition of a two-site, three-site, four-site, and then five-site template produced progressively larger and larger shifts (Fig. 3.5B). Thus, it appeared that the FliG-DNA constructs were binding to the DNA templates with the correct stoichiometry. In addition, two variants of a three-site template strand had been constructed which placed the FliG molecules either directly side-by-side (T3d2) or spaced out by dN10 domains (T3G2) (see Fig. 3.5A). Though with the sites occupied by FliG molecules these two constructs would have the same molecular weight, when run on a native gel the species migrated to different lengths, with the T3d2 species travelling farther than the T3G2 species (Fig. 3.5B). Thus, not only the mass, but to some extent the overall geometry of the constructs could be identified using these gels.

The single, characteristic band produced from the addition of a distinct template to the FliG-DNA conjugates indicated that the DNA-hybridization was working as it should, but no clear-cut evidence could be seen of FliG-FliG interactions induced by the templating process. Two main

predictions of the domain-swap hypothesis should in theory produce distinct signatures on the native gels and neither could be detected. Both runaway polymerization, wherein the domain-swap process stimulates the uncontrolled growth of a FliG polymer, and gap-filling, in which untemplated FliG molecules slot into free spaces between templated molecules, were not visible on the gels, as the bands observed constituted species held together only by DNA-hybridization with no extra weight being added through the interaction of free FliG molecules in solution.

In an attempt to induce these interactions between templated and untemplated FliG molecules, I applied several variations of the templating procedure with the hope that a small tweak in the protein's local environment would be sufficient to stimulate protein-protein binding. Adjusting the salt, pH, the overall concentrations and relative stoichiometries of the components, and several other factors showed no change in the pattern of bands visible on the gels and hence led me to believe that no FliG-FliG interactions were taking place in any of these conditions. I also tested a small variation to the templates in which I made the gaps between FliG binding sites either single- or double-stranded by omitting or including short gap-filling strands with the dN10 sequence (Fig. 3.5C). Single-stranded gaps in theory would be shorter and introduce more flexibility into the system than double-stranded gaps and perhaps these extra degrees of freedom would help FliG assume the correct orientation for binding. The bands on the gels however showed no evidence for FliG-FliG binding in either case and hence led me to believe that more dramatic changes or additions to the system were needed to induce an interaction.

Another possibility however was that the means by which I was searching for interactions, simply looking for extra bands on gels, was not sensitive enough for this system and that different assays were needed. To investigate further, I developed two new methods to search for FliG-FliG interactions and these I describe next.

Native Polyacrylamide Gels Showed No Evidence of FliG-FliG Interactions on Templates

Looking for a more sensitive way to pick out FliG-FliG interactions should they be present in my system incorporating the five-site templates, I expressed and purified another batch of FliG with a single cysteine at its N-terminus and conjugated a maleimide-dye to the protein. Whereas the DNA-labelled FliG possessed a fluorescein fluorophore on its DNA strand, this dye-labelled batch of protein was given a fluorophore at the other end of the visible spectrum, with maximum excitation at 647nm. Thus, I could easily differentiate based on the colour of the label whether the FliG in question was conjugated to a DNA strand or not and hence had two separately identifiable species to work with.

My aim with this new dye-labelled FliG was to add it in excess to solutions containing FliG-DNA bound to various templates and to monitor for the presence of interactions by looking for upward shifts of the dye-labelled FliG band on gels. Previously, when excess FliG had been added it was in the form of excess FliG-DNA which possessed the same fluorophore as those conjugates bound to the binding sites of the template. With an orthogonally-labelled pool of excess FliG which lacked the somewhat bulky DNA tag, I could more accurately determine whether interactions were in fact forming and had simply been missed on previous gels.

Adding this dye-labelled FliG in a four-to-one stoichiometric excess to the DNA-labelled FliG bound to various five-site templates however showed a definitively monomeric band of dye-labelled FliG in all cases (Fig. 3.6A). In addition, the signal from the DNA-labelled FliG on the templates appeared relatively unchanged in the presence or absence of the excess dye-labelled proteins. These two observations clearly indicated that either untemplated FliG molecules were not interacting with templated FliG molecules at all or that the method by which I was testing for this interaction, namely native polyacrylamide gels, was breaking the interactions apart before I could detect them.

Though free-floating FliG could not be observed binding to FliG on a template, one possibility that the results of the previous tests did not exclude was that neighbouring templated FliG molecules

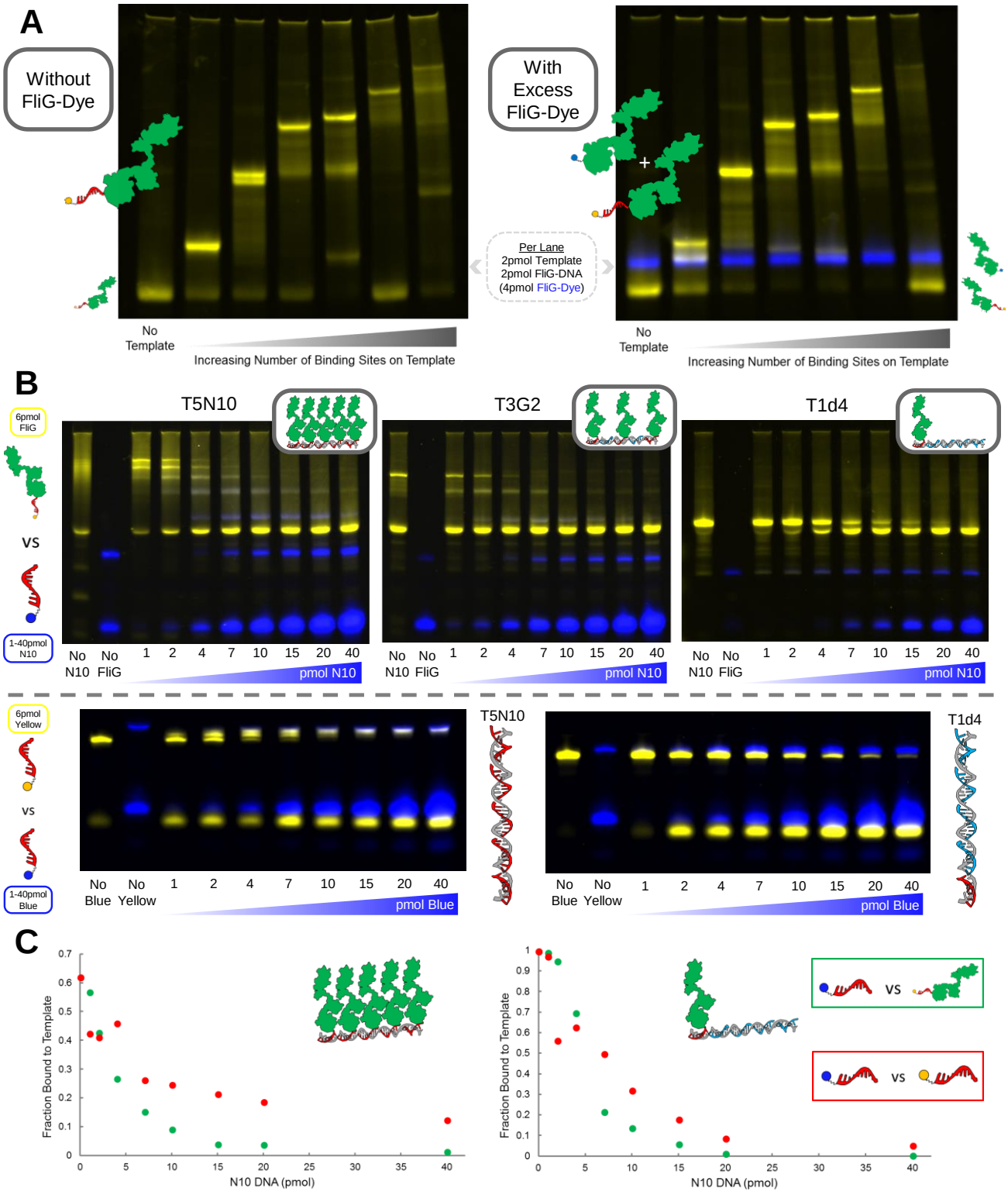


Figure 3.6 | Gels show that FliG-FliG interactions do not occur either off or on the DNA templates. A) Addition of free-floating dye-labelled FliG (blue) in a four-to-one excess over DNA-labelled FliG bound to linear templates (yellow) showed no evidence of interactions between the on-template and off-template species. **B)** FliG-DNA conjugates were assembled on one of three linear templates (yellow) and a dye-labelled strand capable of competing for the same template binding sites was added at higher and higher concentrations (blue). FliG-DNA conjugates were displaced off the template as the competing strand bound to the binding sites. Additionally a control experiment was performed which omitted FliG. One strand was bound to templates (yellow) while another strand was titrated in (blue). **C)** Analysis of the gels showed that there was no significant difference between the amount of competing DNA needed to displace FliG-DNA (green) or just DNA on its own (red). If anything, it appears that less DNA is needed to remove the FliG-DNA conjugates than the unconjugated DNA strand from the template.

were interacting with one another when bound to adjacent sites on the DNA. To test this, I worked on a new assay which looked to pick out signatures of these on-template interactions again using native polyacrylamide gels. For this I used two orthogonally-labelled species as before but in this case, rather than a dye-labelled FliG molecule being added to the mix of templates and FliG-DNA, I added a dye-labelled DNA strand with the same N10 sequence as that found in the DNA tag of the conjugates (Fig. 3.6B). The thought was that the two species, the FliG-DNA and the DNA strand with the N10 sequence, would compete with each other for the same binding sites on the template and, depending on the relative concentrations of the two, would eventually establish an equilibrium with a particular number of FliG-occupied and DNA-occupied sites.

In my experiment, I would allow the FliG-DNA to first bind to the template and establish any FliG-FliG interactions before titrating in the competing N10-DNA strand. Alongside this I would run a control using DNA strands without FliG attached, again titrating in a competing strand with a fluorophore of a different colour. FliG-FliG interactions would in theory make it more difficult for a competing DNA strand to displace a conjugate from the template as the protein would be held in place by both its hybridized DNA tag and its interactions with neighbouring proteins. This would mean it would take a higher concentration of added N10-DNA to remove the conjugate from the template than in the control experiment with just the DNA strands of two different colours. The number of conjugates or DNA strands bound to a template could be read out from the gel at each point in the N10-DNA titration and hence in theory native polyacrylamide gels could detect this signature of on-template FliG-FliG interaction.

In performing this experiment, I chose to test three different templating architectures so that I if I did see signatures of on-template interactions I could determine if these only occurred with a specific arrangement of FliG proteins. The first was a simple one-site template which would have no on-template interactions and hence would act as another control (T1d4) while the second was the full five-site template which would provide all FliG molecules bound with an opportunity to interact with a neighbour (T5N10). The final template tested, incorporating three FliG molecules

spaced out by 10-nucleotide gaps (T3G2), was included in an effort to determine if FliG molecules spaced at such a distance could in fact interact with one another.

Titrating in larger and larger amounts of the competing N10 strand against FliG-DNA conjugates bound to one of the three chosen templates and monitoring the state of the system on gels clearly showed that displacement was occurring, with fewer and fewer conjugates bound to the templates as the N10-DNA concentration increased (Fig. 3.6B). Examining the results quantitatively however, by comparing the displacement levels observed to those seen in the control experiment incorporating two differently-coloured DNA strands, showed that there was no significant increase in the amount of displacing strand needed to remove FliG-DNA from the template as compared to a DNA strand without the FliG protein attached. If anything, it actually seemed to be easier to displace FliG-DNA than a DNA strand on its own (Fig. 3.6C). Thus, it appeared that, as best as could be measured, the FliG molecules bound to the DNA template were not interacting with one another.

Taking in the results of this section as a whole, initial experiments incorporating His-cleaved and monomeric FliG-DNA conjugates showed that they could be bound to DNA templates with the correct, physiological spacing of binding sites in stoichiometries and arrangements defined entirely by the sequence of the templating strand. This was a major milestone in the progress of this project and proved that if using the appropriate protocols and methods, both the number and to some degree the spacing of the FliG proteins bound to a template could be determined. No evidence of runaway polymerization or gap-filling, hallmarks of the domain-swap hypothesis, could be observed however even with more sensitive experiments which looked to detect the incorporation of free-floating, dye-labelled FliG into the templated complexes. However, though binding of free-floating and templated FliG could not be observed, interactions between neighbouring FliG proteins bound to the template could not be discounted. To test for these interactions, an experiment was devised in which a small DNA strand would be titrated into solution, displacing FliG-DNA conjugates from their binding sites. Results from this experiment showed no signatures

of FliG-FliG interactions on the template as the conjugates were displaced at levels of the competing strand similar to, if not smaller, than the DNA-only control.

Thus, somewhat ironically considering the length of time devoted to removing a FliG-FliG interaction in the form of the His-tag-mediated dimerization in the previous section, it appeared that no FliG-FliG interactions of any sort could be observed using the systems presented in this section. What was needed was either a new system or a new method of detection. Pursuing these two possibilities, I worked on both a new system with which I could template the FliG-DNA conjugates (section 3.4) and on developing a different technique for investigating FliG-FliG binding (section 3.5). The search for an interaction between FliG proteins is a continuous theme from now until the end of this work however, and the attempts made in this chapter constitute only the beginning of a much longer endeavour.

3.4 Organizing FliG on an Origami Tile

The work of the previous section indicated that FliG-FliG interactions could not be observed when using short, linear, five-site DNA strands as the structures upon which FliG-DNA conjugates would assemble and when using native polyacrylamide gels as the means by which the state of the system was analyzed. To investigate if either the structures or the methods of analysis that I was using were causing problems, I decided to move away from gels and short linear templates to a new system incorporating a different DNA template and a new technique for visualizing the assembled complexes.

Prior to my joining the project, a researcher involved in the collaboration named Jon Bath had designed a template for FliG assembly which was based off of a **rectangular origami tile**. Work had been done both by Jon and another researcher named Le Liang to assemble and purify the tiles meaning that by the time I joined, a fairly robust protocol had been established which could guarantee pure, well-formed structures. In addition, Le Liang had imaged the tiles using an AFM, and his images consistently showed uniform rectangles bound to the mica surface. Preliminary

attempts at incorporating the FliG protein into this system had given some promising results, with raised regions appearing on the surface of the tiles upon addition of FliG-DNA constructs which used the NTA strategy of conjugation. No sign of FliG rings or any direct evidence of FliG-FliG binding had been observed on these structures however and as the two researchers who developed the system had moved on to other projects, investigations using this system had stopped for the time being.

For me, renewing the work on origami tiles represented a perfect opportunity to test my permanent cysteine-maleimide conjugates in a new setting. Rather than short, linear templates, the origami tile presented a structure which more closely matched the C-ring in overall size and dimensions and which gave the FliG molecules a circular rather than linear framework on which to build upon. Perhaps the energetic cost to build an extra section of untemplated FliG polymer to close up and circularize the short, linear FliG fragments that had assembled on the five-site templates was just too great and rather than take an extended, polymeric form, the templated FliG molecules had assumed the closed, monomeric, and more-energetically-favourable 'cytoplasmic' conformation (see section 1.4). As this section shows, the design of the origami tile template ensured a much lower energetic cost for circularization and hence would in theory be more conducive to the formation of a FliG ring. In addition, visualizing the results of templating using an AFM rather than polyacrylamide gels would shed light on whether the method of analysis was previously preventing detection of FliG-FliG interactions. Perhaps the salts, electric fields, polymer mesh, or some other factor present in the gel-running process broke apart FliG-FliG bonds. Analysis via AFM would provide a completely orthogonal method of detection and could therefore avoid some of the issues that gels might have introduced.

Thus, I focussed my attention on binding the FliG-DNA conjugates I had used in the previous section to this origami tile and visualizing the structures created using AFM. As I will show, several errors in the strands that Jon Bath and Le Liang had been using needed to be fixed before the appropriate template was generated and puzzling out what in fact was wrong took some time. The

correct structure was eventually made however, allowing for the templating of proteins with nanometre precision.

Issues in the Tile Strands Were Resolved, Generating a Suitable Template for the FliG Ring

Jon Bath had designed the origami tile such that it could accommodate a DNA-templated FliG ring of the dimensions seen *in vivo*. The overall structure was rectangular in shape with side lengths of roughly 100nm by 75nm and in order to prevent bending and twisting of the tile, a series of bracing strands had been incorporated near its edges on both front and back faces (schematic in Fig. 3.7A). The surface in the middle of the rectangle therefore presented a flat, uniform region upon which a ring could be templated and thanks to the "breadboard-like" addressability of the origami's grid of staples, FliG binding sites could be placed there with nanometre precision.

In this central region Jon had picked out six of the staples to extend into binding sites for protein-DNA constructs. Arranged in a hexagon, these binding sites traced out the dimensions of a FliG ring, with roughly 50nm separating opposite points of the shape. Jon had repeated this process three more times to pick out three more sets of six staples which formed hexagons on the surface of the tile with smaller and smaller dimensions, reaching roughly 30nm across with the closest-spaced set. Although in theory many different protein-DNA constructs could be bound to these extended-staple binding sites, the work done to date had bound short, linear DNA templates bearing FliG-DNA conjugates (as described in the previous section) to these sites. The thought was that with roughly five FliG molecules per linear template and one template at each of the six binding sites, around 30 FliG molecules could be bound in close proximity on the tile (Fig. 3.7A). As the linear templates were anchored at one end to the origami, the local concentration of the FliG monomers would be high and the hexagonal geometry would be conducive to ring formation. Thus, the design seemed appropriate for stimulating the formation of a circular FliG polymer.

Working with this design, my first task was to confirm that the FliG-DNA conjugates I had created via the cysteine-maleimide reaction could in fact bind to linear templates attached to the origami

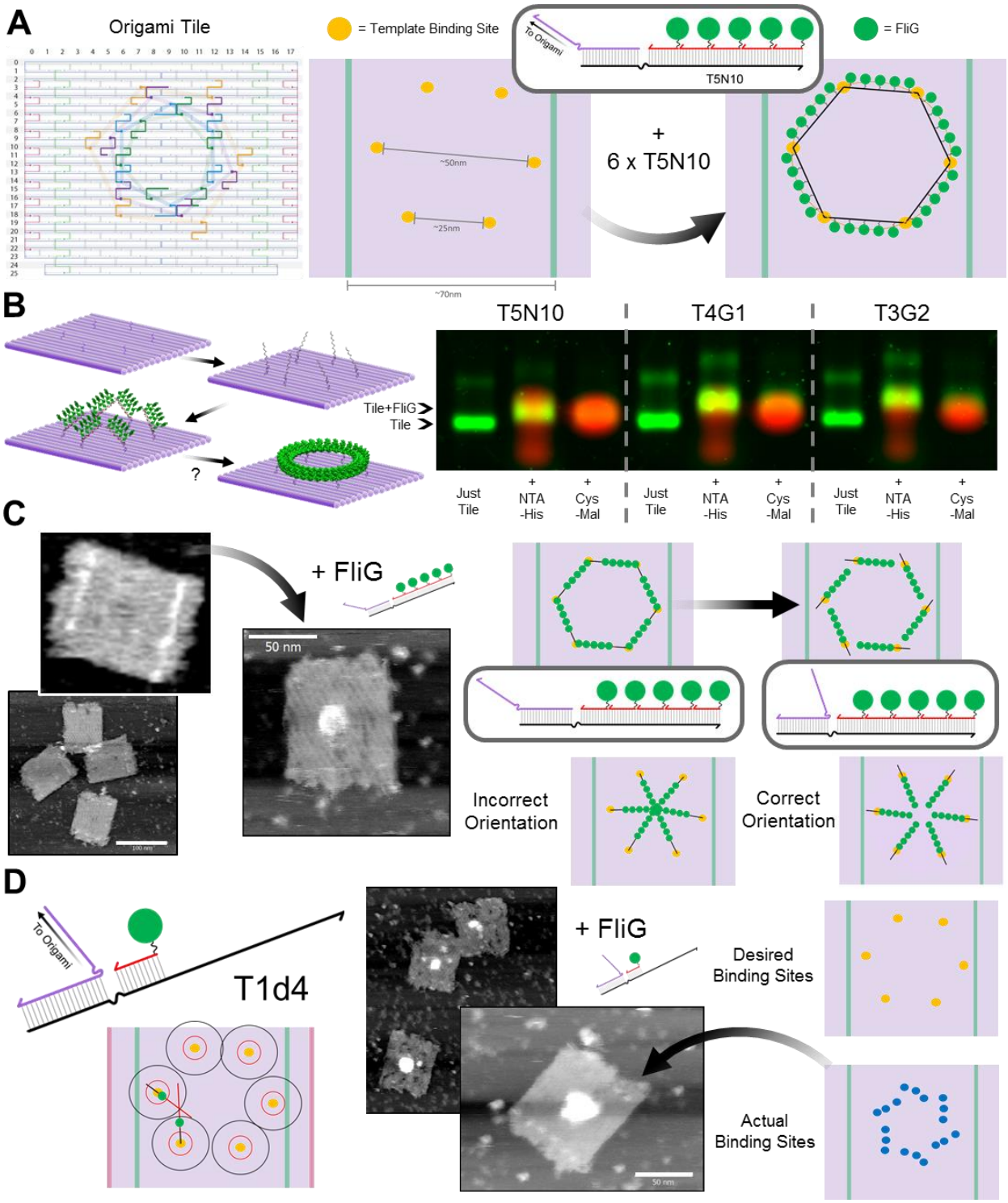


Figure 3.7 | Two major errors had to be corrected with the origami tile before this FliG scaffold could be properly assembled. A) The tile was designed by Jon Bath such that four sets of six hexagons of extended-staple binding sites could be chosen. Attaching five-site templates to these binding sites meant 30 FliG molecules could be bound to the tile. **B)** Placing the proteins in close proximity was hoped to stimulate the formation of a FliG-ring. Agarose gels showed that FliG-DNA conjugates bound to the tiles. **C)** AFM imaging showed well-formed tiles but large, central lumps appeared upon addition of FliG. An error was found in the orientation of tile-binding domains of the linear templates which would allow FliG molecules to meet in the centre of the tile. The lumps remained even after this was fixed. **D)** A single-site template was used in hopes of observing a hexagon of single-protein spots on the surface of the tile. Central lumps were again seen but a review of the DNA found some strands were mislabelled. Rather than the outer six binding sites (yellow), the three inner sets of binding sites (blue) had been incorporated in the template. This explained the central lumps.

tile. To do so I ran a native agarose gel with lanes containing the origami tile bound to one of three different linear templates containing either three, four, or five FliG binding sites (Fig. 3.7B). To these different tile-template complexes I added excess FliG-DNA conjugates which had been created either by the old NTA-DNA strategy or the newer cysteine-maleimide one. The gel showed a definitive upward shift of the origami band upon addition of both varieties of the conjugate. In addition, the dye-label of the conjugates showed that protein and DNA signal overlapped, a strong indication of binding. Very little difference could be observed between the linear templates with three, four, or five binding sites that had been bound to the tile however, indicating that, unlike with smaller structures on polyacrylamide gels, most likely the resolution of the agarose gels was not high enough to give an accurate read-out of the number of FliG proteins bound to the origami.

With the binding of the conjugates to the tiles confirmed, I then set out to visualize the structures using the AFM. For my samples, I chose to make tiles with the most spaced out binding sites to which I would bind fully-filled five-site templates (T5N10). My logic was that creating the largest possible structure would give the best chance at resolving details with the AFM tip such as the hole inside a ring of FliG molecules. Using the annealing protocol and gel-purification procedure developed previously, I was able to produce well-formed tiles and image these using the AFM (Fig. 3.7C). Resolution was such that I could see the brace strands bound at the edges to keep the tile flat as well as the seam of the tile and occasionally the lattice of DNA strands.

Addition of the FliG-DNA conjugates to the tiles however produced some puzzling results. Rather than visualizing a ring or lines of FliG bound to the linear templates on the tiles, large, raised lumps appeared right at the centre of the origami when the conjugates were added (Fig. 3.7C). This was unexpected as with the geometry of binding sites which I had used, even the most distal FliG molecules bound to the T5N10 template should not be able to reach the centre. Looking back at the sequences of the strands I was using, I realized that an error had been made in the domain which connected the linear templates to the extended staples of the origamis. Jon and Le had been using a strand in which the sequence of that domain was reversed to what they thought it was, meaning that there was actually a long double-stranded region of DNA between the binding

site on the tile and the most proximal FliG molecule (Fig. 3.7C). With the extra reach that this architecture provided, the most distal FliG molecules could in fact meet in the centre of the tile and this could explain the large, central lumps that I was observing.

After ordering and receiving new strands with the correct orientation of the template-to-staple linker, I imaged the structures once again only to observe the same large lumps in the centre of the tiles that I had seen previously. Thus, though I had fixed an error, it appeared that this was not the source of the problem. To resolve this, I decided to concentrate on binding single FliG molecules to the six binding sites on the tile rather than look for an entire ring. If I could see six spots corresponding to six FliG molecules at each of these sites then I would know that the template-to-tile and conjugate-to-template linkages were working as they were supposed to and after this was established, I could gradually build larger and more ring-like structures.

For my next experiments then, I attached to the tile templates with a single FliG binding site (T1d4) (Fig. 3.7D). This structure in theory could only bind six FliG molecules, one at each point of the hexagon of extended staples. Though there were some differences observed between the tiles with five-site versus one-site templates, imaging these structures again produced raised lumps at the tile's centre. The lumps were not as large as seen previously and smaller raised spots were also often seen on the tiles (Fig. 3.7D), but nothing close to a hexagon of points was observed. Thus, it appeared that either something was happening with the FliG molecules or that something was still wrong with the origami tile.

I again checked through the sequences of the origami staples and linear templates but could not find any errors as I had previously. Checking through the locations of the staple strands in the 96-well plates they had been delivered in however showed that the identity of the wells had not been correctly documented and that the extended staples which formed the binding sites for the linear templates were in fact in different wells than Jon and Le had noted. For each of these four sets of staples forming four different hexagons of binding sites with different dimensions, there were extended and non-extended versions of the strands present. Thus, if one diameter of hexagon

was chosen, the user would add extended staples corresponding to that particular hexagon and non-extended staples to fill in the sites of the other three, unused hexagons.

What appeared to have happened was that the labels for the extended and non-extended staples had been swapped, meaning that rather than extended staple strands for a specific diameter of hexagon, the inert, non-extended strands were being added. Conversely, the other three hexagons of binding sites which would have been filled with non-extended staples were actually being occupied by extended staples forming active binding sites for the templates. This meant that when I had chosen the outer hexagon of binding sites, the swapping of the labels meant that I was actually adding the three sets of extended staples for the three inner hexagons. This in turn meant that in reality 18 templates could bind to those three sets of six sites all close together near the centre of the tile (Fig. 3.7D). Almost certainly this was the cause of the large raised lumps seen at the centre of the structures during AFM imaging. FliG conjugates on the linear templates (up to 90 conjugates in the case of the T5N10 templates) squeezed together on closely packed binding sites were forming large, densely packed aggregations of proteins which the AFM was visualizing.

Thus, if I wanted to have a hope of observing single FliG molecules arranged in a hexagon and from there work up towards a ring of the proteins, I would need to first add the correct extended staples to the structure. My hope was that simply fixing this error and assembling the tiles as they were designed would produce origami platforms upon which I could see single FliG molecules interacting with one another as I had not seen previously on gels. My next steps then were simply to repeat what I had done previously with the issues involving the strands and sequences now resolved.

Streptavidin but Not FliG Was Visualized Bound at the Correct Sites on the Origami Tile

With the two errors in the origami tile now fixed, I assembled the structure as I had hoped to do previously, with extended staples corresponding to the outer hexagon of binding sites. Both on agarose gels and on the AFM, these corrected tiles looked identical to the previous incorrect

versions, appearing well-folded and the correct shape. I therefore moved on to incorporate FliG bound to single-site templates (T1d4) as I had done before, hoping to visualize a hexagon of spots corresponding to individual FliG molecules. When AFM imaging the structures with these templates added however, I saw no evidence of bound FliG molecules, the tiles appearing exactly as those without the templates. Thinking that perhaps there was just too little FliG to see, I instead incorporated five-site templates (T5N10), hoping that with five times more FliG I would have a better chance of seeing proteins bound to the origami's surface. Again though I saw nothing bound to the tiles. Thinking perhaps the proteins were being lost when I passed the structures through size-exclusion chromatography columns in order to remove excess, unbound FliG, I omitted this purification step and again visualized the tiles. Though the excess FliG could clearly be seen binding to the mica surface, none was observed binding to any region of the tile (Fig. 3.8A), implying that purification was probably not the underlying issue.

To try to pinpoint why I could not see conjugates binding to the tile, I looked for ways in which I could simplify the system. The most obvious alteration I could pick out was to take the linear templates out of the mix through a redesign of the binding sites such that the FliG-DNA conjugates could bind to the extended staples directly. This would remove one breakable DNA-DNA connection between the tile and the protein from the system as well as eliminating any effects that unused portions of the DNA templates might have on FliG binding. I therefore designed new extended staples with a single 10-nucleotide domain which would bind directly to the DNA tag of the FliG-DNA conjugate. Once these strands arrived, polyacrylamide gels confirmed that these new staples did in fact bind to the conjugates, but AFM imaging again demonstrated a distinct lack of FliG proteins visible on the origami tiles (Fig. 3.8A).

Evaluating this simplified system led me to the conclusion that the issue stemmed from one of two factors. The first possibility was that the 10-nucleotide domain linking the FliG-DNA conjugate to the tile was too weak and, though visible on gels, that it was broken by the passing AFM tip. The second possibility was that the FliG protein itself strongly interacted with the mica (Fig. 3.8A) and that it preferentially bound to the imaging surface over the binding sites present on the origami

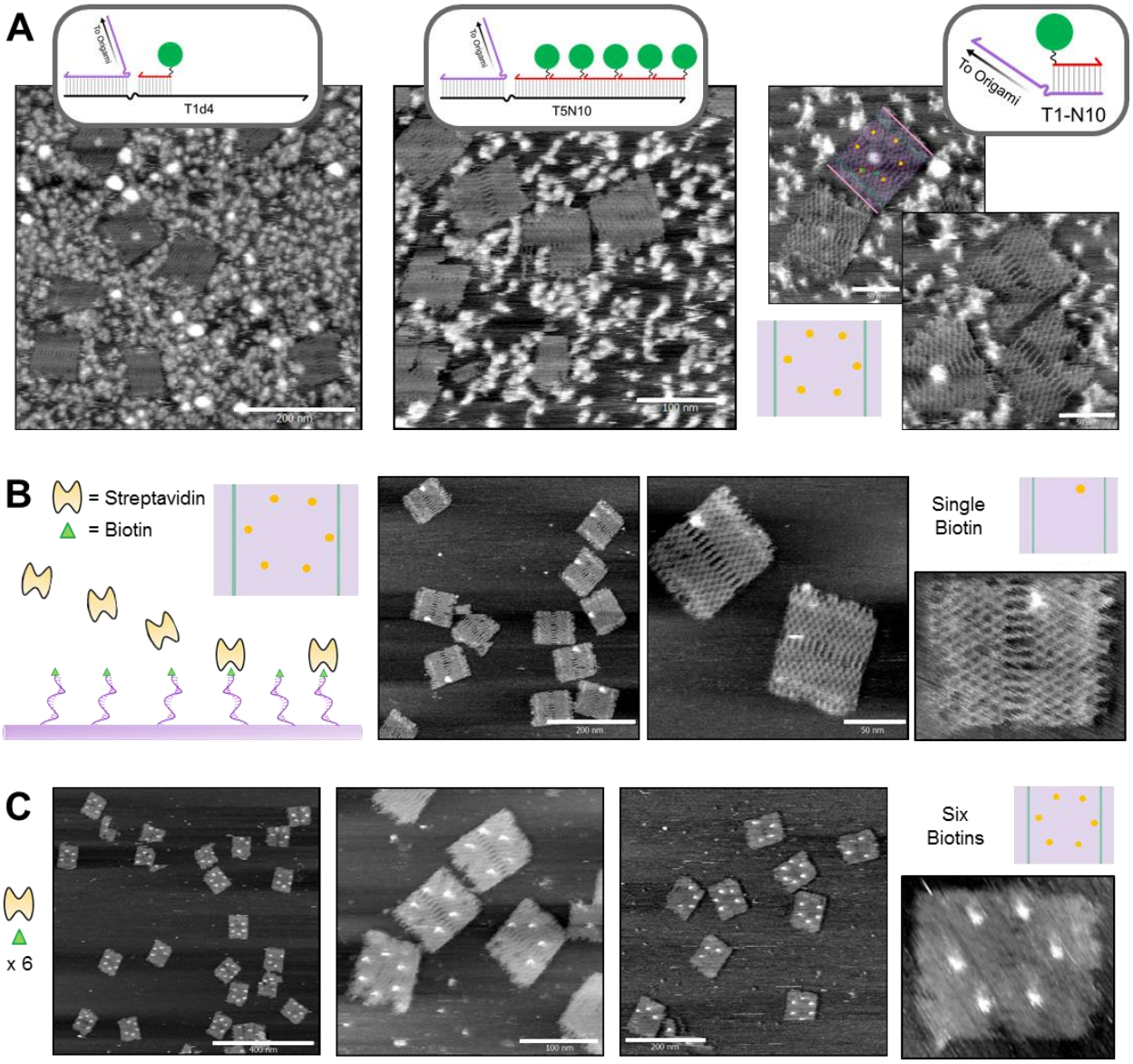


Figure 3.8 | Streptavidin but not FliG proteins could be seen binding to the appropriate sites on an origami tile using the AFM. A) Three different ways of connecting FliG to the template were tried but with all three methods, even with excess FliG left in solution, no binding of the protein could be observed on the tiles. Proteins instead were seen to readily bind to the mica surface. Some spots appearing at the centre of tiles (see rightmost images) were actually proteins bound to the mica which poked up through the seams of the tiles. **B)** To control for any peculiarities in the FliG protein, extended, biotinylated staples were introduced at sites on the tile for streptavidin proteins to bind to. With a single biotinylated staple near the edge of the tile, a single bound streptavidin protein could be seen on the AFM. **C)** Incorporating six biotinylated staples at a hexagon of sites approximately 50nm across produced AFM images showing six streptavidin molecules binding at the locations prescribed by the biotinylated staples.

tile. To test these possibilities, I redesigned the system once again in order to incorporate a different protein with a stronger linkage to the template.

As the protein most commonly used in AFM to demonstrate the nanoscale precision possible with origami structures, I chose **streptavidin** to control for FliG's peculiarities. To attach this protein to the tile I would use its well-characterized binding to biotin, an incredibly strong interaction with dissociation constants measured on the order of $K_d = 10^{-14}$ mol/L [168]. I therefore redesigned the extended staples to include a biotin molecule rather than the 10-nucleotide FliG-DNA binding sequence at their projecting 5' ends (Fig. 3.8B) meaning that I could in theory place six streptavidin molecules in a hexagon on top of the origami tile.

In my first trial with these biotin-presenting tiles, I chose to incorporate an extended biotinylated staple at a single binding site close to the edge of the tile. AFM images of these structures clearly showed a single-molecule-sized spot exactly at the location of that binding site once the streptavidin protein was added (Fig. 3.8B). This indicated that binding of the protein had occurred at the location prescribed by the structure of the tile and that the linkage was stable throughout AFM imaging. To confirm this, I then incorporated biotinylated staples at all six binding sites of the largest hexagon, the same locations that I had previously attempted to attach FliG to. This again showed stable and accurate binding of the streptavidin at the binding sites dictated by the chosen staples, with a hexagon of the molecules visible atop the tile (Fig. 3.8C).

The results of these experiments implied that the sites of the origami tile were in fact addressable and in the correct locations. Thus my previous issues were not due to accessibility of the binding sites but were more likely related to the strength of the FliG-to-tile connection or the propensity for FliG to bind to the mica surface. To try to distinguish between these two possibilities, I attempted to conjugate a biotin molecule to a FliG molecule via the reaction of biotin-maleimide with a single cysteine engineered into the FliG protein. Due to the multivalent nature of streptavidin, my thought was that I could bind these FliG-biotin constructs on top of streptavidin proteins which were already bound to the tiles at the appropriate sites, thereby linking FliG to the origami through a connection that I knew would not break. The conjugation of biotin to FliG was not successful however, as all

attempts resulted in denatured and aggregated FliG stuck in the wells of gels. Thus my thoughts turned to other methods for obtaining a stronger link between FliG and the tile such as extending the DNA binding sequence or conjugating cysteine-FliG to extended staples bearing a maleimide moiety.

Around this time however, while thinking through ways of attaching FliG to the underlying DNA template, I came across several recent publications which changed the direction of my focus. They made me consider an alternative, and perhaps better method by which FliG could be linked to DNA nanostructures and started me off on the topics that constitute Chapter 4 of this work. Thus experiments with the origami tile were put on hold and, as other better systems came to be developed, in the end the tiles were not returned to. As the system of Chapter 5 makes clear, it appears that the connection between the tile and the conjugates was most likely not the issue, leading me to think that problems might have stemmed from a combination of FliG's affinity for the mica and the highly-charged and dense nature of the origami. It could be possible that, due to the stickiness of the FliG protein, the tiles observed on the AFM were all stuck to the mica surface FliG-side-down and that the tightly-packed strands of the origami did not allow any of the proteins on the underside to peek through for the AFM tip to interact with (Fig. 3.8A). This is conjecture however and though time could have been spent with salt and surfactants to try to temper FliG's affinity for the surface, with the improved systems to follow, AFM imaging was able to successfully show FliG bound to DNA nanostructures, making this a mute point.

To reiterate the main points of this section then, I began experiments with the origami tile and AFM imaging after I had failed to detect FliG-FliG interactions using short linear templates and native polyacrylamide gels. Initial experiments gave confusing results until I worked out that there were major errors in both the sequences and labels of some of the key DNA strands of the structure. Once these errors were fixed, I tried to image FliG conjugates binding to the tile in a hexagon of points but only obtained images of apparently blank tiles with no protein attached. Various trials and redesigns of the system still produced images of unadorned origamis so as a control I attempted to visualize streptavidin binding to the tiles via interaction with biotinylated strands. This

was successful and I was able to clearly visualize both a single point and a hexagon of streptavidin molecules bound to the structures. Attempts at binding FliG to the tiles via the biotin-streptavidin interaction were unsuccessful and, as several new discoveries were made in publications around this time, focus shifted away from the tiles to the topics of Chapter 4.

Before those topics are discussed however, one key series of experiments remains to be touched on in this chapter. Just as this section searched for FliG-FliG interactions with a new *in vitro* system and a new technique to detect them, the next section implements new FliG constructs, allowing for new methods to be employed in an effort to detect ARM-ARM interactions and evidence of domain-swap polymerization.

3.5 Disulfide Cross-Linking Experiments with Double-Cysteine FliG

As the previous sections have shown, work with native polyacrylamide gels and AFM imaging produced no clear evidence of FliG-FliG interactions. The domain-swap model for FliG polymerization depends crucially on these interactions and thus, without a reliable way of determining both if and how FliG proteins were contacting one another, I could not truly test the model. Another way of examining the proteins was available however and had been used in the past as one of the main pieces of evidence in support of domain-swap oligomerization.

As a follow up to their work in 2010, which included the first full-length structure of FliG and a model for its assembly into a ring[60], the lab of Lawrence Lee had proposed the domain-swap model of FliG polymerization in a 2016 publication[101]. This publication also included a key set of experiments which provided the most clear-cut pieces of evidence for ARM-ARM polymerization. Protein constructs which contained two cysteines, one near each of the ARM motifs, had been developed by the lab and using SDS-PAGE to look at the behaviour of these constructs *in vivo*, they were able to show disulfide bonds forming between individual FliG proteins, observing ladder-like bands of oligomers on their gels. Comparing the constructs with cysteines near the ARM motifs to control constructs with cysteines elsewhere on the protein showed that the proteins with

ARM-proximal cysteines produced the highest-order oligomers and this was given as evidence that the oligomerizing interactions were mediated by the ARM motifs. In addition, after dividing the bacteria into various fractions, FliG found in the cytoplasmic pool was shown to be monomeric while the FliG proteins isolated with the membranes were shown to produce a ladder of oligomeric bands. This was given as support for the theory that FliG is monomeric until binding to the membrane-bound FliF ring, at which point it oligomerizes into a ring. The work in this experiment was later criticized as it did not control for the differences in FliG concentration between these two fractions[104] but the clear visualization of the oligomeric state of FliG on their gels made me think that it would be a promising technique to try with my own *in vitro* system.

I therefore asked for the same **double-cysteine FliG construct** from the Lee lab that had produced the clearest results in their *in vivo* work. The cysteines of this construct were placed at the edges rather than the centre of the ARM motifs' hydrophobic patches such that any interactions between the patches were not disturbed. The cysteines were close enough to the patches however that if an interaction did occur, a disulfide bond could form across the interface and lock the structure in place (Fig. 3.9A). My thought was to monitor the oligomeric state of this construct *in vitro* using SDS-PAGE and to see if free-floating FliG produced different bands on the gels compared to DNA-template-bound proteins. Should the domain-swap model be correct, one would expect free-floating proteins to be mostly monomeric while template-bound proteins would bind to their neighbours and form long oligomers, seen as ladders on the SDS-PAGE gels. Thus, with this technique I could test both the domain-swap model and whether my *in vitro* system was recapitulating what was thought to be taking place *in vivo*.

As this section shows, I was able to demonstrate oligomerization between FliG molecules using the cysteine-cysteine cross-linking strategy, but obtaining evidence that the interactions seen on the gels were a product of ARM-ARM binding or that they were induced by a DNA-template was more of a challenge. In the end it was difficult to come to any definitive conclusions based on the results observed and cysteine cross-linking was not returned to in any of my later work. That being

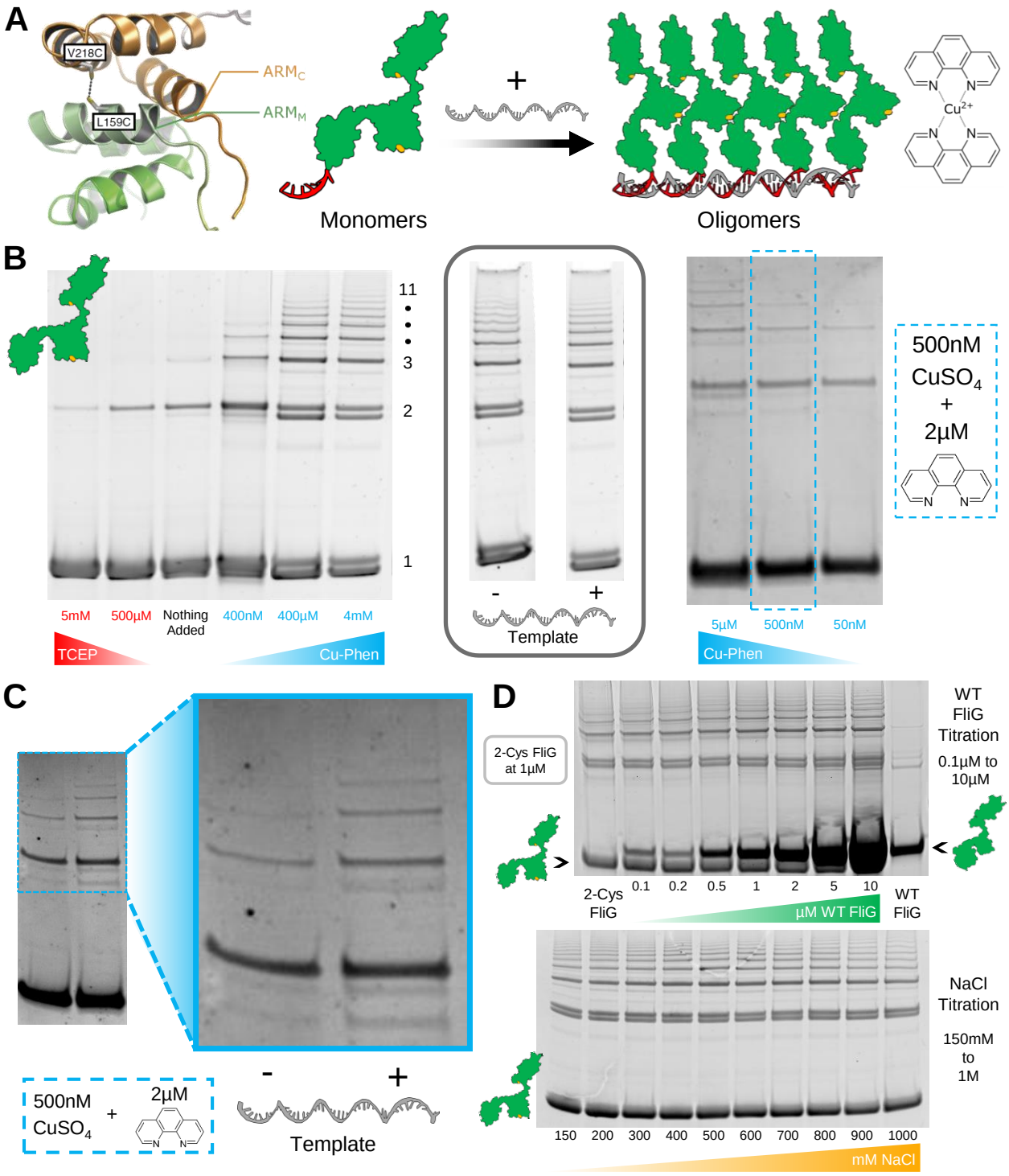


Figure 3.9 | Disulfide cross-linking of double-cysteine FliG did not produce conclusive evidence of FliG oligomerization induced by a DNA template. A) A FliG construct bearing two cysteine mutations, one near each of its ARM motifs (image from (101)), was expressed and purified. The strategy was to add an oxidizing agent like copper-phenanthroline to both free-floating and DNA-templated FliG, hoping to visualize a transition from monomers to oligomers with addition of the template. **B)** Going from reducing to oxidizing conditions produced different weights of oligomers. Initial attempts at cross-linking templated species showed no difference to free-floating FliG. A lower concentration of the oxidizer was found at which oligomers just began to form. **C)** At this 'tipping-point' concentration of oxidizers, a slight increase in the weight of oligomers could be seen upon addition of the DNA template. **D)** Other perturbations to the system produced no observable effect on FliG's oligomeric state. Addition of wild-type FliG was expected to decrease the degree of oligomerization but no decrease was seen. Increasing the ionic strength of the solution is predicted to favour the closed, monomeric form of FliG[101] but adding more and more NaCl produced no change in the oligomeric state.

said, many of the tools developed for this technique proved useful later on and the experiments I present were valuable in giving me a better general understanding of my *in vitro* system.

Cross-Linking Produced No Conclusive Evidence of DNA-Template-Induced Oligomerization

Once the double-cysteine construct was received, I expressed and purified the protein using the standard protocol for FliG (see section 2.2). As mentioned previously, my main aim with this construct was to monitor its oligomeric state using cysteine-cysteine cross-linking and SDS PAGE gels, with the hope that I could see a difference in the state of the proteins upon the introduction of a DNA template (Fig. 3.9A). One downside however was that with this double-cysteine construct, I could not use maleimide-DNA to create a permanent FliG-DNA conjugate as I had previously done (see section 3.1). The maleimide-DNA would conjugate to the cysteines needed for cross-linking and hence would disrupt the experiments. This meant that I instead went back to the conjugation approach using NTA-DNA and the Histidine-tag of FliG. This was not ideal as I had implicated the His-tag in FliG dimerization (section 3.2) and since this realization had generally cleaved off the His-tags before using the proteins. I hoped the overall difference in the oligomeric state between template-bound and unbound proteins would still be present even with dimers in solution but knew that any results obtained would have to be interpreted with the dimers in mind.

My first task then was to translate the cross-linking protocols that the Lee lab had used *in vivo* to my *in vitro* system. The environment in a cell is much different than in a buffered solution in a tube so I needed to calibrate the concentrations of oxidizing agents and reaction times such that I could obtain similar degrees of cross-linking at a given FliG concentration. FliG is thought to be around 500nM in the cytoplasm of a bacterium[101] so this was the concentration that I worked at for most of my experiments, with my first being a simple test of cross-linking in different reducing or oxidizing conditions. Going from reducing conditions with high concentrations of TCEP through the natural state in the buffer to more and more oxidizing conditions with the addition of iodine or copper-phenanthroline, I was able to see a distinct change in the oligomeric state of the protein (Fig. 3.9B). Whereas with a large amount of TCEP the protein was mainly monomeric, at the

highest concentrations of the oxidizers, oligomers ranging up to 11mers could be seen. In addition, double bands could be observed for both monomers and dimers in highly oxidizing solutions which I interpreted as species with either zero or one (in the monomers) or one or two (in the dimers) cysteine-cysteine cross-links.

Thus, I could see FliG oligomers on a gel just as had been shown previously by the Lee lab. What I could not see however was a state where the population of FliG was entirely oligomeric. My theory was that this was due to the fact that I could not obtain as high concentrations of the protein as were present in the FliG rings found within the membrane fraction of the *in vivo* experiments. I hoped however that DNA-templates could do the job of concentrator and hence tried adding short, 5-site templates (T5N10) into the mix. I observed no change with or without the templates however, with the number and intensity of the bands the same in both cases (Fig. 3.9B). I thought this was perhaps due to the highly oxidizing conditions that I had done the experiment in. With so many tightly packed oligomeric bands near the top of the gel, perhaps I simply was not able to see changes with or without the template due to a lack of resolution. Therefore, I did a more fine-grained screen of the concentration of the oxidizers to find the point at which oligomeric bands just became visible on gels (Fig. 3.9B).

With these concentrations and conditions in hand, I then repeated the experiment of adding in the template, but this time did so at an oxidation state which was at the ‘tipping point’ between monomers and oligomers. This produced a visible difference between non-templated and templated lanes (Fig. 3.9C). Upon addition of the template, a greater number of oligomeric bands could be seen (pentamers) than without the template (trimers), and although the images were grainy and the difference was slight, the result was repeatable. In both the cases with and without the template however, monomers were the predominant species and unlike in the *in vivo* case, I could not demonstrate a transition from a completely monomeric to a completely oligomeric population.

Moving forward, I decided to search for conditions in which I could see a clear and convincing transition between monomers and oligomers on the gels, hoping to replicate the results found *in*

vivo. There were a great many dials which I could turn and many conditions to screen however so I tried to focus on the key factors that might play a role inside the bacterium. I tested the concentration of FliG, the ionic strength of the solution, its viscosity, its pH, different stoichiometries of FliG to template, and other factors but none seemed to make any impact on the oligomeric state of FliG visible on the gels (Fig. 3.9D). In addition, I tried titrating in various concentrations of wild type FliG, thinking that interactions between the wild type and the double-cysteine mutants would create oligomers containing both constructs meaning the overall size of assemblies seen on denaturing gels would decrease with more wild type FliG added. No change was observed with increasing wild type FliG concentrations however (Fig. 3.9D). The only factor which seemed to make any difference was the concentration of oxidizer in solution, with more oxidation producing higher-order bands.

Thus, it appeared that the faint and grainy images I had obtained at the 'tipping-point' concentrations of oxidizers were the best that I could produce with my *in vitro* system. And even these I had to interpret with caution. As I lacked control constructs with cysteines placed away from the ARM motifs, I could not definitively say that the interactions which created the oligomers visible on my gels were those predicted by the domain-swap model. It could have been that my FliG proteins were acting as an unstructured polypeptide chain would, simply bumping into neighbouring proteins and forming cross-links at random from non-specific interactions. Therefore, due to a lack of non-ARM-cysteine controls, concerns about His-tag-mediated dimerization, and with no immediate ideas for further manipulations which could more clearly define the effects of the template, I decided to move on to other experiments and did not return to the disulfide cross-linking approach or the double-cysteine mutants in subsequent work.

Taking a step back, my main concerns at this point revolved around the fact that I lacked concrete evidence of the FliG-FliG interactions predicted by the domain-swap model. After modifying a DNA strand with a maleimide group and conjugating this to single-cysteine FliG (3.1), I was able to bind the conjugates to short, linear DNA templates at the appropriate, physiological spacing, a major advance over the NTA-DNA approach. Even with these conjugates bound well to the DNA

template and the issue of His-tag-mediated dimerization resolved (3.2), I was not able to detect FliG-FliG interactions on native polyacrylamide gels. Tests with FliG labelled in two colours as well as with a DNA strand competing for the same binding sites as the conjugates on templates produced no signs of gap-filling, runaway polymerization, or any other signature of FliG-FliG binding either on or off the DNA (3.3). Moving to a different templating system based on an origami tile and a new visualization technique in AFM, I still was unsuccessful in obtaining evidence of FliG-FliG interactions due in part to technical difficulties as, although it could be done with streptavidin, it was difficult to image FliG conjugates bound to the surface of the tile (3.4). In this section (3.5), I made another attempt with a different set of constructs and a new experimental approach to look for evidence of intermolecular FliG binding. Cross-linking double-cysteine mutants in oxidizing conditions produced evidence of oligomers on denaturing gels but it could not be determined if these were the product of ARM-ARM interactions or non-specific binding. Though templating the proteins seemed to produce a slight change in the extent of oligomerization, it was nothing near the difference seen *in vivo*, and the system appeared insensitive to any other perturbations or additives which in theory should have had profound effects.

Thus, without any evidence for FliG-FliG interactions or domain-swap polymerization, I had developed two main theories as to the underlying issue. Either the problem stemmed from the FliG protein itself or from the environment in which it was in. In the first case, my main question was whether FliG was in its native, folded state or in some way denatured. Although the protein appeared monomeric and did not aggregate on native polyacrylamide gels, there could be subtle changes to the configuration of the polypeptide chain as a result of the purification process, the introduced amino acids of the construct, or the conjugation of a DNA strand to the protein which interfered with its functionality. Without going into complicated structural techniques however, there was no simple assay to determine if the protein was in fact in its well-folded, native state. In my *in vitro* system the only reassurance of the protein's stability short of ring-formation would be to note it interacting with other FliG molecules (something I had yet to observe) or one of its binding partners from the C-ring (something I had yet to attempt). Thus it was very possible that the FliG I was working with was not in fact well-folded. As a second option, it was possible that something

was wrong with the environment the FliG protein was in. The *in vitro* system I was using was very different from the cytoplasm of a bacterium and perhaps a key component was absent from the mix which was necessary to aid or to trigger FliG polymerization. It was conceivable that along with a template, FliG needed a specific salt, small molecule, or protein binding partner which I had not provided. The list of possible additives was quite long however and if this was in fact the issue, it was difficult to determine where to begin.

Luckily, around the time that I finished the work which I have just described in this chapter, several key publications came to my attention which gave me an idea of how to proceed. One key addition to the system, which was claimed to both affect the state of FliG and its propensity for oligomerization became my key focus for the work that was to follow and the incorporation of this component into the mix of FliG and DNA forms the backbone of the work presented in Chapter 4.

CHAPTER FOUR

A NEW LAYER

Incorporation of the $FliF_C$ Peptide into the In Vitro System

At this point, I had worked out a reliable protocol for FliG-DNA conjugation and seen the conjugates binding to their prescribed sites on a DNA template. I had also solved the problem of FliG dimerization through cleavage of the His-tag, generating monomeric proteins which behaved, hopefully, like a wild-type protein would *in vivo*. But even with the conjugation and dimerization problems seemingly sorted out, all of my attempts at visualizing FliG-FliG interactions had been unsuccessful. Whether I used short linear DNA templates and native polyacrylamide gels, origami tiles and AFM imaging, or disulfide cross-linking and denaturing gels, I was unable to verify any aspect of the domain-swap model for FliG polymerization and had yet to see any evidence of binding between the ARM motifs. My main hypotheses to explain this lack of polymerization were that either the FliG protein itself was in some way denatured or that a key component was missing from the *in vitro* system which was required for the polymerization to take place.

I thus turned to the literature to search for both a reliable activity assay for FliG and for ideas as to what might be needed alongside the FliG and DNA in my current system. Very little *in vitro* work has been done with FliG in the past and, as I predicted, it appeared the only indication of a functional FliG molecule was its ability to bind to its partners in the C-ring, namely neighbouring FliG proteins, FliM, and the C-terminal domain of FliF (see section 1.4). I began to look more closely at these binding partners and came across some newly published work concerning the interaction between FliF and FliG. The authors showed that binding of the C-terminal domain of FliF ($FliF_C$) to the N-terminal domain of FliG ($FliG_N$) impacted the final, folded conformation of the

FliG molecule[88] and others claimed that this conformational change could possibly increase its propensity to oligomerize with other FliG molecules[91].

It appeared then that I had a prime candidate for inclusion into my system and I was keen to investigate the impacts that this domain of FliF might have on FliG behaviour. Binding of $FliF_C$ to my purified FliG would be a strong indication that the FliG protein was correctly folded and there was the intriguing possibility that this binding might trigger FliG oligomerization. Thus, I was convinced that inclusion of $FliF_C$ was the best way forward and I set out to determine how I might incorporate it into my *in vitro* system.

This chapter centres around $FliF_C$ and my work to study its effects on FliG behaviour *in vitro.* Beginning with the discussion of the design and preparation of a $FliF_C$ construct for *in vitro* use, it then proceeds with the study of how this construct interacts with FliG before moving on to the incorporation of the $FliF_C$ domain with DNA templates. The addition of $FliF_C$ to the mix fits into the general trend of bringing the system closer to that seen with the C-ring *in vivo* and a final section of the chapter deals with attempts at the incorporation of FliM and FliN with that aim in mind. The work presented makes several small but significant advances to the field which I will highlight and the strategy I have employed may be a useful one for other *in vitro* investigations of protein-protein interactions.

4.1 Design and Preparation of the $FliF_C$ Peptide

As discussed in section 1.4, the binding between $FliG_N$ and $FliF_C$ is the strongest interaction found in the cytoplasmic ring, with a K_d measured at around 40nM[90]. This link anchors the C-ring to the membrane and the rest of the BFM, thereby ensuring that the rotation of FliG by the stator units is translated into the rotation of the flagellum. The domain of FliF that is responsible for this interaction however is quite small. Only around 40 of FliF's amino acids extend into the cytoplasm as opposed to over 500 which are membrane-embedded, and it is these cytoplasmic residues which are responsible for FliG binding[169].

Work to characterize the binding between FliF and FliG historically ran into the issue that the FliF protein was very difficult to purify, meaning that no structures of the interface between the two proteins had been published. In the last decade however, several researchers managed to avoid the purification issue by chemically synthesizing a peptide corresponding to the 40 cytoplasmic amino acids of FliF, thereby obtaining a $FliF_C$ domain which could be used in structural studies. This strategy led to the successful crystallization of $FliF_C$-$FliG_N$ complexes from both *T. maritima*[88] and *H. pylori*[91] which showed a large hydrophobic interface between the binding partners and a distinct conformational change in FliG after $FliF_C$ binding. It was claimed that this conformational change might induce interactions between neighbouring $FliG_N$ domains and hence contributed to the formation of the FliG ring. A mechanism was later proposed outlining how the helices of $FliG_N$ might rearrange upon contact with FliF to produce a binding site for a neighbouring FliG molecule[91] though no work had been done in practice to either confirm or reject this hypothesis.

Thus, in the $FliF_C$ peptide I had a promising candidate for inclusion into my system and, through chemical synthesis, I had a strategy for how I might go about obtaining it. This section describes the design of two versions of the $FliF_C$ peptide which could be incorporated into my experiments with FliG and DNA as well as my preliminary attempts at this incorporation. To a great extent this was a step into the unknown as these peptides had not been synthesized or worked with previously and as such it took some time before the proper strategy for their preparation was found.

Both Wild-Type and Single-Cysteine Versions of the $FliF_C$ Peptide Were Designed

Though it seemed a straightforward approach, there was a major difference between my system and the systems used in the structural studies. Whereas I was using proteins from *E. coli* and *Salmonella*, the structural studies used constructs from thermophilic bacteria. The proteins from the two classes of bacteria have the same layout of domains and binding partners but their amino acid composition and, by extension, the amino acids of their binding sites are very different. Thus, the strength with which $FliG_N$ binds to $FliF_C$ in a thermophilic bacterium might be very different from

that in an enteric one and it was not obvious that I would actually be able to observe an interaction as the structural studies had.

Designing a peptide that best matched what had been previously used was my best hope for observing interactions so I searched to find the region of *E. coli* and *Salmonella* FliF that was functionally equivalent to the 38-amino-acid-long peptide used in the studies with *T. maritima*. Aligning the sequences of FliF from the three proteins showed that an identical stretch of 39 amino acids at the protein's C-terminus in both *E. coli* and *Salmonella* corresponded to the region in *T. maritima* that had been shown previously to bind to FliG. I decided that this 39-amino-acid domain would be the best candidate for chemical synthesis and termed it the **$FliF_C$ peptide** (Fig. 4.1A).

Analyzing various parameters of the peptide sequence showed that while the molecular weight of the enteric and thermophilic versions was roughly the same, the net charge was quite different. While the thermophilic version had an isoelectric point of 3.93 and would therefore be quite negatively charged in my slightly alkaline buffer, the enteric version had a pI of 7.53 making it neutral in those same conditions. Looking at predictions for the secondary structure of the peptides on the other hand showed that the two variants matched, both being comprised of two α-helical regions connected by an unstructured segment (Fig. 4.1A). Predictions from the structural studies indicated that the peptide might only form these α-helices upon binding to FliG though and would take a more unstructured arrangement when free in solution[88].

Thus, though the enteric $FliF_C$ peptide had an almost completely different composition of amino acids from the thermophilic variant, many of its parameters, and most importantly its secondary structure, seemed very similar. Both had a single key tryptophan residue at the same site in the structure which the previous studies had shown to be necessary for FliG binding[90] and though no one had shown that this 39-residue domain of FliF was necessary and sufficient for FliG binding in enteric bacteria, the high degree of similarity lent me confidence that it would be.

I therefore decided to have this 39-residue peptide synthesized as my wild-type version of the $FliF_C$ domain, with the main aim of simply adding this to a solution of FliG or DNA-templated FliG

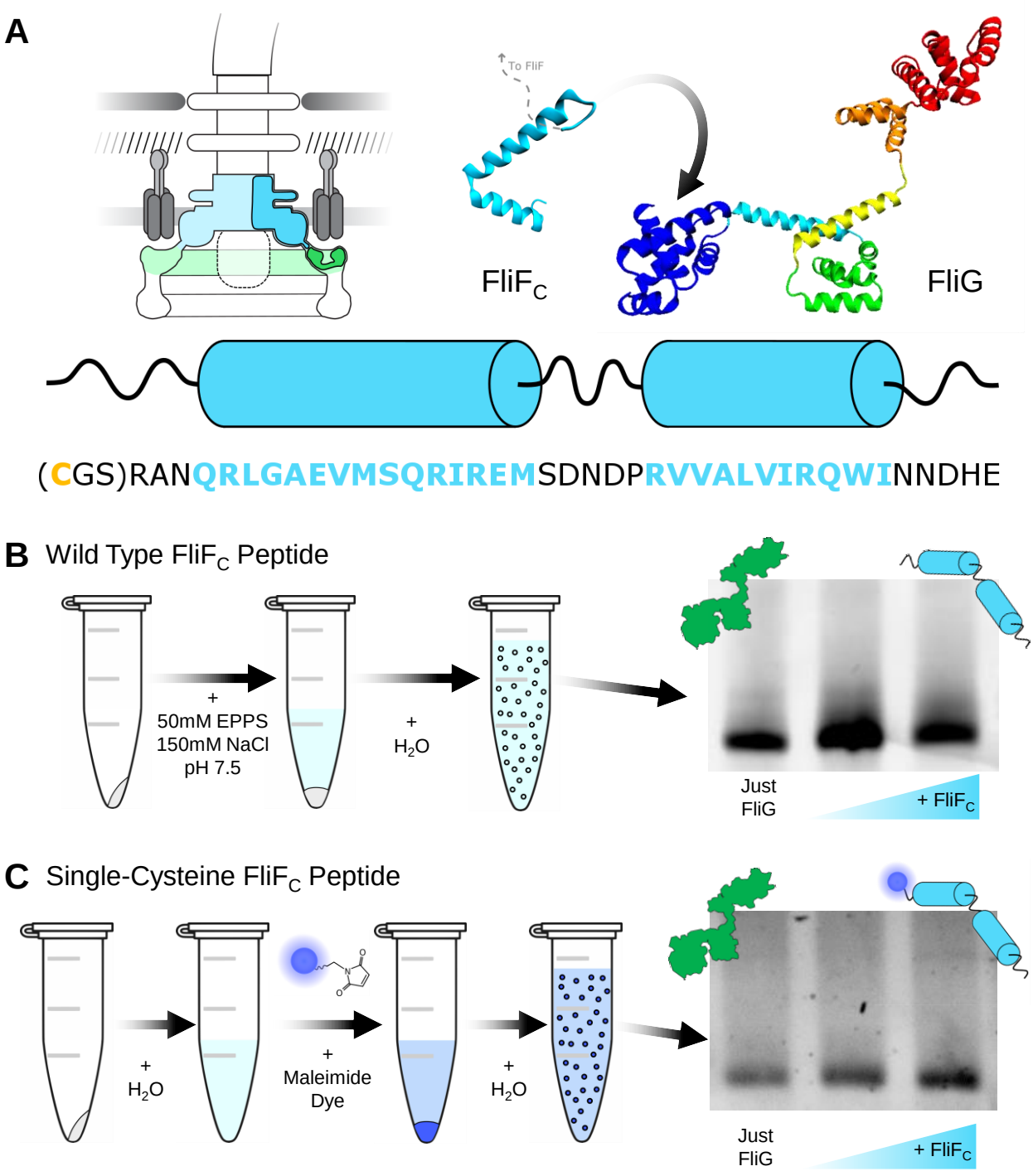


Figure 4.1 | The FliF$_C$ peptide was chemically synthesized but its addition to FliG showed no signs of binding due to issues with solubility. A) Two versions of the cytoplasmic domain of FliF from *E. coli* or *Salmonella* were designed, The first was the 39-amino-acid-long wild-type peptide which was predicted to contain two α-helices (blue barrels) connected by flexible linkers. The second version appended three residues: C, G, and S to the N-terminus for later use in conjugation reactions. **B)** Attempts at dissolving the lyophilized wild-type FliF$_C$ pellet in buffer were unsuccessful. Dilution in water helped to some degree but the solution remained cloudy. Addition of the peptide to FliG in various stoichiometries produced no shifts on a native polyacrylamide gel. **C)** The single-cysteine version of the peptide was successfully dissolved in water. Addition of maleimide-modified dye however produced a bright blue pellet. Further dilution in water gave a cloudy blue solution. Adding this to FliG in various stoichiometries again produced no shifts on native polyacrylamide gels.

and looking for evidence of an interaction between the peptide and FliG or a change in FliG's oligomeric state due to the addition of the peptide. In addition, I designed one variant of the wild-type peptide for use in experiments to follow these preliminary trials. At the N-terminus of this peptide I added three residues: a cysteine to be used for conjugation to other species (fortunately the rest of the sequence was free of cysteines), as wells as two residues (GS) between the cysteine and the rest of the sequence in order to prevent any disturbance to the overall structure that inclusion of a cysteine might cause (Fig. 4.1A).

Both the wild-type and single-cysteine versions of the peptide were ordered and were successfully synthesized and HPLC-purified by a commercial synthesis company. Arriving as lyophilized pellets, my next task was to solubilize the peptides, mix them with FliG, and look for an effect.

Initial Attempts at Observing $FliF_C$-FliG Binding Were Unsuccessful Due to Solubility Issues

Having had the peptides synthesized and purified in the same fashion and by the same company as the structural studies, I again followed their protocol to solubilize the lyophilized pellet of the wild-type $FliF_C$ peptide. In their studies, they had simply dissolved the pellets in their working buffer and had proceeded to use the peptides directly in experiments. Following their lead, I did the same, adding my EPPS buffer to the pellet. Although the pellet did dissolve to some extent, it was very obvious from the cloudy opaqueness of the solution that solubility was an issue with this peptide in my buffer (Fig. 4.1B). Centrifugation showed a large percentage of the total peptide pelleting out of solution and adding various salts, organic solvents, acids, and bases did not seem to have an immediate effect on the solubility.

The one factor which did seem to make a difference however was the overall concentration, as diluting with distilled water led to less pelleting and a slightly clearer solution. Working with this diluted solution, I decided to simply add it to FliG either on or off the DNA templates, hoping to observe a gel shift due to binding or other changes that would indicate an interaction. There was no change however in the height or appearance of either the free-floating or DNA-bound FliG

bands on native gels upon addition of the peptide, even with the peptide in great excess (Fig. 4.1B). This indicated to me that either the peptide did not interact at all with the protein or that binding did not cause an observable shift on the gels. As the peptide was unlabelled, I could not identify where it had run to and thus could not confirm whether or not it co-localized with the FliG band.

It was possible that the peptide was in fact binding to FliG but did not produce a gel shift due to its small size and uncharged nature so I decided to label it in order to track its migration on these gels. To do so, I used the single-cysteine version of the peptide that I had had synthesized, with the goal of attaching a maleimide-conjugated organic dye to its N-terminus. This time, to dissolve the lyophilized pellet of the single-cysteine construct I used distilled water rather than my buffer and obtained a clear, transparent solution even at high concentrations, indicating that the peptide had dissolved (Fig. 4.1C). Taking an aliquot of this solution, I added TCEP and then the maleimide-dye dissolved in DMSO (dimethyl sulfoxide). While the solution remained transparent with addition of the TCEP, addition of the dye immediately turned the solution cloudy and centrifugation showed that a blue pellet had formed (Fig. 4.1C).

The colour of the pellet indicated to me that it contained both the dye and the peptide and implied that the conjugation had been successful but that the solubility of the resulting dye-peptide conjugate was different than that of the peptide on its own. One option was that the hydrophobic nature of the dye caused pelleting while another was that the conjugation had led to a destabilization of the peptide's secondary structure causing it to unfold and aggregate. Either way, there was again an issue with solubility and again all additives that I tried did not decrease the degree of pelleting. Diluting in water showed some improvement in the solubility as it had previously so I decided to mix this diluted solution in with FliG as I had done before.

Again, however, I saw no change in the migration or appearance of either the DNA-bound or unbound FliG bands upon addition of the dye-labelled peptide (Fig. 4.1C). What is more, I could not pick out a dye-peptide band on the gels. The only bands containing the dye signal were those produced by the dye molecule on its own. This indicated to me that most likely the peptide was

either not conjugated to the dye, or not entering the gel due to its forming large, insoluble aggregates. With either explanation though, it appeared that the techniques I was using to look for FliG-$FliF_C$ binding were producing inconclusive results and a different strategy was needed.

I thus began searching for other methods to detect binding that did not require gels and could get around the problems I was having with the peptide. It seemed that the solubility of $FliF_C$ was of great importance and its pelleting was most likely the biggest contributing factor in the failed attempts I had made at observing an interaction. Therefore, I needed to find an assay in which I could use, and somehow monitor, unlabelled peptide. Luckily a wide range of biophysical and biochemical techniques were available to detect and characterize binding between two polypeptides and my next task was to find which I could use to investigate the binding of FliG to $FliF_C$.

4.2 Biophysical Characterization of FliG and $FliF_C$

After the initial unsuccessful attempts at observing binding between FliG and $FliF_C$ on gels, I thought through various explanations for what I had observed. As mentioned in the previous section, the solubility of the $FliF_C$ peptide was a major issue and the lack of binding could simply be due to the peptide being bound in large aggregates and unable to enter the solution. The lack of binding could also be indicative of poorly-folded FliG constructs which did not possess the correct binding pocket for the peptide. A third possibility, which I had tried to test through labelling of the peptide, was that binding was occurring but was not visible on polyacrylamide gels, while a final alternative was that the enteric version of the $FliF_C$ domain simply did not bind to FliG like the thermophilic version did and hence would not be detected no matter which method I tried.

To narrow down these options, I needed to explore different assays and methods which would give me more information about the state and configuration of the FliG constructs I was using and provide a more sensitive read-out of whether the protein was actually binding to the $FliF_C$ peptide. At this point, my only information about FliG was that it ran as a 39kDa monomer would be

expected to run on gels and I had little knowledge concerning its secondary or its tertiary structure. Therefore, I searched for a technique which could provide some finer-grained structural information about the protein but which did not require complex preparatory procedures which placed FliG in environments vastly different from what it experienced in my *in vitro* system, as a technique like X-ray crystallography might.

In the end I settled on circular dichroism (CD) spectroscopy as the first method to experiment with and I begin this section by presenting both the CD spectra of various constructs and the results obtained from experiments conducted using this technique. Following on from this, I present the results from two other complementary assays which I later employed to investigate the binding of FliG to $FliF_C$, with the aim of confirming that the two species did in fact interact. The work that I describe here represents a real turning point in the progress of the overall project, with the results obtained in this section guiding the trajectory of all my work to come. In addition, its main conclusion, though seemingly a small one, is perhaps the most significant piece of information that this project contributes to the field of BFM study and I will highlight its implications as the chapter progresses.

CD Spectroscopy Showed That the Degree of α-Helicity Varies Between FliG Constructs

As just mentioned, I required a technique which would give me some indication of whether or not the FliG constructs I was using were properly folded in the environment in which I had placed them and **circular dichroism (CD) spectroscopy** seemed to be able to provide useful structural information which would allow me to gain some insight into FliG's folded state.

In brief, the circular dichroism of a sample is defined as the difference between the absorption by that sample of left-handed versus right-handed circularly-polarized light[170]. This differential absorption occurs when the sample in question contains chiral groups, groups which are commonly found in many biological macromolecules. Measuring the CD signal of a molecule over a range of wavelengths can build a trace with characteristic peaks and troughs, and the CD

spectrum that is generated can be compared to other spectra from molecules with a known configuration and composition to gain insights into the structure of the sample in question. In the case of proteins, each particular secondary structure of the amino acids gives a distinctive CD trace in the UV range of wavelengths (Fig. 4.2A) and analysis of the overall CD spectrum of a particular sample can give an idea of the percentage of each type of secondary structure present in the protein. In addition, the CD spectrum of a sample can be obtained over a range of different conditions such as temperature or pH and analysis of how the signal changes with these different inputs can determine how and at what point the protein's conformation changes.

Using this technique then, I could in theory gain useful information about the structure of the FliG constructs I was working with. As the full-length crystal structure of at least one species of FliG protein had been solved, the proportions of the various secondary structures that a well-folded protein might be composed of were known and these could be compared to the proportions extracted from the CD spectra observed for the sample in question. If the values for the percentage of α-helices, β-sheets, and random coils obtained from the crystal structure matched those from the experimentally determined spectra, this would be a good indication that the protein being measured was well-folded. Testing various constructs and comparing their spectra would determine if they were all similarly well-folded or if some were denatured to some extent. Additionally, by heating the solution and monitoring the CD signal, I could observe how and when the protein transitioned to a random coil, with a sharp transition indicative of a well-folded protein.

Thus, CD spectroscopy appeared to be a very useful technique to help me answer the questions I had concerning FliG's stability and folded state. The FliG protein itself was also very well-suited to this method as it contained very few aromatic amino acids and was almost entirely α-helical, meaning that its CD spectra would be quite straightforward to obtain and analyze. I therefore prepared five different FliG constructs, along with the $FliF_C$ peptide, for CD analysis with the hope of determining both the percentage of the various secondary structures present and if any of the mutations or conjugations made to the protein detrimentally affected its stability.

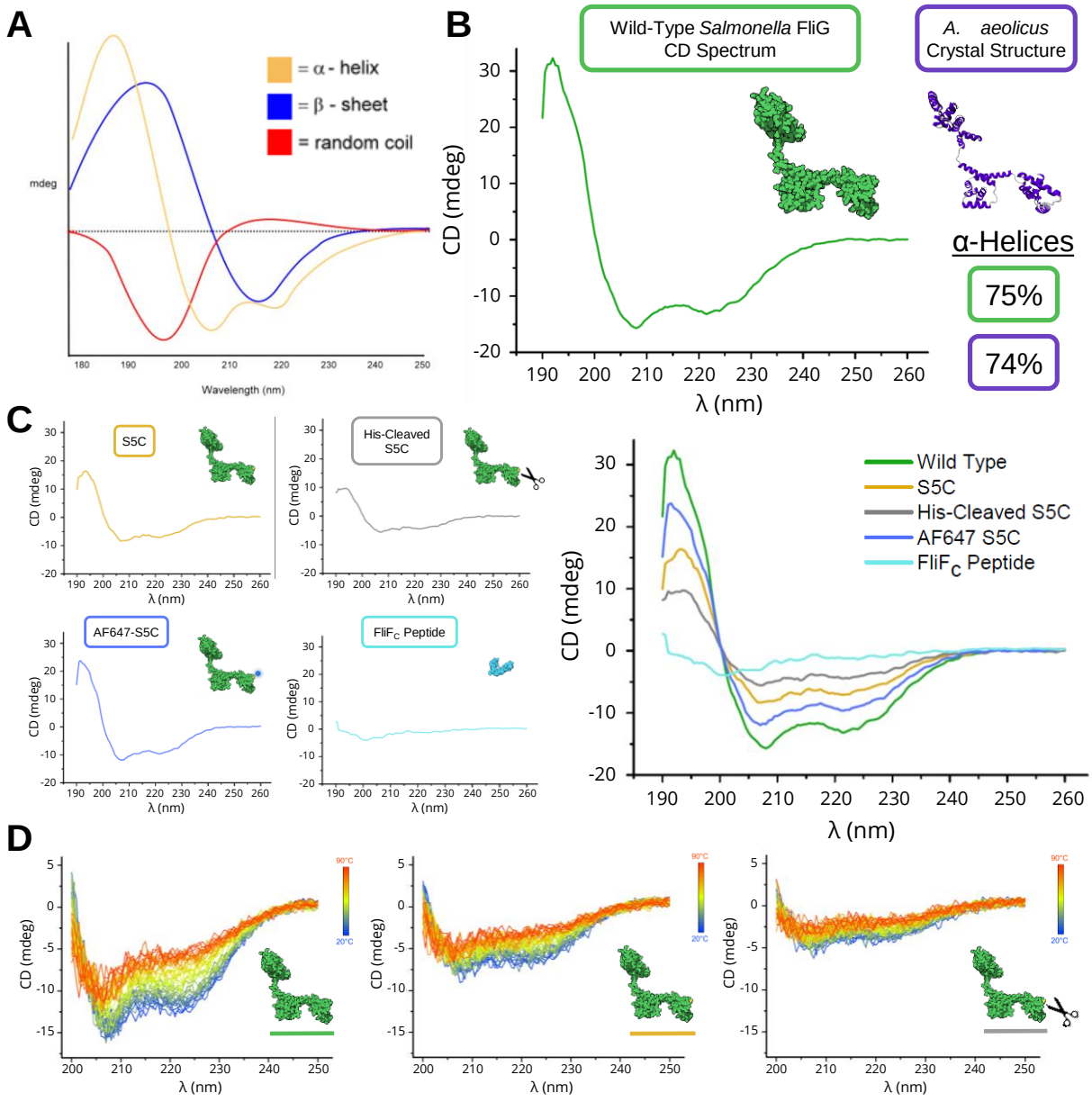


Figure 4.2 | Circular dichroism spectroscopy was used to determine the secondary structure composition of FliG constructs and the $FliF_C$ peptide. A) Each class of protein secondary structure produces a unique CD spectrum in the UV range of wavelengths with characteristic peaks and troughs. Illustration adapted from (171). **B)** CD spectroscopy produced a spectrum for wild-type *Salmonella* FliG which led to the prediction that 75% of the protein was composed of α-helices. The full-length crystal structure from *A. aeolicus* FliG[60] is 74% α-helical[172]. This lends favour to the thought that the wild-type *Salmonella* construct is well-folded. **C)** CD spectra for three other FliG constructs were also obtained. Dye-labelled FliG (AF647-S5C) was shown to be more α-helical than unlabelled single-cysteine FliG (S5C) which in turn was more α-helical than the same construct which had had its His-tag cleaved by a protease (His-Cleaved S5C). All three however showed lower signals than the wild-type, possibly indicating a degree of misfolding. A CD spectrum for the $FliF_C$ peptide was also obtained and produced a trace characteristic of an unstructured, random coil. **D)** CD spectra were taken at temperatures from 20°C to 90°C for three different FliG constructs. All spectra showed a transition from a characteristically α-helical shape to that of a random coil, consistent with heat-induced denaturation. The wild-type FliG construct showed a sharper and more pronounced transition than that seen in either the single-cysteine or His-cleaved single-cysteine constructs.

Analyzing the wild-type FliG protein at 20°C in a low salt buffer gave a CD trace characteristic of a predominantly α-helical protein (Fig. 4.2B). With troughs near 208nm and 222nm and a peak near 193nm, it almost perfectly matched the textbook spectra for α-helices[174] and reassured me that the protein was not completely denatured. Analyzing the data with software to calculate the protein's secondary structure composition given a CD spectra[172] predicted that around 75% of the protein chain was composed of α-helices with the rest being mostly random coil. This was the best indication so far for me that the protein was in its native, folded state as in the full-length crystal structure of FliG from *A. aeolicus*, 74% of the amino acids were structured as α-helices and the remainder were unstructured linkers.

Moving on to the other constructs, I performed the same analysis, obtaining CD spectra for single-cysteine FliG (S5C), single-cysteine FliG with its His-tag cleaved (His-Cleaved S5C), dye-labelled single-cysteine FliG (AF647-S5C), and the $FliF_C$ peptide (Fig. 4.2C). CD spectra of the FliG constructs demonstrated that they were all predominantly α-helical but to less of an extent than in the wild-type protein, with predictions showing that the helices made up a lower percentage of the total polypeptide. Although some differences in the constructs' spectra could perhaps be explained by improperly measured concentrations, it appeared that the mutated and modified versions of FliG were most likely misfolded or unfolded to some degree, though how much this misfolding might have affected their function was hard to say. The $FliF_C$ peptide on the other hand showed no α-helical character, with its trace appearing more like that of a random coil (Fig. 4.2C). This was not entirely surprising as the structural studies which had used thermophilic $FliF_C$ peptides predicted that the domain might take an unstructured form until it was bound to FliG.

In addition to obtaining CD spectra at a single temperature, for three of the constructs I also took CD measurements over a range of temperatures from 20 to 90°C. With this I hoped to see a transition from a characteristically α-helical spectrum to that of a random coil as the protein unfolded with heat. Well-folded, globular proteins tend to show a sharp transition between their native and denatured states but as FliG possessed multiple separate domains, and as each domain might unfold at a slightly different temperature, the transition was likely to be blurred and

I was not confident that a discrete FliG melting point would be easily observed. The CD spectra did show a clear shift however as the temperature increased, with the traces transitioning from an α-helical to a random-coil shape (Fig. 4.2D). In part due to the noisiness of the data, no discrete melting point could be picked for any of the constructs but it was apparent that the sharpest transition between α-helical and unstructured traces occurred with the wild-type construct, as both the single-cysteine (S5C) and the His-cleaved single cysteine (His-cleaved S5C) exhibited smaller and blurrier transitions (Fig. 4.2D).

Thus CD spectroscopy of FliG had confirmed some of my suspicions and supplied me with useful information for the work ahead. Based on what I had observed, I was fairly confident that the wild-type FliG construct was well-folded and presumably functional, whereas the mutated and modified versions of the protein appeared to show some degree of misfolding, at least in the buffers and conditions used for the spectroscopy. The $FliF_C$ peptide on the other hand gave every indication of taking the form of an unstructured, random coil. This lack of structure found in the free-floating, unbound peptide was predicted by the structural studies working with thermophilic $FliF_C$ and thus it did not provide an immediate explanation for the lack of FliG binding that had been so far observed.

I thus had valuable information about the structural state of the constructs I was using but still did not have evidence for FliG-$FliF_C$ binding or a definitive explanation as to why it was not being observed. The CD spectra indicated to me that the best chance I had for observing an interaction between the protein and the peptide would be if I used the unlabelled, wild-type construct of FliG, as it appeared to be well-folded. But without a label on the construct, I was left to find a technique which would be able detect the protein and discern whether or not the peptide was bound to it. I therefore began to search through various biophysical and biochemical methods in the literature and very quickly came upon one which had been used in the first study of thermophilic $FliF_C$ to demonstrate its binding to the $FliG_N$ domain of *T. maritima*[90]. I decided to implement this technique for my system, with the hope that I could demonstrate binding of enteric $FliF_C$ to *Salmonella* FliG.

A Tryptophan Fluorescence Assay Demonstrated One-to-One Binding of FliG to $FliF_C$

The previous study had taken advantage of two useful properties of the $FliF_C$-FliG system in developing an assay which demonstrated binding between the two partners. Firstly, the $FliF_C$ domain contains a single, conserved tryptophan residue which is necessary for FliG binding and which is buried in the hydrophobic cleft of $FliG_N$ when the two interact[90]. Secondly, the FliG protein does not contain a single tryptophan residue in its sequence. Tryptophan is fluorescent in the UV range of wavelengths with its excitation point being around 295nm and its emission spectrum centring around 350nm[175]. The magnitude and position of the peak of this emission spectrum is shifted however depending on the local environment surrounding the tryptophan, with aqueous environments red-shifting and hydrophobic environments blue-shifting the fluorescence. This property lends itself well to applications in the study of protein-protein binding and conformational changes and due to the single, fortunately-placed residue on the $FliF_C$ peptide, tryptophan fluorescence was able to be used to study the binding of $FliF_C$ to FliG.

In their work it was shown that the tryptophan residue on the $FliF_C$ peptide both fluoresced brighter and that the peak of its emission was blue-shifted upon the titration of $FliG_N$ into solution. This could be explained by the peptide interacting with the protein at the predicted site, with the tryptophan residue becoming buried in the hydrophobic environment of the $FliG_N$ binding cleft (Fig. 4.3A). The technique also indicated that the binding was one-to-one, with addition of excess FliG leading to no change in the signal, and was able to give a rough upper bound on the dissociation constant for the interaction. Though I was still uncertain whether the enteric constructs I was using would have as strong and robust interactions as those observed in the thermophilic system, I decided to apply this **tryptophan fluorescence assay** to the $FliF_C$ peptide and wild-type *Salmonella* FliG with the hopes of obtaining similar indications that the two were binding to one another.

I was still concerned about the overall solubility of the peptide so for these experiments I used the unlabelled single-cysteine $FliF_C$ construct, the only version that appeared to have successfully

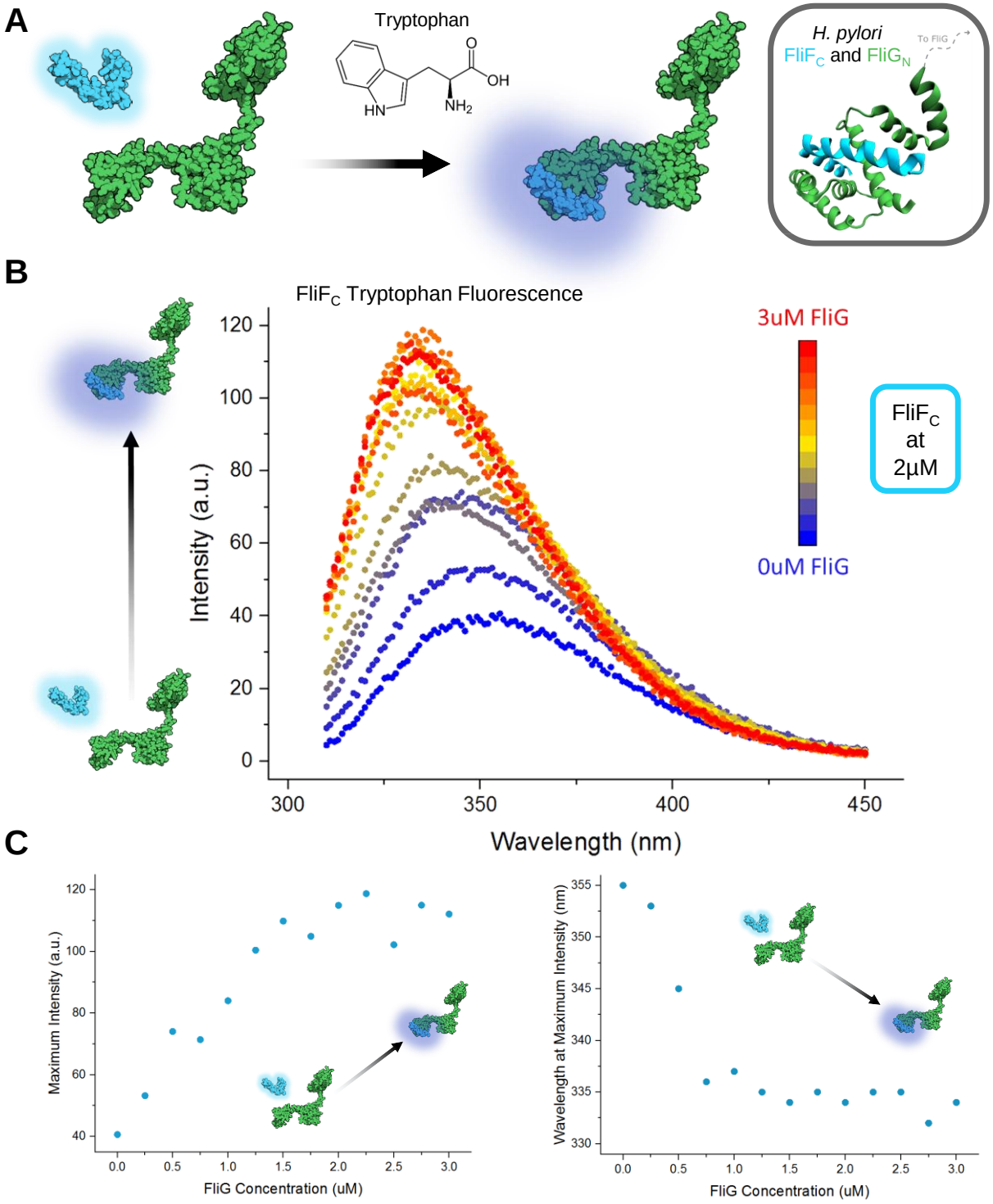


Figure 4.3 | A tryptophan fluorescence assay indicates that the $FliF_C$ peptide binds to wild-type *Salmonella* FliG in a one-to-one manner. A) Previous studies had shown that the single tryptophan on $FliF_C$ fluoresces brighter and the peak of its emission is blue-shifted upon binding to FliG. As can be seen in the crystal structure[91], this change in fluorescence occurs as $FliF_C$ is buried in the hydrophobic cleft of $FliG_N$ upon binding. **B)** Exciting at 295nm and measuring the emission from the tryptophan residue of enteric $FliF_C$ as *Salmonella* FliG was added showed that the signal increased and was blue-shifted as the FliG concentration increased. **C)** The maximum emission of the tryptophan increased three-fold as FliG was added up to a one-to-one stoichiometry and then levelled off as excess FliG was introduced. The wavelength at which the peak emission occurred shifted from 355nm without FliG to 335nm at one-to-one FliG with the trend also levelling off as excess FliG was added.

gone into solution. Keeping the concentration of the peptide at 2μM, I titrated in increasing amounts of wild-type *Salmonella* FliG and measured the fluorescence of the peptide's tryptophan at each point in the titration. When excited at 295nm, the residue gave a very clear signal, and both the maximum intensity of this signal and the location of the peak of the emission spectrum were seen to respond to the addition of FliG (Fig. 4.3B). Increasing the concentration from 0 to 2μM FliG showed an approximately three-fold increase in the tryptophan fluorescence of the peptide, with no further increase seen as excess FliG was added to 3μM (Fig. 4.3C). In addition, the peak of the fluorescence emission was blue-shifted by around 20nm with the addition of FliG, moving from around 355nm in the absence of FliG to 335nm at 2μM FliG (Fig. 4.3C). Again a levelling-off in the trend was seen as excess FliG was added, with no significant change in the peak location occurring above 2μM FliG.

These observations very closely matched those stated in the published results using *T. maritima* constructs and hence led me to believe that the peptide was in fact binding to FliG. I did one set of additional experiments however. To control for the effects that increasing the overall FliG concentration in solution might have on the spectra, I also performed the titration with the same FliG concentrations but rather than adding the protein to the peptide, I added it to free tryptophan residues. Monitoring the fluorescence of these residues showed no change upon addition of FliG, indicating that the mere presence of extra protein in solution did not alter the observed spectra.

This control, coupled with the great similarity between my results and those published for the thermophilic system, convinced me to a great extent that my enteric $FliF_C$ peptide was binding to *Salmonella* FliG. The fact that the trends levelled off as excess protein was added also indicated to me that the binding was one-to-one and specific, with a single peptide interacting with a single FliG. To confirm these conclusions however and to convince myself that binding was in fact taking place, I needed to perform other parallel assays which could more directly demonstrate the interaction. These come in the sections to follow, but before those are discussed I will describe another method which gave further information about the stability of my FliG constructs and which provided another indirect confirmation of $FliF_C$-FliG binding.

A Thermal Shift Assay Indicated That the Stability of FliG Increases upon $FliF_C$ Binding

Though CD spectroscopy had produced some useful insight on the folded state of my FliG constructs and the tryptophan fluorescence assay gave the clearest indication up to this point that the peptide was binding to FliG, I wanted to confirm and extend the observations from these assays through the use of complementary biophysical and biochemical methods. One such method, called a **thermal shift assay** was technically very straightforward, required very little material or preparation, and promised to provide information both on the folded state of my constructs and perhaps add to the evidence of $FliF_C$-FliG binding in my system. I therefore decided to implement this technique with my purified proteins hoping that I could obtain both a fuller picture of their stability and a confirmation that my $FliF_C$ peptide interacted with the FliG protein.

In a similar fashion to the work with tryptophans just described, the thermal shift assay uses fluorescent signatures as a read-out of the folded state of a protein molecule. A special dye called SYPRO Orange is added to the mix which experiences an increase in its quantum yield upon binding to the hydrophobic patches of proteins. This unique characteristic means that as a protein solution is heated to a higher and higher temperature and the individual molecules denature and reveal their hydrophobic interiors, any SYPRO Orange dye in that solution will find more and more locations at which to bind and, upon doing so, its total fluorescence emission will increase (Fig. 4.4A). Hence, with the appropriate samples and conditions, SYPRO Orange fluorescence can be used as a read-out of the folded state of proteins in solution and by heating the mixture and plotting the fluorescence against temperature, one can extract useful parameters from the characteristic curves that are generated such as the melting point of the molecule.

Thus, due to its simplicity and the useful information that it provides, the thermal shift assay has become quite a popular technique and is commonly used to screen for optimal buffer conditions or to compare one construct with another. I looked to employ it in order to obtain a quantitative read-out of the folded state of each of my proteins, comparing the melting profiles from different constructs to obtain an idea of their relative stabilities. In addition, I hoped to determine if the

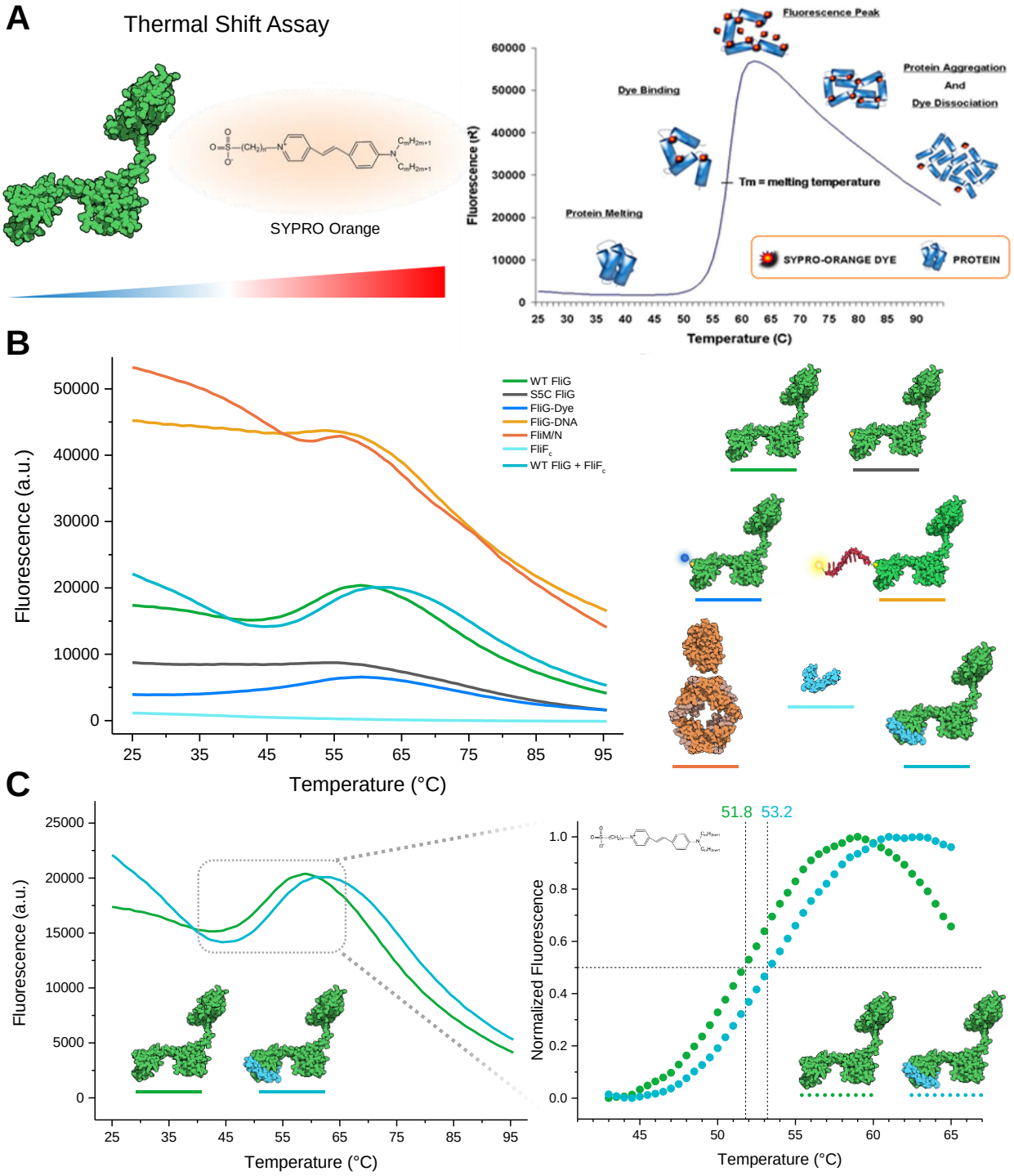


Figure 4.4 | A thermal shift assay demonstrates that C-ring protein constructs vary in stability and that addition of $FliF_C$ to FliG increases the melting point of the protein. A) The thermal shift assay uses the SYPRO Orange dye whose fluorescence increases when bound to the hydrophobic patches of proteins. As the protein is heated, it unfolds, exposing hydrophobic residues to which the dye binds. Once the protein has denatured, the fluorescence reaches its peak and then tails off as heat is further increased and the denatured proteins aggregate, excluding the dye. Illustration from (173). **B)** Performing the thermal shift assay on various C-ring protein constructs gave a variety of curves. Wild-type FliG showed the most distinct transition from folded to unfolded whereas the mutant constructs had broader traces, indicating some degree of initial misfolding. **C)** Addition of the $FliF_C$ peptide to wild-type FliG shifted the entire melting profile to the right and increased the difference between the bottom trough and top peak of the curve in the 'transition zone' of temperatures. The melting temperature was found to be increased by 1.4°C with the addition of the peptide.

addition of the $FliF_C$ peptide to the FliG solution stabilized the proteins and increased their melting temperature as this would be both an indication of binding and a confirmation of the prediction from the previous work with thermophilic constructs that the two species fold around one another, generating a more robust tertiary structure at FliG's N-terminus.

I therefore prepared aliquots of each of the C-ring protein constructs that I had purified and added the SYPRO Orange dye to each. Heating the mixtures from 25°C to 95°C and monitoring the SYPRO Orange fluorescence produced distinct curves for each construct tested. Comparing melting profiles from wild-type FliG (WT FliG), single-cysteine FliG (S5C FliG), dye-labelled FliG (FliG-Dye), DNA-conjugated FliG (FliG-DNA), the $FliF_C$ peptide, as well as the recently-purified FliM-FliN complex (FliM/N, discussed in 4.6) showed a wide range in the intensity of the fluorescence signals and distinct differences in how each construct appeared to transition from its folded to unfolded state (Fig. 4.4B). Whereas some samples produced a curve with a large increase in SYPRO Orange fluorescence over a small range of temperatures, indicative of the sharp transition between the native and the denatured state of a well-folded protein, other constructs showed only a slight increase in fluorescence over a broad range of temperatures. These broader, flatter profiles are characteristic of proteins which are denatured or improperly folded to some degree at the beginning of the temperature ramp, and their presence in my trials indicated that some of my constructs were less well-folded than others.

Comparing the profiles, it seemed that the same overall trends in the stability between the constructs that were seen in the CD measurements were again seen in the thermal shift assay. Both the single-cysteine FliG and the dye-labelled FliG showed broad melting transitions whereas the wild-type FliG exhibited a sharp and distinct change between folded and unfolded states. This matched the observations from CD spectroscopy that wild-type FliG was more α-helical in nature than the two mutant constructs. The two constructs which were tested in this assay but not with CD spectroscopy, DNA-conjugated FliG and the FliM-FliN complex, showed the largest overall fluorescence signals but this most likely is due to the overlap between the spectrums of the dye-label at the 3' end of the DNA and SYPRO Orange in the case of the conjugate and to the many

hydrophobic patches holding together the complex in the case of FliM-FliN. The two samples both seemed to produce broad transitions but the true nature of their unfolding was most likely obscured by the high initial fluorescence. Finally, the $FliF_C$ peptide produced very little signal at all, and due to this low signal, not much could be gleaned about the folded state of the peptide. However, the flat trace seemed to correlate with the previous studies using thermophilic versions of $FliF_C$ and with my CD measurements which indicated that it was unstructured when free in solution.

The most interesting observations from this assay however came when comparing the wild-type FliG trace with that observed for a mixture of the wild-type construct and the $FliF_C$ peptide. As mentioned, the peptide on its own produced practically no signal but when that same quantity of peptide was added to wild-type FliG, it appeared to have quite an effect on the melting profile. With $FliF_C$ added, there was a greater difference between the minimum and the maximum in the transition zone of the curve and the melting temperature was increased by 1.4°C (Fig. 4.4C). The simplest explanation for this result seemed to me that the peptide was binding to FliG and in some way increasing its overall stability.

As mentioned previously, a major finding of the studies using *T. maritima* $FliG_N$ and $FliF_C$ was that binding induces a conformational change in $FliG_N$, with the two components wrapping around one another to jointly produce a tertiary structure much like those seen in the ARM motifs[88]. This joint $FliF_C$-$FliG_N$ domain was predicted to be more stable than either of the two components on their own and helped to explain why the dissociation constant of the complex was so low. The results from the thermal shift assay demonstrating that the FliG melting temperature was increased upon the addition of $FliF_C$ corresponded well with this model for FliF-FliG binding, as according to this theory it would take more energy to denature the stable joint $FliF_C$-$FliG_N$ domain than it would the $FliG_N$ domain on its own.

To bring this section to a close then, in following on from the tryptophan fluorescence studies, I again had a strong indication that the peptide I had had synthesized was in fact binding to the FliG that I had purified. Again the method of detection was an indirect one, but the correlation between its results and those I had obtained previously gave me confidence that I was observing a real

interaction between two binding partners. Indications from the tryptophan fluorescence that the binding was one-to-one fit nicely with previous findings and with current models of BFM assembly which predicted a strong and stable association between the MS-ring and the C-ring, with each and every FliF having a FliG bound to it.

Concerning the stability of my constructs, observations from the thermal shift assay and those from CD spectroscopy showed there was a great deal of variability in the starting folded state from one construct to the next, but both techniques seemed to indicate that the wild-type FliG exhibited a more stable, and most likely more native, folded state than the mutated and modified constructs. This was interesting in light of my earlier results with permanent covalent FliG-DNA conjugates on both short linear templates and origami tiles. I had failed to see any FliG-FliG interactions in those experiments and following the biophysical characterizations of this section I began to think that lack of interaction was perhaps due to some degree of improper folding in the FliG portion of the conjugates.

Following this train of thought, and taking what I had seen from CD spectroscopy, tryptophan fluorescence, and the thermal shift assay, I decided on a new approach for the DNA-templated assembly of the FliG ring. Though I had not yet seen it on gels, I was confident that my enteric $FliF_C$ peptide was binding to wild-type *Salmonella* FliG and it appeared that once bound these two formed a strong and stable complex. Thus I began to think that this natural linkage would perhaps be the best way of incorporating FliG into a DNA-templated system, the method least likely to cause misfolding or denaturation in the protein and therefore most likely to preserve any FliG-FliG interactions. With this in mind, templating the peptide on DNA and using this DNA-templated $FliF_C$ as a ‘binding platform’ for wild-type FliG became my strategy for the *in vitro* assembly of the FliG-ring.

If this strategy proved successful, it would provide definitive proof of $FliF_C$-FliG complex formation, demonstrating for the first time in non-thermophiles that the short 39-amino-acid stretch of FliF dangling down into the cytoplasm was sufficient for FliG binding, a small yet significant contribution to the field. To even attempt this strategy there were technical hurdles to overcome however, as I

needed to both conjugate a DNA strand to the $FliF_C$ peptide and in some way purify the peptide-DNA conjugates. These two challenges are the topics of the next section, with its focus being the production of a pure, permanently linked $FliF_C$-DNA construct.

4.3 Conjugation of a DNA Oligonucleotide to a Cysteine-Modified $FliF_C$ Peptide

As documented in section 4.1, I had previously attempted the conjugation of a maleimide-modified dye molecule to a single-cysteine version of the $FliF_C$ peptide without success. Visualizing the reaction mixtures on gels showed no evidence that any dye-peptide conjugates had been formed. The same maleimide-cysteine reaction had successfully produced FliG-DNA conjugates however and I was not ready to abandon that strategy for the $FliF_C$ peptide system, especially considering that all the components for the reaction were already available to me.

I had the 10-nucleotide-long (N10) maleimide-modified DNA strands that had been used previously in the conjugation and then templating of FliG and decided for the sake of simplicity that I would attempt a reaction between these same strands and the single-cysteine peptide. Though I had yet to observe the peptide on polyacrylamide gels, my hope was that I could use the fluorescence of its tryptophan to detect its presence throughout the conjugation and purification processes and, should the reaction prove successful, I could use the DNA strand as a label for the conjugate on gels, visualizing its location through SYBR Gold staining.

Thus I proceeded with trials looking to conjugate the N10 DNA strand to the $FliF_C$ peptide. This section describes the process and the steps that were undertaken in order to obtain a pure $FliF_C$-DNA conjugate for use in templating experiments. A conjugate of this sort was the key ingredient in the FliG assembly strategy which I was now pursuing and obtaining a homogeneous and well-characterized species was very important for the work that was to come.

HPLC Purification Isolated the Conjugate from Excess Maleimide-DNA and Cysteine-FliF$_C$

As much of the work in obtaining functional reactants and the proper reaction conditions had been done previously, very little preparation was needed for the conjugation of the cysteine-peptide with the maleimide-DNA. Cysteine-FliF$_C$ had been successfully solubilized in water (see 4.1) and was stored in aliquots in the freezer, whereas a great deal of purified maleimide-DNA could be found stored as a frozen pellet. Following a similar procedure to that used during the FliG to DNA conjugation, I added TCEP to thawed aliquots of the single-cysteine peptide to break any disulfide bridges before mixing this with excess maleimide-DNA which had been dissolved in a slightly alkaline buffer (Fig. 4.5A). After a period of incubation at room temperature, I ran aliquots of the reaction mixture on polyacrylamide gels and stained for DNA with SYBR Gold, looking for a shift in the position of the maleimide-DNA band due to the attachment of a peptide at its 5' end.

The reaction was successful from the first attempt as an extra slow-moving SYBR-Gold-stained band appeared on gels upon addition of the peptide (Fig. 4.5B). It was very hard to believe that this was anything else but my desired product and I was confident that I had produced a **FliF$_C$-DNA conjugate**. Significantly, these gels provided the first proof that the peptide could be tracked using PAGE, with the highly-charged DNA acting both as a label upon addition of SYBR Gold and as a migration aid to the relatively uncharged peptide. In addition, the conjugation of the highly hydrophilic DNA strand seemed to clear up any remaining issues relating to the solubility of the peptide, as all solutions with the conjugate appeared transparent and free of pellets.

The issue now arose however of how to purify the product. My initial attempts at purification based on size proved unsuccessful as the peptide appeared to non-specifically adhere to the plastics and resins of the size exclusion columns that I tried. Moving on, I decided to try HPLC purification based on hydrophobicity, imagining that I could separate the hydrophilic DNA from the more hydrophobic peptide and peptide-DNA conjugate using this technique. HPLC had been used by the commercial suppliers in purifying the peptide after synthesis so it seemed logical that the construct would be stable throughout the process. The acetonitrile-containing buffers that were

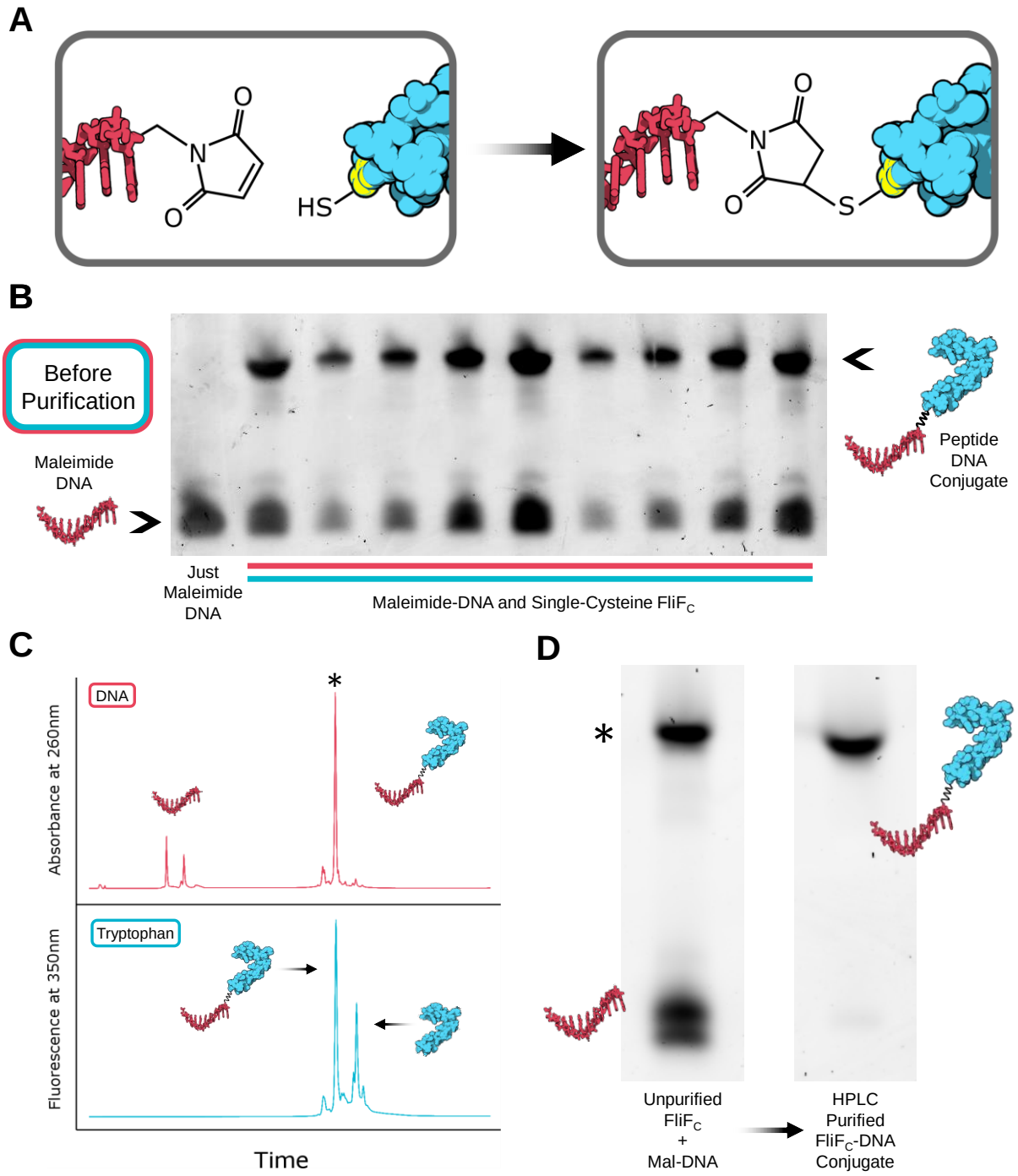


Figure 4.5 | Maleimide-modified DNA was conjugated to a single-cysteine version of the FliF$_C$ peptide and the conjugate was purified using HPLC. A) Using the same reaction scheme as for the FliG-DNA conjugates, a 10-nucleotide-long, maleimide-modified DNA strand was reacted with the thiol of the peptide's N-terminal cysteine to create a permanent thioether bond. **B)** Aliquots from the reaction mix run on a native polyacrylamide gel stained with SYBR Gold showed a distinctive slow-running band not present in controls with only the maleimide-modified DNA. This was found to be the desired peptide-DNA conjugate. **C)** Passing the reaction mix through the HPLC machine and scanning for either DNA absorbance at 260nm or fluorescence from the peptide's tryptophan at 350nm produced distinct and separate peaks for unreacted DNA, unreacted peptide, and the peptide-DNA conjugate (*). **D)** Collecting the HPLC fraction containing the conjugate, vacuum-concentrating its contents, resuspending in water, and running this on a native polyacrylamide gel showed that the process had removed all of the excess DNA.

used would denature the peptide but I reasoned that it was small enough to refold once unfolded and was reassured by the predictions in the literature that $FliF_C$ was in fact unstructured until bound to FliG and hence could transition to its α-helical form even if previously a random coil.

Putting the reaction mix through the HPLC machine produced several clearly separated peaks (Fig. 4.5C). Running the appropriate control samples and scanning for both DNA absorption and tryptophan fluorescence from the peptide showed that this technique was able to resolve the excess DNA from that conjugated to the peptide and even to produce separate peaks containing DNA-conjugated and unconjugated peptide. Thus, by collecting the proper fraction I could solely obtain the $FliF_C$-DNA conjugate, without any of the unreacted DNA or peptide which would interfere with later templating experiments. This was a step better than what I had been able to achieve with the FliG-DNA conjugates, as in that instance, using affinity chromatography or size-exclusion columns I had been unable to remove unconjugated FliG and hence had added a mixture of unconjugated and conjugated protein species to the DNA templates. With this new peptide-based strategy it appeared that I could be assured of a more homogeneous and well-characterized construct for organization on the nanostructures.

Once the peptide-DNA peak was collected however, I still had to find a way of moving the construct from the HPLC buffer that it was in to one better suited for the FliG protein. Previous attempts at mixing FliG with HPLC-purified components had showed that it was sensitive to both the acetonitrile and acetate ions in the buffers of the collected peaks, with the protein irretrievably crashing out of solution. Subjecting the collected conjugate fraction to vacuum concentration was successful in removing the acetonitrile but addition of FliG to the resulting pellet again showed aggregation most likely due to leftover acetate. I therefore attempted to resuspend and buffer exchange the vacuum-concentrated samples using a spin column but, as I had seen previously, the peptide adhered non-specifically to the resins and could not be recovered.

I therefore decided to alter the HPLC buffers that I was using by reducing the concentration of the buffering salts in solution by a factor of ten, hoping that I could still achieve the proper separation of the species and that FliG would be stable in a solution with a small amount of residual acetate.

The traces appeared nearly identical to those obtained with the standard HPLC buffers and the pellets of the vacuum-concentrated fractions appeared visibly less salty and glassier than those obtained previously. Adding an aliquot of wild-type FliG to the conjugates purified using these more minimal buffers showed no signs of pelleting or aggregation and hence assured me that with this method it would be safe to omit the buffer exchanging step.

Running the purified fraction on a gel showed very clearly that HPLC was able to remove all of the unreacted DNA, as only the single slower-running band was observed following purification (Fig. 4.5D). By running unreacted N10 DNA strands on the gel at a range of concentrations and comparing the intensity of the SYBR Gold signal from those bands to that obtained for the conjugate, I could also roughly determine the concentration of the peptide-DNA, which was a useful parameter for any further work in which I would add the conjugates to larger DNA templates.

Thus in a shorter space of time and with considerably less effort than with my conjugation of a DNA strand to FliG, I had been able to obtain a pure $FliF_C$-DNA conjugate. I had used the same reaction scheme but the re-foldable nature of the peptide had allowed me to employ HPLC for purification, a technique which could remove not only excess DNA but unreacted peptide from the solution. For my new *in vitro* assembly strategy then, I could work with purer and better characterized components than I had obtained when using FliG-DNA conjugates. This was a definite advantage to the approach and seemed to add to its chance of success, but the key interaction which the whole scheme depended on, that between FliG and the $FliF_C$ peptide, had only been shown through indirect methods. It had yet to be seen on a gel. The fact that I could now visualize $FliF_C$ via SYBR Gold staining of its DNA-tag was encouraging however and meant that I could track where the conjugate went upon addition of FliG.

The addition of FliG to these conjugates is the focus of section 4.5, but before those experiments were attempted, the conjugates were simply added to and organized on DNA templates. The next section describes these templating experiments.

4.4 Organization of $FliF_C$-DNA Conjugates on Five-Site Linear DNA Templates

Now that the processes of conjugating a DNA strand to the peptide and then purifying the conjugates had been worked out, the next step in the overall assembly strategy was to organize the peptide-DNA constructs on DNA nanostructure templates. To begin with, I chose the simplest and best-characterized templates I had available, the same series of short, linear templates based off of the five-binding-site T5N10 strand which I had used to organize FliG-DNA conjugates (see section 3.3). With these I hoped to obtain polyacrylamide gels similar to those I had run previously in which a ladder of bands formed and I was able to count the number of FliG molecules bound to each of the templates bearing from one to five binding sites. If I could obtain a similar ladder for the peptide-DNA conjugates and count the peptides present in each species, I could be sure that I was adding a template with a particular stoichiometry and arrangement of $FliF_C$ domains to FliG, which was necessary if I was to accurately interpret FliG's behaviour on those templates.

Peptides organized on these templates with the same spacing of binding sites as seen *in vivo* would provide the best mimic so far of the MS-ring. An array of $FliF_C$ domains dangling down for FliG to bind to would mirror what might be seen in the cytoplasm of the bacterium during the assembly phase of the BFM. Following MS-ring formation through the oligomerization of FliF, it is thought that FliG is the next protein to join the assembly, binding to the C-terminal domains of FliF extending into the cytoplasm and thereby linking the forming C-ring to the established structures in the membrane[55]. Through the DNA-templating of $FliF_C$ peptides, I would be able to create a more natural binding platform for FliG than with my previous system using FliG-DNA conjugates and a DNA template. This new layer between the DNA and FliG would ensure that the protein was connected to the underlying scaffold in the closest approximation of its *in vivo* state, hopefully preventing any misfolding that DNA conjugation to the protein might have caused. In addition, binding of FliG to the $FliF_C$ domain had been predicted by some to induce a conformational change which triggered some degree of FliG-FliG interactions[91], and I was keen to see if these interactions could be observed using my *in vitro* system.

Thus, I carried on with the next step in my assembly strategy: organizing $FliF_C$-DNA conjugates on the short, linear templates. I used the same set of DNA template strands that had been employed previously in the search for evidence of the domain-swap hypothesis and I again hoped that with these strands I would be able to find examples of gap-filling or runaway polymerization, reasoning that the previous absence of these behaviours was due to the unnatural nature of the linkage between FliG and its template. I needed to make sure however that, as predicted, there were no interactions between the $FliF_C$ domains themselves and that the conjugates bound stably to the template and at the correct positions and hence began experiments with the aim of producing a clear ladder of bands containing from one to five peptides bound to templates with an increasing number of binding sites.

One to Five $FliF_C$ Peptides Were Arranged on DNA Templates in Various Configurations

For the first DNA templating trials with the peptide-DNA conjugates, I used the same set of six strands as I had used for the FliG-DNA conjugates. Containing from one to five 10-nucleotide-long binding sites, spaced at various distances from one another, they would provide particular spatial arrangements of the peptide which could be used to investigate the domain-swap hypothesis (Fig. 4.6A). With each I had the option of leaving the intervening DNA sequences between the binding sites single-stranded and hence short and flexible, or to add a strand to fill those regions and create a rigid, double-stranded space between the bound peptides.

In order to use these templates effectively, I needed to determine the concentration of the peptide-DNA constructs that I had purified. Though it was not extremely precise, I obtained what in the end was a sufficiently accurate determination of the concentration in two complementary ways. I first compared the magnitude of the tryptophan fluorescence signal seen on the HPLC traces with that from a known amount of unconjugated peptide and then additionally ran a range of volumes of the N10 strand with a known concentration on gels, stained with SYBR Gold, and mapped the intensity of the peptide-DNA band to the trend found for the N10 strands (Fig, 4.6B). Both methods

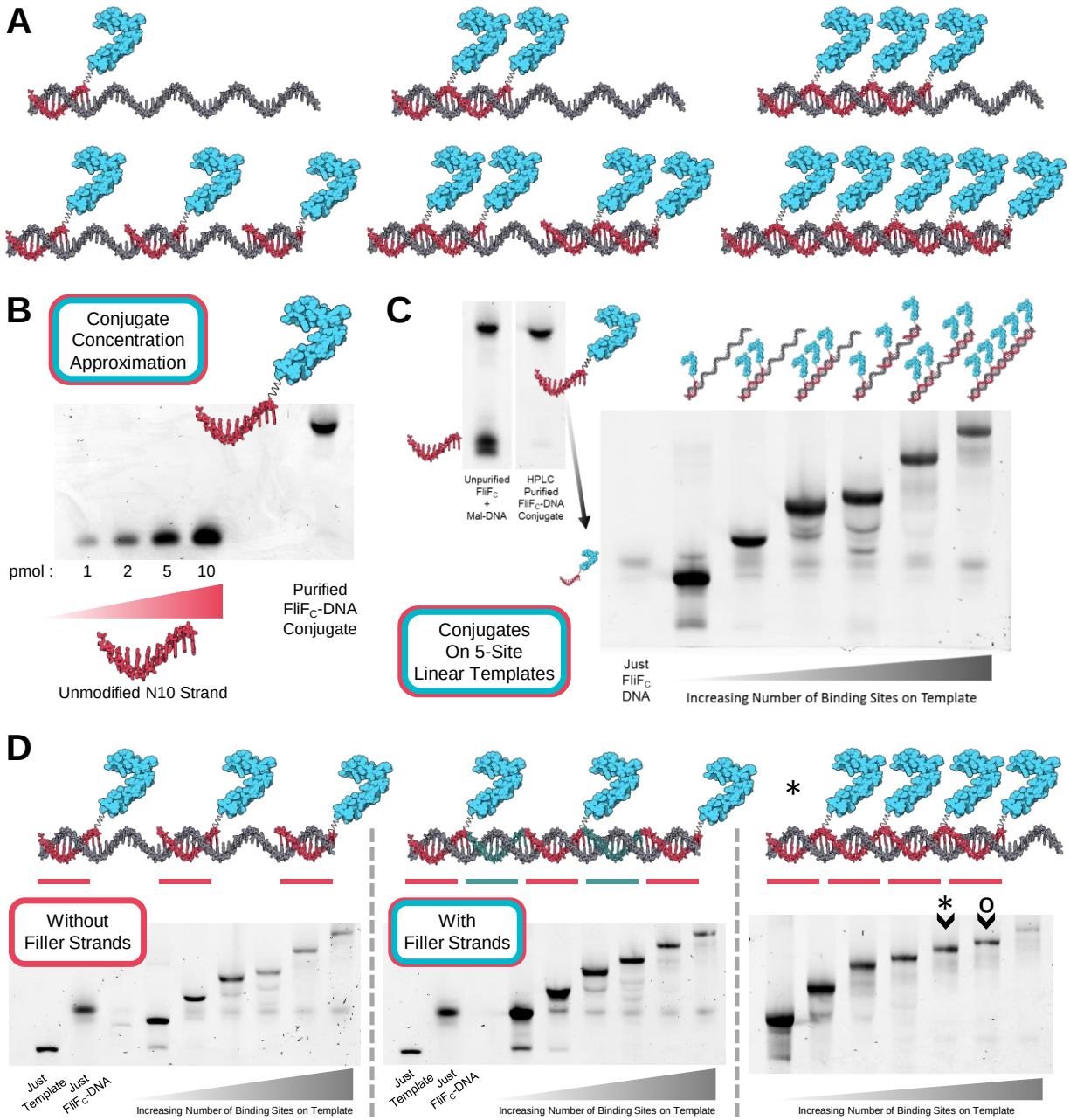


Figure 4.6 | HPLC-purified FliF$_C$-DNA conjugates can be arranged on short linear DNA templates and visualized using native polyacrylamide gels. A) The same set of templates derived from the five-site T5N10 strand as was used for the FliG-DNA conjugates was employed to organize peptide-DNA conjugates. The templates incorporated from one to five 10-nucleotide-long conjugate binding sites which were spaced out at different intervals. **B)** To approximate the concentration of the conjugate once purified, a range of volumes of a DNA strand of known concentration were run on a gel alongside the conjugate. Comparing the intensity of the SYBR Gold staining of the bands gave a fairly accurate estimate of the conjugate concentration. **C)** Mixing excess of the purified FliF$_C$-DNA conjugate with each of the six different templates produced a characteristic ladder of bands on a native polyacrylamide gel. Band migration slowed with more conjugates attached to the template and differences could be seen in the migration of templates with the same number but different spacings of the peptide. **D)** Varying the assembly procedure by adding or removing strands which filled the gaps in the templates between the peptides showed no difference in the banding pattern of the gel (left and centre gels). A new template with four side-by-side sites was added to the set (*) and showed a different migration from a template with two sets of two binding sites (º) (right gel).

tended to agree quite well with one another, enabling me to determine the concentration of the conjugate to within around 1µM.

With this concentration known, I then added peptide-DNA to each of the six templates at a stoichiometric excess to the overall number of binding sites that the templates presented in solution. Incubating the strands and conjugate at room temperature in a magnesium-containing buffer, I then ran the mixtures on a native polyacrylamide gel. After some work with gel percentages, running times, and concentrations I was able to obtain clear evidence that the $FliF_C$-DNA conjugates were binding stably to the templates in the locations prescribed by the sequence of the strands, with a distinct ladder of one to five bound peptides visible on the gels (Fig. 4.6C).

The images of these gels implied that at the spacing given, the peptides were not impeded in any way and could fit one beside the other on the templates as FliG had previously. The gels also showed that given five binding sites, only five peptides will bind, indicating that there was no untemplated recruitment of $FliF_C$ to the structures as might have occurred had the peptides interacted with one another. Thus it appeared that the desired ‘binding platforms’ for FliG could be made simply by mixing the conjugates with the templates.

Small variations in the procedure were also experimented with but produced no substantial differences in the results obtained. Omission of the gap-filling strands showed no change in the number of peptides recruited or any signs of improvement in or hindrance to the templating process (Fig. 4.6D), while annealing the conjugates and templating strands produced no differences that could be observed on the resulting gels. One interesting addition however was a new template containing four sequential binding sites called T4d1 (Fig. 4.6D). This was added for future comparison with the T4G1 template (which bears two sets of two binding sites) as it was thought that the gap-less T4d1 geometry might be more likely to encourage runaway polymerization of FliG. As had been seen for the T3d2 and T3G2 templates, with peptide bound, the two structures migrated different distances on the gel and hence one could differentiate between the two peptide arrangements simply by the location of the bands.

Templating the $FliF_C$-DNA conjugates then proved to be a fairly straightforward process with it clearly shown using native polyacrylamide gels that they could be arranged in various stoichiometries and geometries on short, linear templates. The stage was therefore set for the entrance of FliG, yet the questions remained whether or not the protein would actually bind to these peptides and if the binding of FliG to $FliF_C$ could actually be visualized using gels. The following section investigates these questions and the secondary questions that arose once these were answered. It constitutes the highpoint of this chapter and one of the highpoints of this work as a whole and therefore lends itself to a considerable amount of discussion.

4.5 Binding of the FliG Protein to DNA-Templated $FliF_C$ Peptides

With the peptide-DNA conjugates stably binding to short, linear templates and the resultant structures clearly visualized on native polyacrylamide gels, it now came time to begin on the next step in the assembly strategy: the addition of FliG. The hope was that FliG would bind to the immobilized $FliF_C$ domains as it did in the BFM and that this binding could be observed and investigated using *in vitro* techniques. Previously I had had little success observing binding of the two species using polyacrylamide gels (4.1) but I now presumed that this stemmed from earlier issues with the solubility of the peptide. These issues had been resolved and through techniques such as the tryptophan fluorescence assay and the thermal shift assay, I had been able to obtain evidence that the two species interacted with one another at micromolar and possibly lower concentrations (4.2).

These methods were somewhat indirect but gave me confidence that, should I again look for a band-shift on a gel, I might be able to detect one. There was also the possibility that with the peptides arranged in groups on DNA templates, the FliG protein might bind to them more readily, the templated constructs perhaps exhibiting some degree of cooperative FliG binding as compared to a single, isolated peptide. Cooperativity depended on FliG-FliG interactions however. If one FliG protein was already bound to the templated peptide, the interaction of that bound protein with other FliG proteins would bring them in close proximity to unbound peptides nearby

on the template, increasing their local concentration and thereby encouraging FliG to $FliF_C$ binding. If any FliG-$FliF_C$ binding could be observed on a gel, then signatures of this cooperative binding could also be observed by comparing how FliG interacted with various templates with different stoichiometries and spacing of the peptides. In addition, gels could also very clearly demonstrate evidence of FliG gap-filling or runaway polymerization, key predictions of the domain-swap hypothesis.

Thus, the primary strategy moving forward was to use the same gels that had been used to demonstrate DNA-templating of the peptide to now show binding of FliG to the templated constructs. If that failed, more indirect methods such as the biophysical assays of section 4.2 could be turned to, but the hope was that with the peptide solubility issues now resolved, gels would be sufficient. This section describes the experiments conducted whose aim was to demonstrate the binding of FliG to both free and DNA-template-bound $FliF_C$ peptides and then to characterize the resulting DNA-$FliF_C$-FliG species. The hope was that with the extra $FliF_C$ layer, the templates would more closely mimic the C-ring of the BFM than the designs of Chapter 3, and perhaps this new layer would encourage domain-swapped FliG-FliG interactions. As will be shown however, even with this elaborated system, evidence of the domain-swap hypothesis remained elusive.

Only Wild-Type FliG and Not Mutant Constructs Bound to the $FliF_C$ Peptide

Thus, I began experiments in which I mixed the short, linear templates with $FliF_C$-DNA conjugates and FliG proteins, hoping to observe evidence of self-assembled structures (Fig. 4.7A). Using dye-labelled FliG I could track the position of the protein on gels through the dye's fluorescent signal while tracking in parallel the peptide-DNA conjugates and templates by staining for DNA with SYBR Gold. In this way I would be able to detect if both FliG and the DNA constructs were present in any shifted bands and hence attribute the shift to FliG-$FliF_C$ binding.

As described in the previous section, I mixed the peptide-DNA constructs with six different linear templates and confirmed that the appropriate structures had formed using native polyacrylamide gels. To each of these samples, I then added excess dye-labelled FliG, the dye having been

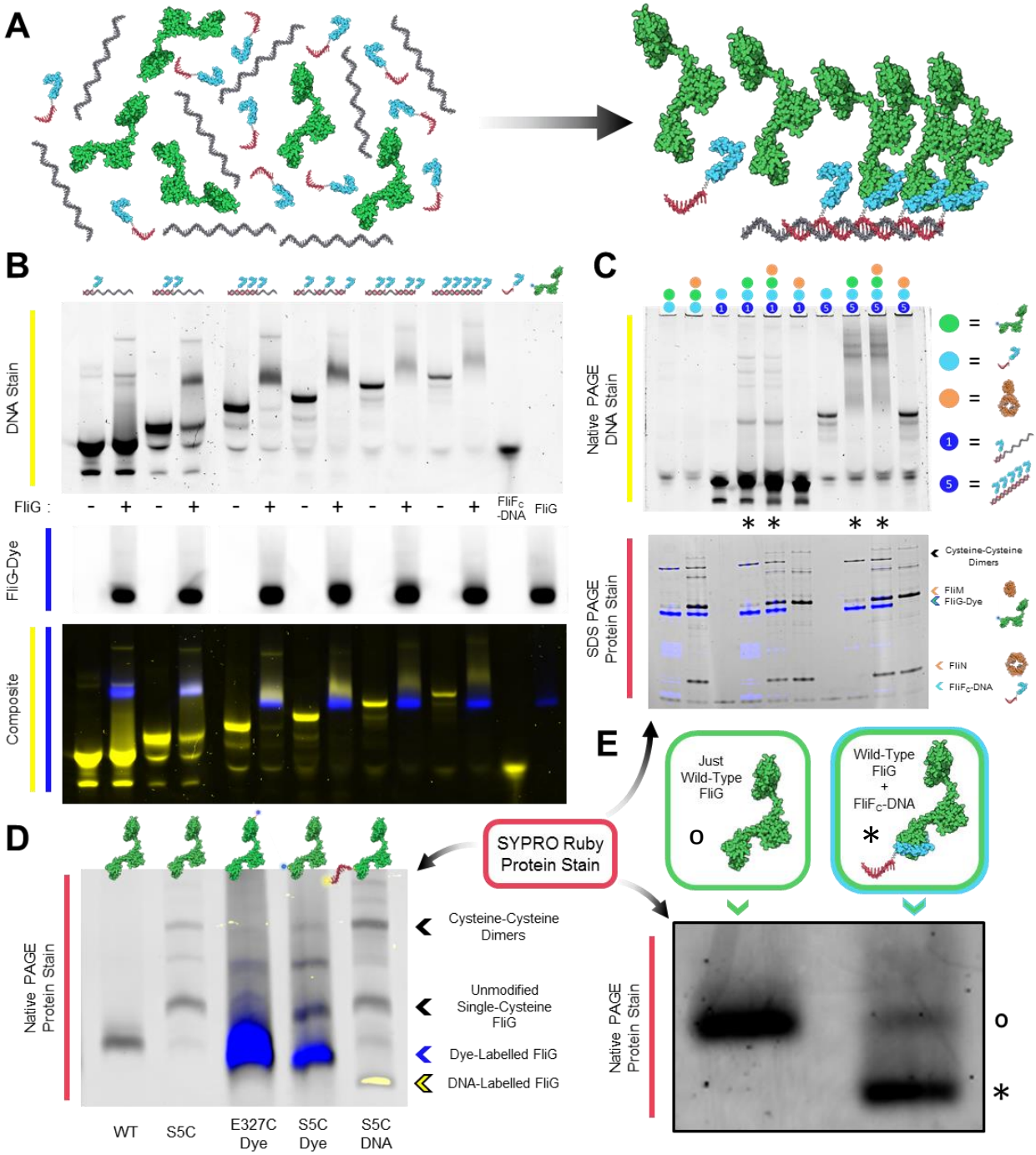


Figure 4.7 | Band-shifts on native polyacrylamide gels seen upon addition of FliG to DNA-templated peptides were found to be caused only by the unmodified fraction of FliG molecules. A) In this strategy for the *in vitro* assembly of a FliG polymer, a $FliF_C$-DNA conjugate is used as an intermediate layer between the DNA template and the FliG protein. $FliF_C$-FliG binding recruits the protein to the structure. **B)** Dye-labelled FliG was added to various peptide-DNA nanostructures and a band shift was observed in the DNA channel. No corresponding shift was seen in the fluorescent signal of the FliG-Dye channel however with the bands appearing as non-interacting monomers. **C)** To prove that the shift only occurred upon addition of FliG, either FliG, FliM-FliN, or both proteins were added to peptide-DNA constructs. Only in lanes where FliG was present did a band-shift occur (*). SDS-PAGE was also conducted to confirm the presence of the added proteins (bottom gel). **D)** Staining for all proteins on a native polyacrylamide gel showed that, apart from the wild-type, the FliG constructs were quite heterogeneous. Bands from cysteine-cysteine dimers and unlabelled FliG implied that conjugation to the dye or DNA molecules had been incomplete. **E)** Adding excess wild-type FliG to the $FliF_C$-DNA conjugate caused a band-shift on a native polyacrylamide gel. The original FliG band (º) was shifted downwards (*) upon complex formation due to the negative charges of the DNA tag.

conjugated to a cysteine engineered near the N-terminus using the maleimide-thiol reaction. After a period of incubation at room temperature, I ran these samples on native polyacrylamide gels, organizing the lanes such that constructs containing a particular template were run together, with samples containing or omitting FliG run side-by-side. This would allow me to very clearly determine if the addition of FliG made any difference to the migration of the templated peptide constructs.

Visualizing the results showed very clearly that samples with and without FliG ran very differently on the gel (Fig. 4.7B). Whereas the lanes with FliG omitted ran as they had previously, showing the single prominent band containing the peptide-DNA construct determined by the sequence of the template, each of the lanes with FliG added produced higher-molecular-weight, DNA-containing bands visible with the SYBR Gold stain. In each case the band was shifted upwards from that of the peptide-DNA constructs alone, and appeared slightly less well-defined, indicating perhaps several nearly-identical species running together or some degree of instability when run on the gel. In addition, there appeared to be a general trend of an increase in the height of this upper band with more and more binding sites present on template, a promising observation should these in fact be FliG-$FliF_C$-DNA constructs. One puzzling and confusing fact stood in the way of me making this assertion however.

Visualizing the fluorescent signal from the dye on the FliG construct showed only a single, monomeric band in each of the lanes with FliG present. In no instance was the fluorescent signal from the dye seen to overlap with the higher-molecular-weight bands (Fig. 4.7B). Thinking the signal was perhaps just too weak to see, I tried the same experiment again with a greater excess of FliG and obtained the same results, while analysis of the intensities of the bands showed that the signal from the dye on FliG was the same whether the protein was run on its own or with the peptide-DNA constructs. These observations were difficult to explain as it appeared that the FliG molecule was somehow affecting the migration of the peptide-DNA structures without being affected in any way itself.

Though puzzling, I came up with a few possible theories to test. First I tried to work out if the FliG was in fact binding to the templated peptides but the structures were falling apart during the running of the gel. Running the samples chilled and at a low voltage showed no changes in the migration of the bands, and the fact that the monomeric FliG-dye band was in no way slowed in the presence of the peptides as compared to the FliG-dye run on its own indicated that the species observed were most likely stable during their movement through the gel's polymer mesh. Thus, gel-induced dissociation could not explain the patterns of the bands. I next tested the possibility that the binding of the FliG to the peptide in some way quenched the fluorescence of the dye on the protein. The dye was situated at the N-terminus of FliG, very near the binding site for $FliF_C$ and it was possible that a change in the dye's local environment following binding could cause a decrease in fluorescence. Monitoring the samples in real time on a fluorimeter as I added peptide or DNA-templated peptides to dye-labelled FliG showed that there was in fact no decrease in fluorescence with the addition of $FliF_C$ meaning that this theory could not explain the lack of FliG-dye signal in the high-molecular-weight bands.

One quite plausible hypothesis was that something other than FliG which was present in the FliG aliquots was responsible for the shift. Though I had been careful to add an equivalent volume of either my protein storage buffer or FliG to the pairs of samples, perhaps some small molecules left over from the purification procedure which were not present in my storage buffer were causing the shifts, or perhaps the mere presence of volume-occupying protein molecules was changing physical parameters of the solution such as its viscosity and altering how it ran on gels. To test this, I thawed aliquots of the FliM-FliN protein complex which I had purified alongside the dye-labelled FliG (as will be described in the next section). In theory these two species would have the same buffer composition and concentrations could be adjusted to obtain the same protein density in both the FliG and FliM-FliN solutions. Thus if a shift was seen upon addition of FliM-FliN, proteins which have never been reported to interact with $FliF_C$, a small molecule in solution or the solution's viscosity would be the most likely explanation for the shifted bands.

Adding FliG, FliM-FliN, or both FliG and FliM-FliN to either T1d4- or T5N10-templated peptides showed quite conclusively however that an unmonitored component in the protein aliquots was not responsible for the previous observations. Addition of FliM-FliN to the templated peptides showed no shifts in the observed bands (Fig. 4.7C). Only when FliG was present in solution did the shifted upper bands appear. Thus it seemed that it was FliG and only FliG that was responsible for the high-molecular-weight species, but why no fluorescent signal from the dye could be observed in the shifted bands was still a mystery.

Around this time, I observed that, contrary to popular belief, the SYPRO Ruby protein stain could be used on native polyacrylamide gels. Previously, I and others had assumed that the denaturant SDS was needed in the buffers in order for the dye to bind properly to the proteins, but several trials had confirmed to me that, though weak, a SYPRO Ruby signal could be obtained for proteins on an SDS-free native gel. I therefore decided to run a set of the FliG constructs that I had made and purified on a native gel and to stain with SYPRO Ruby in order to obtain more information about the state of the proteins in solution. The bands observed were quite informative and provided me with a possible explanation for what was going on in the gels with the templated peptides.

Running wild-type, single-cysteine, two variants of the dye-labelled, and DNA-conjugated FliG on a native gel and staining for total protein with SYPRO Ruby showed that the constructs were more heterogeneous than I had previously assumed (Fig. 4.7D). Although the wild-type FliG ran as a single, monomeric band, all others showed multiple bands above the monomer. Disulfide bridges between cysteines could explain the bands for the single-cysteine construct, whereas the presence of similar bands in the dye- or DNA-conjugated constructs indicated that the labelling of their single cysteines had been incomplete and some fraction of the protein had been left without a dye or a DNA strand and was therefore invisible on the gels unless a protein stain was applied.

This meant that when I had added N-terminally dye-labelled FliG to the templated peptides, I had actually added a substantial amount of unlabelled, single-cysteine FliG alongside the dye-labelled protein. Thinking through the consequences of this I realized that I now had an explanation for

why I had observed a shift in the peptide-DNA band upon addition of FliG but saw no accompanying shift in the dye-labelled FliG band. It appeared to me that only the unlabelled (and therefore invisible) fraction of the FliG protein was interacting with the $FliF_C$ peptides while the dye-labelled fraction did not. The dye had been placed at the N-terminus, very close to the $FliF_C$ binding site, and the conjugation could have caused some degree of local denaturation preventing proper binding site formation. Alternatively, a substantially-sized hydrophobic dye near the binding site could sterically block $FliF_C$ and prevent the interaction. Either way, I was quite convinced that I had finally hit upon an explanation and was eager to now obtain definitive gel-shift evidence of a FliG-$FliF_C$ interaction.

To do so, I chose to work with the wild-type FliG construct. This construct appeared the most homogeneous on the SYPRO-Ruby-stained native gel and was also the one that I had used to obtain evidence of binding through the tryptophan fluorescence and thermal shift assays and which appeared most α-helical with CD spectroscopy. I therefore added excess wild-type FliG to the $FliF_C$-DNA conjugate, incubated at room temperature, and then ran the sample on a native polyacrylamide gel alongside the two components on their own. Staining with SYBR Gold for DNA and then SYPRO Ruby for protein produced a definitive result. Combination of the two components led to the formation of a clear band showing both DNA and protein signal, indicating that it contained both the $FliF_C$-DNA conjugate and the FliG protein and that in all likelihood these two were bound to one another (Fig. 4.7E) (the peptide on its own was too small to be picked out by the protein stain). The band was shifted relative to the bands of both FliG and $FliF_C$-DNA on their own, running slower than the $FliF_C$-DNA band, as explained by an increase in weight, and slightly faster than the FliG band, as explained by an increase in negative charge that the DNA tag would provide.

With this gel-shift evidence I was finally convinced that the $FliF_C$ peptide did in fact bind to my purified FliG protein. This was a big step for the project and a small, yet significant contribution to the field of BFM study. Though a similar conclusion had been drawn for thermophilic bacterial proteins[90,91], using a completely different method **I had now shown for the first time that the 39**

C-terminal amino acids of *Salmonella* FliF were sufficient to bind to *Salmonella* FliG. It had taken some time to demonstrate this interaction, with gel-based puzzles to solve en route, but I was confident that I could now use it in the construction of larger assemblies and hence set out to build and characterize structures made from $FliF_C$, FliG, and DNA templates.

Native PAGE Demonstrated One-to-One Binding Between FliG and DNA-Templated $FliF_C$

Now that I had demonstrated binding of FliG proteins to free-floating peptide-DNA conjugates, the next step was to look for binding of the protein to larger assemblies of DNA-templated peptides. I therefore prepared mixes of the conjugates and six different linear templates as I had done previously, with gels showing the conjugates had bound to the appropriate sites on the templates. My hope was that I could use these templated peptides to search for signatures of FliG domain swap or at the very least for some evidence of FliG-FliG interactions. Cooperative binding of FliG to the templated peptides would be an indication that FliG-FliG interactions were taking place, while gap-filling or runaway polymerization by the FliG proteins would point towards domain swap between neighbours. In all scenarios, FliG-FliG interactions would result in more FliG bound to the template than would be predicted if FliG solely bound to the peptides, and I therefore was searching for any indication that the number of proteins seen on the template was higher than that predicted given a particular stoichiometry and assuming one-to-one binding of FliG to $FliF_C$.

To begin, I added wild-type FliG to peptide-DNA conjugates in a one-to-one ratio, with some mixes containing free-floating conjugate and others containing conjugates bound to a particular template. Running these on a native polyacrylamide gel with lanes alternating between samples with or without FliG, I was able to clearly see upwardly-shifted bands only in the lanes containing FliG (Fig. 4.8A). Moreover, these shifted bands did not all run to the same height but formed a ladder of heavier and heavier species as the number of binding sites on the template increased. This was a very promising sign and indicated to me that FliG was binding to the templated peptides and forming DNA-$FliF_C$-FliG complexes whose size was determined by the number of peptides bound to the DNA template.

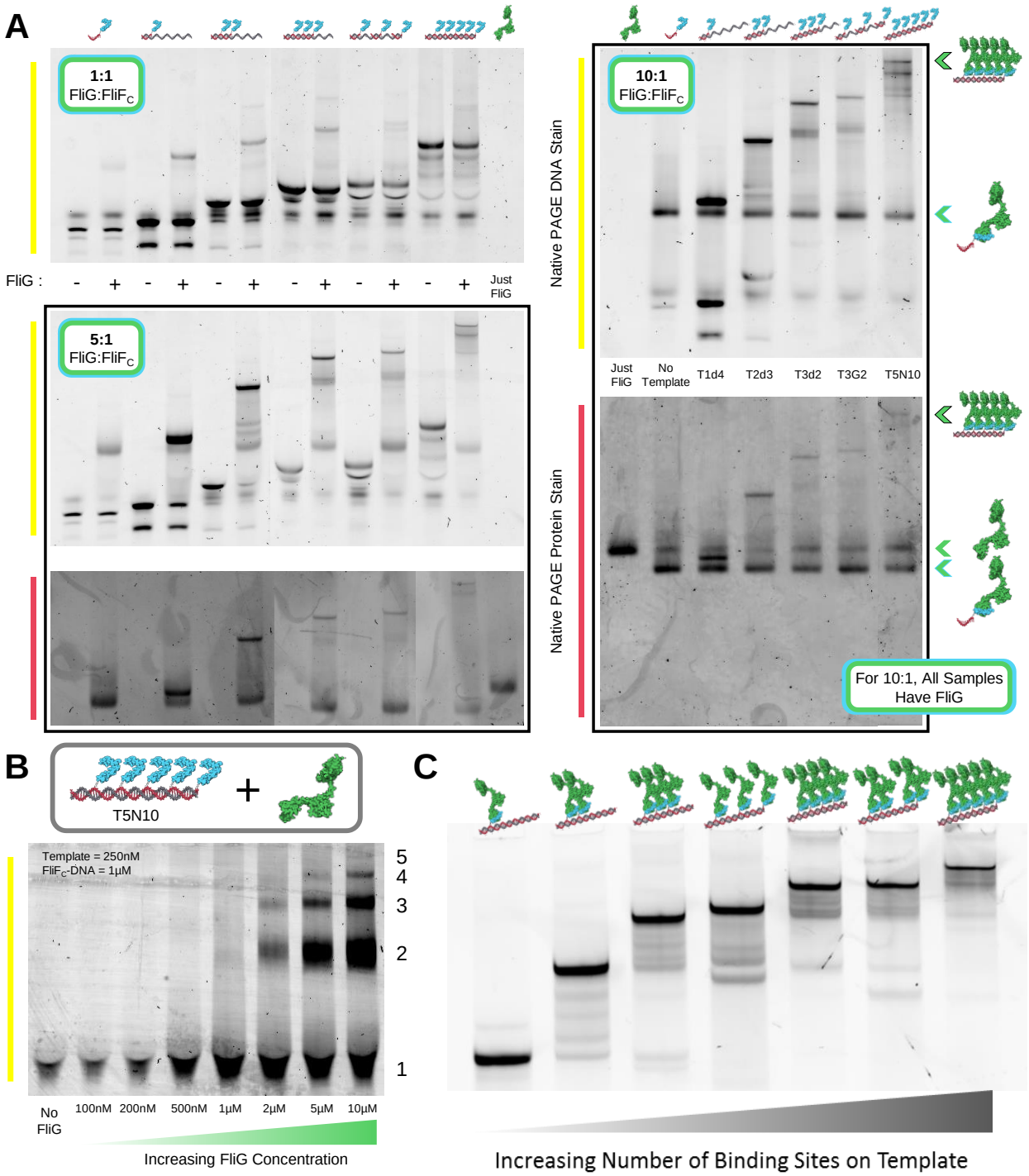


Figure 4.8 | Wild-type FliG binds to DNA-templated peptides in a one-to-one fashion. A) Wild-type FliG was added to both the free-floating $FliF_C$-DNA conjugates and a series of $FliF_C$-template constructs at various stoichiometries. At a 1:1 $FliF_C$-FliG stoichiometry, faint upwardly-shifted bands were observed upon the addition of FliG. These upper bands increased in intensity with a 5:1 and then 10:1 FliG:$FliF_C$ ratio. Staining for protein with SYPRO Ruby showed that these upper bands contained FliG and hence were FliG-$FliF_C$-DNA complexes. **B)** Adding increasing amounts of the FliG protein to a T5N10-peptide construct enabled counting of the sub-stoichiometric bands and confirmed that five FliG proteins were binding to the five peptides in the construct. **C)** Tests for cooperative binding of FliG to the peptide-DNA structures looked to see if constructs with side-by-side binding sites recruited FliG better than those with single or spread out binding sites. Seven different templates were used and solutions mixed keeping the same concentration of FliG, $FliF_C$, and binding sites in each sample, the only difference being how the binding sites were distributed among template strands (for example five sites on five T1d4 template strands or a single T5N10 template strand). No cooperativity was visualized however as all the templates reached a fully occupied state at the same concentration of FliG.

To confirm this however, I needed to obtain some evidence that FliG was present in these upwardly-shifted bands. I therefore stained the gel for total protein using SYPRO Ruby, with the knowledge that the peptide on its own was too small to be picked up by the stain, implying that any bands observed could be attributed to the presence of the FliG protein. The results however were inconclusive. SYPRO Ruby stains native gels less effectively than denaturing, SDS-containing gels and with my one-to-one ratio of FliG to $FliF_C$, it appeared that the protein concentration was too low for the stain to pick it up. Nevertheless, some faint indications of possible bands could just barely be deciphered from the images of the gels making me think that the staining would work given a higher concentration of FliG.

I therefore repeated the experiment but added FliG in a five-to-one ratio to the $FliF_C$ peptide. The resulting gels showed an even clearer ladder of upwardly-shifted bands formed from FliG-containing samples with their migration pattern mirroring the migration pattern of the lower ladder of bands formed by the set of different templated-peptide constructs (Fig. 4.8A). Increasing the FliG-$FliF_C$ ratio to ten FliG molecules per $FliF_C$ peptide showed a yet clearer ladder and hence gave me confidence that these upwardly-shifted bands were **DNA-$FliF_C$-FliG assemblies**. Using SYPRO Ruby, I stained the gels containing the samples with five-to-one and ten-to-one ratios of FliG to $FliF_C$ and clearly saw signal from the FliG protein in the upwardly-shifted bands, thereby confirming my prediction that these were the result of FliG binding to the peptide-DNA structures (Fig. 4.8A).

What I did not see however was any evidence of extra FliG molecules binding to the peptide-DNA complexes. From the shape of the ladder on the gel, it appeared that given, for example, five peptides bound to the DNA template, five FliG molecules and no more were binding to the assembly, even with a great deal of excess FliG in solution. To confirm that this was the case, I performed a titration of FliG against a T5N10-templated peptide construct. As the concentration of FliG was increased, higher and higher bands formed until a fully-filled template was reached (Fig. 4.8B). Counting these sub-stoichiometric bands confirmed that there were indeed five FliG

molecules bound to the template and even with a great excess of FliG molecules in solution, no more than five could be seen to bind.

Therefore, due to the absence of their immediately-obvious signatures on my gels, it appeared that there was no gap-filling or runaway polymerization present in my *in vitro* system and hence with this assembly strategy, no evidence of the domain-swap model for FliG polymerization could be found. There was still a chance however that subtler signatures of FliG-FliG interactions could be observed given the appropriate experiment so I decided to determine if FliG was cooperatively binding to the templated peptides.

If FliG-FliG interactions were occurring during the assembly process, it would in theory be easier for a FliG protein to bind to a site on the template which had an adjacent site with a FliG protein already bound to it. The neighbouring FliG protein would bind to the incoming FliG molecule, hold it in place, and thereby increase the likelihood that it would bind to the peptide at the binding site. Therefore, if one were to have two solutions with the overall concentration of peptides and binding sites the same in each but one with five single-site templates and the other with a single five-site template, cooperative FliG binding would be predicted to produce, at a certain concentration of the protein, a fully-filled five-site template and unfilled single-site templates. Similarly, it would be expected that a template with three FliG-binding sites spaced out from one another (T3G2) would need a higher concentration of FliG to fill its sites than a template with three closely-spaced sites (T3d2). With these predictions and an established method in native PAGE by which the number of occupied sites could easily be determined, I thus had a means by which I could investigate the possibility of cooperative FliG recruitment and, by extension, the presence of FliG-FliG interactions on the template.

Using fluorophore-labelled DNA templates in order to avoid any errors in band intensities due to spurious SYBR Gold staining, I prepared mixes of templates, peptide-DNA conjugates, and wild-type FliG, keeping the concentration of binding sites, peptide-DNA, and FliG constant across trials and changing only the identity of the DNA template. Running these mixes on a gel, I searched for any evidence that templates with side-by-side FliG binding sites recruited the protein more readily

that those with single or spaced out binding sites. Analyzing the bands however showed that there was no clear difference between how well each of the templates recruited FliG, with all of the structures reaching full filling of their binding sites at approximately the same concentrations and stoichiometries of components (Fig. 4.8C).

Thus, even with this new assembly strategy, incorporating the $FliF_C$ peptide and hence a more natural binding platform for the FliG protein, it appeared that FliG-FliG interactions still could not be detected. At this point the questions described at the end of Chapter 3 which originally led to my adoption of this new assembly strategy resurfaced once again. Is there something wrong with my purified FliG protein? Does the template need to more closely mimic the C-ring? Is something more needed in my *in vitro* system?

Due to conclusions draw from the work presented in this chapter, I now had a more complete picture of the state of my FliG constructs and believed that at least the wild-type protein was properly folded. Its α-helical nature, sharp unfolding transition, and most importantly, the fact that it stably bound to the $FliF_C$ peptide indicated to me that the protein was most likely in its native state and hence misfolding was probably not to blame for the lack of FliG-FliG interactions. The other two thoughts, that a new template design was needed or that something additional was required in the system, were options that I had to explore however and though new templates are left for Chapter 5, the addition of one particular component forms the basis for the following section, the final one of this chapter.

4.6 Expression and Purification of the FliM-FliN Complex

Following the work presented in Chapter 3 in which I had failed to observe any FliG-FliG interactions, I had searched through the literature to see if there was any component that I might add to the system which would promote or stimulate the formation of intermolecular FliG bonds. Structural studies with the $FliF_C$ peptide had encouraged me to pursue the incorporation of this

component into my *in vitro* system and the results that I have presented so far in this chapter tell the story of my work towards that goal.

In parallel to the work with the peptide however, I also sought to investigate the addition of another component into my experiments. One recent publication had shown via *in vivo* cross-linking that the protein **FliM** might induce FliG-FliG interactions[104]. This seemed quite plausible as FliM has been shown to bind directly to FliG in the C-ring and it would be relatively straightforward for this binding to induce a conformational change in FliG, opening it up and allowing it to interact with its neighbours, thereby triggering a new phase in C-ring assembly. This seemed an intriguing possibility to investigate, with predictions that could be tested using my DNA templates and purified protein constructs, so I decided that, along with the $FliF_C$ peptide, I would seek to incorporate the FliM protein into my system.

Expression and purification of the FliM protein appeared not to be simple procedures however. FliM is quite unstable on its own and had only been successfully purified as a truncated construct or in a complex with multiple copies of FliN, its other binding partner in the C-ring. In addition, no one had successfully purified *E. coli* FliM either alone or in a complex, meaning that without developing a protocol myself, I could not investigate FliM binding to my *E. coli* FliG constructs. Thankfully however, there was an established procedure for purifying *Salmonella* FliM-FliN complexes and the lab of Lawrence Lee, our collaborator on this project, had both the plasmids and protocol which would allow me to obtain purified *Salmonella* FliM-FliN which I could use in experiments with *Salmonella* FliG.

Therefore, after obtaining the plasmids, I set out to express and purify the *Salmonella* FliM-FliN complex with hopes that I could observe interactions between it and my purified FliG. Beyond this I also wanted to determine if the binding of FliG to FliM would in some way stimulate FliG-FliG interactions, an interesting proposition with ramifications for the C-ring assembly pathway. This section describes my attempts, in the end successful, to obtain purified *Salmonella* FliM-FliN and the experiments I then performed, mixing these proteins with the DNA, FliG, and $FliF_C$ constructs that have been described previously. The end of this section also marks the end of this chapter

and, following the discussion relating to FliM-FliN, I will spend some time highlighting what the work of this chapter has achieved.

No Interactions Could Be Detected Between the Purified FliM-FliN Complex and FliG

The protocol and plasmids for FliM-FliN from our collaborator Lawrence Lee arrived with words of caution that the complex was not as easily purified as FliG and that the proteins had a tendency to aggregate and to stick to the walls of tubes and columns. I therefore followed the prescribed procedure exactly as it was laid out, but after two unsuccessful attempts in which I failed to obtain any protein in my final aliquots, I was left wondering whether the FliM and FliN had actually been expressed and, if so, where they had gone during the purification.

Therefore, during my third attempt at the purification, I took aliquots at each step in the process to find at which point the proteins were being lost. After lysis of the cells, I had spun down to remove cellular debris and then ultracentrifuged to separate the cytosolic from the membrane fractions. Analyzing samples from each of these steps using denaturing SDS-PAGE revealed that there were prominent FliM- and FliN-sized bands present after the lysis which were lost following the ultracentrifugation (Fig. 4.9A). This led me to believe that I was spinning for too long and too fast and that though the FliM-FliN complexes were in fact being expressed, they were being pelleted by the ultracentrifuge and discarded along with the membranes and membrane proteins present in that fraction.

To remedy this issue, I altered the protocol by decreasing the speed and time at which I spun the samples in the ultracentrifuge. The membranes still formed a visible pellet at this slower spin speed, but I was now clearly able to see FliM- and FliN-sized bands present in the supernatant fraction. With the proteins now in solution, I was able to carry out the rest of the purification, pulling out the complex using an IMAC column and the His-tag that had been engineered into FliM. SDS-PAGE of the purified fractions showed three prominent bands corresponding the FliM monomer, the FliN monomer, and what appeared to be the FliN tetramer (Fig. 4.9A). This presence of this

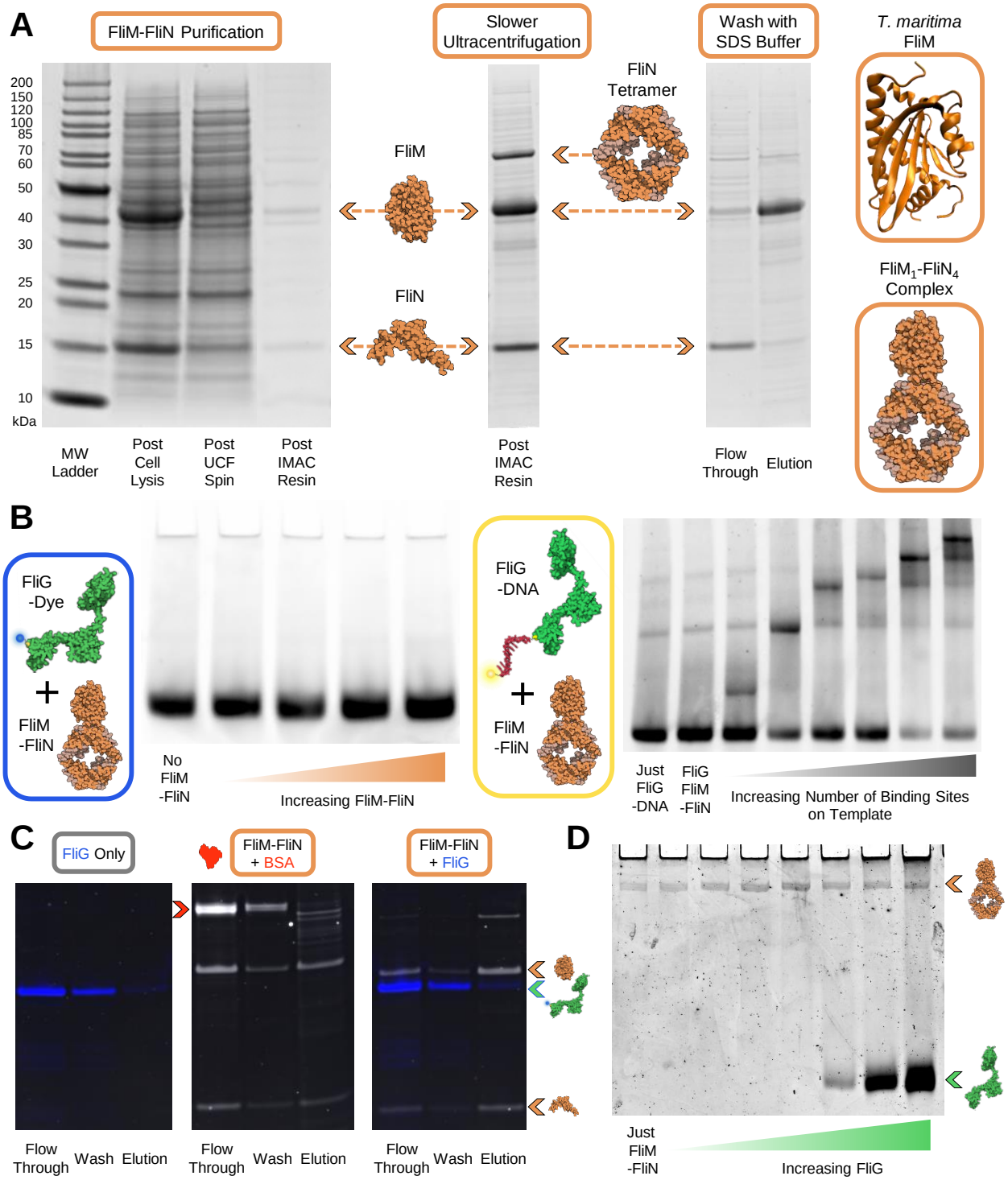


Figure 4.9 | No interactions were observed between purified FliG and the FliM-FliN complex using gel-shift and pull-down assays. A) Initial attempts at purification of FliM-FliN lost the complex during pelleting of the membrane fraction. Slower centrifugation kept the complex in the supernatant. Washing with an SDS-containing buffer showed that FliN was co-purified through its interaction with His-tagged FliM. **B)** Titrating FliM-FliN against both free-floating FliG-dye and template-bound FliG-DNA showed no shifts on a gel. **C)** Passing FliG-dye over resin-bound FliM-FliN and collecting the flow-through, wash, and eluted fractions demonstrated that FliG was not enriched in the eluted fraction as FliN was, indicating that FliG-FliM binding did not occur. Controls with FliG only or BSA proved that non-specific protein-resin or protein-protein interactions were not an issue. **D)** Titrating wild-type FliG against FliM-FliN showed no gel-shifts but demonstrated that the FliM-FliN complex could be visualized on native polyacrylamide gels through SYPRO Ruby staining.

tetramer was not unexpected as FliN monomers have been shown to bind quite strongly to one another in a four-member ring[97], and it was likely that the SDS and heat from the gel was not sufficient to break some of these complexes apart.

Though the molecular weights were just about sufficient to identify the proteins that I had purified as FliM and FliN, I performed one more simple assay to confirm that the FliM band was that of a His-tagged construct and that the presence of the FliN band was due to a non-covalent interaction between FliN and the resin-bound FliM. After washing away unbound components to leave only the FliM-FliN complex on the resin, I then washed through with an SDS-containing buffer and collected the flow-through. Running an aliquot of this flow-through on a gel showed FliN to be the dominant species while running the fraction eluted with imidazole following the SDS wash showed FliM to be the major species there (Fig. 4.9A). This assured me that I had purified the correct construct, pulling out the FliM protein via its His-tag, with the FliN protein co-purifying due to its interaction with FliM.

Taking purified aliquots of the protein complex that I had now assured myself was FliM-FliN, I looked to see whether this complex would interact with FliG by mixing the two together and looking for a band-shift on a native polyacrylamide gel, as I had with FliG and the $FliF_C$ peptide. Due to the fact that it was quite straightforward to visualize, I began by titrating FliM-FliN against a dye-labelled FliG construct and running the mixes on a gel. No shift was observed however as more and more FliM-FliN was added, with the FliG band appearing resolutely monomeric (Fig. 4.9B). Thinking that perhaps FliM-FliN needed some sort of nucleus or cluster of FliG proteins in order for it to bind, I then added the complex to FliG-DNA conjugates which had been organized on various short, linear DNA templates to form the structures described in section 3.3. The bands of these DNA-templated FliG constructs were unmoved upon addition of the FliM-FliN complex however (Fig. 4.9B), indicating that the FliM in the complex was most likely not binding to any of the FliG molecules clustered on the DNA.

One could imagine that the FliM proteins were in fact binding to the FliG proteins but that running the complexes on a gel had broken them apart. The dissociation constant of the interaction

between the two proteins had been measured to be in the micromolar range, making it orders of magnitude weaker than that for FliG-FliF$_C$ binding[95] and thus it was possible that a more sensitive detection method might be needed to pick out the interaction.

I therefore decided to conduct a pull-down assay using the IMAC resin and the His-tagged FliM. With the FliM-FliN complex immobilized on the resin, I would pass a solution of dye-labelled FliG over the resin-bound complexes and collect samples at various points in the washing and then eluting processes, searching for indications that FliG was retained in FliM-containing fractions due to a FliG-FliM interaction. I could then run the samples on a denaturing gel, not having to worry about preserving FliG-FliM bonds, as the mere presence of FliG in an eluted fraction alongside FliM would indicate that the two had bound previously while on the resin.

To control for other confounding interactions which would obscure the results, I set up two tests in parallel to my main trial. The first simply passed FliG through an IMAC resin which had no proteins bound to it. This would allow me to determine the degree to which the FliG protein bound to the resin itself rather than to the FliM protein. It was possible that the His-tags of the FliG construct had not been completely cleaved and that some portion of the constructs would interact with the resin and thus it was important to have this control for comparison to the levels of interaction seen in the main trial. The second control used resin with FliM-FliN bound to it but passed through the BSA protein rather than FliG. BSA does not specifically interact with either FliM or FliN and as such this control would allow me to roughly determine the levels of non-specific protein-protein interactions which might retain proteins on the resin.

Conducting these two controls alongside my main trial, I collected three samples from each and ran them on a denaturing gel. The first sample was from the supernatant obtained after the components had been mixed and incubated and the resin had been pelleted, while the second was taken following the removal of the initial supernatant and a wash with fresh buffer. Proteins in these two fractions would constitute those which had not bound or had bound very weakly to the resin. The third sample collected was taken following incubation of the resin in an imidazole-

containing buffer with its contents therefore constituting the components which had been retained on the resin alongside the His-tagged FliM.

Analyzing the resulting gel showed that the majority of the FliG proteins were to be found in the initial unbound sample, whereas most of the FliM and FliN was to be found in the resin-bound fraction (Fig. 4.9C). The two controls confirmed that there was very little interaction between FliG and the resin itself and that the FliM-FliN complex did not engage in non-specific protein-protein interactions with BSA to any appreciable extent. As FliG was not enriched in the eluted fraction, the pull-down assay confirmed to me that FliG-FliM binding was not occurring in this system at a detectable level and if FliM-FliN was to be effectively incorporated into my *in vitro* system, the constructs and components needed to be re-evaluated.

It was around this time that I engaged in the biophysical characterization of my various BFM protein constructs as described in section 4.2. This demonstrated to me that the FliG proteins I had purified and manipulated varied in their overall stability, α-helicity, and most likely in how well they were folded. As I had used dye-labelled or DNA-labelled FliG constructs in my previous attempts to find evidence of a FliG-FliM interaction, it was possible that misfolding in these proteins had interfered with their binding to FliM. I therefore decided to mix wild-type FliG with FliM-FliN and look for band shifts on a native gel stained with SYPRO Ruby.

Titrating wild-type FliG against FliM-FliN from sub-stoichiometric concentrations until FliG was in a ten-to-one excess and running the mixes on a polyacrylamide gel produced no evidence that the two constructs interacted. FliG remained monomeric and no shift was seen across the lanes of the titration, indicating that even with the well-folded wild-type construct, FliG-FliM binding could not be observed (Fig. 4.9D). One key observation from this gel however was that, using the SYPRO Ruby stain, the FliM-FliN complex could be visualized as a slow-running band near the wells of the gel. Though the binding of FliM to FliN had been inferred indirectly due to their co-purification using IMAC resin, this native gel was a reassuring confirmation that they did indeed form a complex.

Thus, after attempts with several different FliG constructs using both gel-shift and pull-down assays, I had yet to observe binding of FliM to FliG. In parallel to this work, I had been engaged in the experiments with the $FliF_C$ peptide presented earlier in this chapter and as these began to produce results, I shifted my focus away from FliM-FliN. Though I could not explain exactly why I had failed to observe a FliG-FliM interaction, I assumed that my assays were perhaps not sensitive enough to provide evidence of the relatively weak binding. Perhaps performing the pull-down assay using wild-type rather than dye-labelled FliG would have produced different results. Another possibility was that the FliM-FliN construct was denatured or improperly folded in some way. Though co-purification and my final native gel had shown that a complex had formed between the two proteins, perhaps the stoichiometry of FliM and FliN in the complex (thought by most to be one-to-four) was off and, because of this, it had taken on a faulty architecture to which FliG did not bind. The fact that it had a tendency to pellet with the membrane fraction during purification also signalled to me that it was perhaps being expressed at too high a level and that further work with the expression and purification was needed.

In the end I did not return to work with FliM-FliN, though experiments with the complex were far from exhausted. My main objective was still to obtain evidence of FliG-FliG interactions, and though I had not had success with this particular addition to my *in vitro* system, there were other strategies which I was eager to pursue relating to new designs of the DNA template which might encourage interactions between neighbouring FliG proteins. The design and implementation of these new templates constitutes the work presented in Chapter 5, my final chapter of results, but before moving on it is worth now briefly summarizing the highlights of this chapter.

Following unsuccessful attempts using FliG-DNA conjugates to observe FliG-FliG interactions (as described in Chapter 3), I decided to incorporate a new layer into my *in vitro* system: the $FliF_C$ peptide. Recent studies had indicated that this short cytoplasmic region of the FliF protein was necessary and sufficient for binding to FliG in thermophilic bacteria and that binding to $FliF_C$ might encourage FliG polymerization. I therefore designed and ordered a synthetic peptide corresponding the C-terminal region of *E. coli* and *Salmonella* FliF and began experiments to

determine if the peptide bound to my purified FliG proteins (4.1). I had concerns about how well-folded my FliG constructs were however, so I conducted several biophysical assays to obtain information about the structure and stability of my proteins (4.2). CD spectroscopy showed that the degree of α-helicity varied between constructs and that the wild-type exhibited the highest percentage of α-helices, having an overall composition which closely matched that predicted from FliG crystal structures. Using a tryptophan fluorescence assay with this wild-type construct and the $FliF_C$ peptide, I was able to demonstrate one-to-one binding of the two species and then showed using a thermal shift assay that the FliG-$FliF_C$ complex had an increased stability as compared to FliG alone.

Confident that the peptide could bind FliG, I then worked to incorporate it into a new, 'more natural' FliG assembly strategy in which a DNA template would organize $FliF_C$ peptides into architectures with various stoichiometries and spacings, creating different binding platforms for FliG. I first conjugated a single-cysteine version of the peptide to a maleimide-modified DNA strand (4.3) and showed using native polyacrylamide gels that I could arrange these conjugates on short, linear DNA templates (4.4). Incorporating FliG into the system, I was able to demonstrate that the wild-type construct stably bound to the DNA-templated peptides in a one-to-one fashion but could not find any evidence of the assembled FliG proteins interacting with one another (4.5). In an effort to promote FliG-FliG binding, I had also purified the FliM-FliN complex from *Salmonella*, as recent work had suggested that FliG-FliM binding might trigger FliG oligomerization. I was unable to obtain any indication of a FliG-FliM interaction however and addition of the FliM-FliN complex did not obviously trigger FliG-FliG binding (4.6).

Thus, my search for proof of the FliG domain-swap hypothesis continued. The work of this chapter had reassured me of the well-folded state of at least the wild-type FliG construct, and had shown that $FliF_C$ and FliM-FliN, the two most promising additives to my system, had little effect on the overall oligomeric state of FliG. Although there were a great number of other additions and changes I could make to FliG's local environment, no particular additive or perturbation stood out to me as an obvious choice which would result in FliG-FliG binding or FliG-ring formation. There

was one easily-designable and re-configurable component in my *in vitro* system however which begged to be explored. I had done very little with the underlying DNA template up to this point, mostly using short, five-site templates as a way to increase the local concentration of FliG. I therefore decided to pursue the construction of a more C-ring-like template and Chapter 5, my final chapter of results, describes the design and implementation of a DNA nanostructure which fit that bill. With this nanostructure I was able to achieve *in vitro* FliG assemblies which mimicked the *in vivo* organization of the C-ring closer than any of my constructs had previously, and as such, in many ways, this final set of experiments and results constitutes the apex of my work.

CHAPTER FIVE

RING BUILDING

Design and Implementation of a C-Ring-Mimicking DNA Nanostructure

Thus far in my work, as compared to time spent conjugating various components to one another and characterizing different protein constructs, I had done relatively little in the way of designing and building new DNA structures upon which I could assemble FliG. The ease with which DNA can be manipulated and its versatile nature as a nanoscale building material were some of the main benefits to the *in vitro* approach taken by this project and up to this point little had been done to take advantage of the designability of the DNA template. For the most part, I had used short, linear DNA strands based off of a simple stretch of five FliG binding sites, and based on the results I had obtained, it appeared that these linear templates did little more than increase the local concentration of the FliG molecules. There was a case to be made then for the introduction of larger and more complicated DNA nanostructures which might more accurately capture the dimensions and local environments that were to be found in the cytoplasmic ring *in vivo*, and hence might in some way encourage FliG-FliG binding.

One key argument for moving away from short, linear templates stemmed from predictions based off of the domain-swap hypothesis. A linear arrangement of FliG molecules will have, by definition, two free ends, with the terminal FliG molecules at these ends having only a single neighbour (Fig. 5.1A). Should the ARM motifs of the linearly-arranged FliG molecules be bound to one another in an intermolecular fashion, the proteins at the ends of this arrangement, having only a single binding partner, would be left with an unpaired ARM motif, its large hydrophobic surface free and exposed to the aqueous environment. These exposed surfaces would be energetically

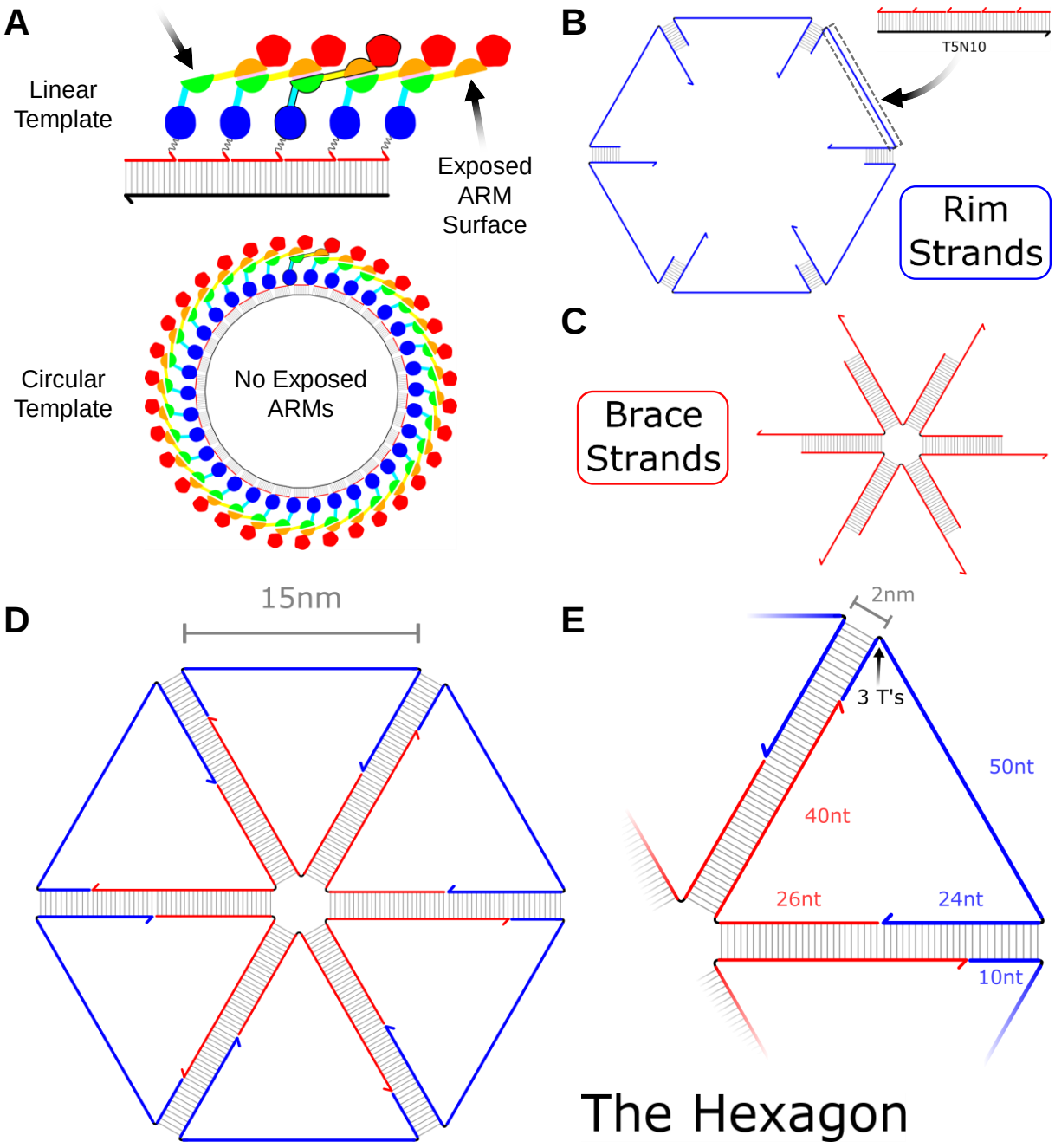


Figure 5.1 | The Hexagon DNA nanostructure is roughly 35nm in diameter and incorporates 30 binding sites for FliF$_C$ and FliG. A) On linear DNA templates such as the T5N10 template shown, FliG molecules at the ends of the structure would each have only one neighbour. This means that if the proteins took their extended, oligomeric form, the terminal molecules would have the hydrophobic patches of one of their ARM motifs exposed to the aqueous environment, an energetically unfavourable state. A circular template by contrast would provide every FliG molecule with two neighbours, allowing all ARM motifs to be hidden. **B)** Using the T5N10 template as a building block, the simplest circular template would link six of these five-site domains together into a ring of 30 binding sites. This head-to-tail linkage strategy was used in the Hexagon nanostructure to bind the rim strands to one another. **C)** To add a degree of rigidity to the Hexagon, six brace strands were incorporated into the design. These formed the centre of the structure, holding the rims in place. **D)** Assembling the 12 unique strands together would produce a hexagonal structure with roughly 15nm-long sides and 30 binding sites around its edges. **E)** The 'spokes' of the Hexagon were 50 nucleotides long, made up of unique domains from both the brace and rim strands. Three thymine bases were placed at the junctions of the structure to provide a degree of flexibility. The T5N10 domains of the rim strands were separated by 2nm gaps, the width of a double helix. This gap could in theory be bridged or filled by extra FliG molecules.

unfavourable, and the domain-swap hypothesis would predict that in a linear arrangement, barring runaway polymerization or spontaneous ring formation, the FliG monomers would most likely be found in the closed state, with their two ARM motifs bound to one another intramolecularly.

In their natural context however, the FliG proteins would assemble on a circular template in the form of the ring of membrane-bound FliF proteins. Such a template would provide each FliG molecule with two neighbouring proteins and, should intermolecular ARM-ARM interactions take place, the circular nature of the template would give the assembled proteins the chance to close up into an oligomeric ring, thereby hiding away all of the hydrophobic surfaces of the ARM motifs through intermolecular interactions (Fig. 5.1A).

In theory then, it was possible that the reason I had not observed any FliG-FliG interactions up until this point was that my short, linear templates favoured the closed, compact conformation of the FliG monomers, and that more than this simple five-protein nucleus was needed to stimulate the formation of a FliG-ring. Perhaps what I required was a more C-ring-like template, providing greater aid to the FliG monomers during ring formation as they transitioned from their closed, cytoplasmic state to the interlocked, oligomeric form found in the C-ring.

This seemed a reasonable possibility and would explain why even with the addition of $FliF_C$ and the FliM-FliN complex into the system, I had not observed any FliG-FliG interactions. Therefore, with my time on the project drawing to a close, I resolved to design and implement a C-ring-like DNA nanostructure, with the hopes that a circular DNA template would provide the appropriate impetus to encourage FliG oligomerization. This chapter describes both the design process and the characterization of this DNA nanostructure before moving on to the templating of both the $FliF_C$ peptide and the FliG protein on its circularly-arranged binding sites. It concludes with several images obtained using AFM microscopy which constitute the closest this work comes to the construction of a C-ring *in vitro*. Hence this chapter forms a fitting end to the presentation of my results as a whole.

5.1 Design and Characterization of the Hexagon DNA Nanostructure

For the sake of both efficiency and simplicity, in designing a circular, ring-like template, I looked to incorporate into the structure the same sequences and 10-nucleotide binding sites that I had characterized and successfully used to template proteins and peptides on the linear, five-site strands. I had taken a similar approach when using the DNA origami tile (see section 3.4), as for that structure I had merely bound six T5N10 strands via one of their ends to extended staples projecting from the tile's surface. This technique did not prove successful however, and in a sense it was not surprising. Binding the five-site templates to the tile only at one end created little more than an aggregation of six linear templates, with nothing apart from an increase in the local concentration of the FliG molecules to encourage ring formation. To really stimulate circularization, a truly circular template had to be built where each FliG binding site was nestled between two other FliG binding sites.

Techniques for creating circular DNA structures with the dimensions seen in the C-ring had been published, but they tended to use a great number of unique DNA strands[176], or required the painstaking and low-yield process of ligating strands together[144]. I wished to incorporate the same 10-nucleotide binding sites I had already worked with, and therefore aimed to produce a simple design which could organize those binding sites in a roughly circular manner without ligation or a great number of accessory strands which I would have to design and characterize. Extending the concept introduced in the origami tile of placing five-site linear DNA templates beside one another, I decided to pursue a design based off the head-to-tail linkage of strands bearing five FliG binding sites. Designing unique sequences for each of the domains which connected the five-site linear templates would allow me to control how many of the strands linked up and with linking domains at both of their ends, I could design the sequences to ensure that the strands formed a circular assembly.

This section explains the design of the C-ring-mimicking DNA nanostructure that I created and goes through the experiments that were undertaken in order to characterize the structure and determine whether or not it had assembled correctly. The design is simple but it proved to be quite

straightforward to characterize and, as the final section shows, quite effective in organizing $FliF_C$ and FliG.

The Hexagon DNA Nanostructure is 35nm in Diameter and Presents 30 FliG Binding Sites

To mimic the dimensions and stoichiometries of the membrane-embedded ring of FliF proteins to which FliG binds during the assembly of the BFM, I was looking to design a radially-symmetric structure roughly 30 to 50nm in diameter which possessed from 26 to 34 binding sites for the FliG protein. Taking as my building blocks the 50-nucleotide sequences of the T5N10 templates that I had used previously, I therefore sought to arrange these regions, approximately 15nm in length, containing five FliG binding sites, into a C-ring-like formation. The most straightforward way of doing this was to attach the five-site regions to one another in a head-to-tail fashion, making a chain of binding sites which could then loop back on itself to form a circular assembly. To match the specifications of the FliF ring, I decided to build a structure containing six of these five-site regions connected via linking domains into a circle roughly 35nm in diameter and possessing 30 binding sites. The decision to use six five-site domains meant that the structure I began to design possessed six-fold symmetry. Therefore, I took to calling my design **the Hexagon** and in the work to follow I will use this name to refer to the nanostructure I created.

Although the most basic way of obtaining 30 side-by-side binding sites would be simply to bind six strands with the T5N10 sequence to one another in a circular fashion (Fig. 5.1B), I was doubtful of whether such a flexible, predominantly single-stranded template would be able to effectively capture the constraints imposed by the FliF ring *in vivo*. Therefore, in designing the structure, I decided to incorporate six more strands, bringing the total to 12, which could reinforce or brace the binding-site-containing strands and thereby add a degree of rigidity to the assembly. These supporting strands, found at the centre of the structure, I termed the **brace strands** (Fig. 5.1C), while the strands bearing the FliG binding sites were arranged around the outside and called the **rim strands** (Fig. 5.1B). When arranged with the braces in the middle and the rims around the

outside, the twelve strands together resembled a wagon wheel with six spokes radiating from the centre towards the outer edges (Fig. 5.1D).

Using this overall geometry, I then determined the lengths of various domains and the sequences of the strands with the aid of software which could predict the secondary structures and complexes that would form between several input strands in a solution at a given temperature and ionic strength[177]. Although I wished for the structure to be as robust as possible, the length of the domains connecting the strands, and hence the stability of the duplex which would form, was constrained by both the dimensions of the assembly and the length of the oligonucleotides which could be cheaply synthesized by the commercial DNA supplier. This meant that only ten nucleotides joined each rim strand to one another while the braces were quite stably connected by 26-nucleotide-long duplexes. When all 12 strands were pooled in one pot however, the combination of brace-brace, rim-rim, and brace-rim duplexes were enough to hold the Hexagon together even at 37°C, the highest temperature I would feasibly go to in any experiments involving FliG.

Therefore, it seemed that I had a reasonably stable design for templating FliG. Each brace strand was unique but possessed three domains: two for interacting with adjacent brace strands and one for binding to a rim strand. Moreover, the six unique rim strands all had four domains: two for contacting adjacent rim strands, one for binding to a brace strand, and a common T5N10 domain at their centre for binding five of the 10-nucleotide-long strands conjugated to $FliF_C$ peptides (Fig. 5.1E). At certain junction points of the structure I also added a stretch of three unpaired thymine bases, thereby incorporating a degree of flexibility to the joints and ensuring that steric clashes would not hamper the assembly of the Hexagon. This flexibility meant the structure could bend and twist and even fold up on itself when its 12 component strands were added on their own, but upon addition of strands which bound to the rims and made them double-stranded (such as those conjugated to $FliF_C$), the structure would most likely flatten out, its braces taking a more splayed and planar configuration.

One downside to this design however was that linking six strands together rather than using one long, ligated, circular strand meant that each five-binding-site domain was separated by around 2nm, the width of a DNA double helix. These 2nm gaps were not ideal but in theory the unstructured linker between the DNA tag and the peptide in my conjugates would provide sufficient flexibility for the FliG proteins adjacent to the gaps to span them and there was also the prediction from the domain-swap hypothesis that excess, untemplated FliG molecules might arrive at those locations and fill the gaps in the FliG oligomer that was forming. In any case, this design would provide a much stronger impetus for ring-formation than a single linear five-site template would and its simplicity meant I was eager to proceed with its assembly.

Therefore, I ordered the 12 strands that I had designed and began to think of techniques I could use to determine whether or not they assembled into my desired structure. My hope was that I could confirm in some way that the Hexagon had formed correctly before purifying the correctly-formed assembly and then adding to this the $FliF_C$-DNA conjugate and wild-type FliG protein. To a great extent I followed this plan, and while the addition of $FliF_C$ and FliG is left for section 5.2, the work that I describe next focusses on how the assembly of the Hexagon was characterized and a purification method worked out.

Gels and AFM Showed That the Hexagon Nanostructure Assembled as Designed

In my attempts to determine if the Hexagon assembled as I had designed it to, I decided that I should characterize each of the components separately on my way to building up the final structure. Thus, I began with the rim strands and proceeded to investigate their assembly using native polyacrylamide gels. I was able to show that each of the rim-rim interactions between pairs of adjacent rim strands formed correctly by observing clear dimer bands on the gels when each pair was mixed. With this confirmed, I then added the rims together one by one, obtaining solutions containing from two to six rim strands. I added each new rim to solution in slight excess to the rims in the previous mix meaning that I would be able to see both the complete and the sub-stoichiometric complexes that formed. Annealing from 95°C and then running these mixes on a gel produced the ladder of bands that I had hoped for, with species present containing from one

to six rims (Fig. 5.2A). In addition, when all six strands were mixed, the highest band on the gel ran slower than would be predicted for a simple linear chain of six strands and though other methods would be needed to confirm this, it implied that the rims had circularized as I hoped they would.

Following this same procedure, I then characterized the six brace strands found at the centre of the Hexagon. The pairwise interactions between adjacent braces again appeared to all form correctly, and a clear ladder of bands could be seen on the gels, with species containing from one to six braces (Fig. 5.2A). This again implied to me that the strands assembled as designed and when combined with the similar results for the rims, led me to the conclusion that the individual units of the structure, the rims and braces, most likely formed correctly.

Building from this, I next looked to see if the Hexagon would assemble properly when I added the braces and rims together. Starting from the structure containing six brace strands, I added rim strands in one by one, annealed, and ran the mixes on a gel. This produced a clear ladder of bands showing the sequential binding of the rims to the six-brace structure (Fig. 5.2B). Complexes containing the braces and from one to six rims could be clearly identified but the gel also made apparent the fact that the Hexagon structure, especially if $FliF_C$ and FliG were to be added to it, was too large to reliably characterize using polyacrylamide gels. The full mix of six braces and six rims produced a band right at the well of the gel, meaning that resolving any heavier structures would not be possible. If I was to characterize any assemblies larger than the Hexagon I would need to move to a method which could differentiate between species with high molecular weights.

For the next step in my characterization then, I used native agarose gels. These have a larger pore size than polyacrylamide gels do and are therefore better at separating larger species. I wanted to determine if, once the Hexagon was assembled, the binding sites on the currently single-stranded rim strands could in fact be accessed by the 10-nucleotide (N10) strands that were conjugated to the peptide and hence I added in an excess of those same N10 strands to the Hexagon. As I predicted the shift might be a small one, I added in N10 strands which bore a fluorophore at their 5’ end rather than a peptide, in order to visualize whether or not the strands

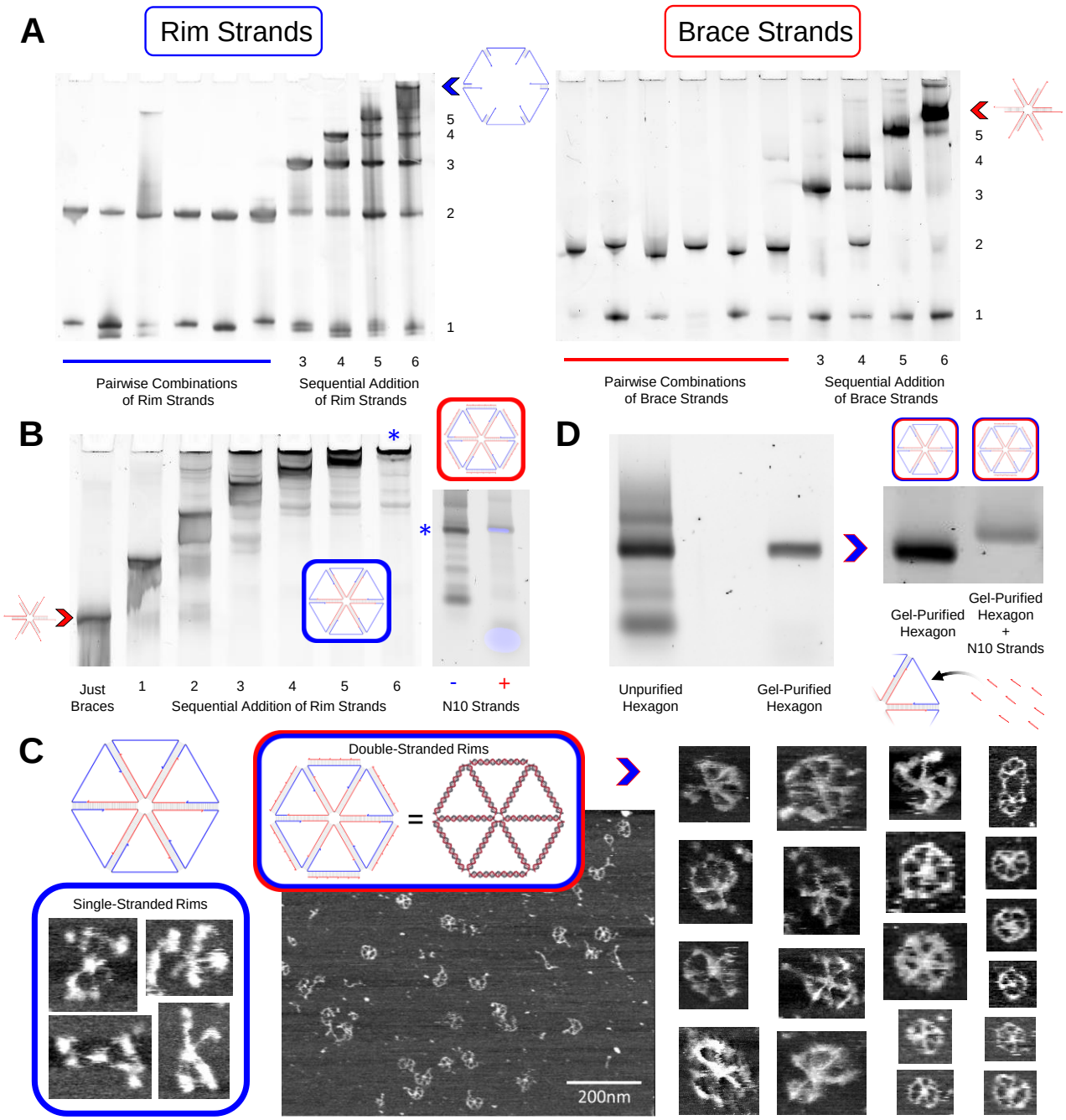


Figure 5.2 | The 12 DNA strands of the Hexagon assemble into the designed 'wagon-wheel' structure. **A)** To investigate the assembly of the six rim strands, each pair of adjacent rims was mixed together and dimer bands were seen in all cases on a native polyacrylamide gel. Sequential addition from 2 to 6 rim strands produced a ladder of bands which proved the six strands assembled. The highest band on the gel was slower than that predicted for a linear chain of six strands indicating that the structure had circularized. A similar process with the brace strands confirmed that all six strands assembled into a complex. **B)** Sequential addition of rim strands to a six-brace structure produced a ladder of bands on a native polyacrylamide gel. Addition of dye-labelled, rim-binding N10 strands to the complete set of 12 Hexagon strands (*) produced a shifted band containing the fluorescent N10 signal on an agarose gel. **C)** AFM imaging of the 12 Hexagon strands with single-stranded rims showed highly-flexible structures containing spoke-like braces. Addition of the rim-binding N10 strands produced much more uniform structures with the designed, 'wagon-wheel' appearance. When fully-formed, they possessed six evenly-spaced brace strand spokes connected by the encircling rims. Incomplete structures and the occasional dimer were seen, implying that purification was needed. **D)** The Hexagon was purified by cutting the appropriate band from an agarose gel and extracting the structure-containing liquid. Gels showed the technique gave a single, pure species to which the N10 strand could bind.

had in fact been incorporated. Running the Hexagon and the Hexagon with N10 strands added side-by-side on a native agarose gel showed that there was in fact a noticeable shift with the addition of N10 strands and that this shifted band contained the fluorophore signal (Fig. 5.2B). This proved that the binding sites on the rims of the Hexagon were in fact accessible, at the very least to unconjugated N10 strands, and led me to believe that I might have success once the time came to assemble the $FliF_C$-DNA conjugates on the structure.

Though these gels had provided evidence that assemblies with the correct weights for the given mix of strands had formed, they did not prove that the strands connected as they were meant to or that the structures which formed actually resembled the Hexagon that I had designed. To determine this, I needed to implement a different method for characterizing the structures so, as I had for the origami tile, I turned to AFM.

On the AFM I imaged two different assemblies. The first was the mixture of six brace strand and six rim strands without any other strands added which I called the Hexagon and which I had observed forming a band right at the well of native polyacrylamide gels. These had large single-stranded regions along their rims and I expected that there would be a great deal of flexibility in the structures, meaning that it might be quite difficult to pick out any of their defining features. Due to these concerns, I also prepared a second sample which contained the six braces and six rims of the Hexagon but to which I also added an excess of the rim-binding N10 strand in order to make the edges of the structure double-stranded and therefore more rigid. I hoped that with all sides double-stranded in the second sample I would be able to visualize a more planar and untwisted Hexagon which would enable me to more clearly determine whether or not the structure assembled as I had designed it to.

In general, the images I obtained of the two samples confirmed my predictions. Without the rim-binding N10 strands, most of the DNA observed form clumps on the mica surface, indicative of a flexible and relatively unstructured assembly. There were however recognizable features of the Hexagon present, as the spoke-like structures of the braces could be seen in many cases (Fig. 5.2C). The spokes were not evenly spread out from one another but I hoped that this was due to

the lack of double-stranded rims between them and that the addition of the N10 strands would remedy this.

Imaging the second sample with the double-stranded rims confirmed that addition of N10s did remedy this problem and proved fairly definitively that the Hexagon assembled as I had designed it to. Planar hexagonal structures could be seen on the mica with the correct dimensions and six evenly-spaced spokes emanating from their centre, producing the wagon-wheel-like geometry I had designed (Fig. 5.2C). This was reassuring and gave me confidence that with my design I had a reasonable chance of templating a C-ring-like assembly of FliG molecules. I did see incompletely-formed Hexagons however and though this was most likely due to my addition of certain strands in excess to others, I reasoned that in order to obtain any decipherable images when $FliF_C$ and FliG were added, I would need to find a way of purifying away incompletely-assembled structures from the desired Hexagons.

There were several possible ways that I could go about obtaining purified Hexagon structures but the most straightforward for me was through gel extraction. I had already characterized the Hexagons using gels and knew where the properly-assembled structures migrated to. Simply by cutting out the appropriate band, squeezing out the Hexagon-containing liquid, and exchanging into the desired buffer, I could obtain purified structures ready to be imaged by AFM or combined with $FliF_C$ and FliG. I therefore pursued this approach, running an agarose gel with the Hexagon as I had done previously and cutting out the correct band. Squeezing the liquid from this band and running this alongside an unpurified Hexagon sample confirmed that gel extraction was successful in obtaining a single purified species (Fig. 5.2D). In addition, adding this purified sample to excess N10 strands showed a convincing upward shift in the band produced, indicating that the purified structure could bind to added components.

Thus, gel extraction provided a way forward for producing homogeneous samples of six-sided, six-spoked Hexagon assemblies which, at the very least, were capable of binding to the N10 DNA sequence that would be conjugated to the peptide. The design of the structure was a simple one, using only 12 unique DNA strands, but gels and AFM had shown that it assembled correctly and

that its dimensions and the spacing of its binding sites roughly matched those seen for the C-ring of the BFM. The next step then was the addition of the $FliF_C$-DNA and FliG to the structure, with the hope both that the peptide and protein would stably bind around the rim of the Hexagon, and that, once bound, FliG-FliG interactions might occur and be detected.

5.2 Assembly of $FliF_C$ and FliG on the Hexagon DNA Nanostructure

With the Hexagon assembling correctly, I set out to incorporate both the $FliF_C$-DNA conjugate and FliG into the system. The hope was that simply combining the Hexagon strands, the $FliF_C$-DNA conjugates, and the FliG proteins in one solution would result in a self-assembled structure which would promote the formation of FliG-FliG bonds (Fig. 5.3A). To investigate the progress and final state of the assembly, I had both gels and AFM at my disposal, two tools which were particularly useful in confirming that the Hexagon structure on its own was well-formed and which I was confident could also be used to characterize the full protein-DNA assembly.

I therefore proceeded to mix all of the components together, attempting to assemble $FliF_C$ and then FliG on the Hexagon DNA nanostructure. Some promising gel-shifts were observed and interesting AFM images were acquired during the characterization of the complexes but, as I will show, the ever-elusive FliG-FliG interaction could not be detected. Thus, with this being my final section of results, I was unable in the end to obtain evidence of intermolecular FliG bonds or of the domain-swap hypothesis in general. Discussion of why the interactions were not observed is reserved for the final chapter of this work, while in this section I stay close to the results, presenting what, in the end, was the nearest I came to obtaining a DNA-templated FliG ring.

An Excess of $FliF_C$-DNA and FliG is Needed to Fully Fill the Hexagon's Binding Sites

As I had had success in the past templating the HPLC-purified $FliF_C$-DNA conjugates and wild-type FliG on short, linear templates, I chose these constructs rather than the FliG-DNA conjugates as the species that I would add to the Hexagon nanostructure. For a preliminary test, I mixed and

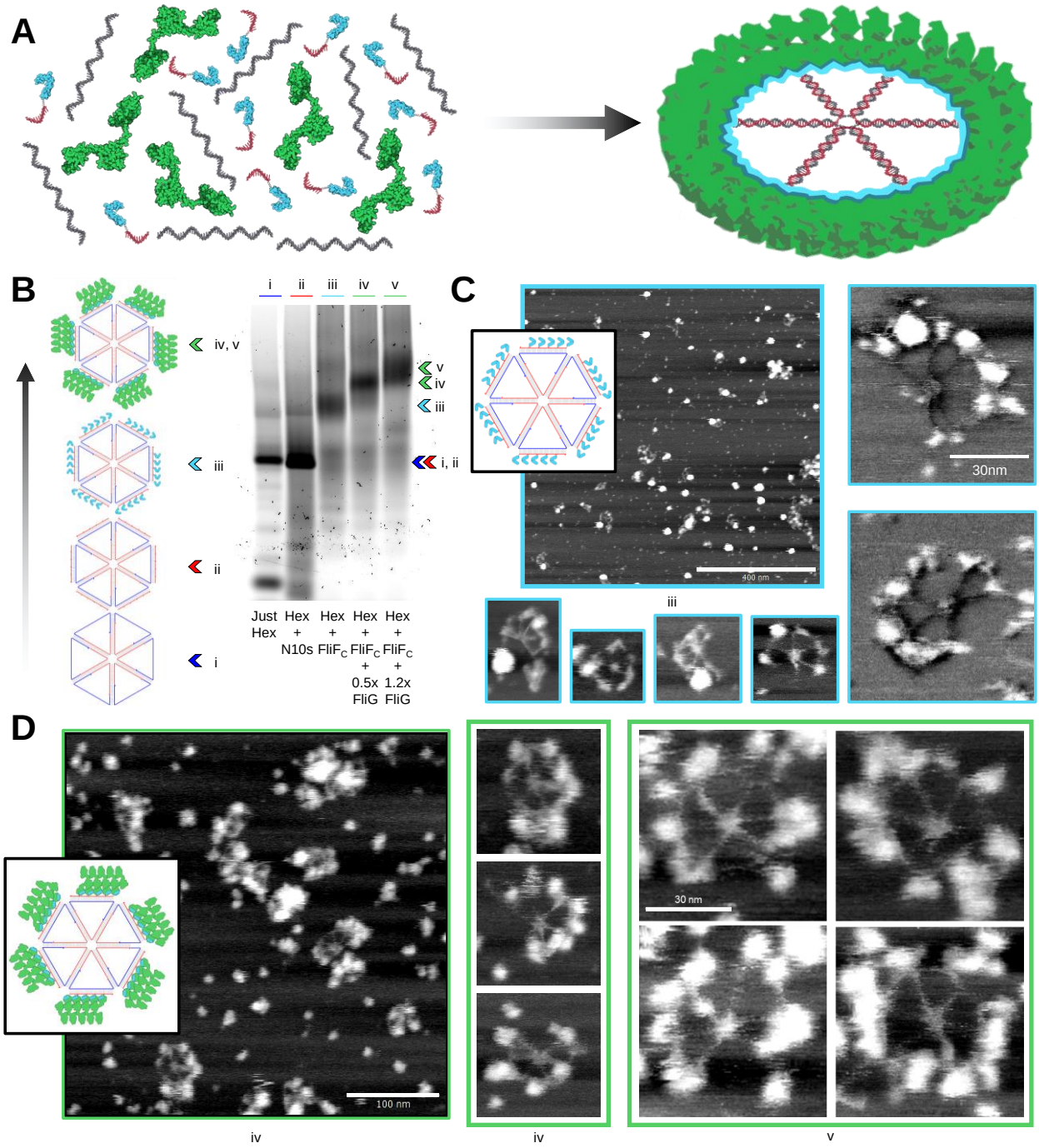


Figure 5.3 | FliF$_C$-DNA and FliG, added in a slight stoichiometric excess to the binding sites present, assembled as punctate spots on the rims of unpurified Hexagon templates. A) Simply by mixing the strands for the Hexagon, the FliF$_C$-DNA conjugate, and FliG, I hoped to obtain a self-assembled, DNA-templated FliG ring. **B)** Adding the components one by one to unpurified Hexagons showed clear shifts in the bands observed on an agarose gel, with the addition of FliF$_C$, sub-stoichiometric (0.5x) FliG, and then a slight excess (1.2x) of FliG producing larger and larger species. **C)** AFM imaging of the FliF$_C$-DNA conjugate bound to the Hexagon showed large clumps of the peptide forming both on and off the structure. Binding did appear to take place at the correct sites around the rim however. **D)** Addition of FliG led to the disappearance of the large clumps, leaving single-protein-sized spots visible. With sub-stoichiometric FliG (left and middle panels) the protein was seen to bind around the edges of the structure, presumably at the appropriate sites on the rim strands. Protein density was observed concentrated at the vertices of the structure however, indicating that binding sites at these locations were perhaps easier to access than others along the rim. Imaging samples with a slight excess of FliG (right panel) showed a greater density of single proteins on the surface and a greater occupancy of sites around the rim of the structure. Punctate spots rather than a continuous ring were observed however, indicating that neighbouring FliG proteins were most likely not interacting with one another.

annealed the 12 Hexagon strands and, without purification of the Hexagon, added to this the peptide-DNA in a slight stoichiometric excess to the binding sites. To aliquots from this peptide-Hexagon mix, I then added FliG either sub-stoichiometrically or in a slight stoichiometric excess to the $FliF_C$. After leaving the mixtures to incubate for a period of time at room temperature, I ran the various samples on an agarose gel to determine where the complexes ran to. As I had not purified the Hexagon from unincorporated strands or done anything to remove excess $FliF_C$ or FliG, there were several bands and smears in each lane of the gel, indicating a heterogeneous mix of species, but the main band in each lane showed a trend of shifting upwards as more components were added (Fig. 5.3B). With $FliF_C$, sub-stoichiometric FliG, and then an excess of FliG added to the mix, the observed band was seen to migrate increasingly slowly, indicating that the peptide and protein constructs were in fact binding to the Hexagon.

To confirm this however, I needed to look at the assemblies using another method which would allow me to determine the geometry and configurations of the protein-DNA complexes. As I had for characterizing the Hexagon on its own, I again turned to AFM to provide me with this structural detail. Images of just the Hexagon showed, as I have described in the previous section, flexible assemblies of spokes which flattened out into wagon-wheel-like shapes as the rims were made double-stranded. Incomplete assemblies and unincorporated strands could be seen on the surface however which, along with the extra bands and smears seen on the gel, were a result of the fact that the structure had not been purified.

Moving on, I next imaged the complex formed between the Hexagon and the peptide-DNA conjugate without any FliG added. This produced some interesting results. The peptide had been added in a slight stoichiometric excess to the binding sites of the DNA structure so ideally I would be able to see the rims of the Hexagon filled by a continuous, raised region of density corresponding to the bound peptide. Instead, what I observed were patchy distributions of the peptide around the rim of the structure, with each patch corresponding to a punctate spot much larger than would be predicted for a single peptide (Fig. 5.3C). These spots were also seen on the mica surface without any accompanying DNA structure, indicating that they were probably not

template-induced but a feature of the peptide-DNA conjugates themselves. Though difficult to explain, the presence of these clumps correlated well with the tendency of the peptide to stick to tubes and resins and perhaps constituted a clustering of their hydrophobic interfaces which occurred when FliG, their usual binding partner, was not present.

Though still unsure about the peptide-Hexagon structures that had been observed, I next imaged the complexes that had formed when the Hexagon, $FliF_C$, and FliG were added together. Looking first at those mixes with sub-stoichiometric FliG, I was able to see that the large clumps present when FliG was absent had now disappeared to leave only single-protein-sized spots on the surface. In addition, proteins bound to the binding sites on the Hexagon's rim could clearly be made out, although these again formed punctate spots which in this case tended to cluster around the vertices of the Hexagon (Fig. 5.3D). I reasoned that this tendency was perhaps accounted for by the fact that it might be easier for the peptide-DNA conjugates or FliG proteins to bind to the structure near the 2nm gap at the ends of the spokes, as these areas would be free of any other bound species which might sterically hinder an incoming molecule.

For my final sample I again looked at a complete mix of the Hexagon, $FliF_C$, and FliG but here FliG was in a slight stoichiometric excess to the $FliF_C$-DNA conjugates. These images also showed the characteristic single-protein-sized spots on the surface but focussing on the Hexagon assemblies showed that there was a greater filling of the binding sites around the rims (Fig. 5.3D). The density was still mostly seen as individual, punctate spots but here the spots were closer together and approaching a continuous ring around the outside of the structure. The greater filling seen with this slight excess as compared to the sub-stoichiometric FliG concentrations was an encouraging trend and gave me hope that with just a touch more FliG added in solution I might be able to observe a continuous FliG ring. This was dependent on the formation of stable and observable FliG-FliG interactions however which, up to this point, I had no proof of. So though it appeared that a greater excess of $FliF_C$-DNA and FliG was needed to achieve full occupancy of the Hexagon's binding sites, there was no evidence to suggest that even with all the sites filled, a domain-swapped or otherwise connected ring would form.

Several Techniques Were Used in an Attempt to Remove Unbound $FliF_C$-DNA and FliG

Following the acquisition of the AFM images just described, I decided to determine how great an excess in the $FliF_C$ and FliG stoichiometries was needed in order for the binding sites of the Hexagon to be fully filled. To accomplish this, I returned to polyacrylamide gels and used a simpler DNA template: a single rim strand from the Hexagon. This strand was very similar to the five-site linear template I had used earlier, merely incorporating two extra domains on either side of the T5N10 sequence for linking it to the other rims of the structure, and this meant that I could use conditions for running the gel similar to those I had used in the work of Chapter 4.

I therefore added an increasing amount of $FliF_C$-DNA to the rim strand to find which ratio of conjugates to binding sites was sufficient to obtain complete occupancy of the template. In my AFM experiments I had used only a slight (1.2x) excess, but running the mixes on a native polyacrylamide gel showed that only after around a 2x and definitely by a 5x excess were the sites completely filled (Fig. 5.4A). Testing the ratio of FliG to $FliF_C$ that was needed, I used a 2:1 stoichiometry of $FliF_C$-DNA to binding sites and added FliG in stoichiometries of 1:1, 2:1, and 5:1 to the peptide. Running the gel demonstrated that only with the protein in 5x excess to the peptide could a completely-filled rim be achieved (Fig. 5.4A) and therefore confirmed that the slight (1.2x) excess of protein to peptide that I had used for my AFM experiments was insufficient for complete assembly on the Hexagon.

In a sense this was reassuring, as it suggested that simply increasing the concentrations of $FliF_C$-DNA and FliG would be enough to achieve a fully-bound template. Increasing the concentrations came with some issues however, mostly relating to how the excess, unbound peptide and protein could be removed following the assembly process. If these excess species were left in solution, they would carpet the mica surface during AFM imaging and make it impossible to pick out any structural features of the assemblies that formed on the Hexagons. Thus I began to work through various methods which could be used to isolate the peptide-and-protein-bound Hexagons from

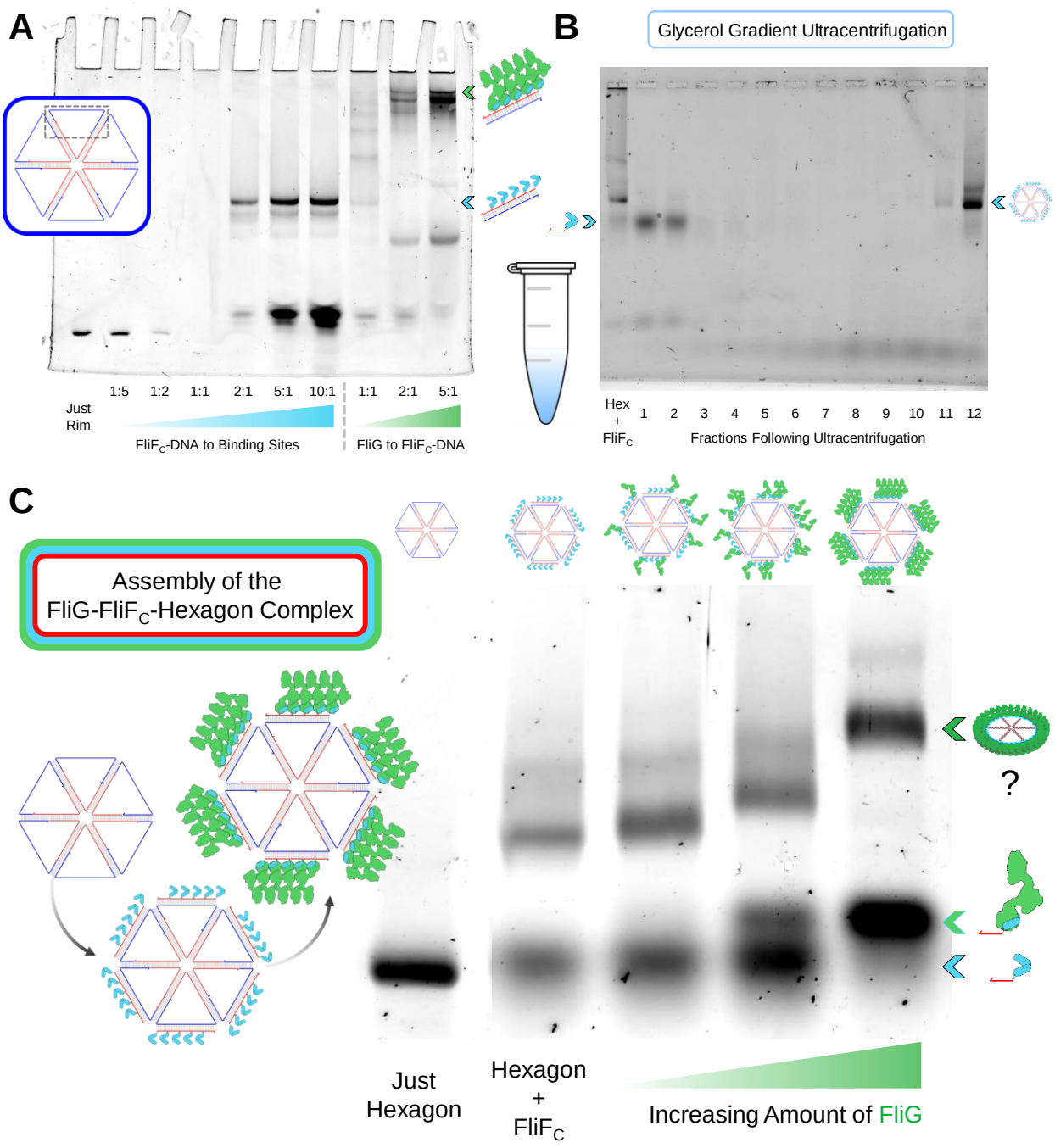


Figure 5.4 | Excess FliF$_C$-DNA and FliG were needed for full filling of the Hexagon's binding sites but a purification method could not be found which removed unbound peptide and protein. **A)** Using a single rim from the Hexagon, FliF$_C$-DNA and then FliG were titrated in to find at which point all binding sites became occupied. A 5x excess of both produced full filling of the sites. **B)** A density gradient was created by layering solutions containing varying percentages of glycerol on top of one another. Applying the peptide-Hexagon complex to the top of the gradient and ultracentrifuging separated the components by their density. Free, unbound peptide-DNA could be seen in fractions 1 and 2 at the top of the gradient, while the peptide-Hexagon complex was found almost entirely in fraction 12. **C)** To assemble the FliG-FliF$_C$-Hexagon complex, the Hexagon nanostructure was first purified by gel-extraction. To this purified structure was added the FliF$_C$-DNA conjugate in a two-fold excess to the number of binding sites present on the Hexagon, This produced a substantial gel-shift. FliG was then added in increasing concentrations with 1:1, 2:1, and finally 5:1 stoichiometries between the protein and the peptide. The dominant band visible shifted upwards with each addition of FliG, indicating that the protein bound to the structure. By a 5:1 stoichiometry of protein to peptide, all free FliF$_C$ appeared to be bound to the FliG, indicating that by this point in the titration, all binding sites on the Hexagon most likely had been occupied.

the excess, free-floating species, with the hope that I could image the purified assemblies using AFM.

Gel extraction had worked very well for removing incomplete assemblies and unbound strands from the fully-formed Hexagon, so I decided to try the same technique for the purification of peptide-Hexagon constructs. I opted against attempting to extract the full protein-peptide-Hexagon complex as I was concerned that running FliG through the gel and then squeezing it out of the cut-out band might be too much for the protein to handle and lead to misfolding. So instead, I cut out the band that formed following the addition of only $FliF_C$-DNA to gel-purified Hexagons, squeezed out the solution as I had done previously and ran an aliquot of this on a gel. None of the constructs appeared to have been extracted from the gel however, or, if they had, they had been lost somewhere during the extraction process, as there was no evidence of any signal in the lane with that sample. It was highly likely that the sticky nature of the peptide contributed to the failure of the technique, as it had a tendency to bind non-specifically to tubes and may have also stuck to the gel matrix. Thus, the success I had achieved through gel extraction of the Hexagon could not be replicated with the peptide-Hexagon complex and I had to pursue other purification techniques.

A common technique for purifying DNA origami structures from excess staples involves precipitating the complexes using PEG (polyethylene glycol) such that the fully-formed origami enters the pellet, while the smaller staples can be removed with the supernatant. I reasoned that this process might be useful with my system for separating unbound $FliF_C$ from the peptide-Hexagon complexes and hence I applied the standard PEG precipitation procedure to a solution containing the peptide-Hexagon assemblies. I found that the structures did not pellet with the PEG however and remained entirely in the supernatant along with the unbound $FliF_C$-DNA. This was not entirely surprising as, with only 12 strands, the Hexagon had a much smaller molecular weight than the standard origami and was not perfectly suited for the technique.

The next technique I tried was glycerol gradient ultracentrifugation. In this method, solutions with varying percentages of glycerol are layered on top of one another to create a gradient of density from the bottom to the top of a tube. The sample is applied to the top and then ultracentrifuged,

with denser components migrating to the denser solutions near the bottom of the tube and the less-dense components remaining near the top. I predicted that the peptide-Hexagon complex would be denser than a single peptide-DNA conjugate and hence would travel farther down the gradient. By collecting fractions one by one from the top to the bottom of the gradient I could hopefully separate the excess $FliF_C$-DNA from the $FliF_C$-Hexagon complex and thereby purify it.

Applying my sample to the top of the gradient, and spinning in the ultracentrifuge, I then collected 12 fractions from the top to the bottom of the tube. Running these fractions on an agarose gel I could see immediately that the separation had been successful. While the excess peptide-DNA could be found near the top of the gradient in fractions one and two, the peptide-Hexagon complex was found almost entirely in the final fraction 12 (Fig. 5.4B). This was very encouraging so I decided to image the isolated peptide-Hexagon complexes both with and without FliG added using the AFM. Although gels of the samples showed that FliG did in fact bind to the glycerol-gradient-purified peptide-Hexagon complexes, I was unable to obtain clear images of the samples using the AFM, most likely due to the large amount of glycerol present in solution during imaging. Attempts to remove the glycerol via buffer exchange were unsuccessful as the peptide-Hexagon complexes bound to the tubes and columns used and thus, after several attempts, no clear structural information could be obtained for samples purified using glycerol gradient ultracentrifugation.

With my time on the project drawing to a close and without another immediately obvious purification method, I ran one final gel to look at the assembly of the entire FliG-$FliF_C$-Hexagon complex with excess components added. Using gel-purified Hexagons for my DNA template gave the clearest images I had yet obtained of the assembly process, as adding more and more FliG led to slower and slower migration of the major species present in each of the lanes (Fig 5.4C). Upon reaching a 5x excess of FliG to $FliF_C$, a dark and well-defined band had formed containing the Hexagon, $FliF_C$, and FliG and it appeared that all the binding sites had been saturated, as lower bands showed all the peptides present were bound to FliG. It was tempting to think then that this dark, well-defined band represented the fully-assembled protein-DNA complex that I had

designed and perhaps, though this was stretch of the imagination, that the FliG proteins were bound around the Hexagons' rims and thereby perfectly positioned for intermolecular interactions to take place. Without a way of purifying away excess peptide and protein, I could not AFM image the sample however as previous attempts with protein concentrations this high had led to complete carpeting of the mica surface. Therefore, the exact nature of the complex seen on the gel remained a mystery and, in the end, I was unable to provide any evidence of a DNA-templated FliG ring.

Looking back then at this final chapter of results, I had aimed to investigate whether a circular rather than a linear template could induce polymerization of templated FliG molecules. This might be expected based on the domain-swap model of FliG polymerization, as a circular template would provide every sequestered FliG molecule with two neighbouring FliG molecules which could bind to the hydrophobic patches of its ARM motifs, thereby concealing them from the aqueous solution and creating an energetically-favourable interlocked configuration. To test this, I had designed and characterized a simple DNA nanostructure made from 12 unique strands which I called the Hexagon (5.1). It was to take a 'wagon-wheel-like' geometry, being roughly 35nm in diameter, with six spokes emanating from its centre which served to hold the FliG-binding-site-containing rims in place. Using both gels and AFM I had been able to verify that the Hexagon structures assembled as I had designed them to, with clear images of undecorated wagon-wheels on the mica surface.

I then proceeded to incorporate both the $FliF_C$-DNA conjugates and the FliG protein with the assembled Hexagons (5.2). Agarose gels showed that both the peptide and the protein could bind onto the structure and AFM imaging revealed the assembly of $FliF_C$ and FliG around the rims of the Hexagon as designed. The assemblies were patchy however and no FliG-FliG interactions could be detected, so I attempted to raise the relative concentrations of $FliF_C$-DNA and FliG with respect to the Hexagon in order for the peptide and protein to be in a high stoichiometric excess. To successfully image assemblies with these high $FliF_C$-DNA and FliG concentrations however, I needed to in some way remove the unbound, excess components and I next attempted various purification strategies which I hoped would allow me to isolate only the protein-peptide-Hexagon complex. Gel extraction and PEG precipitation were both unsuccessful due in part to the peptide's

propensity to stick non-specifically to surfaces, while glycerol gradient ultracentrifugation separated the desired species from unbound components but led to problems during AFM imaging due to the high viscosity of the solutions obtained. In the end gels seemed to show that a Hexagon nanostructure with its binding sites fully-filled could be assembled through the addition of excess $FliF_C$-DNA and FliG, but without a reliable method of purification, the complexes could not be imaged.

Thus, my time on the project ended and the results just presented were the last that I obtained. Though I had been unable to provide evidence of FliG-FliG interactions or the domain-swap hypothesis, I had learned a great deal about the C-ring proteins and made considerable strides towards the development of an *in vitro* system which could successfully template the formation of a FliG ring. In drawing things to a close, the final chapter of this work brings together the main conclusions from each of the preceding chapters and ends by looking ahead to possible extensions to the system and directions that could be taken with the work in the future.

CHAPTER SIX

CONCLUSIONS

To bring this work to a close, it is worthwhile now to take in all the results together, and to highlight the main themes and conclusions that can be drawn from the project as a whole. Therefore, in this final chapter I will first condense and summarize my work, before emphasizing its main themes, and then finally I will describe some of the future directions that could be taken with this project.

6.1 A Brief Summary

I began this work with an overview of the landscape in which it sits and an introduction of the problem at hand. In the context of proteins and protein complexes, the **bacterial flagellar motor** is one of the most remarkable and dynamic assemblies found in nature, with the coordinated motion of its components powering directed bacterial locomotion. The structure is made out of hundreds of individual proteins arranged into functional sub-complexes, each responsible for a different facet of the assembly, rotation, or switching of the motor. One key sub-complex is the **cytoplasmic ring** which coordinates both the torque generation by the stator units and the switching of the rotation direction through signals from the chemotactic machinery. The central component of this ring is **FliG**, a multi-domain protein which is necessary for both BFM assembly and bacterial motility. Its C-terminal domain bears the site of torque generation[60], its middle domain receives inputs from the chemotactic signalling pathway in order to coordinate switching[87],

while its N-terminal domain anchors the C-ring to the membrane through its interaction with the membrane-embedded FliF protein[89]. During the assembly of the motor, FliG is one of the first components to join and forms an oligomeric ring with approximately 26 to 34 subunits[56]. One recent hypothesis states that this FliG ring assembles through a polymerization process involving the **ARM motifs** of the molecule, in which intermolecular interactions are formed as the ARM_M motif of one molecule interacts with the ARM_C motif of its neighbour[101]. This **domain-swap polymerization** carries on until a fully interlocked ring of FliG proteins forms which can change conformation in a coordinated manner between a CW- and CCW-rotating state. Testing this domain-swap model of FliG polymerization was the main goal of this work, with the aim being to provide evidence of ARM-ARM FliG interactions in an *in vitro* setting using a **DNA nanostructure** to template the proteins into an oligomer-favouring organization.

With the problem introduced, the presentation of my results began in Chapter 2 which centred around the use of **NTA-modified DNA** to bind to His-tagged FliG and thereby attach it to an underlying DNA template. DNA strands were modified with three NTA groups following a procedure developed previously in the lab[153] and these were shown to bind to purified His-tagged FliG. There were concerns about the stability of this interaction however and other conjugation methods were sought. I implemented one strategy which used an NTA-modified strand to guide another strand bearing an activated ester close to a specific lysine on the protein for a **DNA-templated conjugation reaction**[162]. In the end this technique was successful but at very low yields and with questionable site-specificity. The main finding that arose from this work however was that NTA-Histidine interactions were stable and could be visualized using **native polyacrylamide gels**, a much more straightforward technique than the TIRF microscopy that had been used previously to confirm an interaction. To take advantage of this finding, a **pentakis-NTA strand** was created which bound more stably to the 10x His-tag of FliG than the previous tris-NTA strand had, and complexes of this strand, FliG, and various **linear DNA templates** were run on native gels in order to characterize the complexes that formed. While FliG molecules spaced apart on 20-nucleotide-long binding sites could be seen to fully occupy a three-site template, attempts

at binding five conjugates on closer-packed, 10-nucleotide-long binding sites were unsuccessful, most likely due to the bulk of the pentakis-NTA groups. A new conjugation strategy was then sought which could permanently link FliG to its DNA tag using a compact chemistry which would not hinder the binding of FliG to templates with closely-packed binding sites.

Chapter 3 focussed on the implementation of one such compact linkage, formed between a **maleimide-modified DNA strand** and a FliG construct bearing a single N-terminal cysteine[157]. Once a purification strategy for the modified strand was worked out, the maleimide-DNA was shown to stably bind the **single-cysteine FliG** proteins, but attempts at arranging the conjugates on linear DNA templates were hindered due to the presence of FliG-FliG dimers. A great deal of time was spent in trying to figure out the cause of this interaction, with cysteine-cysteine and domain-swapped dimers eventually ruled out. In the end it turned out the species seen on gels which was preventing proper templating was in fact a **His-tag-mediated FliG dimer** which could only be broken apart through the two-step process of incubating at 37°C and then with a His-tag-cleaving protease. Once the FliG was returned to its native, monomeric state, it was quite straightforward to conjugate it to a maleimide-modified strand and then to template these conjugates on short, **linear DNA templates**. Due to the smaller size of the linkage, five FliG-DNA conjugates could be arranged side-by-side on a template bearing 10-nucleotide-long binding sites and, with native polyacrylamide gels, both the stoichiometry and to some degree the spacing of the bound conjugates could be determined. Evidence of interactions between FliG molecules could not be observed however. Tests with two colours to differentially label templated and free-floating FliG gave no sign of an interaction between the two species while an experiment in which FliG-DNA conjugates and unconjugated DNA strands competed for the same sites on the template produced no signatures of interactions between FliG proteins bound to the template. Thinking that perhaps the simplicity of the linear template was to blame for the lack of interactions, a new structure based on a **DNA origami tile** was used to template a greater number of FliG molecules. Even after several issues with the origami were resolved however, FliG could not be imaged bound to the tile's surface. Clear images of streptavidin bound via a biotin to the tile were obtained

however, implying that it was either the linkage between the FliG and the tile or some property of the FliG protein itself that caused the issues. To round out the chapter, one final method to search for FliG-FliG interactions was employed using **double-cysteine mutants for cross-linking.** Although cross-linked oligomers were observed, they were not induced by the DNA template and no proof could be obtained that they were held together by domain-swapped ARM-ARM interactions. I therefore began to think that my *in vitro* set-up required a different template or perhaps a new component added to the system which would trigger FliG oligomerization.

One candidate that recent studies had indicated might induce FliG-FliG interactions was introduced into the *in vitro* system in Chapter 4, where I worked on incorporating the cytoplasmic domain of FliF called **FliF$_C$**[88,91]. I designed and had synthesized a peptide corresponding to the FliF$_C$ domain from *Salmonella* and *E. coli* and, while working through issues with its solubility, looked to see if it would bind to my purified FliG protein. Initial trials showed that it did not bind to a dye-labelled FliG construct and I began to question how well-folded my purified proteins were. I therefore performed several biophysical assays to characterize the folded state of several FliG constructs, thinking that misfolding in FliG might be the reason that I had not observed FliG-FliG interactions. **CD spectroscopy** indicated that the wild-type FliG protein was most likely well-folded whereas other constructs showed varying degrees of α-helicity. A **tryptophan fluorescence assay** using the wild-type protein and the FliF$_C$ peptide indirectly demonstrated that the two species bound to one another in a one-to-one fashion, while a **thermal shift assay** using the SYPRO Orange dye showed that binding of FliF$_C$ to FliG increased the stability of the protein. Taking in these results, I decided to pursue a templating strategy in which I would use **DNA-templated FliF$_C$ peptides** as a binding platform for wild-type FliG proteins, with the hope that binding of FliG to the templated peptide would spur on FliG oligomerization. I therefore conjugated a maleimide-DNA strand to a single-cysteine version of the FliF$_C$ peptide and was able to purify the resulting conjugate using HPLC. Templating the **peptide-DNA conjugates** on various short, linear DNA templates with up to five sites, I confirmed, using native polyacrylamide gels, that they assembled as they were designed to. I then added FliG to the template-bound peptides and,

following some initially puzzling results, found that only **wild-type FliG proteins** and not mutant constructs were able to bind to the $FliF_C$ peptide. This binding was the first proof that the 39 C-terminal amino acids of FliF from an enteric bacterium were sufficient to bind to *Salmonella* FliG, a small but significant contribution to the field. Using the wild-type protein then, native PAGE confirmed that it bound to the templated peptides in a one-to-one fashion, but the gels obtained still did not produce any indication of a FliG-FliG interaction. I therefore sought to incorporate another component, the **FliM-FliN complex**, into the system with the hopes that it would stimulate FliG oligomerization[104]. After working through issues with the purification, the complex was isolated and added to both free-floating and templated FliG but no FliM-FliG or FliG-FliG interactions were found either via gel-shifts or in a pull-down assay. I therefore decided to investigate whether a more circular DNA template was needed in order to stimulate FliG oligomerization and set about designing a new DNA nanostructure.

Chapter 5 discussed the design of this new DNA nanostructure and its use in templating a more C-ring-like geometry of $FliF_C$ and FliG. Called **the Hexagon**, it was composed of 12 unique strands which assembled into a wagon-wheel shape roughly 35nm in diameter with 30 binding sites for $FliF_C$ and FliG arranged around its outer rim. Both gels and AFM imaging showed that the structure assembled as it was designed to, while gel extraction was successful in isolating the complete Hexagon from incomplete assemblies and unincorporated strands. Adding $FliF_C$-DNA and then wild type FliG to the nanostructure produced band-shifts on an agarose gel indicating that they formed a large **protein-peptide-Hexagon complex**, while AFM imaging showed that peptide and protein bound around the rims of the structure at the appropriate locations. The templated $FliF_C$ and FliG formed spots and patches rather than a continuous ring of density however, indicating that the gap-filling or runaway polymerization predicted by the domain-swap hypothesis was not occurring. Experiments with a single rim strand of the structure showed that a greater excess of $FliF_C$ and FliG was needed to fill all the binding sites of the Hexagon and I hoped that simply increasing the peptide and protein concentrations might stimulate ring formation. With such high concentrations of the components, I could not successfully image the complexes by AFM however

so I needed to find a way to isolate the protein-peptide-Hexagon structures from the unbound, excess components. Trials with gel extraction, PEG precipitation, and glycerol gradient ultracentrifugation were all unsuccessful in the end and though a clear band, most likely corresponding to the Hexagon with fully-occupied sites, could be seen on an agarose gel, it was unable to be purified and imaged. Hence, in the end no evidence of FliG-FliG interactions or the domain swap hypothesis was obtained, with questions remaining as to why the templated constructs did not appear to bind to one another.

6.2 The Key Lessons

In looking back over the work just summarized, several patterns emerge relating to the experiments conducted and the results obtained and two major phases stand out in the development of the project. The first deals with conjugation, how to graft a FliG molecule to an underlying DNA template, while the second deals with the search for FliG-FliG interactions and evidence for the domain-swap hypothesis. Here I will briefly highlight the results obtained in the context of these two major phases and draw conclusions relating to the final outcomes of the project.

The conjugation phase of my work went through three main stages as I searched for an appropriate way in which to attach FliG to DNA. My first attempts with NTA-modified strands were successful in the sense that FliG could be attached to a DNA strand and organized on a template, but the bulk of the NTA-Histidine linkage prevented the proteins from being arranged with the spacing seen in the C-ring. The technique had the advantage of being site-specific as the DNA tag bound only to the Histidine-tag engineered into the protein and it was also quite useful that the linkage could be broken and reformed simply through the addition of imidazole or nickel. Additionally, the overall strength of the linkage could be tuned somewhat by incorporating a greater or lesser number of NTA groups on the DNA strand, as I showed that five NTA groups bound more stably than three to FliG's 10x His-tag. Thus the technique had many benefits but

was simply not conducive to the geometric constraints of the *in vitro* set-up that I was working with.

The next stage in my work with conjugation centred on the reaction between maleimide-DNA and cysteine-FliG. Again this linkage was site-specific with a one-to-one stoichiometry due to the convenient fact that FliG naturally does not contain any cysteine residues and this meant that the reaction could be targeted to a single cysteine engineered at the desired location. In contrast to the NTA-Histidine linkage, this reaction produced a more permanent connection which was insensitive to any of the standard small molecules commonly included in aqueous buffers. If I needed to remove FliG from the underlying DNA template, I would therefore have to remove the entire FliG-DNA conjugate through DNA-based techniques such as strand displacement. Though this was not a major issue, the method seemed to have a more substantial drawback in that it appeared that single-cysteine FliG mutants and the FliG-DNA constructs made from these mutants were not as stable as the wild-type protein. CD spectroscopy indicated that they exhibited a smaller percentage of α-helix than was expected while the thermal shift assay produced broad transitions indicative of a somewhat misfolded protein. It was perhaps possible that the close proximity of the DNA strand to the N-terminal amino acids had changed the local environment there and caused some degree of misfolding at that location. So though using this technique I was able to template the conjugates with the spacing seen in the C-ring, I figured that a method which more closely matched the binding of FliG to its template *in vivo* would give me a better chance of observing more natural FliG behaviour.

Thus, in the final stage of my work dealing with conjugation, I incorporated the region of FliF that FliG naturally binds to during the assembly of the BFM in bacteria[89]. Using the wild-type protein ensured that I was working with a construct whose folding was closest to that of the FliG found *in vivo* and I hoped this would help me with my goal of visualizing native FliG-FliG interactions. Once issues such as the solubility of the $FliF_C$ peptide were worked out, the binding between protein and peptide was found to be very strong and dissociation was never an issue in the experiments conducted. The binding was naturally one-to-one and site-specific, meaning I did not have to worry

that the FliG protein was being arranged on the template in an improper orientation. In conjugating the DNA to the peptide rather than the protein, I did not have to be as concerned about secondary structure or misfolding. Structural studies seemed to indicate that the peptide was able to unfold and fold much better than FliG could and the fact that peptide-protein binding was observed assured me that the functionality of the peptide was not impaired by its conjugation to a DNA strand. Thus, this third stage in the conjugation phase of the project became the last one, as the previous issues stemming from strategies which negatively affected either the templating process or the folded state of FliG were now resolved by using this 'natural' way of organizing FliG on the DNA nanostructures.

Though it is difficult to generalize the lessons learnt during the conjugation phase of the project, some of the conclusions I came to may be useful in other systems which seek to conjugate DNA strands to different proteins of interest. With FliG it was beneficial to add in a layer in between the DNA template and the protein in order to make a more natural 'binding platform' for FliG assembly. This meant that the DNA strand did not have to be conjugated to FliG directly but instead could be attached to an intermediate species, the $FliF_C$ peptide, which was less likely to be detrimentally affected by the conjugation. In other systems seeking to template proteins or other molecules, it may also be beneficial to seek out a robust component which a DNA strand could be reliably conjugated to, and to use this component rather than the DNA itself to bind the molecule of interest and bring it to its place on the underlying DNA scaffold.

The other main lesson from this phase of the work was that, rather than being a minor technical consideration, the conjugation strategy that is used to join the molecule of interest to its template can have a dramatic effect on both the templating process and the state of the templated molecule. Though general thoughts such as whether the reaction will produce a site-specific or non-specific linkage, whether it will lead to either a single or multiple DNA strands being bound, and whether the linkage will be permanent or reversible are commonly taken into consideration when choosing a conjugation strategy, I learned from my work that it is also useful to question whether or not the steric bulk of the linkage will affect the templating, and whether the reaction itself or manipulations

of the components to enable the reaction might negatively affect the integrity of the molecule of interest. With this in mind, my main conclusion from this phase of work is that which conjugation strategy to use requires serious thought, as in most cases it presents the first major hurdle to overcome when seeking to template components using a DNA scaffold and until the correct strategy is found the project cannot proceed to its main aims.

The conclusions from the second phase of the work, the search for FliG-FliG interactions, are mostly relevant to the assembly pathway of the BFM, but some more general lessons can be drawn as well. The first is the fact that Histidine-tags can be more interfering in protein work than generally thought. Many biochemical and even structural studies investigating proteins leave the His-tags attached during their experiments[90] and although FliG may be a special case, it would not be surprising if some conclusions drawn by these studies were affected in some way by the actions of the His-tag. The His-tag-mediated dimerization seen in my work has not been reported previously in other systems and is most likely an oddity of the FliG protein, perhaps due to the close proximity between the His-tag used and the FliF binding site. Nevertheless, it seems that when studying protein behaviour, removal of the His-tags should not be overlooked in the overall expression and purification procedure.

Focussing on FliG and the C-ring of the BFM specifically, the main conclusion of this work is that with the *in vitro* system used, no FliG-FliG interactions could be detected. For the early work in which I used FliG-DNA conjugates, such as in the two-colour and competition experiments of section 3.3, the lack of interactions could possibly be explained as the product of some degree of misfolding in the FliG constructs. CD spectroscopy and the thermal shift assay gave every indication that the wild-type FliG construct was in its folded, native state however and in the work of Chapters 4 and 5, the absence of FliG-FliG binding is most likely not to be explained by any degree of denaturation. The one-to-one binding between FliG and $FliF_C$ in these chapters was a strong indication that the protein was well-folded, as it fit with the leading picture of the C-ring. In this theory, each FliF molecule binds a single FliG[89] and though some variations on this idea predict more FliG than FliF proteins in the C-ring as a whole, the extra FliG proteins are thought

to interact only with other FliG proteins and not with the underlying FliF ring itself[79]. Thus five peptides on a five-site template would be expected to interact with five FliG proteins, and my gels showed they did. Another theory however predicted that binding of FliG to FliF would induce a conformational change which would trigger FliG-FliG oligomerization[91]. My results showed no change in FliG's oligomeric state whether it was on its own, bound to $FliF_C$ in solution, or to template-bound $FliF_C$-DNA conjugates. In all cases, if the $FliF_C$ peptide was present, FliG bound only to it and nothing else, including neighbouring FliG molecules.

There are several possible ways in which these results can be explained. The first is that the FliG molecules are not in fact interacting with one another and that something in the *in vitro* set-up is amiss and therefore preventing the interactions from occurring. It could be that even the Hexagon structure was not close enough to the geometries and architecture of the *in vivo* template formed by FliF proteins and that for interactions to occur, a different diameter or stiffness or a greater continuity in the binding sites was needed. Another alternative is that something is wrong with the aqueous environment in which FliG sits, with some parameter such as the pH or ionic strength not as it is in the bacterium's cytoplasm or some small molecule or ion missing which is necessary for intermolecular FliG binding to take place. One final possibility is that a key binding partner of FliG, which in some way triggers polymerization, must be added to the solution before a ring can form. Although the FliM-FliN complex was added, it may be that the complex was over-expressed and improperly folded, meaning that with a cleaner sample, FliM-triggered FliG polymerization might be observed. It could also be that a different, weaker-binding and less well-known protein is needed, but listing the possible candidates and then including options for various templates, conditions, and small molecule additives shows that the possibilities and combinations are nearly endless.

Another major option is that the FliG proteins are in fact interacting with one another but that the techniques being used are not sensitive or gentle enough to observe them. Many observations were made using native polyacrylamide gels and it is possible that the electric fields, salts, or the cross-linked matrix of the gel itself in some way can disrupt FliG-FliG interactions. The AFM

images obtained also do not clearly show any FliG oligomerization but this is perhaps not surprising as there would have to be an unmistakeable visual signal, like a runaway polymer or a fully formed ring, to definitively conclude that the proteins were engaging in intermolecular binding. Other more indirect techniques for detecting protein-protein binding such as through surface plasmon resonance (SPR) might be able to provide evidence of an interaction that gels and AFM missed, but most would not give any indication of the stoichiometry or geometry of the proteins and thus a great deal of work would be needed to definitively demonstrate the DNA-templated formation of a FliG ring with such indirect measures. Proving the interactions were based on domain-swapping of the ARM motifs would be even more difficult as it would need a finer-grained level of detail which in turn would require more sophisticated structural techniques.

The last way in which these results can be explained is that some aspect of the domain-swap hypothesis is incorrect and that the signatures of interactions I was searching for, such as gap-filling and runaway polymerization, do not occur with the FliG protein. As I noted in the introduction, the key piece of evidence supporting template-induced oligomerization of FliG *in vivo* is now discredited, as differences in the oligomer state between motor-bound and cytoplasmic FliG seen using double-cysteine cross-linking mutants[101] can be simply explained by the dilution of the cytoplasmic fraction[104]. As my results also show no signs of template-induced oligomerization, more thought should perhaps be given to the possibility that FliG rings do not take on an interlocked, domain-swapped configuration. Perhaps the ARM motifs of each FliG monomer interact with one another in a compact, intramolecular fashion and the interaction between neighbouring FliG molecules is via a weaker, more breakable binding site such as those proposed to be revealed at the N-terminal domain upon the binding of $FliF_C$ to FliG[91]. Testing for this non-domain-swapped FliG architecture would take finer-grained structural techniques than those used in this thesis but if evidence were found that the contacts between neighbouring FliG proteins are in fact relatively transient and unstable, it could help to explain the results of my work.

Thus, the lack of evidence of FliG-FliG interactions either stemmed from a real absence of FliG oligomers in solution or from the use of techniques which were not sensitive enough to detect

them. The list of possible additives and alterations which might induce FliG polymerization appears prohibitively long, while the other techniques that might be used to pick out FliG-FliG interactions are for the most part indirect and provide little structural information from which ring formation could be inferred. That being said, though there might not be a clear way forward for the project as a whole, several loose ends and intriguing possibilities are still left to be tied up and tested, and in the final section, I will describe the work that could still be undertaken to further investigate this *in vitro* system.

6.3 Future Directions

The most obvious continuations for this project stem from the unfinished stories left by the experimental systems I have already worked with. Working from what has already been established, tying up these loose ends could be a straightforward way to advance this work and to come to a more definite conclusion as to the *in vitro* behaviour of the FliG protein.

The most promising leads to pick up on come from Chapter 5 concerning my work with the Hexagon DNA nanostructure. The unsolved question I was left with there was how to purify the protein-peptide-Hexagon complex. Agarose gels seemed to indicate that I had succeeded in obtaining a structure in which all the binding sites for $FliF_C$ and FliG were filled, but I would need to use the AFM in order to determine whether or not 30 bound FliG proteins, approximately organized in a circle roughly 35nm in diameter, would form a ring. Unpurified samples with the concentrations of $FliF_C$ and FliG that I had added would form a dense protein carpet on the mica surface meaning that no structural information could be picked out. Purification was therefore essential in order to determine the final form of the assemblies, but attempts with various techniques had led to the loss of the sample.

More elaborate purification strategies could be tested such as affinity purification using streptavidin-coated beads. It would be quite straightforward to incorporate a biotin-labelled strand into the Hexagon structure, and using this to bind a protein-peptide-Hexagon assembly to the

beads would allow for the excess protein and peptide to be washed away. Elution and then AFM imaging of the eluted complexes would reveal if the purification had been successful and if any FliG rings had formed.

Another possibility would be to avoid purification altogether and instead to search for a way in which to bind only Hexagon-containing assemblies and not free proteins and peptides to the mica surface for AFM imaging. Perhaps a certain coating or additive with a specific charge or hydrophobicity could be included which encouraged only DNA and not amino acids to attach to the imaging surface. The hope would be that this altered surface would not in any way impact the interactions between FliG molecules but could provide a way for the complexes assembled on the Hexagons to be visualized without being obscured by a carpet of unbound molecules.

Aside from continuing the work of Chapter 5, there are still possibilities to investigate with the FliM-FliN complex. The purification and expression of the complex presented a challenge and in the end only after reducing the speed of the centrifugation step could I isolate the FliM-FliN that had previously pelleted with the membranes. It is possible that these purified complexes had been overexpressed and formed inclusion bodies in the bacteria in which the constituent proteins could have been denatured to some extent. Although native polyacrylamide gels demonstrated that the FliM and FliN proteins still formed a single complex, perhaps the binding sites for FliG on the FliM proteins had been misfolded and this was why I had not observed FliG-FliM binding.

To further this work, the complexes would have to be expressed and purified again using a lower level of the inducer molecule, with the hope that they would remain in their non-pelleting, cytoplasmic form as FliG does. These newly-purified complexes could then be added to both free-floating FliG as well as FliG bound to linear templates as had been done with the previous batch. They could also be added to Hexagon-templated FliG, something that has yet to be tried, to see if binding between FliM and FliG or FliM-induced FliG polymerization could be observed, most likely through the use of agarose gels or a pull-down assay. If an interaction was observed, AFM imaging could be conducted to determine if larger, C-ring-like assemblies had formed.

Working though outstanding issues with the Hexagon and the FliM-FliN complex would be the most direct continuation of the experiments I have presented but there are also several extensions to the work which would take a little more development and preparation but could be useful in investigating FliG's behaviour.

One key unexplored idea is to use variants of the FliG protein developed by the lab of Lawrence Lee which introduce mutations in the hydrophobic patches of either the ARM_M or ARM_C motif and prevent them from forming an ARM-ARM complex. These mutations discourage the molecule from taking its closed form in which the ARM motifs interact intramolecularly, and instead encourage the molecule to take its extended form, more amenable to oligomerization. One idea would be to attach one of these mutants to DNA templates and to use it as a nucleus or starting point for FliG oligomerization. Binding wild-type FliG molecules beside this mutant with an accessible ARM motif could lower the energy barrier for runaway polymerization or FliG-ring formation. A second idea would be to use both the ARM_M-mutated construct and the ARM_C-mutated construct, placing one at either end of a linear DNA template. Providing both a starting and ending point for an oligomer via the free ARM motifs of the mutants could encourage neighbouring FliG proteins to interact and make gap-filling a more likely prospect. Either runaway polymerization or gap-filling could be observed on a gel or on the AFM and would provide clear evidence of the domain-swap model for FliG polymerization.

Aside from these new mutants, other visualization techniques could be explored to look at the systems already developed. Although the conditions during gel-running or the moving tip of an AFM might disturb any FliG-FliG interactions that are present, perhaps cryo-electron microscopy of the protein-peptide-Hexagon structures could preserve any interactions present and produce higher resolution images of the complexes. A great deal of work would have to be done in learning this or other such new techniques and preparing the samples for them, but it might be worthwhile if the FliG rings could be observed.

Thus, there are quite a few possible continuations and extensions to the work presented which might help to test the validity of the domain-swap hypothesis, and to bring the goal of the DNA-

templated assembly of the bacterial flagellar motor's cytoplasmic ring closer to fruition. More broadly however, the most intriguing direction that could be taken with this system is perhaps to apply the same DNA-templating approach to other protein complexes. As mentioned in the introduction, work has just begun in this area, with groups studying proteins like α-hemolysin[144], histones[142], and nucleoporins[143] using DNA nanostructures. These proteins naturally interact with one another without a template, and hence the templating of a FliG ring would be a new step in the development of the technique. On a more general level however, taking advantage of DNA nanotechnology to spatially organize proteins seems like a promising way for a great many questions concerning protein function to be answered and it would not be surprising to see this approach adopted on a much wider scale in the coming years.

MATERIALS

BFM PROTEIN SEQUENCES

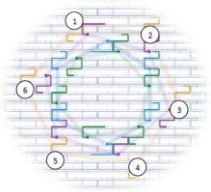
E.coli FliG	MGHHHHHHHHHHLEVLFQGPEFSNLTGTDKSVILLMTIGEDRAAEVFKHLSQREVQTLSA AMANVTQISNKQLTDVLAEFEQEAEQFAALNINANDYLRSVLVKALGEERAASLLEDILETR DTASGIETLNFMEPQSAADLIRDEHPQIIATILVHLKRAQAADILALFDERLRHDVMLRIATFG GVQPAALAELTEVLNGLLDGQNLKRSKMGGVRTAAEIINLMKTQQEEAVITAVREFDGELA QKIIDEMFLFENLVDVDDRSIQRLLQEVDSESLLIALKGAEQPLREKFLRNMSQRAADILRD DLANRGPVRLSQVENEQKAILLIVRRLAETGEMVIGSGEDTYV
Salmonella FliG	MGHHHHHHHHHHLEVLFQGPEFSNLSGTDKSVILLMTIGEDRAAEVFKHLSTREVQALST AMANVRQISNKQLTDVLSEFEQEAEQFAALNINANEYLRSVLVKALGEERASSLLEDILETR DTTSGIETLNFMEPQSAADLIRDEHPQIIATILVHLKRSQAADILALFDERLRHDVMLRIATFG GVQPAALAELTEVLNGLLDGQNLKRSKMGGVRTAAEIINLMKTQQEEAVITAVREFDGELA QKIIDEMFLFENLVDVDDRSIQRLLQEVDSESLLIALKGAEPPLREKFLRNMSQRAADILRD DLANRGPVRLSQVENEQKAILLIVRRLAETGEMVIGSGEDTYV
E.coli $FliG_{nm}$	MGHHHHHHHHHHLEVLFQGPEFSNLTGTDKSVILLMTIGEDRAAEVFKHLSQREVQTLSA AMANVTQISNKQLTDVLAEFEQEAEQFAALNINANDYLRSVLVKALGEERAASLLEDILETR DTASGIETLNFMEPQSAADLIRDEHPQIIATILVHLKRAQAADILALFDERLRHDVMLRIATFG GVQPAALAELTEVLNGLLDGQNLK
$FliF_C$	RANQRLGAEVMSQRIREMSDNDPRVVALVIRQWINNDHE
Cysteine $FliF_C$	CGSRANQRLGAEVMSQRIREMSDNDPRVVALVIRQWINNDHE
Salmonella FliM	MGHHHHHHHHHHSSGHIDDDDKHMGDSILSQAEIDALLNGDSDTKDEPTPGIASDSDIRP YDPNTQRRVVRERLQALEIINERFARQFRMGLFNLLRRSPDITVGAIRIQPYHEFARNLPVP TNLNLIHLKPLRGTGLVVFSPSLVFIAVDNLFGGDGRFPTKVEGREFTHTEQRVINRMLKLA LEGYSDAWKAINPLEVEYVRSEMQVKFTNITTSPNDIVVNTPFHVEIGNLTGEFNICLPFSMI EPLRELLVNPPLENSRHEDQNWRDNLVRQVQHSELELVANFADIPLRLSQILKLKPGDVLP IEKPDRIIAHVDGVPVLTSQYGTVNGQYALRVEHLINPILNSLNEEQPK
Salmonella FliN	MSDMNNPSDENTGALDDLWADALNEQKATTTKSAADAVFQQLGGGDVSGAMQDIDLIMD IPVKLTVELGRTRMTIKELLRLTQGSVVALDGLAGEPLDILINGYLIAQGEVVVVADKYGVRI TDIITPSERMRRLSR

Engineered residues in red, mutation sites **blue**.

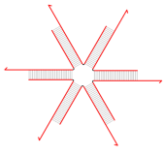
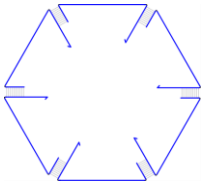
PURIFIED PROTEIN CONSTRUCTS

Protein	Species	Mutant	Length (a.a.)	MW (kDa)	Notes
FliG	*E. coli*	WT	352	39.3	
	E. coli	E327C	352	39.3	Dye-Label
	E. coli	L159C/V218C	352	39.3	Cross-Linking
	Salmonella	WT	352	39.4	
	Salmonella	S5C	352	39.4	Dye/DNA-Label
	Salmonella	E327C	352	39.4	Dye-Label
$FliG_{nm}$	*E. coli*	L159C	210	23.3	Cross-Linking
$FliF_C$	*E. coli/Salmonella*	WT	39	4.6	
	E. coli/Salmonella	Cysteine	42	4.8	DNA-Label
FliM	*Salmonella*	WT	357	40.6	Complex
FliN	*Salmonella*	WT	137	14.8	

DNA SEQUENCES

	Strand	Sequence	Length (nt)	Notes
DTPC	Guide	NH_2-NH_2-NH_2-CCGTCCTGCCTCTAATCTAA	20	NTA-Modified
	Reactive	GGCAGGACGG-NH_2	10	NHS-Ester
	Displacer	TTAGATTAGAGGCAGGACGG	20	
Templating	Pentakis Guide	NH_2-NH_2-NH_2-NH_2-NH_2-CCGTCCTGCCTCTAATCTAA	20	NTA-Modified
	T3Guide	CGCCTGAAGCTCTGGGCGTTTTAGATTAGAGGCAGGACGGTTAGATTAGAGGCAGGACGGTTAGATTAGAGGCAGGACGG	80	3-Site Guide Template
	N10 Handle	NH_2-GGCAGGACGG	10	Maleimide-Modified
	N10 Handle + Dye	NH_2-GGCAGGACGG-**FAM**	10	Maleimide-Modified
	dN10	GGCGAGTAGG	10	Gap-Filler
	T5N10	GTTATTGTCTCTGTCCTGTTTCCGTCCTGCCCCGTCCTGCCCCGTCCTGCCCCGTCCTGCCCCGTCCTGCC	71	5-Site Template
	T4G1	GTTATTGTCTCTGTCCTGTTTCCGTCCTGCCCCGTCCTGCCCCTACTCGCCCCGTCCTGCCCCGTCCTGCC	71	4-Site Template with Gap
	T4d1	GTTATTGTCTCTGTCCTGTTTCCGTCCTGCCCCGTCCTGCCCCGTCCTGCCCCGTCCTGCCCCTACTCGCC	71	4-Site Template
	T3G2	GTTATTGTCTCTGTCCTGTTTCCGTCCTGCCCCTACTCGCCCCGTCCTGCCCCTACTCGCCCCGTCCTGCC	71	3-Site Template with Gaps
	T3d2	GTTATTGTCTCTGTCCTGTTTCCGTCCTGCCCCGTCCTGCCCCGTCCTGCCCCTACTCGCCCCCTACTCGCC	71	3-Site Template
	T2d3	GTTATTGTCTCTGTCCTGTTTCCGTCCTGCCCCGTCCTGCCCCTACTCGCCCCTACTCGCCCCCTACTCGCC	71	2-Site Template
	T1d4	GTTATTGTCTCTGTCCTGTTTCCGTCCTGCCCCTACTCGCCCCTACTCGCCCCCTACTCGCCCCCTACTCGCC	71	1-Site Template
	T1-5 Label	AAACAGGACAGAGACAATAAC-**FAM**	21	Template Label
Origami Tile	Ring Staple 1	CAGGACAGAGACAATAAC**TTT**CCAAGCGCGCCTGATAAATTGTGCCCTGACG	52	Extended Staples to Bind to Template Strands Bearing FliG-DNA Conjugates 
	Ring Staple 2	CAGGACAGAGACAATAAC**TTT**CCGCCACCTGATATTCACAAACAACCCCCTGC	53	
	Ring Staple 3	CAGGACAGAGACAATAAC**TTT**TTCCTTATACCTCCCGACTTGCGGAGCCTAA	52	
	Ring Staple 4	CAGGACAGAGACAATAAC**TTT**TGGATTATACAAAGAAACCACCAGAATCAACA	53	
	Ring Staple 5	CAGGACAGAGACAATAAC**TTT**GCCTCTTCCCGGCACCGCTTCTGGTCTGGCCT	53	
	Ring Staple 6	CAGGACAGAGACAATAAC**TTT**GCGGATGGTGCATCAAAAAGATTAGATAGCGT	53	
	Ring Staple 1 – N10	CCGTCCTGCC**TTT**CCAAGCGCGCCTGATAAATTGTGCCCTGACG	44	Binds Single Conjugate
	Ring Staple 1 – Biotin	Biotin-**TTT**CCAAGCGCGCCTGATAAATTGTGCCCTGACG	34	Biotinylated Staples to Bind to Streptavidin Proteins
	Ring Staple 2 – Biotin	Biotin-**TTT**CCGCCACCTGATATTCACAAACAACCCCCTGC	35	
	Ring Staple 3- Biotin	Biotin-**TTT**TTCCTTATACCTCCCGACTTGCGGAGCCTAA	34	
	Ring Staple 4- Biotin	Biotin-**TTT**TGGATTATACAAAGAAACCACCAGAATCAACA	35	
	Ring Staple 5- Biotin	Biotin-**TTT**GCCTCTTCCCGGCACCGCTTCTGGTCTGGCCT	35	
	Ring Staple 6- Biotin	Biotin-**TTT**GCGGATGGTGCATCAAAAAGATTAGATAGCGT	35	
	For the sequences of the scaffold strand and other staple strands, see (178).			

DNA SEQUENCES (CONTINUED)

	Strand	Sequence	Length (nt)	Notes
The Hexagon	BraceAB	CAGTGGCCGCTGAACACTCCACGCGC**TTT**GCGCGCACCAGCTGCTCC CTGGCACGGCGGTCGTTCTCGC	69	
	BraceBC	CGAGCGCAGAGCGAGTCGGCCGTCGC**TTT**GCGCGTGGAGTGTTCAGC GGCCACTGCGGACAGGACTCAC	69	
	BraceCD	CTTGCACCATCGAGAACCGGTGGCGC**TTT**GCGACGGCCGACTCGCTC TGCGCTCGCAGAGTCCATGCGC	69	
	BraceDE	CAGCCTGCATGCCAGTGAGGGTGCGC**TTT**GCGCCACCGGTTCTCGAT GGTGCAAGCGGTGCACCCAGAG	69	
	BraceEF	CCCAGGTGCGAGGCTGGAGTGCCCGC**TTT**GCGCACCCTCACTGGCAT GCAGGCTGCCAGTGGAGCACTC	69	
	BraceFA	CGTGCCAGGGAGCAGCTGGTGCGCGC**TTT**GCGGGCACTCCAGCCTCG CACCTGGGCAGACTGGCTAGCG	69	
	RimAB	CAGACAGCGC**TTT**CCGTCCTGCCCCGTCCTGCCCCGTCCTGCCCCGT CCTGCCCCGTCCTGCC**TTT**GCCGCACCACGTGAGTCCTGTCCG	90	
	RimBC	GTGGTGCGGC**TTT**CCGTCCTGCCCCGTCCTGCCCCGTCCTGCCCCGT CCTGCCCCGTCCTGCC**TTT**GTGCCGCTTCGCGCATGGACTCTG	90	
	RimCD	GAAGCGGCAC**TTT**CCGTCCTGCCCCGTCCTGCCCCGTCCTGCCCCGT CCTGCCCCGTCCTGCC**TTT**CGGCCGACACCTCTGGGTGCACCG	90	
	RimDE	GTGTCGGCCG**TTT**CCGTCCTGCCCCGTCCTGCCCCGTCCTGCCCCGT CCTGCCCCGTCCTGCC**TTT**CGCGAGCACCGAGTGCTCCACTGG	90	
	RimEF	GGTGCTCGCG**TTT**CCGTCCTGCCCCGTCCTGCCCCGTCCTGCCCCGT CCTGCCCCGTCCTGCC**TTT**CGCGCCTGCGCGCTAGCCAGTCTG	90	
	RimFA	CGCAGGCGCG**TTT**CCGTCCTGCCCCGTCCTGCCCCGTCCTGCCCCGT CCTGCCCCGTCCTGCC**TTT**GCGCTGTCTGGCGAGAACGACCGC	90	
	N10	GGCAGGACGG	10	Makes Rims Double-Stranded

A NOTE ON BUFFERS

For most of the work presented in this thesis, a single standard buffer was used. Referred to simply as the **EPPS Buffer**, the composition is as follows:

- 50mM EPPS (4-(2-Hydroxyethyl)-1-piperazinepropanesulfonic acid)
- 150mM NaCl
- pH 7.5 (using concentrated NaOH).

Small additions were made to the buffer for certain experiments. For example, when storing purified protein aliquots,10% v/v glycerol was added to the solution. When folding origami tiles, or binding conjugates to templates,10mM $MgCl_2$ was added.

Other buffers required for specific techniques (such as AFM or CD spectroscopy) were made as needed and are mentioned throughout the methods which follow.

METHODS

The methods described are listed chronologically as they appear in the main text. For easier reference they are also divided by the chapter in which they are first used although some are reused in later chapters following their first appearance.

CHAPTER TWO

NTA Modification of DNA

Tris-aminated DNA (80μM, 100μL) was reacted with N-succinimidyl 3-(2-pyridyldithio)propionate (SPDP) in dimethylsulfoxide (DMSO, 100mM, 25μL) and trimethylamine (TEA, 0.75μL) for one hour at room temperature. The solution was then diluted with HEPES (800μL, 500mM, pH 7.5) and excess reagents removed using Amicon Ultra centrifugal filters (MWCO 3000, 14100xg for 30 minutes). The retentate was collected and treated with tris(2-carboxyethyl)phosphine (TCEP, 150mM, in 500μl 500mM HEPES buffer, pH 7.5) for 15 minutes at room temperature. The excess reagents were removed using Amicon Ultra centrifugal filters (MWCO 3000, 14100xg for 30 minutes). The retentate was washed twice with HEPES (50mM, pH 7.5), before it was reacted for one hour at room temperature with NTA maleimide (Dojindo Molecular Technologies, 70mM, in 10μl 25mM HEPES, pH 7.5) in a final volume of approximately 40μL. The DNA was precipitated by adding sodium acetate (3M, 6μL, pH 5.2), glycogen (20mg/mL, 1μL), and cold ethanol (100%, 100μL) to the reaction mix, incubating at -80°C for 15 minutes, then centrifuging (4°C, 20000xg for 45 minutes). The pellet was stored at -20°C for later use.

Functionality Assay for the NTA Modification Using TIRF Microscopy

DNA modified using the above protocol (20μM, 1μL) was incubated with nickel sulfate (500μM, 2μL) in sodium phosphate buffer (10mM, 17μL) for one hour at room temperature and then passed through a size exclusion column (Bio-Rad Micro Bio Spin 6, 1000g, 2 minutes) to remove excess nickel. To the nickel-incubated NTA-DNA was then added a biotinylated complementary strand (5μM, 1μL), biotin (5μM, 1μL), and purified FliG protein labelled with an Alexa-647 dye (14μM, 2.14μL) and the mixture left for one hour at room temperature. The mix was then diluted with phosphate buffer (10mM, 216μL) and pipetted into the space between a microscope slide and a coverslip functionalized with polyethylene glycol-biotin incubated with streptavidin. Unbound species were washed out of the chamber before imaging for FliG with a 633nm laser set to TIRF. Images were collected with an Andor CCD camera. Control experiments were performed as above but with a Cy5 fluorophore labelled DNA strand in place of the NTA-DNA and FliG.

Plasmid Amplification

Plasmids received from the lab of Lawrence Lee on filter paper were suspended in buffer and 1pg to 100ng of the DNA was transformed into competent bacterial cells (NEB5-α, 50μL) by heat shock following the manufacturer's protocol. Transformed bacteria were plated on LB (Luria broth)-agar with the appropriate antibiotic and incubated overnight at 37°C. Single colonies were inoculated in 3mL of LB with antibiotic and grown for 8 hours at 37°C before 300μL of this starter culture was added to 100mL of LB with antibiotic. These cultures were left to grow overnight shaking at 250rpm at 37°C. Plasmids were extracted from the bacterial cells through the use of the protocol and reagents supplied in Qiagen Plasmid Maxi Kits. The isopropanol-precipitated pellet containing the plasmid was stored at -20°C and dissolved in TE (Tris-EDTA) buffer when required for protein expression.

FliG Expression

Around 100ng of plasmid was transformed into NEB T7 Express competent *E. coli* cells (50μL) through heat shock following the manufacturer's instructions. The cells were plated on LB-agar with the appropriate antibiotic and left to grow at 37°C overnight. 25mL of LB with antibiotic was then inoculated with a single colony and grown overnight at 37°C. 10mL of this starter culture was then used to inoculate 2L of LB with antibiotic and grown at 37°C to an optical density (OD_{600}) of 0.3-0.4. The temperature was then decreased to 18°C until an OD of 0.6-0.8 was reached, at which point protein expression was induced through the addition of IPTG (Isopropyl β-D-1-thiogalactopyranoside) to 400μM. The cells were left at 18°C overnight for protein expression and their OD measured the following morning. They were then harvested via centrifugation (4000xg, 20min, 4°C) and the cell pellet resuspended in Lysis Buffer (20mM Tris, 100mM

NaCl, pH 7.5, 10mL per 1L at 1 OD). The suspended cells were then broken apart using a Dounce homogenizer before being lysed using a French press. Cellular debris was removed through centrifugation (10000xg, 20mins, 4°C) and the cytoplasmic fraction was separated from membranes by spinning in the ultracentrifuge (440000xg, 1hr, 4°C). The cytoplasm in the supernatant was flash-frozen in liquid nitrogen and stored at -80°C until further use.

FliG Purification and Dye-Labelling

50mL aliquots of clarified lysate were thawed on ice before imidazole and TCEP were added to 10mM and 200µM respectively. 400µL of IMAC resin, pre-incubated with nickel and washed with the lysis buffer, was then mixed with the lysate and left rotating (2hrs, 4°C) to allow the His-tagged FliG to bind the resin. The resin was then pelleted in the centrifuge (500xg, 5mins, 4°C) and the supernatant removed. The resin was then transferred to an empty gravity flow column and washed first with the lysis buffer containing 300mM NaCl and 20mM imidazole and then with lysis buffer containing 300mM NaCl and no imidazole. If the construct was to be dye-labelled, an aliquot of the resin was removed and run on an SDS-PAGE gel (Bolt Bis-Tris, 4-12%) and compared with a series of BSA samples of known concentration through Coomassie staining to obtain an idea of the amount of protein present. TCEP was then added to a final concentration of 200µM, and the dye added to the protein in a 4:1 stoichiometric excess. The resin was left overnight at 4°C in a Falcon tube and excess dye removed by three rounds of pelleting the resin, washing with lysis buffer, and removing the supernatant (500xg, 5mins, 4°C). The resin was then transferred to a gravity flow column and the protein was eluted from the column through the addition of 1.5mL of lysis buffer with 500mM imidazole and 10mM EDTA added. If required, the eluted protein was buffer exchanged into the appropriate buffer using NAP-10 gravity columns. The protein was then aliquoted and glycerol added to 10% v/v before being flash-frozen in liquid nitrogen and stored at -80°C for further use. An aliquot was reserved for use in a BCA Assay, in which the concentration of the protein was determined by following the manufacturer's protocol.

Reactive Strand Modification

To a solution of DNA (15nmol) in water (75µL) was added the cross-linker disuccinimidyl suberate (DSS) in DMSO (50mM, 75µL), MeCN (75µL) and TEA (1µL). The mixture was shaken for 30 minutes at room temperature. The DNA was then passed through a size exclusion column (Bio-Rad Micro Bio Spin 6, 1000xg, 2 minutes) to remove excess DSS and then precipitated by adding sodium acetate (3M, 28µL, pH 5.2), glycogen (20mg/mL, 1µL), and cold ethanol (100%, 565µL) to the reaction mix, incubating at -80°C for 15 minutes, then centrifuging (4°C, 20 000xg for 1 hour). The pellet was resuspended in 40µL of EPPS reaction buffer (50mM EPPS, 400mM NaCl, 0.02%v/v Tween-20, pH 7.7) and used immediately for the conjugation reaction and functionality assay.

DNA-Templated Conjugation Reaction

Reactions were performed by mixing the guide strand (0.25µM) with FliG (0.25µM), NiCl2 (0.25µM), and freshly prepared reactive strand (0.25µM) in EPPS buffer (50mM, pH 7.7) containing NaCl (400mM) and Tween-20 (0.02 v/v%). The mixture was left to react overnight at room temperature. After the overnight reaction, EDTA (2.5mM) and the displacer strand complementary to the full sequence of the guide strand (2 molar equivalents) was added and left to displace the reacting ON at room temperature for 30 minutes. Larger reaction volumes were then concentrated using Amicon Ultra centrifugal filters (MWCO 3000, 14100xg for 30 minutes) and washed with water (using the same filters) before PAGE analysis and/or purification. Aliquots not purified or used in PAGE analysis were kept at -20°C for storage.

Templating Gels with NTA-DNA

Either tris-NTA-modified or pentakis-NTA-modified DNA was mixed, to a final concentration of 1µM, with T3Guide (330nM), $NiCl_2$ (20µM), and dye-labelled FliG (3µM) in EPPS Buffer with 10mM $MgCl_2$ added. The mix was incubated for one hour at 4°C and an aliquot with 1-3pmol FliG then loaded on a native polyacrylamide gel (6%, 19:1 acrylamide:bisacrylamide, in 1xTAE buffer). Adequate separation of the components was achieved by running at 200V for 30-60 minutes at 4°C. The gel was stained with SYBR Gold following the manufacturer's instructions and then imaged for the dye-labelled protein signal and the stained DNA using a gel scanner with the appropriate filters.

Maleimide Modification of DNA

An aminated DNA strand dissolved in EPPS Buffer at pH 7.2 was mixed with the sulfo-SMCC cross-linker solubilized in DMF such that the final concentration of the DNA was 500μM and the sulfo-SMCC was in a 100:1 stoichiometric excess to the DNA. The reaction was left shaking at room temperature for 30mins before buffer exchanging into MilliQ water using a Micro Bio-Spin column. The modified DNA was then purified via RP-HPLC using a C_{18} column (Buffer A: 0.1M TEAA, 5% v/v MeCN, Buffer B: 0.1M TEAA, 70% v/v MeCN). The peak corresponding to the maleimide-modified DNA was collected and vacuum-concentrated to a pellet. The pellet was stored at -20°C for further use.

Maleimide to Cysteine Conjugation Reaction

Pellets of maleimide-modified DNA were dissolved and then exchanged into EPPS Buffer using Micro Bio-Spin columns. Aliquots of single-cysteine FliG were thawed and TCEP was added to a final concentration of 5mM. The DNA was added to the protein in a 6:1 stoichiometric excess and the reaction left for 4hrs at 4°C. Excess, unreacted DNA was removed either by passing the mixture through Microspin G-50 spin columns or by mixing with nickel-loaded IMAC resin, pelleting, removing the supernatant, and then eluting with imidazole. The final concentration of the conjugate was determined roughly by running on a native polyacrylamide gel alongside DNA standards of known concentrations, and more precisely through the use of a BCA assay.

Removal of Histidine-Tags from FliG Constructs

Aliquots of FliG constructs were thawed and then incubated at 37°C for 30mins in a thermocycler. The PreScission Protease was thawed and, once the FliG incubation period was complete, was added to the protein at 1μL of protease for every 20μL of FliG. The cleavage reaction was left to take place at room temperature overnight. A glutathione agarose resin to bind the protease's GST-tag was added following the reaction and the whole mixture then passed through a G-50 spin column which separated the resin-bound protease from the soluble His-cleaved FliG. Aliquots of the His-cleaved proteins were flash-frozen in liquid nitrogen and stored at -80°C for further use. Their final concentration was determined using a BCA assay and the extent of His-tag cleavage was measured using an SDS-PAGE gel capable of separating His-cleaved and non-His-cleaved proteins by size.

Templating Gels with Maleimide-Cysteine FliG-DNA Conjugates

FliG-DNA conjugates were mixed, to a final concentration of 1μM, with the appropriate variant of the five-site T5N10 template in EPPS Buffer with 10mM $MgCl_2$ added such that there were 3 conjugates present for every binding site in solution. If desired, the gap-filling dN10 strand was added in a 1.5:1 stoichiometry to its binding sites on the template. The mix was incubated for one hour at 4°C and an aliquot with 1-3pmol FliG then loaded on a native polyacrylamide gel (6%, 19:1 acrylamide:bisacrylamide, in 1xTAE buffer). Adequate separation of the components was achieved by running at 200V for 30-60 minutes at 4°C. The gel was stained with SYBR Gold if needed and then imaged for either the fluorescent labels on the DNA tag of the conjugate or the SYBR-Gold-stained DNA using a gel scanner with the appropriate filters.

Origami Tile Assembly and Purification

In a 1xTE buffer with 12.5mM $MgCl_2$, the M13mp18 scaffold strand was added to the appropriate staple strands and extended ring staple strands in a 10:1 excess. These were annealed from 96°C to 20°C in a thermocycler decreasing in temperature at 1°C/min and then stored at 4°C. An aliquot of the origami was run on an agarose gel (0.75%, 1xTAE+12.5mM $MgCl_2$, 80V, 90mins) alongside M13mp18 scaffold strand to determine if the structure had folded. If folded, the full volume was then loaded on an agarose gel for gel extraction purification. Following electrophoresis (0.75% agarose, 1xTAE+12.5mM $MgCl_2$, 80V, 90mins), the gel was stained with SYBR Gold and the folded origami band excised from the gel using a scalpel and a UV transilluminator. Liquid was squeezed out of the gel slice manually, collected, and then an aliquot run on an agarose gel to confirm the purification had been successful. Excess template and FliG-DNA conjugates were then added to these purified tiles if required.

AFM Imaging of the Origami Tile

Origami tiles or tiles with protein bound were diluted to 5nM in imaging buffer (1xTE, 12.5mM $MgCl_2$, 150mM NaCl, 4mM $NiCl_2$) and 5µL was pipetted onto a disc of freshly-cleaved mica and left for 5mins to allow the tiles to bind to the surface. An extra 95µL of imaging buffer was then added to cover the sample and submerge the Bruker AC40 AFM probe and the tiles were then imaged using a Cypher S AFM in tapping mode in liquid.

Double-Cysteine Cross-Linking of FliG

Double-cysteine FliG mutants were purified as described previously, and mixed, if the experiment required it, with nickel, NTA-modified N10 strands, and five-site templates, using the concentrations and incubations described earlier. Standard cross-linking was carried out with FliG at 500nM or 1µM. The reaction involved adding $CuSO_4$ to 1,10-phenanthroline and then adding these combined components to FliG for final concentrations of 400µM $CuSO_4$ and 1.6mM 1,10-phenanthroline. The reaction was left for 30 seconds at room temperature before quenching with 20mM N-ethylmaleimide. Aliquots from the reaction were then loaded on denaturing SDS-PAGE gels (Bolt Bis-Tris, 4-12%), with the desired separation of bands achieved after 30-60mins running at 200V at room temperature. Gel were imaged after staining for total protein with SYPRO Ruby following the manufacturer's instructions using a gel scanner with the appropriate filters. Some cross-linking reactions were performed with 200µM iodine rather than copper-phenanthroline as the oxidizer, but followed the same protocol aside from this alteration.

CHAPTER FOUR

Circular Dichroism Spectroscopy of C-Ring Proteins

Protein samples for CD spectroscopy were prepared in low salt PBS buffer (8mM K_2HPO_4, 2mM KH_2PO_4, 20mM NaCl) to a concentration of 0.1mg/mL. 200µL of each sample was loaded in a quartz cuvette with 1mm path length and CD spectra were taken from 190-260nm at 20°C using a Jasco J-815 CD spectrophotometer. The baseline was set using the low salt PBS buffer and data collected every 0.5nm. For the temperature ramp experiments, CD spectra of the samples were acquired from 200-250nm every 2°C as the solution was heated from 20°C to 90°C. The percent of each secondary structure present was calculated by inputting the acquired spectra into the K2D3 web server.

Tryptophan Fluorescence Assay for FliG-FliF$_C$ Binding

Mixtures of wild-type FliG and single-cysteine FliF$_C$ were prepared in EPPS Buffer such that FliF$_C$ was at 2µM and FliG ranged from 0-3µM at 250nM increments. The samples were left for 2hrs at room temperature and then loaded in a quartz cuvette and their emission from 310-450nm scanned by a Cary Eclipse fluorimeter after excitation at 295nm. Plotted data are an average of the values from three scans and two samples at each concentration of FliG. Control experiments using free residual tryptophan were conducted following the same protocol but with 2µM tryptophan in place of FliF$_C$.

Thermal Shift Assay for Protein Stability

2µg of each C-ring protein construct was mixed with 5µL of 50x SYPRO Orange dye and diluted to 50µL in EPPS Buffer. Samples were placed in a quantitative PCR machine, SYPRO Orange fluorophores were excited at 492nm, and their emission was monitored at 568nm every 0.5°C as the solution was heated from 25°C to 95°C at 1°C/min. Melting temperature of the constructs was determined by plotting the derivative with respect to temperature of the curves obtained and finding the minimum value of this curve. This usually corresponded to a point halfway between the peak and trough of the curve in melting transition phase. A blank sample with SYPRO Orange on its own in EPPS Buffer was also run but gave negligible signal.

Peptide to DNA Conjugation and Purification

Single-cysteine FliF$_C$ peptide (50nmol, 100µM final concentration) was mixed in a 2:1 stoichiometric excess to maleimide-modified DNA (25nmol, 50µM final concentration) in EPPS Buffer with TCEP added in to a final concentration of 1mM. The reaction was left for 2hrs at room temperature at which time an aliquot was taken and run on a native polyacrylamide gel and stained with SYBR Gold to determine the yield of the conjugation. The mixture was then purified via RP-HPLC using a C_{18} column (Buffer A: 10mM TEAA, 5% v/v MeCN, Buffer B: 10mM TEAA, 70% v/v MeCN). The peak containing the peptide-DNA conjugate was collected, vacuum-concentrated to a pellet, and stored at -20°C for later use.

Templating Gels with Peptide-DNA and Wild-Type FliG

Peptide-DNA conjugates were mixed, to a final concentration of 1μM, with the appropriate variant of the five-site T5N10 template in EPPS Buffer with 10mM $MgCl_2$ added such that there were 1.25 conjugates present for every binding site in solution. If desired, the gap-filling dN10 strand was added in a 1.5:1 stoichiometry to its binding sites on the template. If FliG was required, it was also added at this time in the appropriate ratio to $FliF_C$. The mix was incubated for two hours at 4°C and an aliquot with 1-3pmol $FliF_C$ then loaded on a native polyacrylamide gel (6%, 19:1 acrylamide:bisacrylamide, in 1xTAE buffer). Adequate separation of the components was achieved by running at 200V for 30-60 minutes at 4°C. The gel was stained with SYBR Gold and then imaged for total DNA before being stained with SYPRO Ruby and imaged for total protein using an Amersham Typhoon gel scanner with the appropriate filters.

FliM-FliN Expression and Purification

The expression and purification procedure for FliM-FliN follows that described for FliG with the following difference. Following expression of the construct by the bacterial cells and lysis using the French press chamber, the solution was centrifuged at 18000xg for 1hr to remove cellular debris and most membranes. No ultracentrifugation step was included. The supernatant was flash-frozen and later purified using IMAC resin as described for the FliG constructs.

CHAPTER FIVE

Design of the Hexagon DNA Nanostructure

The length and sequences of the 12 strands making up the Hexagon structure were finalized after an iterative cycle of designing and testing various architectures using the NUPACK web server. Sequences were designed such that they did not interfere with conjugates binding to the T5N10 sequence and such that they remained in the appropriate duplex at least to 40°C. Due to computational restraints, the actual 12 strands had to be concatenated into a smaller number of longer sequences in order to be tested by the software. These were split into the 12 strands described in the text due to restrictions from the DNA synthesis company on the length of strands that could be cheaply synthesized.

Hexagon Assembly and Purification

In EPPS Buffer with 10mM $MgCl_2$, 1.0pmol of each of the six brace strands was added to 1.2pmol of each of the six rim strands giving final concentrations of the strands as 1μM and 1.2μM respectively. These were annealed from 96°C to 20°C in a thermocycler decreasing in temperature at 1°C/min and then stored at 4°C. To purify, the full volume was then loaded on an agarose gel for gel extraction purification. Following electrophoresis (2% agarose, 1xTAE, 80V, 2hrs), the gel was stained with SYBR Gold and the complete Hexagon band excised from the gel using a scalpel and a UV transilluminator. Liquid was squeezed out of the gel slice by placing it in an empty Micro Bio-Spin column inside a collection tube and centrifuging at 1000xg for 2mins. The liquid was collected from the tube and an aliquot run on an agarose gel to confirm the purification had been successful. The purified Hexagon was stored at 4°C for later use but on most occasions was used immediately in subsequent experiments, with $FliF_C$ and FliG being added directly to the purified structure.

AFM Imaging of the Hexagon with $FliF_C$ and FliG

Hexagon structures or Hexagons with protein bound were diluted to 5nM in EPPS Buffer with 10mM $MgCl_2$ and 4mM $NiCl_2$ and 5μL of this was pipetted onto a disc of freshly-cleaved mica and left for 5mins to allow the Hexagons to bind to the surface. An extra 95μL of imaging buffer was then added to cover the sample and submerge the Bruker AC40 AFM probe and the complexes were then imaged using a Cypher S AFM in tapping mode in liquid.

Glycerol Gradient Ultracentrifugation of the Peptide-Hexagon Complex

A glycerol gradient was made in an ultracentrifuge tube by manually layering 80μL each of EPPS Buffer solutions containing from 45% to 15% glycerol at 5% intervals using a gel-loading tip. This was left upright overnight to form a semi-continuous gradient. 50μL of the peptide-Hexagon complex in EPPS Buffer with 20% glycerol was layered on top of the gradient which was then spun at 285000xg for 3hrs at 4°C in a Beckmann ultracentrifuge. The gradient was carefully removed from the centrifuge and 50μL fractions were pipetted from the top to the bottom using a gel-loading tip. These fractions were then run on an agarose gel (2% agarose, 1xTAE, 80V, 2hrs) to determine where the complexes and free peptide-DNA had migrated to.

REFERENCES

1. What is the macromolecular composition of the cell? Available at: http://book.bionumbers.org/what-is-the-macromolecular-composition-of-the-cell/. (Accessed: 14th June 2018)
2. Goodsell, D. S. Escherichia coli. *Biochem. Mol. Biol. Educ.* **37,** 325–332 (2009).
3. Gilbert, W. ORIGIN OF LIFE - THE RNA WORLD. *Nature* **319**, 618-618, doi:10.1038/319618a0 (1986).
4. Darwin, C. On the Origin of Species by Means of Natural Selection. (Murray, 1859).
5. Ponomarenko, E. A. *et al.* The Size of the Human Proteome: The Width and Depth. *Int. J. Anal. Chem.* **2016,** 1–6 (2016).
6. Sullivan, L. H. *The tall office building artistically considered*. (J.B. Lippincott Co., 1896).
7. Alberts, B. *et al.* Protein Function. (2002).
8. PDB-101: Learn: Flyers, Posters & Other Resources. Available at: https://pdb101.rcsb.org/learn/flyers-posters-and-other-resources. (Accessed: 13th June 2018)
9. Lodish, H. F. *Molecular cell biology*. (W.H. Freeman, 2000).
10. Kovacs, N. A., Petrov, A. S., Lanier, K. A. & Williams, L. D. Frozen in Time: The History of Proteins. *Mol. Biol. Evol.* **34,** 1252–1260 (2017).
11. Beltrao, P. & Serrano, L. Specificity and Evolvability in Eukaryotic Protein Interaction Networks. *PLoS Comput. Biol.* **3,** e25 (2007).
12. Gerstein, M., Lesk, A. M. & Chothia, C. Structural mechanisms for domain movements in proteins. *Biochemistry* **33,** 6739–49 (1994).
13. What are the time scales for diffusion in cells? Available at: http://book.bionumbers.org/what-are-the-time-scales-for-diffusion-in-cells/. (Accessed: 16th June 2018)
14. Levy, E. D., Pereira-Leal, J. B., Chothia, C. & Teichmann, S. A. 3D complex: a structural classification of protein complexes. *PLoS Comput. Biol.* **2,** e155 (2006).
15. PDB-101: Molecule of the Month By Date. Available at: https://pdb101.rcsb.org/motm/motm-by-date. (Accessed: 18th June 2018)
16. Pröschel, M., Detsch, R., Boccaccini, A. R. & Sonnewald, U. Engineering of Metabolic Pathways by Artificial Enzyme Channels. *Front. Bioeng. Biotechnol.* **3,** 168 (2015).
17. Good, M. C., Zalatan, J. G. & Lim, W. A. Scaffold proteins: hubs for controlling the flow of cellular information. *Science* **332,** 680–6 (2011).
18. How big is the “average” protein? Available at: http://book.bionumbers.org/how-big-is-the-average-protein/. (Accessed: 18th May 2018)
19. Warringer, J. & Blomberg, A. Evolutionary constraints on yeast protein size. *BMC Evol. Biol.* **6,** 61 (2006).
20. Zandi, R., Reguera, D., Bruinsma, R. F., Gelbart, W. M. & Rudnick, J. Origin of icosahedral symmetry in viruses. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 15556-15560, doi:10.1073/pnas.0405844101 (2004).
21. Goodsell, D. S. & Olson, A. J. Structural Symmetry and Protein Function. *Annu. Rev. Biophys. Biomol. Struct.* **29,** 105–153 (2000).
22. How many reactions do enzymes carry out each second? Available at: http://book.bionumbers.org/how-many-reactions-do-enzymes-carry-out-each-second/. (Accessed: 20th June 2018)
23. Bonetti, D., Troilo, F., Brunori, M., Longhi, S. & Gianni, S. How Robust Is the Mechanism of Folding-Upon-Binding for an Intrinsically Disordered Protein? *Biophys. J.* **114,** 1889–1894 (2018).
24. Marsh, J. A. et al. Protein Complexes Are under Evolutionary Selection to Assemble via Ordered Pathways. *Cell* **153**, 461-470, doi:10.1016/j.cell.2013.02.044 (2013).
25. Zacharias, M. Accounting for conformational changes during protein-protein docking. *Curr. Opin. Struct. Biol.* **20**, 180-186, doi:10.1016/j.sbi.2010.02.001 (2010).
26. Toyama, B. H. *et al.* Identification of Long-Lived Proteins Reveals Exceptional Stability of Essential Cellular Structures. *Cell* **154**, 971-982, doi:10.1016/j.cell.2013.07.037 (2013).
27. Sobott, F., Benesch, J. L. P., Vierling, E. & Robinson, C. V. Subunit exchange of multimeric protein complexes - Real-time monitoring of subunit exchange between small heat shock proteins by using electrospray mass spectrometry. *J. Biol. Chem.* **277**, 38921-38929, doi:10.1074/jbc.M206060200 (2002).
28. Marsh, J. A. & Teichmann, S. A. Structure, Dynamics, Assembly, and Evolution of Protein Complexes. *Annu. Rev. Biochem.* **84,** 551–575 (2015).
29. Monod, J., Wyman, J. & Changeux, J. P. ON NATURE OF ALLOSTERIC TRANSITIONS - A PLAUSIBLE MODEL. *J. Mol. Biol.* **12**, 88-&, doi:10.1016/s0022-2836(65)80285-6 (1965).
30. Bloom, L. B. Loading clamps for DNA replication and repair. *DNA Repair (Amst).* **8,** 570–8 (2009).
31. Lomakin, A. *et al.* Kinetic theory of fibrillogenesis of amyloid beta-protein. *Proc. Natl. Acad. Sci. U. S. A.* **94,** 7942–7 (1997).
32. Kojima, S. & Blair, D. F. The Bacterial Flagellar Motor: Structure and Function of a Complex Molecular Machine. in *International review of cytology* **233,** 93–134 (2004).
33. Lozupone, C. A. & Knight, R. Global patterns in bacterial diversity. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 11436–40 (2007).
34. Jarrell, K. F. & McBride, M. J. The surprisingly diverse ways that prokaryotes move. *Nat. Rev. Microbiol.* **6,** 466–476 (2008).

35. Namba, K., Yamashita, I. & Vonderviszt, F. Structure of the core and central channel of bacterial flagella. *Nature* **342,** 648–654 (1989).
36. Berg, H. C. The rotary motor of bacterial flagella. *Annu. Rev. Biochem.* **72**, 19-54, doi:10.1146/annurev.biochem.72.121801.161737 (2003).
37. Berg, J. M., Tymoczko, J. L. & Stryer, L. A Rotary Motor Drives Bacterial Motion. (2002).
38. Tomoyasu, T. *et al.* The ClpXP ATP-Dependent Protease Regulates Flagellum Synthesis in Salmonella enterica Serovar Typhimurium. *J. Bacteriol.* **184,** 645–653 (2002).
39. Magariyama, Y. *et al.* Very fast flagellar rotation. *Nature* **371,** 752–752 (1994).
40. How fast do cells move? Available at: http://book.bionumbers.org/how-fast-do-cells-move/. (Accessed: 24th June 2018)
41. Macnab, R. M. & Aizawa, S. I. BACTERIAL MOTILITY AND THE BACTERIAL FLAGELLAR MOTOR. *Annual Review of Biophysics and Bioengineering* **13**, 51-83 (1984).
42. Armitage, J. P. & Macnab, R. M. Unidirectional, intermittent rotation of the flagellum of Rhodobacter sphaeroides. *J. Bacteriol.* **169,** 514–8 (1987).
43. Sowa, Y. & Berry, R. M. Bacterial flagellar motor. *Q. Rev. Biophys.* **41,** 103–132 (2008).
44. Armitage, J. P., Pitta, T. P., Vigeant, M. A., Packer, H. L. & Ford, R. M. Transformations in flagellar structure of Rhodobacter sphaeroides and possible relationship to changes in swimming speed. *J. Bacteriol.* **181,** 4825–33 (1999).
45. Xie, L., Altindal, T., Chattopadhyay, S. & Wu, X.-L. From the Cover: Bacterial flagellum as a propeller and as a rudder for efficient chemotaxis. *Proc. Natl. Acad. Sci. U. S. A.* **108,** 2246–51 (2011).
46. Blair, D. F. How Bacteria Sense and Swim. *Annu. Rev. Microbiol.* **49,** 489–520 (1995).
47. Larsen, S. H., Reader, R. W., Kort, E. N., Tso, W.-W. & Adler, J. Change in direction of flagellar rotation is the basis of the chemotactic response in Escherichia coli. *Nature* **249,** 74–77 (1974).
48. Berg, H. C. Bacterial behaviour. *Nature* **254,** 389–392 (1975).
49. Wadhams, G. H. & Armitage, J. P. Making sense of it all: Bacterial chemotaxis. *Nat. Rev. Mol. Cell Biol.* **5**, 1024-1037, doi:10.1038/nrm1524 (2004).
50. Welch, M., Oosawa, K., Aizawa, S. & Eisenbach, M. Phosphorylation-dependent binding of a signal molecule to the flagellar switch of bacteria. *Proc. Natl. Acad. Sci. U. S. A.* **90,** 8787–91 (1993).
51. Clegg, D. O. & Koshland, D. E. The role of a signaling protein in bacterial sensing: behavioral effects of increased gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **81,** 5056–60 (1984).
52. Mesibov, R., Ordal, G. W. & Adler, J. The range of attractant concentrations for bacterial chemotaxis and the threshold and size of response over this range. Weber law and related phenomena. *J. Gen. Physiol.* **62,** 203–23 (1973).
53. Berg, H. C. The rotary motor of bacterial flagella. *Annu. Rev. Biochem.* **72**, 19-54, doi:10.1146/annurev.biochem.72.121801.161737 (2003).
54. Thomas, D. R., Francis, N. R., Xu, C. & DeRosier, D. J. The three-dimensional structure of the flagellar rotor from a clockwise-locked mutant of Salmonella enterica serovar Typhimurium. *J. Bacteriol.* **188,** 7039–48 (2006).
55. Macnab, R. M. How bacteria assemble flagella. *Annu. Rev. Microbiol.* **57**, 77-100, doi:10.1146/annurev.micro.57.030502.090832 (2003).
56. McCarter, L. L. MotX, the channel component of the sodium-type flagellar motor. *J. Bacteriol.* **176,** 5988–98 (1994).
57. McCarter, L. L. MotY, a component of the sodium-type flagellar motor. *J. Bacteriol.* **176,** 4219–25 (1994).
58. Braun, T. F., Al-Mawsawi, L. Q., Kojima, S. & Blair, D. F. Arrangement of Core Membrane Segments in the MotA/MotB Proton-Channel Complex of *Escherichia coli* . *Biochemistry* **43,** 35–45 (2004).
59. Mandadapu, K. K., Nirody, J. A., Berry, R. M. & Oster, G. Mechanics of torque generation in the bacterial flagellar motor. *Proc. Natl. Acad. Sci. U. S. A.* **112**, E4381-E4389, doi:10.1073/pnas.1501734112 (2015).
60. Lee, L. K., Ginsburg, M. A., Crovace, C., Donohoe, M. & Stock, D. Structure of the torque ring of the flagellar motor and the molecular basis for rotational switching. *Nature* **466,** 996–1000 (2010).
61. Bai, F. *et al.* Conformational Spread as a Mechanism for Cooperativity in the Bacterial Flagellar Switch. *Science* **327**, 685-689, doi:10.1126/science.1182105 (2010).
62. Kojima, S., Nonoyama, N., Takekawa, N., Fukuoka, H. & Homma, M. Mutations Targeting the C-Terminal Domain of FliG Can Disrupt Motor Assembly in the Na+-Driven Flagella of Vibrio alginolyticus. *J. Mol. Biol.* **414**, 62-74, doi:10.1016/j.jmb.2011.09.019 (2011).
63. Miyanoiri, Y. *et al.* Structural and Functional Analysis of the C-Terminal Region of FliG, an Essential Motor Component of Vibrio Na+-Driven Flagella. *Structure* **25**, 1540-+, doi:10.1016/j.str.2017.08.010 (2017).
64. Dyer, C. M., Vartanian, A. S., Zhou, H. & Dahlquist, F. W. A molecular mechanism of bacterial flagellar motor switching. *J. Mol. Biol.* **388,** 71–84 (2009).
65. Diepold, A. & Armitage, J. P. Type III secretion systems: the bacterial flagellum and the injectisome. *Philos. Trans. R. Soc. B Biol. Sci.* **370,** 20150020 (2015).
66. Abrusci, P. *et al.* Architecture of the major component of the type III secretion system export apparatus. *Nat. Struct. Mol. Biol.* **20,** 99–104 (2013).
67. Homma, M., Kutsukake, K., Hasebe, M., Iino, T. & Macnab, R. M. FlgB, FlgC, FlgF and FlgG - A family of structurally related proteins in the flagellar basal body of *Salmonella-typhimurium*. *J. Mol. Biol.* **211**, 465-477, doi:10.1016/0022-2836(90)90365-s (1990).

68. Stallmeyer, M. J., Aizawa, S., Macnab, R. M. & DeRosier, D. J. Image reconstruction of the flagellar basal body of Salmonella typhimurium. *J. Mol. Biol.* **205,** 519–28 (1989).
69. Macnab, R. M. Bacterial flagella rotating in bundles: a study in helical geometry. *Proc. Natl. Acad. Sci. U. S. A.* **74,** 221–5 (1977).
70. Erhardt, M. *et al.* The role of the FliK molecular ruler in hook-length control in *Salmonella enterica*. *Mol. Microbiol.* **75,** 1272–1284 (2010).
71. Kuwajima, G. *et al.* Export of an N-terminal fragment of Escherichia coli flagellin by a flagellum-specific pathway. *Proc. Natl. Acad. Sci.* **86,** 4953–4957 (1989).
72. Samatey, F. A. *et al.* Structure of the bacterial flagellar protofilament and implications for a switch for supercoiling. *Nature* **410,** 331–337 (2001).
73. Macnab, R. M. & Ornston, M. K. Normal to curly flagellar transitions and their role in bacterial tumbling - stabilization of an alternative quaternary structure by mechanical force. *J. Mol. Biol.* **112**, 1-30, doi:10.1016/s0022-2836(77)80153-8 (1977).
74. Armstrong, J. B. & Adler, J. Location of genes for motility and chemotaxis on the Escherichia coli genetic map. *J. Bacteriol.* **97,** 156–61 (1969).
75. Francis, N. R., Sosinsky, G. E., Thomas, D. & DeRosier, D. J. Isolation, Characterization and Structure of Bacterial Flagellar Motors Containing the Switch Complex. *J. Mol. Biol.* **235,** 1261–1270 (1994).
76. Chen, S. Y. *et al.* Structural diversity of bacterial flagellar motors. *Embo J.* **30**, 2972-2981, doi:10.1038/emboj.2011.186 (2011).
77. Beeby, M. *et al.* 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 (2016).
78. Francis, N. R., Irikura, V. M., Yamaguchi, S., DeRosier, D. J. & Macnab, R. M. Localization of the Salmonella typhimurium flagellar switch protein FliG to the cytoplasmic M-ring face of the basal body. *Proc. Natl. Acad. Sci. U. S. A.* **89,** 6304–8 (1992).
79. Thomas, D. R., Morgan, D. G. & DeRosier, D. J. Rotational symmetry of the C ring and a mechanism for the flagellar rotary motor. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 10134–9 (1999).
80. Dean, G. E., Aizawa, S. I. & Macnab, R. M. flaAII (motC, cheV) of Salmonella typhimurium is a structural gene involved in energization and switching of the flagellar motor. *J. Bacteriol.* **154,** 84–91 (1983).
81. Yamaguchi, S. *et al.* Genetic-evidence for a switching and energy-transducing complex in the flagellar motor of S*almonella-typhimurium*. *J. Bacteriol.* **168**, 1172-1179, doi:10.1128/jb.168.3.1172-1179.1986 (1986).
82. Thomas, D., Morgan, D. G. & DeRosier, D. J. Structures of Bacterial Flagellar Motors from Two FliF-FliG Gene Fusion Mutants. *J. Bacteriol.* **183,** 6404–6412 (2001).
83. Berg, H. C. The rotary motor of bacterial flagella. *Annu. Rev. Biochem.* **72**, 19-54, doi:10.1146/annurev.biochem.72.121801.161737 (2003).
84. Lloyd, S. A., Whitby, F. G., Blair, D. F. & Hill, C. P. Structure of the C-terminal domain of FliG, a component of the rotor in the bacterial flagellar motor. *Nature* **400,** 472–475 (1999).
85. Brown, P. N., Hill, C. P. & Blair, D. F. Crystal structure of the middle and C-terminal domains of the flagellar rotor protein FliG. *EMBO J.* **21,** 3225–3234 (2002).
86. Lam, K. H. *et al.* Multiple Conformations of the FliG C-Terminal Domain Provide Insight into Flagellar Motor Switching. *Structure* **20**, 315-325, doi:10.1016/j.str.2011.11.020 (2012).
87. Paul, K., Gonzalez-Bonet, G., Bilwes, A. M., Crane, B. R. & Blair, D. Architecture of the flagellar rotor. *Embo J.* **30,** 2962–2971 (2011).
88. Lynch, M. J. *et al.* 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 (2017).
89. Kihara, M., Miller, G. U. & Macnab, R. M. Deletion analysis of the flagellar switch protein FliG of Salmonella. *J. Bacteriol.* **182,** 3022–8 (2000).
90. Levenson, R., Zhou, H. & Dahlquist, F. W. Structural Insights into the Interaction between the Bacterial Flagellar Motor Proteins FliF and FliG. *Biochemistry* **51,** 5052–5060 (2012).
91. Xue, C. *et al.* 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 (2018).
92. Ogawa, R., Abe-Yoshizumi, R., Kishi, T., Homma, M. & Kojima, S. Interaction of the C-terminal tail of FliF with FliG from the Na+-driven flagellar motor of Vibrio alginolyticus. *J. Bacteriol.* **197,** 63–72 (2015).
93. Peifer, M., Berg, S. & Reynolds, A. B. A repeating amino acid motif shared by proteins with diverse cellular roles. *Cell* **76,** 789–91 (1994).
94. Park, S.-Y., Lowder, B., Bilwes, A. M., Blair, D. F. & Crane, B. R. Structure of FliM provides insight into assembly of the switch complex in the bacterial flagella motor. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 11886–91 (2006).
95. Lam, K.-H. *et al.* Structural basis of FliG-FliM interaction in *Helicobacter pylori*. *Mol. Microbiol.* **88,** 798–812 (2013).
96. Sarkar, M. K., Paul, K. & Blair, D. 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,** 9370–5 (2010).
97. Brown, P. N., Mathews, M. A. A., Joss, L. A., Hill, C. P. & Blair, D. F. Crystal structure of the flagellar rotor protein FliN from Thermotoga maritima. *J. Bacteriol.* **187,** 2890–902 (2005).
98. Ma, Q., Sowa, Y., Baker, M. A. B. & Bai, F. Bacterial Flagellar Motor Switch in Response to CheY-P Regulation and Motor Structural Alterations. *Biophys. J.* **110**, 1411-1420, doi:10.1016/j.bpj.2016.02.023 (2016).

99. Delalez, N. J. *et al.* Signal-dependent turnover of the bacterial flagellar switch protein FliM. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 11347–51 (2010).
100. Delalez, N. J., Berry, R. M. & Armitage, J. P. Stoichiometry and turnover of the bacterial flagellar switch protein FliN. *MBio* **5,** e01216-14 (2014).
101. Baker, M. A. B. *et al.* Domain-swap polymerization drives the self-assembly of the bacterial flagellar motor. *Nat. Struct. Mol. Biol.* (2016). doi:10.1038/nsmb.3172
102. Yorimitsu, T., Mimaki, A., Yakushi, T. & Homma, M. The conserved charged residues of the C-terminal region of FliG, a rotor component of the Na+-driven flagellar motor. *J. Mol. Biol.* **334,** 567–83 (2003).
103. Takekawa, N. *et al.* The tetrameric MotA complex as the core of the flagellar motor stator from hyperthermophilic bacterium. *Sci. Rep.* **6,** 31526 (2016).
104. Kim, E. A. *et al.* Biogenesis of the Flagellar Switch Complex in Escherichia coli: Formation of Sub-Complexes Independently of the Basal-Body MS-Ring. *J. Mol. Biol.* **429**, 2353-2359, doi:10.1016/j.jmb.2017.06.006 (2017).
105. Sircar, R. *et al.* Assembly States of FliM and FliG within the Flagellar Switch Complex. *J. Mol. Biol.* **427**, 867-886, doi:10.1016/j.jmb.2014.12.009 (2015).
106. Kim, E. A. *et al.* Architecture of the Flagellar Switch Complex of Escherichia coli: Conformational Plasticity of FliG and Implications for Adaptive Remodeling. *J. Mol. Biol.* **429**, 1305-1320, doi:10.1016/j.jmb.2017.02.014 (2017).
107. Bennett, M. J., Schlunegger, M. P. & Eisenberg, D. 3D domain swapping: A mechanism for oligomer assembly. *Protein Sci.* **4,** 2455–2468 (1995).
108. Crestfield, A. M., Stein, W. H. & Moore, S. On the aggregation of bovine pancreatic ribonuclease. *Arch. Biochem. Biophys.* **Suppl 1,** 217–22 (1962).
109. Liu, Y. & Eisenberg, D. 3D domain swapping: As domains continue to swap. *Protein Sci.* **11**, 1285-1299, doi:10.1110/ps.0201402 (2002).
110. Huber, A. H., Nelson, W. J. & Weis, W. I. Three-dimensional structure of the armadillo repeat region of beta-catenin. *Cell* **90,** 871–882 (1997).
111. Minamino, T. *et al.* Structural insight into the rotational switching mechanism of the bacterial flagellar motor. *Plos Biol.* **9,** e1000616–e1000616 (2011).
112. Kane, R. S., Takayama, S., Ostuni, E., Ingber, D. E. & Whitesides, G. M. Patterning proteins and cells using soft lithography. *Biomaterials* **20**, 2363-2376, doi:10.1016/s0142-9612(99)00165-9 (1999).
113. Qin, D., Xia, Y. N. & Whitesides, G. M. Soft lithography for micro- and nanoscale patterning. *Nat. Protoc.* **5**, 491-502, doi:10.1038/nprot.2009.234 (2010).
114. Watson, J. D. & Crick, F. H. C. Molecular structure of nucleic acids - a structure for deoxyribose nucleic acid. *Nature* **171**, 737-738, doi:10.1038/171737a0 (1953).
115. Crick, F. H. C. & Watson, J. D. The Complementary Structure of Deoxyribonucleic Acid. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **223,** 80–96 (1954).
116. Seeman, N. C. Nucleic-acid junctions and lattices. *J. Theor. Biol.* **99,** 237–247 (1982).
117. Yurke, B., Turberfield, A. J., Mills, A. P., Simmel, F. C. & Neumann, J. L. A DNA-fuelled molecular machine made of DNA. *Nature* **406,** 605–608 (2000).
118. Goodman, R. P., Berry, R. M. & Turberfield, A. J. The single-step synthesis of a DNA tetrahedron. *Chem. Commun.*, 1372-1373, doi:10.1039/b402293a (2004).
119. Rothemund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440,** 297–302 (2006).
120. Douglas, S. M. *et al.* Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459,** 414–418 (2009).
121. Dietz, H., Douglas, S. M. & Shih, W. M. Folding DNA into twisted and curved nanoscale shapes. *Science* **325,** 725–30 (2009).
122. Ke, Y. G., Ong, L. L., Shih, W. M. & Yin, P. Three-Dimensional Structures Self-Assembled from DNA Bricks. *Science* **338**, 1177-1183, doi:10.1126/science.1227268 (2012).
123. Benson, E. *et al.* DNA rendering of polyhedral meshes at the nanoscale. *Nature* **523,** 441–444 (2015).
124. Seelig, G., Soloveichik, D., Zhang, D. Y. & Winfree, E. Enzyme-free nucleic acid logic circuits. *Science* **314**, 1585-1588, doi:10.1126/science.1132493 (2006).
125. Shin, J.-S. & Pierce, N. A. A synthetic DNA walker for molecular transport. *J. Am. Chem. Soc.* **126**, 10834–10835 (2004).
126. Douglas, S. M. *et al.* Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nucleic Acids Res.* **37**, 5001-5006, doi:10.1093/nar/gkp436 (2009).
127. Wagenbauer, K. F., Sigl, C. & Dietz, H. Gigadalton-scale shape-programmable DNA assemblies. *Nature* **552,** 78–83 (2017).
128. Langecker, M. *et al.* Synthetic Lipid Membrane Channels Formed by Designed DNA Nanostructures. *Science* **338**, 932-936, doi:10.1126/science.1225624 (2012).
129. Douglas, S. M., Bachelet, I. & Church, G. M. A Logic-Gated Nanorobot for Targeted Transport of Molecular Payloads. *Science* **335**, 831-834, doi:10.1126/science.1214081 (2012).
130. Mirkin, C. A., Letsinger, R. L., Mucic, R. C. & Storhoff, J. J. A DNA-based method for rationally assembling nanoparticles into macroscopic materials. *Nature* **382**, 607-609, doi:10.1038/382607a0 (1996).
131. Sharma, J., Chhabra, R., Liu, Y., Ke, Y. G. & Yan, H. DNA-templated self-assembly of two-dimensional and periodical gold nanoparticle arrays. *Angew. Chem.-Int. Edit.* **45**, 730-735, doi:10.1002/anie.200503208 (2006).

132. Storhoff, J. J. *et al.* What controls the optical properties of DNA-linked gold nanoparticle assemblies? *J. Am. Chem. Soc.* **122**, 4640-4650, doi:10.1021/ja993825l (2000).
133. Maune, H. T. *et al.* Self-assembly of carbon nanotubes into two-dimensional geometries using DNA origami templates. *Nat. Nanotechnol.* **5,** 61–66 (2010).
134. Jungmann, R. *et al.* Multiplexed 3D cellular super-resolution imaging with DNA-PAINT and Exchange-PAINT. *Nat. Methods* **11**, 313-U292, doi:10.1038/nmeth.2835 (2014).
135. Schreiber, R. *et al.* Hierarchical assembly of metal nanoparticles, quantum dots and organic dyes using DNA origami scaffolds. *Nat. Nanotechnol.* **9,** 74–78 (2014).
136. Bayrak, T. *et al.* DNA-Mold Templated Assembly of Conductive Gold Nanowires. *Nano Lett.* **18,** 2116–2123 (2018).
137. Stains, C. I., Porter, J. R., Ooi, A. T., Segal, D. J. & Ghosh, I. DNA sequence-enabled reassembly of the green fluorescent protein. *J. Am. Chem. Soc.* **127**, 10782-10783, doi:10.1021/ja051969w (2005).
138. Fu, J. L., Liu, M. H., Liu, Y., Woodbury, N. W. & Yan, H. Interenzyme Substrate Diffusion for an Enzyme Cascade Organized on Spatially Addressable DNA Nanostructures. *J. Am. Chem. Soc.* **134**, 5516-5519, doi:10.1021/ja300897h (2012).
139. Fu, J. *et al.* Multi-enzyme complexes on DNA scaffolds capable of substrate channelling with an artificial swinging arm. *Nat. Nanotechnol.* **9,** 531–536 (2014).
140. Derr, N. D. *et al.* Tug-of-War in Motor Protein Ensembles Revealed with a Programmable DNA Origami Scaffold. *Science* **338**, 662-665, doi:10.1126/science.1226734 (2012).
141. Wollman, A. J. M., Sanchez-Cano, C., Carstairs, H. M. J., Cross, R. A. & Turberfield, A. J. Transport and self-organization across different length scales powered by motor proteins and programmed by DNA. *Nat. Nanotechnol.* **9,** 44–47 (2014).
142. Funke, J. J. *et al.* Uncovering the forces between nucleosomes using DNA origami. *Sci. Adv.* **2,** e1600974–e1600974 (2016).
143. Fisher, P. D. E. *et al.* A Programmable DNA Origami Platform for Organizing Intrinsically Disordered Nucleoporins within Nanopore Confinement. *ACS Nano* **12,** 1508–1518 (2018).
144. Henning-Knechtel, A., Knechtel, J. & Magzoub, M. DNA-assisted oligomerization of pore-forming toxin monomers into precisely-controlled protein channels. *Nucleic Acids Res.* **45,** 12057–12068 (2017).
145. Shen, W., Bruist, M. F., Goodman, S. D. & Seeman, N. C. A Protein-Driven DNA Device That Measures the Excess Binding Energy of Proteins That Distort DNA. *Angew. Chemie Int. Ed.* **43,** 4750–4752 (2004).
146. Endo, M., Katsuda, Y., Hidaka, K. & Sugiyama, H. Regulation of DNA Methylation Using Different Tensions of Double Strands Constructed in a Defined DNA Nanostructure. *J. Am. Chem. Soc.* **132,** 1592–1597 (2010).
147. Liu, Y. *et al.* Influence of Histone Tails and H4 Tail Acetylations on Nucleosome–Nucleosome Interactions. *J. Mol. Biol.* **414,** 749–764 (2011).
148. Frey, S., Richter, R. P. & Goerlich, D. FG-rich repeats of nuclear pore proteins form a three-dimensional meshwork with hydrogel-like properties. *Science* **314**, 815-817, doi:10.1126/science.1132516 (2006).
149. Song, L. Z. *et al.* Structure of staphylococcal alpha-hemolysin, a heptameric transmembrane pore. *Science* **274**, 1859-1866, doi:10.1126/science.274.5294.1859 (1996).
150. Yang, X. H. *et al.* Flagella Overexpression Attenuates Salmonella Pathogenesis. *PLoS One* **7**, 8, doi:10.1371/journal.pone.0046828 (2012).
151. Moriarty, T. J. *et al.* Real-time high resolution 3D imaging of the Lyme disease spirochete adhering to and escaping from the vasculature of a living host. *PLoS Pathog.* **4**, 13, doi:10.1371/journal.ppat.1000090 (2008).
152. Nain, S. & Sharma, N. N. Propulsion of an artificial nanoswimmer: a comprehensive review. *Front. Life Sci.* **8**, 2-17, doi:10.1080/21553769.2014.962103 (2015).
153. Goodman, R. P. *et al.* A facile method for reversibly linking a recombinant protein to DNA. *ChemBioChem* **10,** 1551–1557 (2009).
154. Janknecht, R. *et al.* Rapid and efficient purification of native histidine-tagged protein expressed by recombinant vaccinia virus. *Proc. Natl. Acad. Sci.* **88,** 8972–8976 (1991).
155. Tusk, S. E. Templated self-assembly of the bacterial flagellar motor torque ring in vitro. (University of Oxford, 2016).
156. Knecht, S., Ricklin, D., Eberle, A. N. & Ernst, B. Oligohis-tags: mechanisms of binding to Ni2+-NTA surfaces. *J. Mol. Recognit.* **22,** 270–279 (2009).
157. Chilkoti, A., Chen, G., Stayton, P. S. & Hoffman, A. S. Site-Specific Conjugation of a Temperature-Sensitive Polymer to a Genetically Engineered Protein. *Bioconjug. Chem.* **5,** 504–507 (1994).
158. Deiters, A. *et al.* Adding amino acids with novel reactivity to the genetic code of *Saccharomyces cerevisiae*. *J. Am. Chem. Soc.* **125,** 11782–11783 (2003).
159. Keppler, A. *et al.* A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat. Biotechnol.* **21**, 86-89, doi:10.1038/nbt765 (2003).
160. Los, G. V. *et al.* HaloTag: A novel protein labeling technology for cell imaging and protein analysis. *ACS Chem. Biol.* **3,** 373–382 (2008).
161. Zakeri, B. *et al.* Peptide tag forming a rapid covalent bond to a protein, through engineering a bacterial adhesin. *Proc. Natl. Acad. Sci.* **109,** E690–E697 (2012).
162. Rosen, C. B. *et al.* Template-directed covalent conjugation of DNA to native antibodies, transferrin and other metal-binding proteins. *Nat. Chem.* **6,** 804–809 (2014).

163. Cuatrecasas, P. & Parikh, I. Adsorbents for affinity chromatography. Use of N-hydroxysuccinimide esters of agarose. *Biochemistry* **11,** 2291–2299 (1972).
164. Shu, D., Zhang, H., Jin, J. & Guo, P. Counting of six pRNAs of phi29 DNA-packaging motor with customized single-molecule dual-view system. *EMBO J.* **26,** 527–537 (2007).
165. Smyth, D. G., Blumenfeld, O. O. & Konigsberg, W. Reactions of N-ethylmaleimide with peptides and amino acids. *Biochem. J.* **91,** 589–95 (1964).
166. Hermanson, G. T. *Bioconjugate techniques*. (Academic press, 2013).
167. Edgcomb, S. P. & Murphy, K. P. Variability in the pKa of histidine side-chains correlates with burial within proteins. *Proteins Struct. Funct. Genet.* **49,** 1–6 (2002).
168. Green, N. M. Avidin and streptavidin. *Methods Enzymol.* **184,** 51–67 (1990).
169. Grünenfelder, B., Gehrig, S. & Jenal, U. Role of the cytoplasmic C terminus of the FliF motor protein in flagellar assembly and rotation. *J. Bacteriol.* **185,** 1624–33 (2003).
170. Greenfield, N. J. & Fasman, G. D. Computed circular dichroism spectra for the evaluation of protein conformation. *Biochemistry* **8,** 4108–4116 (1969).
171. Circular Dichroism (CD) Spectroscopy. Available at: http://www.proteinchemist.com/cd/cdspec.html. (Accessed: 15th August 2018)
172. Louis-Jeune, C., Andrade-Navarro, M. A. & Perez-Iratxeta, C. Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. *Proteins Struct. Funct. Bioinforma.* **80,** 374–381 (2012).
173. Thermal Shift Assay Diagram. Available at: https://commons.wikimedia.org/wiki/File:Thermal_Shift_Assay_diagram.svg. (Accessed: 17th August 2018)
174. Holzwarth, G. & Doty, P. The ultraviolet circular dichroism of polypeptides. *J. Am. Chem. Soc.* **87,** 218–28 (1965).
175. Burstein, E. A., Vedenkina, N. S. & Ivkova, M. N. Fluorescence and the location of tryptophan residues in protein molecules. *Photochem. Photobiol.* **18,** 263–279 (1973).
176. Spruijt, E., Tusk, S. E. & Bayley, H. DNA scaffolds support stable and uniform peptide nanopores. *Nat. Nanotechnol.* **13,** 739–745 (2018).
177. Zadeh, J. N. *et al.* NUPACK: Analysis and design of nucleic acid systems. *J. Comput. Chem.* **32,** 170–173 (2011).
178. Baker, M. A. B. *et al.* Dimensions and Global Twist of Single-Layer DNA Origami Measured by Small-Angle X-ray Scattering. ACS Nano **12**, 5791–5799 (2018).