Characterisation and analysis of polar growth proteins within rod-shaped bacterial species

Alec Jack Cameron Hutchinson

A thesis submitted for the degree of Masters by Research

University of East Anglia

School of Biological Sciences

January 2023

This copy of the thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that use of any information derived there from must be in accordance with current UK Copyright Law. In addition, any quotation or extract must include full attribution.

Abstract:

Polar growth is the main mechanism used by most rod-shaped bacteria to drive growth within their life cycle. Known phylum that use polar growth include the Gram-positive Actinobacteria and the Gram-negative Pseudomonadota. Many of the Gram-positive bacteria utilise a homologue of the large DivIVA family as a vital protein in their growth mechanism due to its innate ability to localise to the pole of the cell. Whilst DivIVA naturally accumulates at the pole, the partner proteins of DivIVA that build up the growth mechanism are recruited and regulated to drive functional polar growth, though the range of partner proteins of DivIVA vary by species. A well-documented example of this the Tip Organising Centre (TIPOC) in the Actinobacteria *Streptomyces coelicolor*. *S. coelicolor* is a GC-rich, soil-dwelling filamentous bacteria thoroughly researched for its vast capacity for secondary metabolite production. *S. coelicolor* grows through polar growth at the tip of its filamentous hyphae utilising the TIPOC: a large multi-protein complex. This complex is constructed from three smaller protein complexes (DivIVA, Scy and FilP) and is known for its ability to interact with other important cellular mechanisms such as the ParAB system (used for chromosome segregation). However, unlike Gram-positive bacteria, many of the Gram-negative bacteria lack a DivIVA homologue and a common mechanism across these species is yet to be found. Though a new protein called Growth Pole Ring (GPR) has been identified recently in *Agrobacterium tumefaciens* which affects polar growth.

In this study, we have characterised a new potential partner protein of the TIPOC encoded by the previously researched gene *SCO5569*. *SCO5569* has been designated *dia* due to its ability to affect the determined hyphal diameter during active polar growth. We used: computational analysis of Dia and other homologs to infer structural properties and functions of the protein, *in vivo* BACTH assays to test for direct protein:protein interactions and *in vitro* SDS-PAGE to analyse the oligomerisation of Dia. The computational analysis revealed Dia is highly prevalent in many Gram-positive Actinobacteria species and a few Gram-negative phyla. Dia and its homologs all share a conserved region between residues 35-144 which was inferred to possess the function of DivIVA and ATP synthase subunit B domains. We indicated that Dia is capable of self-interaction along with direct interaction with DivIVA, Scy, FilP, ParB and ParH. These interactions suggest that Dia may be novel components of both the TIPOC and ParAB system of *S. coelicolor*. The isolated and overexpressed oligomerised Dia likely forms a dimer which was resistant to high temperature (95°C) and β-mercaptoethanol. Alongside investigating *dia*, we investigated possible homologues of GPR as alternatives to DivIVA for Gram-negative rod-shaped bacteria that possess polar growth. We identified an extensive range of GPR homologs throughout Gram-negative bacteria. We analysed and localised a protein in the α-proteobacteia *Labrenzia aggregata* which showed signs of similarity, though very distantly related. The *L. aggregata* GPR homologs appeared to retain a conserved region of GPR between residues 278-1815 which possessed domains of GPR, some linked to its role in polar growth. We designated this protein Lcy and used *in vivo* fluorescent microscopy to find that Lcy localises at the pole of *L. aggregata* throughout its life cycle including during cell division. This study has helped to reveal a new potential TIPOC component which may directly interact with the ParAB system. This suggests a new link between the mechanisms that drive polar growth and chromosome segregation in bacteria. Finally, we have demonstrated that, as an alternative to DivIVA, homologs of GPR are more prevalent in Gram-negative bacteria, even distantly related species. This distantly related homolog appeared to retain the function for polar growth.

Access Condition and Agreement

Each deposit in UEA Digital Repository is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the Data Collections is not permitted, except that material may be duplicated by you for your research use or for educational purposes in electronic or print form. You must obtain permission from the copyright holder, usually the author, for any other use. Exceptions only apply where a deposit may be explicitly provided under a stated licence, such as a Creative Commons licence or Open Government licence.

Electronic or print copies may not be offered, whether for sale or otherwise to anyone, unless explicitly stated under a Creative Commons or Open Government license. Unauthorised reproduction, editing or reformatting for resale purposes is explicitly prohibited (except where approved by the copyright holder themselves) and UEA reserves the right to take immediate ‘take down’ action on behalf of the copyright and/or rights holder if this Access condition of the UEA Digital Repository is breached. Any material in this database has been supplied on the understanding that it is copyright material and that no quotation from the material may be published without proper acknowledgement.

Table of Contents:

Acknowledgements:

I would like to thank my supervisor Gabriella Kelemen for her support and guidance throughout my research in the laboratory and writing period. To my amazing friends and colleagues Daniel Moye and Mashael Alanazi for their support through the year. Finally, thank you to all the undergraduate students, especially Stefan Harper who helped with parts of this project.

List of Abbreviations:

BACTH assay: Bacterial Adenylate Cyclase-based Two-Hybrid assay

BLAST: Basic Local Alignment Search Tool

GPR: Growth Pole Ring

kDa: Kilodalton

lcy: labrenzia cytoskeletal protein

NMR: Nuclear Magnetic Resonance spectroscopy

SALPs: SsgA-like proteins

scy: streptomyces cytoskeletal protein

SDS-PAGE: Sodium Dodecyl Sulfate–Polyacrylamide Gel Electrophoresis

SMC: Structural Maintenance of Chromosomes

TEM: Transmission Electron Microscopy

TIPOC: Tip Organising Centre

Chapter 1:

Introduction

The current speculation for the global estimate of prokaryote species on Earth is potentially 1.6 million (Louca *et al*., 2019). Although, previous speculations have predicted up to a possible 1 trillion microbial species worldwide (Locey & Lennon, 2016). Contrary to this vast number of species only 20,000 species have been sequenced and documented (Parte *et al*., 2020). The extent to which genomes have been analysed is limited. Even with some of our most documented species, *Streptomyces coelicolor* for example, only about 54% of the complete genome has been analysed, with these analysed genes have been assigned to clusters with Orthology (Heinsch *et al*., 2019). With only just over half of the genome analysed and assigned a function, this demonstrates the lack of understanding within even the most well studied microbes. These gaps in the knowledge only grow with the identification or reclassification of new species. This can be seen in with reclassification of 3 Ferric-reducing bacteria between the related genus of *Geobacter* to *Geomonas* (Xu *et al*., 2019) and 3 species from the *Stappia* genus to *Labrenzia* genus (Biebl *et al*., 2007) due to insight provided by the 16S rRNA sequencing and chemotaxonomic analysis. Nevertheless, within the most documented species, there is understanding as to role and function certain clusters of Orthology. For example, in the *S. coelicolor* genome, only 0.5% of genes have been identified to be involved in the cell growth and division (Heinsch *et al*., 2019), but the proteins and mechanisms identified (the TIPOC and ParAB systems) form quite complex and sensitive systems some of which have components shared across many different species through the homology of the components. The homology of these shared systems that control important aspects of bacterial life cycle, such as growth and cell division, are pivotal in investigating the unknown functions of genes in new, reclassified or existing genomes to expand our knowledge.

1.1 Cell division:

All bacteria use cell division to produce genetically identical offspring. For most bacterial species, binary fission is utilised where the growing cell divides to create to identical daughter cells each with a single, identical chromosome. For most bacteria, the division event occurs after the parent cell has reached optimal growth, duplicated and segregated the chromosome. With this simple event being highly prevalent, there must be an efficient system in place to ensure the genetic material is passed on as quickly as possible.

1.1.1 The driver of cell division - FtsZ:

For efficient cell division, most bacterial species utilise a homologue of FtsZ which share a common structure. The overall structure of the FtsZ homologues possessed 2 main domains: a GTP-ase at the N terminus and a tubulin-like loop domain at the C terminus. Both domains were connected by a central helix (Löwe & Amos, 1998). This common structure was found to possess remarkable similarity to the eukaryotic protein tubulin been though there was a severe lack of homology between these proteins (Nogales *et al*., 1998). The structural similarity was clearest around residues for GTP binding and hydrolysis (Erickson *et al*., 2007).

The most widely known example of FtsZ driven cell division is within *Escherichia coli*. The cytokinesis of the cell is broken up into three key stages: Initiation of FtsZ polymerisation, Maturation of the FtsZ and Constriction of the divisome (de Boer, Crossley & Rothfield, 1992). Within the initiation stage, FtsZ monomers localise at the future site of division and begin to polymerise into a ring-like structure called a Z-ring (Bi & Lutkenhaus, 1991). This polymerisation is regulated through the conversion of GTP to GDP where FtsZ filamentation is triggered when GTP is bound (Erickson, 1998). After a short period of time, the Z-ring enters the next stage to mature the structure. The Z-ring begins to recruit essential downstream divisome proteins to begin formation of the divisome. Once the divisome is complete, the divisome begins to constrict the cytoplasm, with septal peptidoglycan synthesis occurring eventually dividing the bacteria into 2 separate daughter cells (Gerding *et al*., 2009). Interestingly, before the constriction of the divisome, the polymerised Z-ring is capable of applying a constrictive force to the cell wall. The constrictive action of FtsZ was demonstrated in liposomes which contained modified FtsZ which had the membrane-targeting sequence of MinD attached and in the presence of GTP. The modified FtsZ was able to form

membrane-associated rings within liposomes, with signs of indentation on the walls, though indentation formation was dependent on the addition of GTP. (Jiménez *et al*., 2011; Osawa, Anderson & Erickson, 2008).

1.1.2 Stabilisation and maturation of the Z ring:

Although FtsZ can naturally self-assemble, there are different systems across bacterial species for the regulation of the assembly of Z-rings. They are split into 2 different types: positive and negative regulation. Positive regulation consists of proteins which promote the formation of the Z-ring whilst negative regulation is characterised by proteins which inhibit the formation of Z-rings over chromosomes and near to the poles of the cell (Huang, Durand-Heredia & Janakiraman, 2013). Positive regulation helps Z-ring formation through proteins which help stabilise the Z-ring and anchor the forming Z-ring at the future division site. One of the best studied examples of this is the use of the trans-membrane proteins FtsA and ZipA within the Gram-negative bacteria: *E. coli* (Pichoff & Lutkenhaus, 2002). Homologues of these 2 transmembrane proteins (FtsA and ZipA) are found across other species but in varying degrees. FtsA is found abundant across many bacterial species whereas ZipA homologues are only found across Gram-negative bacterium.

FtsA homologues have been found to be the most widely conserved member of the divisome (Haeusser & Margolin, 2016) and is found to be similar to actin in both structure and being capable forming protofilaments (Szwedziak *et al*.2012). The exact function FtsA plays within the divisome is currently uncertain due to the current evidence from different bacterial models. The current *E. coli* model shows that *ftsA* mutants which are incapable of polymerisation actually gain a new function, which is being tolerant to the loss of ZipA. In turn, this suggests that polymerised FtsA and monomeric FtsA compete with each other, where monomeric FtsA is used to recruit the downstream divisome proteins and the polymerised FtsA doesn't (Pichoff *et al*., 2012). However, *Bacillus subtilis* mutant strains of FtsA have shown that polymerisation of FtsA is required for recruitment of downstream divisome proteins through a loss of function (Szwedziak *et al*., 2012). ZipA in the current *E. coli* model appears to share a similar function to FtsA. ZipA can interact with the polymerised FtsZ ring (through its C-terminus) and tether it to the membrane through an embedded N-terminus domain. Collectively this helps to anchor the Z-ring in place (Ohashi *et al*., 2002; Pichoff & Lutkenhaus, 2002). However, ZipA

mutants are still capable of forming Z-rings though less frequently (Hale & de Boer, 1999).

As mentioned before ZipA homologues are not found in Gram-positive bacteria and instead a different protein (SepF) has been shown to perform the role of ZipA instead. This has been shown through direct interaction of SepF and FtsZ and self-interaction within a Yeast two-hybrid assay, alongside *sepF* mutants showing abnormal septa formation among other cell defects (Hamoen *et al*., 2006). The overlap of function is reaffirmed through SepF overexpression in *ftsA* mutants correcting problems with division and *sepF-ftsA* double mutants in *B. subtilis* unable to form Z-rings and being lethal (Ishikawa *et al*., 2006). Since SepF seems to shares the same function as ZipA, it is unsurprising that SepF is required for correct septum formation with deformed septum forming in null mutants (Hamoen *et al*., 2006; Ishikawa *et al*., 2006) and consequently SepF localises at the division site, with this interaction being dependent on FtsZ (Hamoen *et al*., 2006). SepF appears to promote the polymerisation of FtsZ (Gündoğdu *et al*., 2011). This polymerisation effect has been documented through Transmission Electron Microscopy (TEM) of SepF and FtsZ *in vitro*. SepF appears to polymerise into large ring-like structures of approximately 50 nm. FtsZ is also capable of polymerising in the presence of GTP, but with the addition of SepF larger tubular structures formed with a diameter of 48 nm (which is remarkably similar to the polymerised SepF). The continued incubation of the polymerising FtsZ and SepF lead to larger more complex structures forming. SepF rings were found to bind multiple FtsZ protofilaments and aid the formation of the Z-ring (Gündoğdu *et al*., 2011).

With understanding of how the Z-ring is formed and tethered to the membrane, the next step in the process of forming the divisome is the maturation of the Z-ring. Within *E. coli* and *B. subtilis* the process of maturation occurs in similar yet different ways: with *E. coli* requiring each divisome protein being recruited in a specific fashion whereas *B. subtilis* the recruited divisome proteins can be recruited independently of each other (Harry, Monahan & Thompson, 2006). Within both bacteria, one of the localised proteins are homologues of FtsK (first protein for *E. coli*) which is a DNA translocase (Margolin, 2000). FtsK homologues in *E. coli* are known to recruit the next protein to mature the Z-ring, whilst also utilising their C-terminus to has been connected to interacting with the sister chromosomes and transporting them into their respective daughter cells (Bigot *et al*., 2004; Aussel *et al*., 2002), thus, tying FtsK to both cell division and chromosome segregation. The

homologue for FtsK in *B. subtilis* is called SpoIIIE has currently been shown to be involved in DNA translocation during sporulation (Sharp & Pogliano, 2003).

The next stages in the Z-ring maturation process differ between *E. coli* and *B. subtilis*. As previously mentioned, how each divisome protein is recruited varies in addition to differences between the homologues of divisome proteins. The system within *E. coli* recruits FtsQ, FtsL and FtsB following the accumulation of FtsK (Figure 1-1). These three recruited proteins form a larger complex (called the FtsQLB complex) which forms prior to the localisation of the septum and was identified through co-immunoprecipitation (Buddelmeijer & Beckwith, 2004). These proteins are similar in structure: possessing a large periplasmic domain, a short cytoplasmic domain and a transmembrane region. The functions of FtsQ and FtsB are unknown, but FtsQ possess an α-domain with similarities to polypeptide transport domains (van den Ent *et al*., 2008). FtsB on the other hand, is known to be essential for cells and null mutants are lethal (Buddelmeijer & Beckwith, 2004). The functions of FtsL have been inferred by data from *B. subtilis*. Upon the recruitment of the FtsQLB complex in *E. coli*, the next required divisome protein accumulates called FtsW. FtsW has been found to be involved in the transport of lipid-linked precursors for peptidoglycan synthesis (Mohammadi *et al*., 2011). The *ftsW* downstream gene, *ftsI*, is recruited next and is known to help finish the synthesis of new peptidoglycan and was identified via binding of penicillins (Goffin & Ghuysen, 1998). FtsN is next in the sequence with a ring-like localisation pattern near to the septum and late in the division cycle. Additionally, *ftsN* null mutants are lethal making FtsN a necessary gene within the divisome (Dai, Xu & Lutkenhaus, 1993). Finally, AmiC is a known amidase recruited after AmiC cleaves the murein crosslinks holding the septum cell wall together. This subsequently degrades the septum cell wall and eventually leads to the separation of the daughter cells (Weiss, 2004).

In *B. subtilis*, the homologues of FtsQ and FtsW are called DivIB and DivIC respectively and are known to still form the same protein complex as *E. coli* with FtsL (Figure 1-1). These proteins possess a similar structure to their *E. coli* homologues and in a similar fashion to FtsQ, the function of DivIB is still unknown and it possess the same α-domain. A suggested function of FtsL is to regulate the constriction of Z-rings through the regulation of recruiting similar constricting divisome proteins (Kawai & Ogasawara, 2006). The function of DivIC is still unknown, much like its *E. coli* homologue, but has been shown to be vital for cell survival through null mutants (Levin & Losick, 1994). Finally, PBP2B (a homologue

of FtsL) has also been shown to be involved in the final stages of peptidoglycan synthesis (Goffin & Ghuysen, 1998).

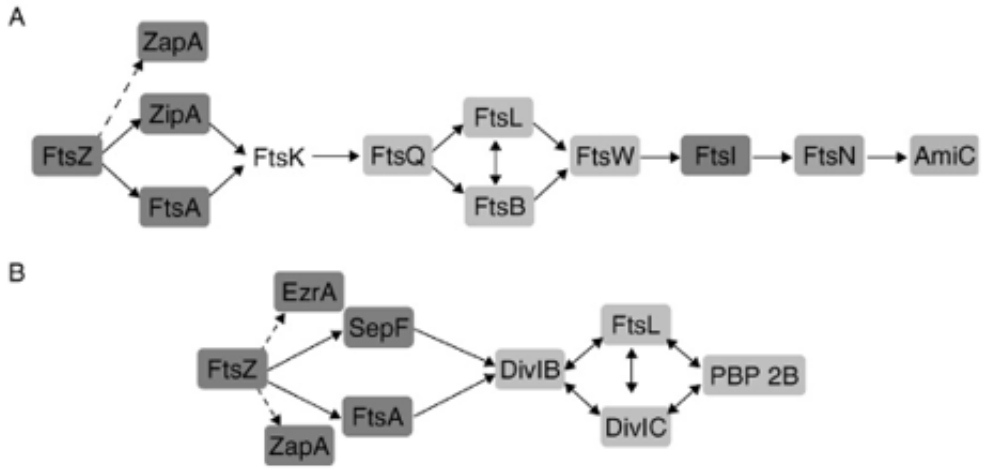


Figure 1-1. The different components need to stabilise and mature the Z-ring. The Z-ring is stabilised for both *E. coli* and *B. subtilis* through recruitment of FtsZ, FtsA, ZipA or SepF. The maturation of both rings is slightly different. (A) For *E. coli*, the recruitment of each required protein is dependent on the previous protein leading to a linear progression of the maturation. (B) For *B. subtilis*, the components need for maturation are recruited interdependently of each other leading to the maturation occurring as a single step. Taken from (Harry, Monahan & Thompson, 2006).

1.2 Other systems for successful bacterial division:

Although the Z-rings are critical for cell division to occur within reproducing bacteria, they are not the only process required for genetically identical offspring to be produced. For identical offspring to be produced before cell division occurs, complete copies of the bacterial chromosome must be transferred to the offspring (known as chromosome segregation) and the correct formation of the septa to prevent cutting of the chromosome. The coordination and regulation of these two mechanisms are key for efficient reproduction, with chromosome segregation occurring before septa positioning. So far, a large multi-species superfamily of proteins has been attributed to determining chromosome segregation and septa positioning: the ParA/MinD superfamily. This superfamily can be broken down into two subgroups which control one of the 2 required processes: the ParA subgroup (controls chromosome segregation) and the MinD subgroup (determines the septa position). Although these subgroups are grouped together, how each homologue mechanically works across species can be vastly different (Lutkenhaus, 2012). However, the common characteristic between these subgroups is the occurrence of a deviant walker ATPase motif within a nucleotide-binding P-loop close to the N-

terminal of each homologue (Koonin, 1993). For each subgroup, there is a common system within most bacteria (ParA has the ParAB system and MinD has the MinDCE system) though they differ in mechanics unsurprisingly.

1.2.1 Controlled septum positioning through the MinD system:

The MinD system is utilised for organising a necessary event of successful cell division (septum positioning) to produce the correct daughter cells. The mechanics of MinD can vary across different species to meet their required needs which is typically to produce offspring with the same cell shape and size as the parent. To meet these requirements, the mechanics of the MinD system heavily controlled and function of proteins within the MinD system are highly conserved. For most dividing bacteria, the correct offspring produced would identical and equal sized progeny which is determined by the septum forming in the mid cell during division. The mid cell is typically identified in the process by spatial cues of the bacteria such as the location of the chromosome or the poles of the cell. Before we can discuss the different molecular or structural cues bacterial species utilise, we first have to appreciate how the site of division for septum positioning is formed and regulated. As mentioned prior, we explained the regulation and role of FtsZ in septum site positioning. FtsZ is a known marker of future septum sites (Bi & Lutkenhaus, 1991) and the GTPase function of FtsZ is critical for regulating its assembly and disassembly as large filaments. The polymerisation of FtsZ into Z-rings occurs in the presence of GTP but depolymerisation occurs when the GTPase function is upregulated (Erickson, 1998). Through spatially regulating the GTPase activity of FtsZ, the position of Z-ring formation can be directed to the desired section of the cell. The MinD system is capable of regulating this activity and was originally identified within *E. coli* due to increased levels of anucleate cells formed within null mutants (Adler *et al*., 1967). The MinD system contains three main proteins: MinD, MinC and MinE and the system is also known as the MinDCE system (Figure 1-2). The MinDCE system has been accredited to controlling FtsZ placement as null mutants of the system have been seen with Z-rings forming closer to the poles and mini-cells forming which lack a chromosome (anucleate) (de Boer, Crossley & Rothfield, 1992). The MinDCE system facilitates this regulation through the formation of the MinD-MinC-ATP complex. The MinD complex anchors the complex to the cell pole and the C-terminus of MinC prevents lateral polymerisation of FtsZ. Moreover, FtsA and ZipA are actively competing with MinC C-terminus to bind to FtsZ to anchor the polymers to the cell membrane (Dajkovic *et al*., 2008;

Shen & Lutkenhaus, 2009). Oscillations of MinD between the poles of the cell allows the complex to form at high concentration at cell poles and low concentration at the mid-cell (Dajkovic *et al*., 2008) which in turn, focuses FtsZ to be anchored and polymerised at the mid-cell (the desired division site). The oscillation of MinD is influenced through ATP, where ATP-bound dimeric MinD settle at a pole and continue to accumulate along the cell membrane towards the mid-cell (Lackner, Raskin & do Boer, 2003; Hu, Gogol & Lutkenhaus, 2002). MinE begins to form a ring assembly at the edge of the dimeric MinD-ATP section and alters the regulation of the complex. MinE competes with MinC for binding to the MinD-ATP complex which reduces formation of any MinD-MinC complexes alongside disassembling any current complexes. In addition, MinE increases the ATPase activity of MinD to convert the dimer into released MinD-ATP monomers (Hu & Lutkenhaus, 2001). This process is continued until all dimer MinD-ATP has been disassembled from the pole, whereupon the MinE assembly disassembles itself (Park *et al*., 2012; Park *et al*., 2011). The released MinD-ATP monomers diffuse away from the MinE assembly and are converted back into dimeric MinD-ATP (through nucleotide exchange of ADP and ATP) and localise to the opposite pole of the cell and the process repeats. The reasons for the pole-to-pole MinD oscillation are still unknown. In the *E. coli* genome, no ParA homologues or chromosome segregation mechanism has been found (Livny, Yamaichi & Waldor, 2007). Although, a *parS* site near the *oriC* of *E. coli* has been identified (called *migS*) (Yamaichi & Niki, 2004). With MinD possessing a similar function to ParA, it is possible that MinD helps with the segregation of chromosomes instead of ParA. This is further supported by the fact that MinD has been found to bind directly to DNA in the presence of ATP. The MinD oscillation patterns has been seen to help to transport and anchor the replicated chromosomes to a pole through binding of MinD during cell division (Di Ventura *et al*., 2013).

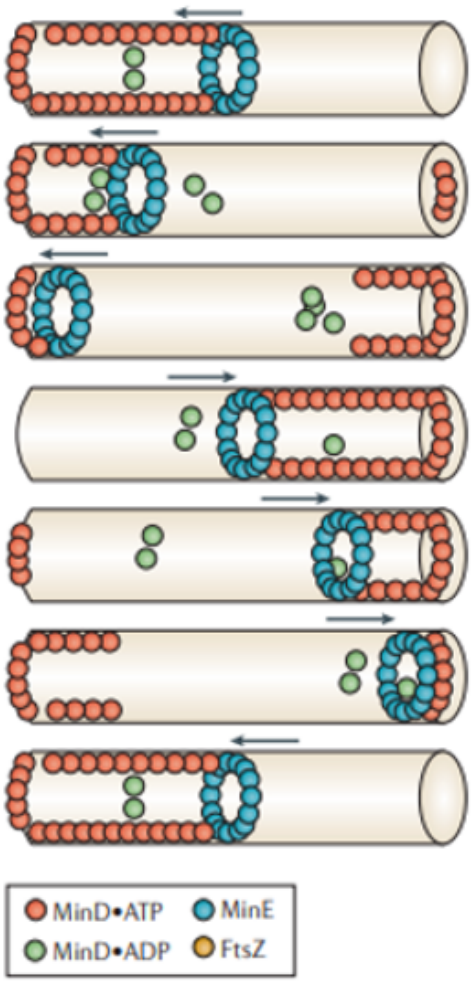


Figure 1-2. Oscillation patterns of *E. coli* MinD. ATP-bound dimeric MinD accumulates at a pole of the cell (red) and continues to grow towards the mid-cell. This eventually stimulates the formation of a MinE ring (blue) which upregulates the ATPase activity of MinD. ATP is converted to ADP, causing the disassembly of the MinD dimer to MinD-ADP monomers (green) which diffuse towards to the opposite pole. Once the MinE ring has disassembled all dimeric MinD, it begins to disassemble as well. Nucleotide exchange of ADP with ATP occurs to stimulate the formation of new ATP-bound dimeric MinD which accumulates at the opposite pole. This process is repeated again. Taken from (Lenz & Søgaard-Andersen, 2011).

As for *B. subtilis*, the localisation of MinD, to the poles and newly formed division septa is controlled through interactions with DivIVA and an intermediary protein MinJ, though DivIVA is required to localise first (Cha & Stewart, 1997; Marston *et al*., 1998; Patrick & Kearns, 2008) (Figure 1-3). This is due to innate attraction of DivIVA to the negative curvature of the cell membrane (Ramamurthi & Losick, 2009). The innate attraction of DivIVA has also been shown to be a fundamental reason DivIVA localises to the site of division in the centre of the cell (due to the newly synthesised septa has a negative curvature) (Eswaramoorthy *et al*., 2011). Alongside MinD, homologues of MinC are also found within *B. subtilis* which perform an almost identical function to their *E. coli* counterparts: they inhibit FtsZ polymerisation through ATP-dependent localisation to MinD at the pole (Marston & Errington, 1999; Karoui & Errington, 2001). The localisation of MinD and MinC to the poles helps to promote FtsZ rings to form in the midcell. However, since *B. subtilis* lacks a MinE homologue, a regulatory mechanism for the MinDC complex is yet to be identified.

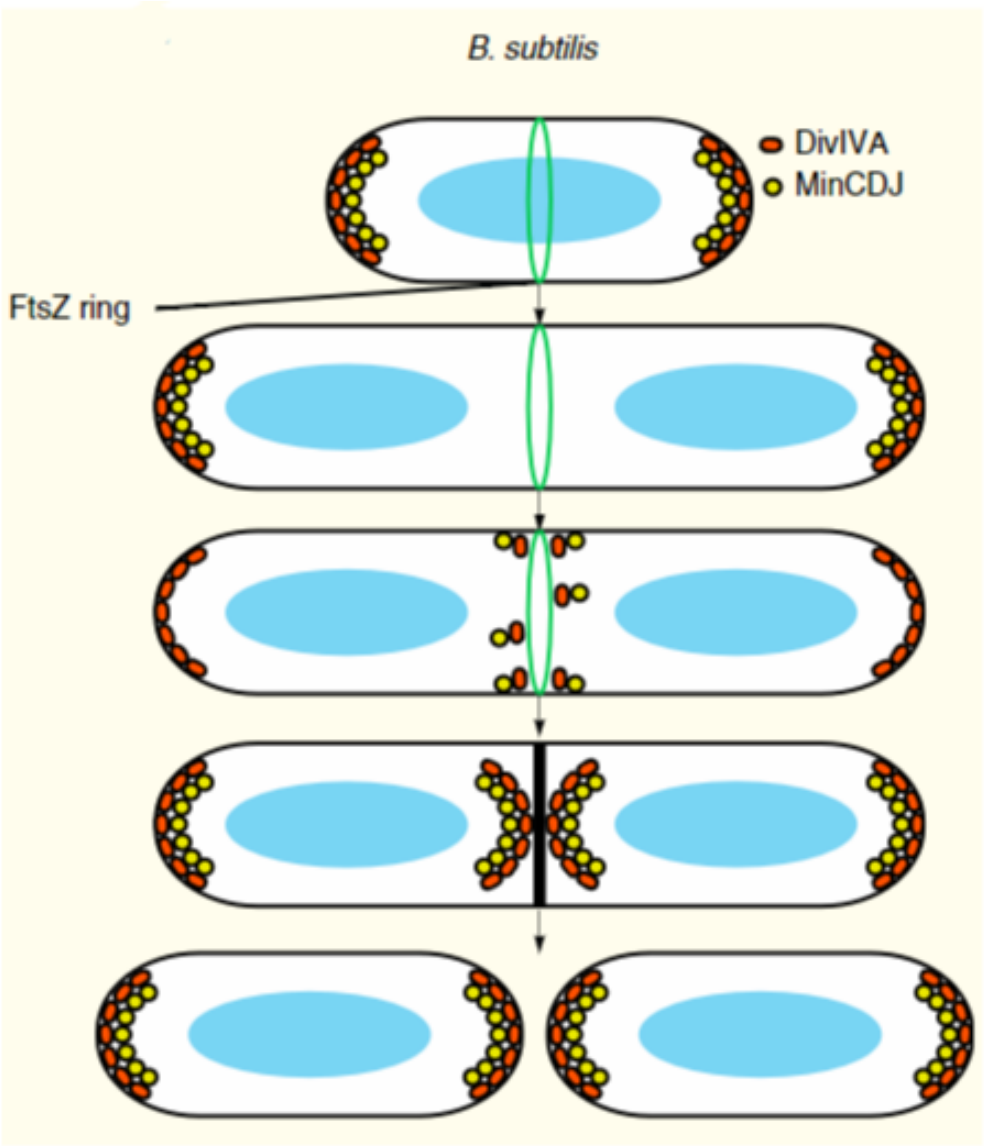


Figure 1-3. Localisation of MinD in *B. subtilis*. DivIVA (red) localises along the negative curvature of the poles of the cell and recruits MinJ (yellow), as an intermediate, to recruit MinD (yellow). MinC (yellow) is recruited by MinD and prevents the polymerisation of FtsZ at the poles. FtsZ localises and polymerises into a Z-ring (green) to form the division septa. With formation of the division septa, DiviVA, MinJ and MinD are recruited in turn to the division septa to prevent further division sites forming. Taken from (Rowlett & Margolin, 2013).

1.2.2 Chromosome segregation – the ParAB system:

Since we have now investigated the variations in mechanics and roles the MinD proteins play across bacterial species for septum positioning, we can begin to consider the functions and systems contained within the other subgroup of the superfamily: ParA. The ParAB system in *E. coli* (also known as the Type 1 plasmid partitioning system) was the first known example for the ParA subgroup of the ParA/MinD superfamily. The major components of the system are ParA (a weak Walker box ATPase), DNA-binding protein ParB homologue (can be known as

SpoB) and a cis-acting centromeric DNA site (known either as *parS* or *parC*) (Davis, Martin & Austin, 1992; Watanabe *et al*., 1989; Ebersbach & Gerdes, 2001). The characteristics of the proteins involved are ParA can dimerise when bound to ATP which regulates the oligomerisation of ParA, and when the ATPase function is activated, monomeric ParA is released along with ADP. This ATP function is triggered by interactions of ParA with ParB and confirmed with ATP-binding ParA mutants incapable of interacting with ParB. ParB, on the other hand, is a Helix-Turn-Helix DNA binding protein which binds to the *parS* DNA sites situated around the replication origins of either the plasmid or chromosome. As mentioned before, the mechanics of this system can vary across species, with the Type 1 plasmid partitioning system of *E. coli* functioning through oscillations of ParA to consistently localise and retain the plasmid to the chromosome. The most well-documented example is the *E. coli* plasmid pB171 system (Figure 1-4). This system is characterised by ParB binding to the *parS* site of the pB171 plasmid (Ringgaard *et al*., 2007) whilst ParA bind to one end of the chromosome through recruiting ATP (Leonard, Butler & Löwe, 2005). The activated ParA (has ATP bound) begins to polymerise and form a filament along the chromosome. The filament continues to extend until contact with ParB of the ParB-*parS* complex whereupon the ParA ATPase function is activated, and ATP is hydrolysed to ADP (Gerdes, Howard & Szardenings, 2010). Consequently, this causes the dismantlement of the ParA residue from the filament, releasing a ParA monomer. Structurally, this results in the filament contracting allowing another activated ParA to bind to the ParB-*parS* complex and pulling the complex. This process is repeated until the plasmid complex is located at the opposite end of the chromosome. Simultaneously, as the ParA monomers are released, new ParA monomers bind to the opposite end of the chromosome with ATP, forming a new ParA filament. This oscillation of ParA continues to shift the plasmid along the chromosome from end to end (Ebersbach & Gerdes, 2001; Ebersbach *et al*., 2006). This continual oscillation has been observed with high resolution microscopy and if multiple plasmids are involved an equilibrium between the plasmids occurs, where all the plasmids are evenly distributed across the chromosome and can all be retained in identical offspring (Ringgaard *et al*., 2009).

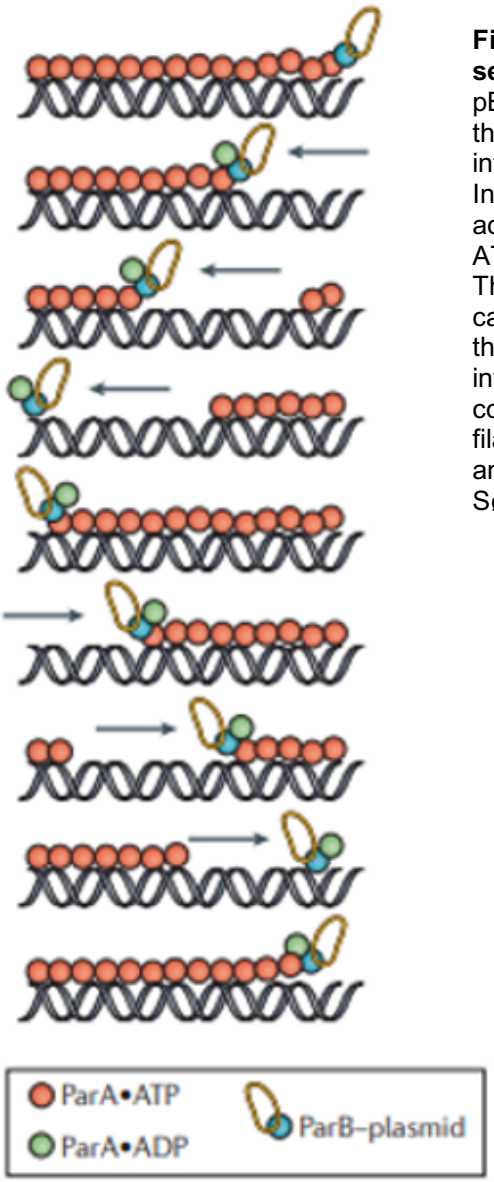


Figure 1-4. Osciliation of ParA for the segregation of the *E. coli* pB171 plasmid. The pB171 plasmid is tethered to the chromosome via the formation of a ParB-*parS* complex (blue) interacting with an ATP-bound ParA filament (red). Interaction of ParB and ATP-bound ParA, activates the ATPase activity of ParA, converting ATP into ADP to for ADP-bound ParA (green). This disassembles the ParA filament which turn causes it to contract and pull the plasmid towards the opposite end of the chromosome. This interaction continues until the ParA filament is completely disassembled. Meanwhile, a new ParA filament forms at the other end of the chromosome and the process is repeated. Taken from (Lenz & Søgaard-Andersen, 2011).

Another well-documented ParAB system is found within the crescent-shaped bacteria *Clautobacter crescentus*. The same interactions between ParA, ParB and *parS* occur as in the ParB-*parS* complex forms and ParA polymerises/depolymerises in the presence/hydrolysis of ATP respectively (Ptacin *et al*., 2010). However, the overall ParAB mechanism within *C. crescentus* does not function through oscillation (Figure 1-5). The use of other partner proteins is required for the system to function: PopZ and TipN. PopZ localises at the old pole (where the cell division event to create this cell occurred) and forms a polymeric complex that directly binds to the ParB-*parS* complex to anchor the chromosome to the old pole. It also recruits temporal regulators such as CckA to control the cell cycle progression. (Ebersbach *et al*., 2008). *popZ* null mutants appeared to show

cell division defects and a lack of polar organisation (Bowman *et al*., 2008). On the other hand, TipN is a coiled-coil protein vital for the polarity of the cell at the new pole. *tipN* null mutants exhibited polarity defects which included mislocalisation of the division machinery (Lam, Schofield & Jacobs-Wagner, 2006). Through disruptions of either of these polar proteins, the processes involved in cell division are impaired. The original model for chromosome segregation within *C. crescentus* was the original chromosome was anchored to one pole via PopZ interacting with the ParB-*parS* complex. Then TipN localised at the new pole, recruited ParA and in an ATP-dependent manner, polymerisation of ParA occurs. The ParA filament continues to extend until interacting with the ParB-*parS* complex of the replicated chromosome and the ParA-ParB mechanism would relocate the chromosome to the opposite pole (Ptacin *et al*., 2010). However, further research revealed that instead PopZ releases the ParB-*parS* complex during chromosome replication and PopZ monomers localise to both poles of the cell. The adjoined chromosomes, at the centromeres, drift apart. Meanwhile, ParA is recruited and polymerised at both poles, in the presence of ATP, with both filaments binding to the ParB-*parS* complex of each chromosome. The same ParA-ParB mechanism pulls each chromosome to opposite poles (Ptacin *et al*., 2014).

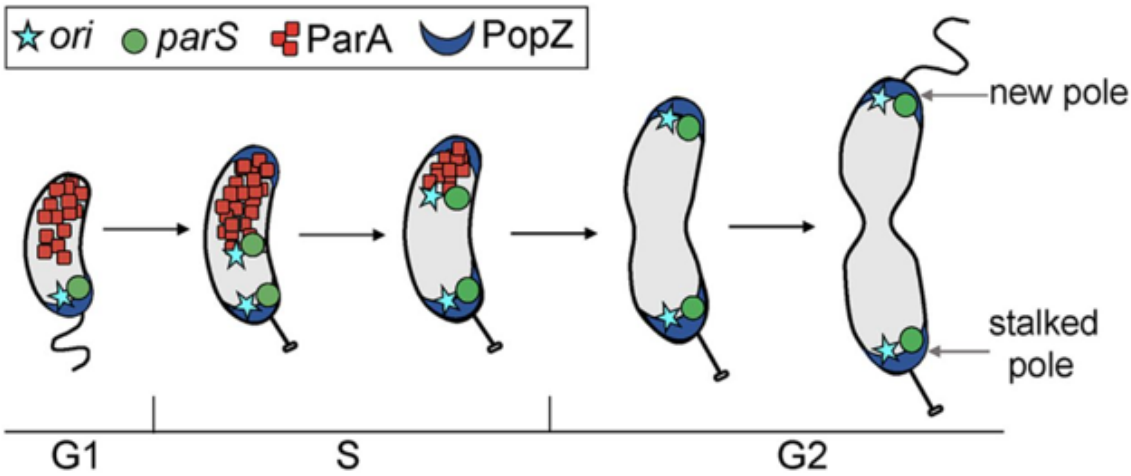


Figure 1-5. Chromosome segregation in *C. crescentus*. The chromosome is tethered to a pole during most of the life cycle of *C. crescentus*, through a ParB-*parS* complex (green dot and star) binding to PopZ – a membrane tethering protein (blue crescent). Once the chromosome is replicated, ParA filaments (red squares) accumulates from the opposite pole and adjoin to the replicated chromosome at the ParB-*parS* complex. The ParA filament retracts and pulls the replicated chromosome to the opposite pole for tethering through ParB and PopZ interacting. At this point, cell division can occur. Taken from (Meléndez, Menikpurage & Mera, 2019).

Finally, another unique ParAB model is found within *B. subtilis*. The *B. subtilis* system utilises ParA and ParB homologues (Soj and Spo0J respectively) which possess the shared biochemical functions across the subgroup, but Soj and Spo0J have extra partner proteins and capacities within the system. An example of this the inability of sporulation seen in *spo0J* null mutants an Soj acting as a suppressor of Spo0J (Ireton, Gunther & Grossman, 1994). Null mutants of both Soj and Spo0J have shown chromosomal segregation defects in both forms of division: sporulation and vegetative division which can lead to cells being anucleate. These defects highlight their role in chromosome segregation (Sharpe & Errington, 1996; Ireton, Gunther & Grossman, 1994). Alongside the defects in the cell life cycle, the direct interactions of Soj dimerising when bound to ATP and binding to DNA (Hester & Luthkenhaus, 2007), the upregulation of Soj ATPase activity when interacting with Spo0J (Scholefield *et al*., 2011) and the formation of the ParB-*parS* complex near the *oriC* (Lin & Grossman, 1998; Breier & Grossman, 2007) have been confirmed. But, as mentioned before, Soj and Spo0J have a range of different capacities which are especially prevalent with Spo0J. As stated before, anucleate cells form in s*po0J* null mutants due to the lack of chromosome segregation. This is due to Spo0J utilising other partner proteins to organise the chromosome for segregation: SMC (Structural Maintenance of Chromosomes), ScpA and ScpB. SMC is recruited to Spo0J, which has formed a ParB-*parS* complex around the *oriC* of the chromosome and lack of SMC localisation leads to defects in the chromosome segregation (Sullivan, Marquis & Rudner, 2009; Gruber & Errington, 2009). The ScpA and ScpB proteins form a complex with SMC (as the SMC-ScpA-ScpB complex) around the origin and *scpA* and *scpB* null mutants also demonstrated defects in chromosome segregation (Soppa *et al*., 2002). A unique role to Soj is the dimeric complex appears to trigger overinitiation of DNA replication though direct interaction between the dimer and DnaA (the initiator protein for DNA replication) is yet to seen (Murray & Errington, 2008). The overinitiation may be triggered by Soj allowing DnaA greater access to the *oriC* through conformational changes in the chromosome. The other function of Soj is directly binding with DnaA in its monomeric form. This interaction has been deemed to be a part of a larger system called the Sda checkpoint which regulates the change to sporulation (Murray & Errington, 2008). The steps within the checkpoint are the dimeric Soj enhances DNA replication which delays the change to sporulation. Interaction of Spo0J activates the ATPase of Soj and disassembles the complex. The monomeric Soj interacts with DnaA to form a new complex and inhibit DNA replication. Collectively, this acts as a switch from DNA replication to

sporulation and helps to explain the lack of sporulation seen in the *spo0J* null mutant.

1.3 Bacterial growth:

With a throughout understanding of the shared bacterial systems for cell division, we can now focus on the common systems of bacterial growth. Lateral growth is commonly driven by homologues of MreB, with these homologues first being identified in *B. subtilis* (Levin *et al*., 1992) and *E. coli* (Wachi *et al*., 1987). In *E. coli*, MreB, is found within an operon with several other genes: MreC and MreD which have been suggested to form a complex together (MreBCD complex) through their known interactions (Kruse, Bork-Jensen & Gerdes, 2005). These proteins have been connected to determining cell shape and knocking them prevents the formation of the rod shape (Wachi *et al*., 1989; Doi *et al*., 1988). Like MreB, homologues of MreC and MreD are found within other bacteria including *B. subtilis* (Levin *et al*., 1992; Varley & Stewart, 1992). Along with the MreBCD complex, other homologues of MreB have been found in *B. subtilis* called Mbl (MreB-like) and MreBH (MreB Homologue) (Jones, Carballido-López & Errington, 2001; Carballido-López *et al*., 2006). These homologues have been shown to influence cell shape but through slightly different ways. Mbl, alongside MreB, was found to form filaments which bind to the cell membrane and follow a helical pattern from pole to pole. The filaments were identified through GFP-tagged fusions (Mbl) and Immunofluorescence (MreB). The pattern of localisation has been related to eukaryotic actin with having a similar cytoskeletal role (Jones, Carballido-López & Errington, 2001). This similarity was later confirmed when the structure of MreB was solved (van den Ent *et al*., 2014). In addition, these homologues have been shown to influence the physiology of *B. subtilis* differently (through null mutants) and interact with hydrolytic enzymes (CwlO and LytE) involved in peptidoglycan synthesis (Defeu Soufo & Graumann, 2006; Domínguez-Cuevas *et al*., 2013). In conjunction to Mbl, MreBH was found to co-localise with the other MreB isoforms in the helical pattern and interact with LytE (a cell wall hydrolase). Mutants of MreBH or LytE showed defects in the morphology of the cell wall (Carballido-López *et al*., 2006). Filaments of these three homologues appear to follow a helical localisation pattern, whereupon they interact with different cell wall synthesis machinery and contribute to their regulation (Daniel & Errington, 2003). However, the pattern of this model has contested (Figure 1-6). Previous research has suggested that MreB localises in patches along the cell membrane and the motion of MreB is controlled

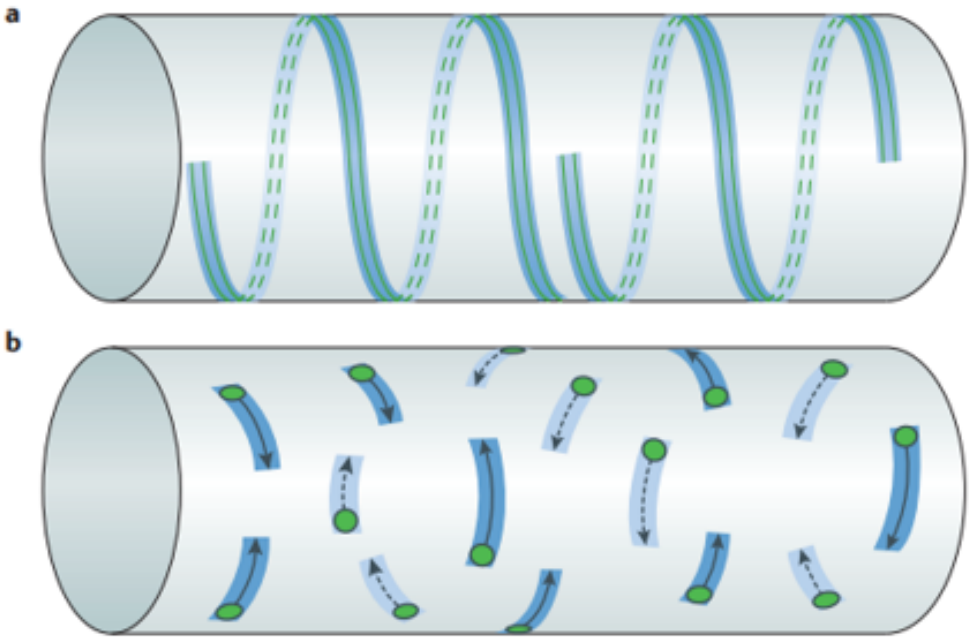


Figure 1-6. The 2 considered models for effect of MreB on lateral growth and shape determination. (a) The accepted model for MreB localisation. MreB formed actin-like filaments (solid and dashed lines) which spiralled along the cell wall. The helical arrangement of the filaments caused peptidoglycan to be inserted into the cell wall in a helical fashion (blue shading). (b) The contesting model for MreB localisation. MreB localises as distinct patches along the cell wall (green patches) which rotate along the circumference of the cell wall (arrows show direction) and dependent of the peptidoglycan synthesis machinery (blue shading). Taken from (Errington, 2015).

by the cell wall synthesis machinery, rather than a long MreB filaments (Domínguez-Escobar *et al*., 2011; Garner *et al*., 2011; Swulius *et al*., 2011). There has been other recent data showing MreB forms antiparallel filaments that interact with the cell membrane *in vitro* (van den Ent *et al*., 2014; Salje *et al*., 2011) which supports the helical model initially suggested. From all this data, the current model stands that MreB homologues assemble into filaments which naturally accumulate to the membrane of the cell. The associated filaments continue to extend and follow a cylindrical path along the membrane. These ever-growing filaments recruit multiple cell wall synthesis complexes and begin synthesis of new peptidoglycan (Errington, 2015).

1.3.1 Polar growth:

For many bacterial species, lateral growth is utilised for expansion of the cell, though another form of growth (polar growth) is used instead. Key phylum of bacteria known to use the polar growth mechanism are Gram-negative Pseudomonadota (such as *Agrobacterium* and *Rhizobium*) (Brown *et al*., 2012) and Gram-positive Actinobacteria (Daniel & Errington, 2003). DivIVA was the first protein tied to driving polar growth; with discovery of the protein in *B. subtilis* although *B. subtilis* utilises DivIVA at the poles for division, with the partner proteins: MinJ and MinD (Cha & Stewart, 1997; Patrick & Kearns, 2008). Since then, homologues of DivIVA have been found across many different species including model organisms Actinobacteria such as *Streptomyces coelicolor* (Flärdh, 2003) and *Mycobacterium tuberculosis*. The DivIVA homologue within *M. tuberculosis* is known as Wag31 (Kang *et al*., 2008). With DivIVA homologues found across many Gram-positive bacterial species, it is unsurprising that these homologues share similar functions even though they are used for different mechanics (Growth or Cell division). A shared characteristic is that DivIVA homologues appear to localise at the pole of the cell (Flärdh, 2003; Marston *et al*., 1998). Within generated *divIVA* null mutants of *Corynebacterium glutamicum*, DivIVA homologues of *S. coelicolor* managed to make the mutant viable whilst homologues from *B. subtilis* were unable to (Letek *et al*., 2008). This suggests enough variance in the homologues that their maybe evolutionary divergence, though if this is the case the characteristics of the homologues could vary by species. An example of this is the *M. tuberculosis* DivIVA homologue Wag31 can directly interact with the penicillin-binding protein 3 (PBP3) to protect it from oxidative stress-induced cleavage (Mukherjee *et al*., 2009). Although Gram-positive bacteria that grow via polar growth have been found to possess DivIVA, a homologue of DivIVA within Gram-negative bacteria utilising polar growth is yet to be found. A common mechanism within these bacteria has not been identified to date (Oliva *et al*., 2010).

Even though a DivIVA homologue in Gram-negative bacteria is yet to be found, some Gram-negative species have been identified using polar growth, nonetheless. This is especially the case within the *Rhizobiaceae* family, with various species known to use polar growth such as the *Agrobacterium tumefaciens* and *Sinorhizobium meliloti*. These species both utilise unipolar growth from a new pole which demonstrates the inherited characteristics of this variant of polar growth (Brown *et al*., 2012). Within these species, their models of growth share similar components to well-documented cell division systems from both Gram-positive and

Gram-negative bacteria. *S. meliloti* is known to utilise PodJ which is used to help determine the cell polarity and *A. tumefaciens* possess FtsA and FtsZ homologues which are seen to localise at the growth pole especially (Cameron, Zupan & Zambryski, 2015). These models acquisition of universally, highly conserved cell division proteins as components demonstrates the unique evolution of these models. Out of these two species, the *A. tumefaciens* polar growth model has been documented in further detail than *S. meliloti*. In a similar fashion to the *S. meliloti* model, *A. tumefaciens* uses PodJ, alongside PopZ to help determine cell polarity for polar growth. Time-lapse and super-resolution microscopy revealed that a homolog of PodJ localises at the old pole during growth. The homolog of PopZ was found to solely accumulate at the growing pole even in newly formed offspring (Grangeon *et al*., 2015). Deletion of PodJ resulted in abnormal behaviours like branching, growth poles that failed to transition and long cells that couldn't divide (Anderson-Furgeson *et al*., 2016), highlighting the importance of PodJ in maintaining unipolar growth through determining cell polarity. Alongside these repurposed cell division proteins, the *A. tumefaciens* model has been found to use a unique protein complex to coordinate polar growth known as the Growth Pole Ring (GPR). GPR forms a hexameric ring structure at the growing pole of the cell and is largely constructed of continuous α-helices, with homologs found across many Rhizobiales (Zupan *et al*., 2019). GPR were also seen localising at the midcell prior to cell division and at the growth poles of newly formed sibling cells (Zupan *et al*., 2019). The deletion of the GPR monomer C-terminus and human apolipoprotein A-IV like coiled-coil domains lead to severe defects in cell morphology, with very little complementation to cells lacking GPR (Zupan *et al*., 2021), demonstrating the role of the PGR in organising peptidoglycan synthesis. Recently one of these homologs of GPR from the Rhizobiales Order has been identified and analysed. The GPR homolog (RgsE) has been found in *S. meliloti* which is needed for maintaining polar growth and rod cell shape. Deletion of RgsE caused a defect in the cell morphology but the deletion strain was capable of growing and dividing at a slow rate (Krol *et al*., 2020). Both *A. tumefaciens* and *S. meliloti* appear to utilise a distinct protein complex to help organise the peptidoglycan/membrane synthesis during polar growth as an alternative to DivIVA, with some components recruited by the complex (such as FtsZ and FtsA) being homologues found across various species. This helps highlight the fact that other alternative protein complexes can used for different segments of polar growth mechanisms within species.

1.4 Streptomyces coelicolor:

After outlining the required proteins for many different mechanisms need within the life cycles of bacteria, we will now focus on the bacterium *S. coelicolor*. A well-documented Actinobacteria bacteria of filamentous polar growth and the main model of organism of this study. *S. coelicolor* is GC rich, soil-dwelling, Gram-positive bacteria known for its vast secondary metabolite production. *S. coelicolor* and the rest of its phylum differs from most bacteria, not only due to its filamentous structure, but also due to its possession of linear chromosomes (Lin *et al*., 1993). The 8 Mbp linear chromosome of *S. coelicolor* possess 7,825 genes, containing 20 estimated clusters for secondary metabolites. The chromosome consists of a central region which contains vital genes for the primary metabolites and the origin of replication (*oriC*). The central region is then flanked by two terminal arms which contain the genes for secondary metabolites. In conjunction to the large chromosome, approximately 12.5% of the genes within are predicted to encode for regulatory functions such as putative sigma factors and DNA binding proteins (Bentley *et al*., 2002). Collectively this highlights the vast complexity of the genome and life cycle of *S. coelicolor*.

1.4.1 S. coelicolor life cycle:

As stated previously, *S. coelicolor* grows via filamentous growth which has led to it possessing a complex life cycle remarkably similar to that of fungi. This life cycle can be roughly broken down into 4 different stages: Germination, Vegetative growth, Aerial growth and Sporulation (Figure 1-7). The first stage within the life cycle begins with the germination of a single uni-genomic spore production of a single or two germ tubes which is initiated under suitable conditions (Jyothikumar *et al*., 2008). The germ tubes continue to grow and branch through polar growth at the apical tip of the hyphae to form a vegetative mycelium (known as the Vegetative stage). The hyphal network formed are used to acquire nutrients from the surroundings and the network branching events occur along the lateral wall of the hyphae. Interestingly, the hyphal branching events have been shown to have an increased rate of DNA replication which is proportional to the rate of growth (Flärdh, 2003). During the vegetative growth stage, division of the hyphae is suspended where cross-walls are formed through partial septation occurs instead. Since growth only occurs at the tip of the hyphae, hyphal branching is required for *S. coelicolor* to grow at an ever-increasing rate through increasing the number of tips. The growing vegetative stage of *S. coelicolor* is commonly described as possessing a bald

phenotype when grown on nutrient agar. When nutrients are lacking, *S. coelicolor* switches from the vegetative stage to the growth of aerial hyphae into the air, away from the media (aerial growth stage). The change in nutrients causes a signalling cascade to trigger the switch (Kelemen & Buttner, 1998). The growth of the aerial hyphae results in long, unbranching hyphae with multiple genomes throughout (characterised as white, fuzzy colonies) which eventually progresses into the sporulation stage. Upon switching to the sporulation stage, evenly spaced septa are formed to compartmentalise single chromosomes of the hyphae (Flärdh & Buttner, 2009). The developing spore chain continues to develop until mature whereupon it is characterised by the production of a grey pigment (Kelemen *et al*., 1998) and signalling the end of the life.

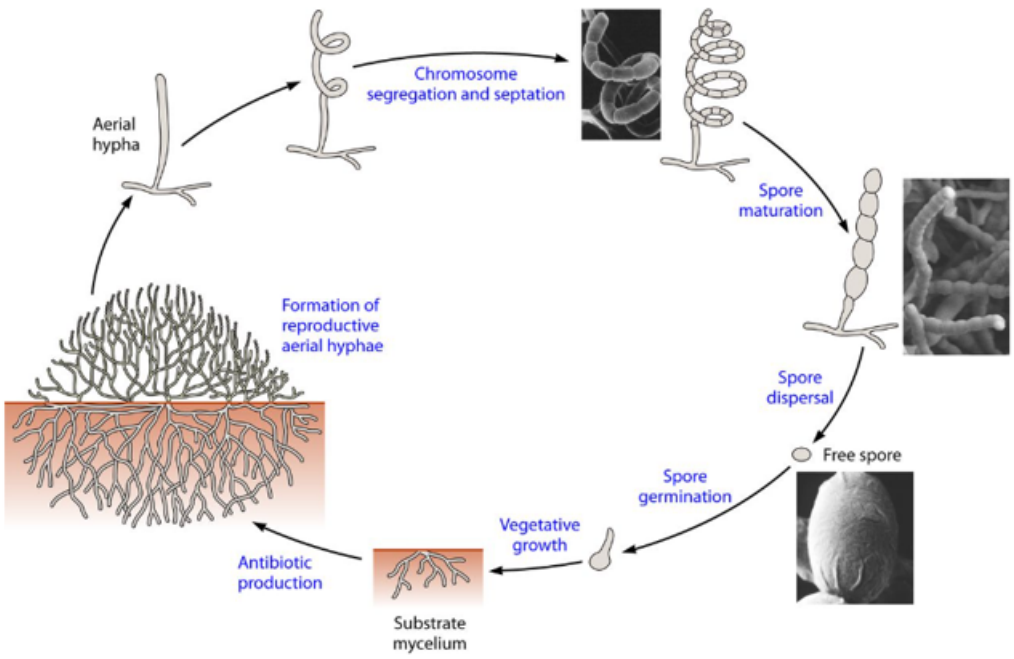


Figure 1-7. The life cycle of *S. coelicolor*. The life cycle begins with a free spore, which under desirable conditions, germinates a germ tube. The germ tube is grown and branched into the surrounding area to form a mycelial network for acquisition of nutrients. Upon insufficient nutrients being available, a network of aerial hyphae is constructed into the air. Upon adequate aerial hyphae growth, the aerial hyphae are transitioned from polar growth to the formation of septa to mark future spores. The spores are developed, through segregation of the chromosome and septation, to be dispersed for the cycle to begin again. Taken from (Barka *et al*., 2015).

1.4.2 S. coelicolor cell wall composition and synthesis:

With *S. coelicolor* being a Gram-positive bacteria, its cell wall is composed of alternating *N*-acetlyglucoamine and *N*-acetylmuramic acid residues which are adjoined through the use of short peptides containing both D- and L- amino acids. For the cross-linking of the short peptides, the D-Alanyl-D-Alanine motif has been identified as a highly conserved substrate which is crucial for the crosslinking enzymes (Egan *et al*., 2017). With this being the case, *S. coelicolor* has been shown to contain a range of glycoproteins including a D-Alanyl-D-Alanine carboxypeptidase. Null mutants of this glycoprotein have demonstrated increased levels of sensitivity to cell wall-binding antibiotics including Vanacomycin (Keenan *et al*., 2019). Although this is an important motif used in cell wall synthesis, it can be replaced with other similar substrates. These alternative substrates are usually used when the cell wall is not developing properly such as when the antibiotic Vanacomycin is used. The substrates D-Alanyl-D-Lactate and D-Alanyl-D-Serine can be substituted into the cell wall and bound to Lipid II instead of D-Alanyl-D-Alanine to allow for lower levels of resistance. D-Alanyl-D-Lactate lacks one hydrogen bond when bound whereas D-Alanyl-D-Serine contains an additional hydroxyl group when bound compared to D-Alanyl-D-Alanine. These differences in bonds allow for lower levels of drug resistance (Stogios & Savchenko, 2020). *S. coelicolor* has been shown to possess and express a cluster of seven genes (*vanSRJKHAX*) which help with the synthesis of one of these alternative cell wall substrates. The *vanHAX* genes of the cluster are responsible for reprogramming cell wall synthesis to produce precursors with the D-Alanyl-D-Lactate motif which induce resistance to Vancomycin. Null mutant strains of *vanHAX* had arrested growth at one tenth the concentration of Vancomycin to inhibit growth of the Wildtype strain (Hong *et al*., 2004). VanS, in the cluster, appears to act as a phosphatase which negatively regulates VanR-acetyl phosphate complex needed for the expression of the other *van* genes (Hutchings, Hong & Buttner, 2006). Additionally, cell wall synthesis can be altered indirectly by other proteins which control the site of synthesis. The *S. coelicolor* Mre protein complex has been inferred to form a spore wall synthesis complex which when knocked out produced for spores with increased sensitivity to Vanacomycin (Kleinschnitz *et al*., 2011).

1.4.3 S. coelicolor polar growth – The TIP Organising Centre (TIPOC):

S. coelicolor utilises a unique structure to drive polar growth within its hyphae. This structure is constructed from multiple protein complexes and is called the Tip Organising Centre (TIPOC) (Figure 1-8). This structure is present at the apical tip of growing hyphae at all times (Holmes *et al*., 2013; Flärdh *et al*., 2012). Three major proteins are required to form this complex and are known to influence growth and branching of the hyphae. These three proteins are: DivIVA, FilP and Scy (*Streptomyces* cytoskeletal protein) and these proteins are known to share a similar coiled-coil structure which helps give rise to their known domains. DivIVA is known to contain 2 heptad coiled-coil domains adjoined by a lengthy linker section but FilP and Scy possesses a different coiled-coil format. The structure of FilP and Scy is characterised by N-terminal heptad coiled-coils and a C-terminal 51mer coiled-coils, both connected by a linker (Walshaw, Gillespie & Kelemen, 2010). The shared heptad coiled-coils possess 2 non-polar residues (at positions 1 and 4) and 2 polar residues (at positions 5 and 7) which encourage the formation of an electrostatically formed homodimer from 2 helices (Mason & Arndt, 2004). The C-terminal 51mer coiled-coils are constructed from several repeating hendecad (11mer) regions which possess hydrophobic residues (at positions 1, 4 and 8). This leads to a standard series of repeats which is 7, 11, 11, 11, 11, 51 (Walshaw, Gillespie & Kelemen, 2010).

As stated previously, DivIVA homologues are found across many different species and utilised for polar growth at sites with negative curvature. The *S. coelicolor* DivIVA homologue structurally has been found to contain 2 vital domains which have been identified in altering two key functions of DivIVA: branching of the hyphae and oligomerisation of the protein. The 22 amino acids N-terminal section of DivIVA has been demonstrated to be critical for establishing new DivIVA foci along the hyphal lateral wall for new branches. This characteristic was highlighted through N-terminal deletion mutants alongside Egfp tagged fusions. For the oligomerisation of DivIVA, a coiled-coil domain (between 202 – 309 amino acids) was found to be vital for DivIVA monomers to form the higher order assemblies needed for polar growth (Wang *et al*., 2009). Alongside these functions, the N-terminal of DivIVA has also been linked to the natural behaviour of DivIVA to localise along negatively curved cell membranes through computer simulations. The simulations suggested that DivIVA N-terminal section uses 6 specific amino acid side chains (4 from arginine residues, 2 from leucine residues) to adsorb to phospholipids in the cell membrane through salt bridges, though these bonds are temporary, and probability

influenced by salt concentration and pH. Alongside this, the convex helical structure of polymerised DivIVA was believed to influence the formation of the negative curvature of the hyphal tip (Jurásek, Flärdh & Vácha, 2020). As the pole of *S. coelicolor* hyphae possesses the largest negative curvature within the hyphae, it is unsurprising DivIVA accumulates at these poles and polar growth is driven from these sites. The branching of hyphae is not only influenced by DivIVA structural properties, the expression of DivIVA heavily influences this function. High DivIVA expression resulted in swollen tips and hyperbranching of the hyphae, whilst low expression caused a short, hyphal tips which where disformed and possessed branches closer to the tip (Flärdh, 2003). The influence of DivIVA on branching has been further characterised through time-lapse microscopy. DivIVA was found to localise at the future sites for new tips before branches had started to form. This was seen in overexpressed strains where hyperbranching sites were outlined as distinct foci (Hempel *et al*., 2008). A suggested mechanism for DivIVA forming new sites for branches is the Tip-focused splitting of localised DivIVA which localises to the cell wall and accumulates into new branch sites. This mechanism was shown experimentally where time-lapse microscopy of DivIVA-Egfp should clumps of DivIVA-Egfp breaking away from the tip foci and transitioning to future branch sites lower down the hyphae (Richards *et al*., 2012). Though a problem with this mechanism is it only explains the formation of branches near the tip. Another regulatory mechanism of DivIVA is through phosphorylation/dephosphorylation by AfsK and SppA. The AfsK phosphorylation appears to cause the disassembly of the DivIVA larger structure, with *asfK* knockouts characterised by a reduction in branching (Hempel *et al*., 2012; Saalbach *et al*., 2013). The DivIVA dephosphorylation via SppA has been confirmed through both *in vivo* and *in vitro* experiments. Within these experiments, SppA appears to help regulate and maintain polar growth of *S. coelicolor* through preventing the dephosphorylation of DivIVA by Asfk (Passot, Cantlay & Flärdh, 2022). The dephosphorylation of DivIVA by SspA influences both the branching of hyphae and tip extension by polar growth in *S. coelicolor* as *sspA* null mutants demonstrated frequent and spontaneous growth arrests. This led to issues in maintaining tip extension. However, the a*sfk-sspA* double mutant averted some of the phenotypic effects of SspA confirming both AsfK and SspA target DiviVA (Passot, Cantlay & Flärdh, 2022) and act together as part of a regulatory mechanism.

Originally, FilP was discovered within *S. coelicolor* through a previously known homologue (Avicel binding protein - AbpS) and the search for crescentin

homologue within the genome. The homologue of FilP was identified in *Streptomyces reticuli* due to possessing affinity to a crystalline form of cellulose (called avicel) hence its name (Walter, Wellmann & Schrempf, 1998). The *S. coelicolor* homologue was identified and analysed due to possessing a similarity to crescentin and was labelled Filamentous Intermediate-like protein (FilP). The structure of FilP consists several domains: a non-coiled head, a heptad coiled repeat (coil 1), a linker, a 51-mer coiled repeat (coil 2) and a non-coil tail (Alcock, Unpublished). Fluorescent monitoring of FilP revealed it localised around the apical tip hyphal curvatures as long cables (Bagchi *et al*., 2008). Moreover, recent studies have shown that FilP also assembles behind the apical tip and long the hyphae as a larger network (Fuchino *et al*., 2013). These long branching networks are capable of self-assembling *in vitro* and *in vivo* and directly interacts with DivIVA to establish a gradient of FilP away from the apical tip (Fuchino *et al*., 2013). The self-assembling networks have been identified to form rope-like and striated structures *in vitro* which has been capable of altering the morphology of *E. coli* cells *in vivo*. The self-assembling capacity of FilP was linked to the C-terminus of its second coil subdomain through expression of FilP variants *in vivo* (Alcock, Unpublished). Atomic force microscopy of *filP* knockouts has showed reduced cell wall rigidity at the hyphal tips of *S. coelicolor* (Bagchi *et al*., 2008). Together this suggests that FilP may play a role in the growth mechanics of *S. coelicolor* as well as providing structural support for the hyphae.

The last major protein involved in the TIPOC is Scy, which is known for its direct role in accumulating the other major TIPOC proteins together and consequently regulating their localisation to coordinate polar growth. Scy is a large protein (1326 amino acids) which is rich in alanine and glutamine and highly conserved across *Streptomyces*. Scy appears to localise at the apical tip and directly interacts with DivIVA and FilP. Pelleting assays and BACTH assays of Scy with either DivIVA or FilP confirmed these interactions, while null mutants of *scy* were found to reduce growth and generate smaller colonies to the wildtype. Moreover, overexpression of Scy caused hyperbranching confirming its role in the placement of branch sites. Alongside hyperbranching, Scy overexpression also resulted in altered localisation of DivIVA simultaneously (similar to overexpression of DivIVA). As these effects of Scy on DivIVA helps to confirm the direct interaction between them and their roles in coordinating and driving polar growth at the tip (Holmes *et al*., 2013). Even though Scy is highly important in regulating growth and branching at the apical tip, there are signs it influences spore formation within *S.*

coelicolor. *scy* mutants should signs of uneven and irregular spore sizes, branched aerial hyphae and abnormal chromosome segregation. Since aerial hyphae grow via polar growth, defects in growth may have directly altered spore formation or affected a mechanic coupling the two processes (Ditkowski *et al*., 2010).

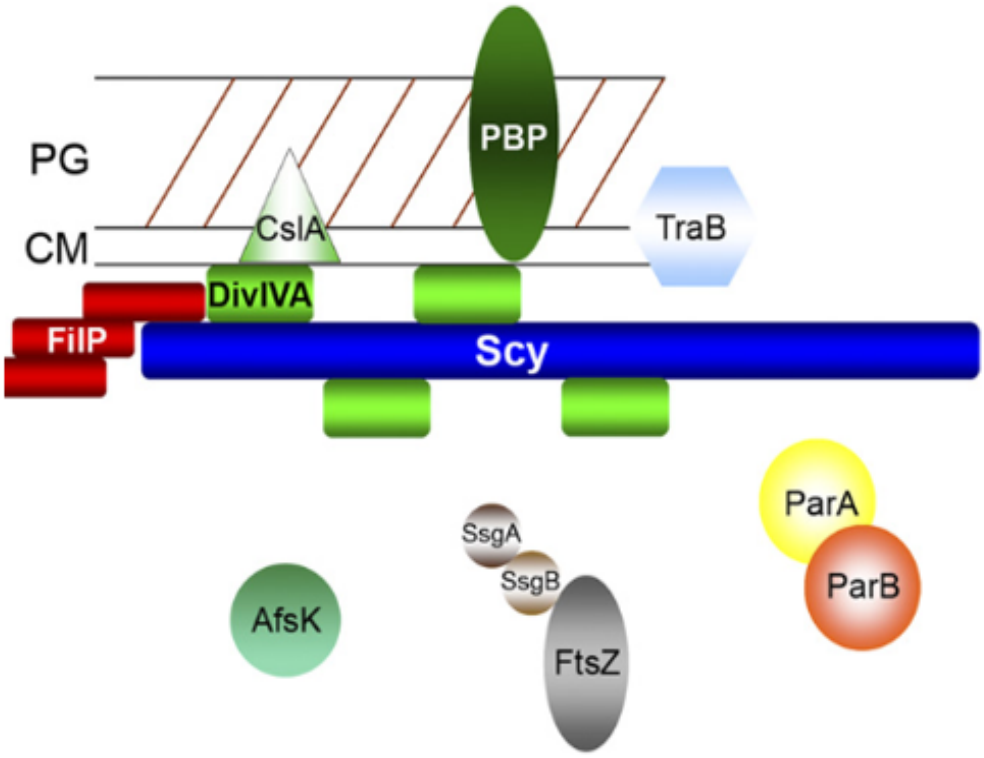


Figure 1-8. The structure of the TIPOC in *S. coelicolor*. The three main proteins that construct the TIPOC (DivIVA, FilP and Scy) help localise the assembly to the cell membrane at the hyphal tip to coordinate polar growth. Other proteins for regulation of components of the TIPOC or for additional interactions between the TIPOC and other systems, can be seen surrounding the assembly. Taken from (Holmes *et al*., 2013).

1.4.4 Cell division of S. coelicolor through FtsZ:

FtsZ is critical for most bacteria, in the production of septa during cell division, to produce identical offspring. However, *S. coelicolor* is capable of functioning without FtsZ even when the gene is knocked out, though its functions are limited. *ftsZ* null mutant strains are characterised by their white phenotype, lack of sporulation and reduced growth. The reduced growth is caused by the lack of compartmentalisation as the vegetative hyphae grow. Compartmentalisation is caused by FtsZ forming cross-walls in the periodically as the hyphae grow to provide structural support. The lack of structural support makes the hyphae more susceptible to breaking. If hyphae break, they seal themselves, with growth being

temporarily arrested. These mutant strains are also incapable of forming the sporulation septa for compartmentalising single chromosomes. As the septa can't form, division of the aerial hyphae never occurs and the strain fails to produce uni-genomic spores. This mutant essentially prevents the transition into the sporulation life cycle stage of *S. coelicolor* alongside reducing vegetative growth as a whole (McCormick *et al*., 1994).

For controlling the formation of FtsZ rings throughout the life cycle of *S. coelicolor*, three different promoters (P1, P2 and P3) are expressed at everchanging levels throughout the life cycle. P1 has increased expressed during vegetative growth and whilst P2 is solely upregulated before sporulation takes place. P3 is expressed continuously throughout the life cycle (Flärdh *et al*., 2000). With these different rates of expression of FtsZ at different stages, FtsZ can localise into the desired patterns necessary to form the FtsZ rings for that stage. Monitored vegetative growth through time-lapse microscopy showed the FtsZ rings forming the cross-walls through a spiral pattern of intermediates (Jyothikumar *et al*., 2008). During sporulation, an FtsZ ring localises to the base of the aerial hyphae to form a septum, separating the hyphae for compartmentalisation into spores (Kwak *et al*., 2001). The formation of the basal septum limits the enhanced FtsZ expression to the aerial hyphae which in turn promotes FtsZ ring formation. The FtsZ monomers form spiral-like intermediates into a dispersed pattern, similar to vegetative growth. The intermediates condense into helical filaments and eventually FtsZ rings periodically along the hyphae to mark the sites for septum positioning (Grantcharova, Lustig & Flärdh, 2005).

1.4.5 The regulation of FtsZ within S. coelicolor:

As we have talked about previously, positive regulation is a common system used to regulate where FtsZ localises within the cell. *S. coelicolor* is known to possess this such system, though the genes *ssgA* and *ssgB* (which are a part of the SALPs family). Both of these genes were originally identified and shown to influence the events of sporulation through null mutants. These mutants prevented the onset of sporulation, demonstrating their involvement in the regulation of cell division via sporulation (van Wezel *et al*., 2000; Keijser *et al*., 2003), with recent fluorescent microscopy studies have revealed details of how SsgA and SsgB regulate the events prior to sporulation (Willemse *et al*., 2011). The first step in the process begins with the SsgA naturally accumulating at the hyphal tip and at evenly spaced

positions down the hyphae (Young aerial stage). The SsgA foci along the hyphae, recruit SsgB and are fixed to the cell wall in an alternating fashion (Early division stage). SsgB begins to naturally accumulate on the opposite cell wall to the SsgA-SsgB complexes. At the same time, FtsZ forms a spiral-like filament along the hyphae, connecting to the SsgA-SsgB complexes (Early division stage). Following this, FtsZ is recruited to the SsgB foci (Pre-division foci stage) and both FtsZ and SsgB form ladder-like rings across the hyphae (Z-rings stage) (Willemse *et al*., 2011).

The properties of the SsgA and SsgB inferred from this model are: SsgA can localise independently of the other proteins involved and SsgB is involved in the correct localisation of FtsZ and the formation of the Z-rings. This function of SsgA was confirmed in *ftsZ* null mutants, with SsgA localisation unhindered. Interestingly, FtsZ was unable to correctly localise in *ssgA* null mutants confirming the reliance of correct localisation of SsgA. The property of SsgB was also confirmed in *ssgA* null mutants, with SsgB localising along the hyphal wall but not in the correct positions and unable to form rings that flank the Z-rings. *In vivo* FtsZ:SsgB interactions were confirmed through *E. coli* BACTH assays and *S. coelicolor* Forster resonance energy transfers. Finally, electron microscopy of *in vitro* FtsZ:SsgB interactions showed at FtsZ filament length was increased with SsgB present. (Willemse *et al*., 2011).

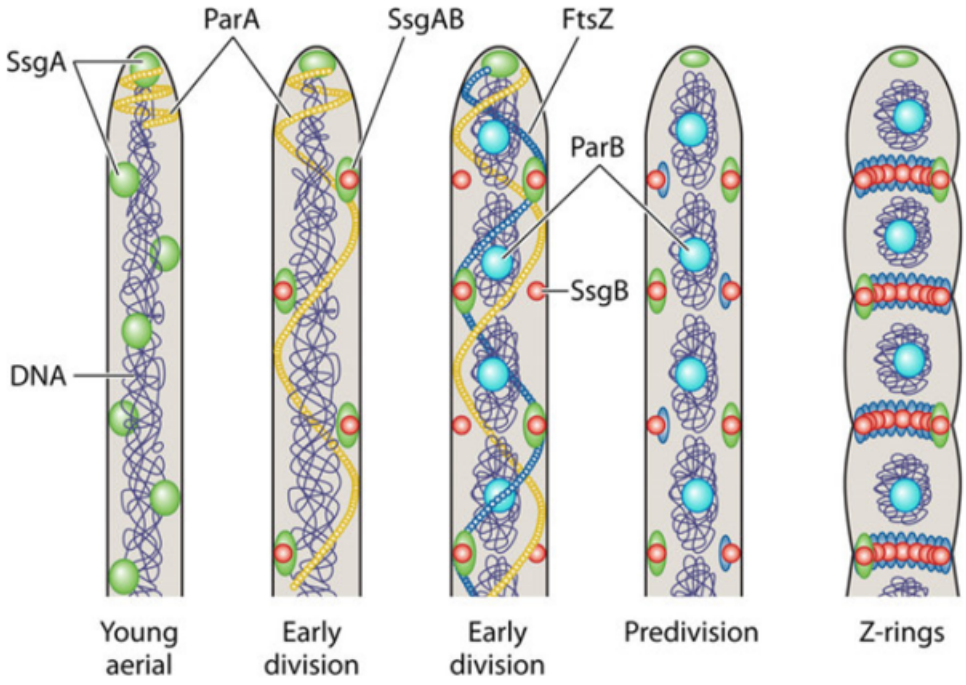


Figure 1-9. Positive regulation of FtsZ ring formation via the SsgA-SsgB complex in *S. coelicolor*. SsgA (green) form along the cell membrane at distinct, alternating foci within the young aerial hyphae, with ParA accumulated behind the SsgA foci located at the pole. Initially in early division, SsgB (red) is recruited to the SsgA foci (forming the SsgA-SsgB complex) and subsequently to the cell membrane opposite the complex. FtsZ forms a spiral-like filament running the length of the hyphae, bound to the alternating SsgA-SsgB complexes. In predivision, FtsZ shifts form a spiral-like filament to being tethered to the cell membrane by SsgB, opposite the SsgA-SsgB complex foci. The ParA filament has been completely dismantled. Following predivision, rings of SsgB form between opposite foci and are flanked by FtsZ rings to determine septation sites. Taken from (Barka *et al*., 2015).

1.4.6 Wider interactions of the TIPOC of S. coelicolor:

A link between polar growth and sporulation has been shown whereby Scy directly interacts with ParA. Scy was found to directly interact with ParA both *in vivo* and *in vitro*. *In vivo* BACTH assays and elution of His-Scy-ParA complexes demonstrated this, alongside *in vivo* co-localisation immunofluorescence studies of Scy-mCherry coupled with ParA-Egfp and Scy-Egfp coupled with ParA. Additionally, the signals for Scy were predominately seen in shorter aerial hyphae (that were still growing) and the ParA signals were seen in longer aerial hyphae (possibly preparing for sporulation) suggesting that both proteins only localise together for a brief time, possibly at a change between polar growth and sporulation. (Ditkowski *et al*., 2013).

The interaction between Scy and ParA is dynamic, with each protein regulating the polymerisation of the other. The regulation upon each other is determined by the quantity of both proteins, where the protein in higher

concentration depolymerises the other. This dynamic was first observed in *scy* null mutants where ParA localised at the hyphal tip, mirroring the wildtype phenotype, but overexpression of Scy did allow for ParA to localise at the tip and no ParA filaments were seen extending down the hyphae. This was further confirmed through *in vitro* pelleting assays and dynamic light scattering that, increased levels of Scy, inhibited ParA polymerisation and vice versa, but only in the presence of ATP (Ditkowski *et al*., 2013). With these interactions being the case, this dynamic is key for the change between polar growth and chromosome segregation. High levels of Scy limit ParA polymerisation at the tip so a single chromosome is anchored. When the process of sporulation is about to begin, ParA production is upregulated resulting in the disassembly of Scy and the TIPOC. Finally, ParA can polymerise throughout the hyphae to begin chromosome segregation. In addition to the Scy-ParA interaction, the relation between chromosome localisation and the TIPOC has been revealed through time-lapse microscopy. During active vegetative growth, ParA is found to only localise at the hyphal tip (with the TIPOC) whilst ParB is found bound to every chromosome throughout the hyphae. Due to the ParA-ParB interaction, a single chromosome can consistently be anchored to the hyphal tip during active growth and when chromosome replication occurs, ParA rebinds to one of the chromosome copies. This relation allows for a single chromosome to always be present at the hyphal tip during most of the life cycle of *S. coelicolor* for effective and responsive polar growth (Kois-Ostrowska *et al*., 2016).

Another interaction between a TIPOC component (DivIVA) and the ParB-*parS* complexes of *S. coelicolor* has been seen within various *in vivo* experiments. Fluorescently tagged DivIVA and ParB were found to localise and interact at the hyphal tip when coexpressed in *S. coelicolor*. This direct interaction was confirmed in a LexA bacterial two-hybrid system. (Donovan *et al*., 2012). This demonstrates that DivIVA holds a second function where it helps to recruit ParB to the cell pole and doing so likely helps anchors a chromosome to the region of active polar growth.

1.4.7 S. coelicolor chromosome segregation – ParAB system:

The genus of *Streptomyces* is drastically different in terms of its life cycle and genetic arrangement to most bacteria, with the genus is known to possess an exceptionally large linear genome (8 million bp roughly) (Lin *et al*., 1993). Coupled with a complex reproductive system through which 100s of identical spores are produced, each limited to containing a single chromosome. Due to the vast production of offspring, many chromosomes are replicated and segregated simultaneously which requires a more complex chromosome segregation system.

Like the other chromosome segregation systems (seen previously in *E. coli* and *B. subtilis*), *S. coelicolor* utilises the ParAB system with the main components being ParA (at locus *SCO3886*), ParB (at locus *SCO3887*) and the *parS* sites. A difference though is that there are 20 *parS* sites throughout the chromosome which ParB monomers form individual ParB-*parS* complexes with during chromosome segregation (Jakimowicz, Chater & Zakrzewska-Czerwi´nska, 2002). The function of the ParAB system was also characterised by null mutants of ParA and ParB resulting in spore chains with multiple anucleate spores. Δ*parA* strains were found to contain approximately 26% anucleate spores (Jakimowicz *et al*., 2007), Δ*parB* and Δ*parAB* strains containing 13% anucleate spores (Kim *et al*., 2000).

Just like the other ParAB systems, ParA is the first to localise, seen within the vegetative and growing aerial hyphae of *S. coelicolor*. Through immunolocalization, ParA was seen forming helices along the hyphae before FtsZ and substantial helices were seen when FtsZ started to localise. When the Z-rings had fully formed into their ladder pattern, the ParA signals could no longer be seen. In addition, ParA was found to mediate the formation of the ParB complexes through ATP-bound dimeric ParA and enhance their affinity to *parS* in *in vitro* studies (Jakimowicz *et al*., 2007). ParB was visualised instead by an Egfp-tagged fusion which localised in both vegetative and aerial hyphae. The ParB-Egfp signals were mainly seen near the tip of vegetative hyphae, though weak signals were seen throughout further down. The aerial hyphae, on the other hand, had bright ParB-Egfp signals seen repeatedly throughout (Jakimowicz *et al*., 2005) which would be bound to the bound to the *parS* sites within the chromosome (Jakimowicz, Chater & Zakrzewska-Czerwi´nska, 2002). In addition, time-lapse microscopy of FtsZ-YPET in ParAB mutants of *Streptomyces venezuelae* highlighted the influence the ParAB system has FtsZ ring formation. The mutant strains showed FtsZ rings forming earlier and persisting longer in the hyphae than their wildtype counterparts. This

however didn't impact the formation of the mature spores, or the time required to form (Donczew *et al*., 2016).

Finally, unlike previous ParAB systems, ParB is capable to two unique functions in *S. coelicolor* to enhance chromosome segregation. Firstly, deacetylation of ParB enhances the binding of ParB to the *parS* within the centromere of the chromosome. (Li *et al*., 2020). Secondly, ParB can recruit TopA (Topoisomerase A I) for remodelling of the topology of the DNA within the chromosome during sporulation. Currently TopA is the only known topoisomerase homologue in *S. coelicolor* (Champoux, 2001; Szafan *et al*., 2013). Null mutants of *topA* exhibit uneven ParB localisation and pre-spore compartments, with the null mutant causing lethal phenotype for the strain (Szafan *et al*., 2013). Collectively, the deacetylation of ParB and the modelling of the topology of the chromosomes appear to be crucial for correct chromosome segregation.

1.5 Potential TIPOC component of S. coelicolor:

As speculated previously, there are many proteins yet to be classified within *S. coelicolor* which may be involved in the polar growth mechanism. One such protein may be *SCO5569* which has been given the temporary name *dia*. Preliminary analysis of secondary structure of Dia from database entries appear to identify a coiled-coil region forming between residues 75 – 250 and an unstructured C-terminus (Ireland, Unpublished). As highlighted earlier, coiled-coil structures are common within both the N- and C-termini of the polar growth proteins (DivIVA, Scy and FilP) of *S. coelicolor* and are vital for their polar growth function. With this being the case, the theorised coiled-coil within Dia may allow for a similar function as the polar growth proteins and involvement within the TIPOC. To confirm if Dia is involved within this mechanism, previous research of *dia* null mutants has shown that Dia has an ability to influence the hyphal structure: both vegetative and aerial. During the vegetative growth stage, the determined diameter of the hyphae has been found fluctuate drastically throughout (Ireland, Unpublished). As the diameter of the hyphae is determined at the pole during active polar growth, Dia appears to influence the TIPOC mechanisms, directly or indirectly. In addition, the *dia* knockouts produced aerial hyphae containing sporulation septa that were unevenly distributed. This resulted in the formation of uneven compartments, some containing multiple chromosomes (Ireland, Unpublished). With the size of forming spores regulated through the FtsZ and ParAB systems of *S. coelicolor*, Dia may be able to

interact with these mechanisms as well. These 2 mechanics of *S. coelicolor* are regulated by two main systems: the TIPOC and the ParAB system which Dia seems to possibly influence or be a part of a novel component. To investigate if Dia is involved with these systems, localisation of a single crossover event using a DiaEgfp construct found that Dia accumulates behind the hyphal tip and at various points further down the hyphae (Ireland, Unpublished). We also confirmed this pattern again within vegetative hyphae in previous research I conducted using a double crossover version of DiaEgfp (Hutchinson, Unpublished). We also noticed that the DiaEgfp localisation pattern mirrored the DivIVAEgfp localisation pattern.

1.6 Aims:

DivIVA and its homologues are prevalent across many rod-shaped bacteria and are key for determining the process of polar growth. DivIVA recruits the necessary partner proteins to guarantee the successful initiation of polar growth and help regulate the mechanism. Through the key interactions of these different proteins, maintained polar growth can occur which is vital for cells to survive and thrive. However, it has been identified that some rod-shaped bacteria (especially Gram-negative bacteria) that appear to grow from polar growth do not possess a DivIVA homologue. With this being the case, other key proteins must be utilised to allow for polar growth to occur. *Streptomyces* possesses a complex life cycle where the constant growth and branching of many compartmentalised hyphae is necessary for the finding food sources and eventually the beginning of reproduction through sporulation. There are many proteins for determining and regulating polar growth with the most vital proteins being: DivIVA, FilP and Scy have been well documented forming a larger complex called the TIPOC to achieve these functions. DivIVA has been showing to initiate polar growth through recruiting cell wall synthesis machinery, while FilP has been shown to form large self-assembling structures that provide structural support throughout the hyphae. Scy appears to accumulate FilP and DivIVA together to localise the growth and structural support at the apical tip, coordinating polar growth. Alongside regulating polar growth, Scy and DivIVA has been shown to play a role in chromosome segregation through interaction with ParA and ParB. Given the complexity of the TIPOC and its use throughout the life cycle of *Streptomyces*, alongside the ParAB system, it is likely there are other key proteins involved in the TIPOC to help regulate aspects of the structure. A possible novel component is the previously identified gene *SCO5569*, which had been designated *dia* for its apparent affect to alter the width of hyphae as the cell grew. This study

plans to further investigate the function of Dia via determining the structural properties of the protein and if this protein could be a new novel component of the TIPOC or ParAB systems. We will determine the properties of Dia through analysis of its structure within computational models and the isolation of the protein whilst overexpressed. To determine if Dia is a novel component of the TIPOC or ParAB, we will explore possible partner proteins of Dia within the TIPOC and ParAB and test for direct interactions with Dia. Collectively, this should provide new understanding into proteins that influence polar growth and chromosome segregation within *Streptomyces* and may expand the current model used that underpins polar growth and chromosome segregation within the genus. Alongside this, this study will investigate the occurrence of rod-shaped bacteria, that grow by polar growth whilst not possessing a DivIVA homologue. We plan to search for homologues of other known polar growth proteins within their genome and isolate the selected proteins' localisation pattern through fluorescent microscopy for comparison to current polar growth models. Through this, we hope to start identifying possible alternative proteins that underpin polar growth across species to provide insight for these DivIVA-lacking rod-shaped bacteria.

Chapter 2:

Materials and methods

2.1 Bacterial strains and plasmids:

Details about the different bacterial strains and genetic constructs used during this study (Table 2-1 to 2-2).

Table 2-1 – *E. coli* strains used in this study.

Strain or plasmid	Relevant characteristic/description	Genotype	Source or reference
Escherichia coli K12			
DH5α	*E. coli cloning strain*	*F^- endAI hsdRJ7 (r-, mit) supE44 thi-J A-recAI gyrA96 relAI deoR A(lacZYA-argF)- U169 480dlacZAM1*	(Grant *et al*., 1990)
BW25113	*E. coli recombination strain*	Δ(*araD-araB*)*567* Δ(*rhaD-rhaB*)*568 ΔlacZ4787* (:rrnB-3) *hsdR514 rph-1*	(Grenier *et al*., 2014)
BL21 DE3	*E. coli overexpression strain*	*F^-, dcm, ompT, lon, $hsdS_B(r_B^- m_B^-)$, gal, λ (DE3), pLysS(cm^R)*	(Fu, Lin & Cen, 2007)
Rosetta DE3	*E. coli overexpression strain*	*F^- ompT $hsdS_B(r_B^- m_B^-)$ gal dcm lacY1 λ(DE3) pRARE (Cam^R)*	(Fu, Lin & Cen, 2007)
BTH101	*E. coli expression strain*	*F^-, cya-99, araD139, galE15, galK16, rpsL1 (Str^r), hsdR2, mcrA1, mcrB1*	(Karimova, Ullmann & Ladant, 2000)
HB101	*E. coli helper strain*	*F^-, rk^-, mk^-, leu^-, pro^-, $thia^-$, $lacZ^-$, rec A^-, endo 1^-, Str^R*	(Phadnis & Das, 1987)
Labrenzia aggregata LZB033			
L. aggregata	*Wildtype L. aggregata*	*Wildtype genome*	(Curson *et al*., 2017)

Table 2-2 – Plasmid/Cosmid DNA used in this study.

Strain or plasmid	Genotype	Source or reference
Cosmids		
7A1	Cosmid carrying the *dia* (*SCO5569*) region of the *S. coelicolor* chromosome.	(Redenbach *et al*., 1996)
Plasmids		
pUT18c	pUC19 derivative with T18 domain upstream of Multiple cloning site (MCS).	(Karimova, Ullmann & Ladant, 2000)
pUT18	pUC19 derivative with T18 domain downstream of MCS.	(Karimova, Ullmann & Ladant, 2000)
pKT25	pSU40 derivative with T25 domain upstream of MCS.	(Karimova, Ullmann & Ladant, 2000)
pKNT25	pSU40 derivative with T25 domain downstream of MCS.	(Karimova, Ullmann & Ladant, 2000)
pUT18c-zip	pUT18C with leucine zipper	(Karimova, Ullmann & Ladant, 2000)
pKT25-zip	pKT25 with leucine zipper	(Karimova, Ullmann & Ladant, 2000)
pUT18c-Dia	pUT18c with *dia*	This work
pUT18-Dia	pUT18 with *dia*	This work
pKT25-Dia	pKT25 with *dia*	This work
pKNT25-Dia	pKNT25 with *dia*	This work
pUT18c-DivIVA	pUT18c with *divIVA*	Kelemen lab, University of East Anglia
pKT25-DivIVA	pKT25 with *divIVA*	Kelemen lab, University of East Anglia
pUT18c-Scy	pUT18c with *scy*	Kelemen lab, University of East Anglia
pKT25-Scy	pKT25 with *scy*	Kelemen lab, University of East Anglia
pUT18c-FilP	pUT18c with *filP*	Kelemen lab, University of East Anglia

pUT18-FilP	pUT18 with *filP*	Kelemen lab, University of East Anglia
pKT25-FilP	pKT25 with *filP*	Kelemen lab, University of East Anglia
pKNT25-FilP	pKNT25 with *filP*	Kelemen lab, University of East Anglia
pUT18c-ParA	pUT18c with *parA*	Kelemen lab, University of East Anglia
pUT18-ParA	pUT18 with *parA*	Kelemen lab, University of East Anglia
pKT25-ParA	pKT25 with *parA*	Kelemen lab, University of East Anglia
pKNT25-ParA	pKNT25 with *parA*	Kelemen lab, University of East Anglia
pUT18c-ParB	pUT18c with *parB*	Kelemen lab, University of East Anglia
pKNT25-ParB	pKNT25 with *parB*	Kelemen lab, University of East Anglia
pUT18c-ParH	pUT18c with *parH*	Kelemen lab, University of East Anglia
pUT18-ParH	pUT18 with *parH*	Kelemen lab, University of East Anglia
pKT25-ParH	pKT25 with *parH*	Kelemen lab, University of East Anglia
pKNT25-ParH	pKNT25 with *parH*	Kelemen lab, University of East Anglia

pET28a	*ori pBR322, T7 Promoter, His•Tag coding sequence, lacI, kan, ori f1*	Novagen
pET28a-Dia	pET28a with *dia*	This work
pUC18-PromEgfpApra		Kelemen lab, University of East Anglia
pUC18-GlymEgfpApra	pUC18 with a GlyEgfpApra cassette	Kelemen lab, University of East Anglia
pUC18-PromEgfpSpec	pUC18 with a ProEgfpSpec cassette	This work
pUC18-GlymEgfpSpec	pUC18 with a GlyEgfpSpec cassette	This work
pUC18-PromCherryApra	pUC18 with a PromCherryApra cassette	This work
pUC18-GlymCherryApra	pUC18 with a GlymCherryApra cassette	Kelemen lab, University of East Anglia
pUC18-PromCherrySpec	pUC18 with a PromCherrySpec cassette	This work
pUC18-GlymCherrySpec	pUC18 with a GlymCherrySpec cassette	This work
pK18mobsacB	pK18 derivative that confers sucrose sensitivity	(Schäfer *et al*., 1994)
pK18mobsacB-Lcy	pK18mobsacB with Lcy	This work
pK18mobsacB-LcyProEgfpApra	pK18mobsacB with LcyProEgfpApra	This work
pRK2013	pRK2 derivative with the replicon of ColEI	(Figurski & Helinski, 1979)

2.2 Medias:

The components for each media needed to grow the different *E. coli* and *L. aggregata* strains used within this study (Table 2-3 to 2-4).

2.2.1 Solid media

Table 2-3 – Components of the solid media used within this study.

Type of media	Use	Components	Reference
Lennox Broth (LB)	Used to grow *E. coli* strains.	Bacto-tryptone: 16 g Yeast extract: 8 g NaCl: 8 g Glucose: 1.6 g dH_2O: up to 1600 ml	(Kieser *et al*., 2000)
Yeast Tryptone Sea Salts (YTSS)	Used to grow *L. aggregata* strains.	Bacto-tryptone: 2.5 g Yeast extract: 4 g Sea salts: 20 g dH_2O: up to 1000 ml	(González *et al*., 1996)

Procedure:

The LB ingredients above were dissolved into the dH_2O whilst 4 g of agar was measured and poured into separate 500 ml Duran bottles. 400 ml aliquots of the dissolved media were dispended into separate Duran bottles, mixed and autoclaved.

The YTSS ingredients above were dissolved into the dH_2O whilst 6 g of agar was measured and poured into separate 500 ml Duran bottles. 300 ml aliquots of the dissolved media were dispended into separate Duran bottles, mixed and autoclaved.

For YTSS + 10% Sucrose, the same procedure was performed but 30 g of Sucrose was measured, poured into each Duran bottle and dissolved into the 300 ml aliquots. Then 6 g of agar was measured and added to each aliquoted YTSS mixture, mixed and autoclaved.

2.2.2 Liquid media

Table 2-4 – Components of the liquid media used within this study.

Type of media	Use	Components	Reference
Lennox Broth (LB)	Used to grow *E. coli* strains.	Bacto-tryptone: 10 g Yeast extract: 5 g NaCl: 5 g Glucose: 1 g dH_2O: up to 1000 ml	(Kieser *et al*., 2000)
Super Optimal Broth (SOB)	Used to grow transformed *E. coli* BW25113 whilst recombining DNA	Bacto-tryptone: 10 g Yeast extract: 2.5 g NaCl: 0.29 g KCl: 0.093 g $MgCl_2$: 1.02 g $MgSO_4$: 1.23 g dH_2O: up to 500 ml	Kelemen lab, University of East Anglia
Yeast Tryptone Sea Salts (YTSS)	Used to grow *L. aggregata* strains.	Bacto-tryptone: 2.5 g Yeast extract: 4 g Sea salts: 20 g dH_2O: up to 1000 ml	(González *et al*., 1996)

Procedure:

The ingredients for LB, SOB and YTSS above were dissolved into the dH_2O. Either 10 ml or 50 ml aliquots of the dissolved media was dispensed into separate 10 ml universals or 250 ml conical flasks respectively.

2.3 Bacterial growth conditions and storage:

All *E. coli* and *L. aggregata* strains were grown in either Solid or Liquid required media (Table 2-3 to 2-4), containing the appropriate compounds (Table 2-5 to 2-7), at 30°C or 37°C (depending on following procedures).

Table 2-5 – Concentrations of antibiotics used in this study.

Antibiotic	Stock concentration (mg/ml)	Final concentration (µg/ml)
Kanamycin	100	50
Ampicillin	100	50
Apramycin	100	50

Table 2-6 – Concentrations of chromogenic substrates and overexpression-inducing compounds used in this study.

Compound	Stock concentration	Final concentration
X-gal	20 mg/ml	40 µg/ml
IPTG	1 M	0.2 mM

Table 2-7 – Concentrations of growth-inhibiting substrates and recombination-inducing compounds used in this study.

Compound	Stock concentration	Final concentration
Arabinose	1 M	100000 mM

2.4 General Molecular Biology Methods:

2.4.1 PCR

Used for the generation of altered *dia* DNA fragments through High-fidelity PCR. Two similar variants of the *dia* gene were generated: a wildtype copy and a variant lacking the stop codon. The genes were called *dia* and *diaUT.*

All PCR programs were performed using a Bio-Rad T100 Thermal Cycler and Phusion High-Fidelity DNA Polymerase (Thermofisher). The components, programs and primer sequences for each PCR can be seen below (Table 2-8 to 2-19).

Table 2-8 – PCR Master mix components used to generate the *dia* and *diaUT* fragments in this study.

Components	Volume (µl)
dH_2O	16.5
5 x GC buffer	10
dNTPs (1.25 mM each)	10
50% DMSO	5
$MgCl_2$ (25 mM)	5
7A1 cosmid template DNA	1
Thermo Fischer phusion DNA polymerase	0.5
50 pmole/µl forward primer	1
50 pmole/µl reverse primer	1

Table 2-9 – PCR Master Mix components used to isolate the end section of the *Icy* gene in this study.

Components	Volume (µl)
dH_2O	25
5 x GC buffer	10
dNTPs (1.25 mM each)	10
$MgCl_2$ (25 mM)	2
L. aggregata chromosomal DNA	0.5
Thermo Fischer phusion DNA polymerase	0.5
50 pmole/µl forward primer	1
50 pmole/µl reverse primer	1

Table 2-10 – PCR Master Mix components used to generate and extend the Spec cassette in this study.

Components	Volume (µl)
dH_2O	21.5
5 x GC buffer	10
dNTPs (1.25 mM each)	10
50% DMSO	3
$MgCl_2$ (25 mM)	2
Spec/Strep cassette template DNA	1
Thermo Fischer phusion DNA polymerase	0.5
50 pmole/µl forward primer	1
50 pmole/µl reverse primer	1

Table 2-11 – PCR Master Mix components used to extend the ProEgfpApra cassette in this study.

Components	Volume (µl)
dH_2O	22
5 x GC buffer	10
dNTPs (1.25 mM each)	10
50% DMSO	3
$MgCl_2$ (25 mM)	2
ProEgfpApra DNA gel extract	0.5
Thermo Fischer phusion DNA polymerase	0.5
50 pmole/µl forward primer	1
50 pmole/µl reverse primer	1

Table 2-12 – PCR conditions used to generate the *dia* and *diaUT* fragments in this study.

Temperature (°C)	Duration	
98	1 min	
98	30 sec	Repeat 10 times
55	30 sec	
72	1 min	
98	20 sec	Repeat 15 times
72	90 sec	
72	5 min	
20	5 min	

Table 2-13 – PCR conditions used to isolate the end section of the *Icy* gene in this study.

Temperature (°C)	Duration	
98	2 min	
98	30 sec	Repeat 20 times
65	30 sec	
72	1 min	
72	5 min	
20	5 min	

Table 2-14 – PCR conditions used to generate and extend the Spec cassette in this study.

Temperature (°C)	Duration	
98	1 min	
98	30 sec	Repeat 10 times
55	30 sec	
72	26 sec	
98	30 sec	Repeat 17 times
72	56 sec	
72	5 min	
20	5 min	

Table 2-15 – PCR conditions used to extend the ProEgfpApra cassette in this study.

Temperature (°C)	Duration	
98	1 min	
98	30 sec	Repeat 9 times
55	30 sec	
72	1 min	
98	20 sec	Repeat 14 times
72	90 sec	
72	5 min	
20	5 min	

Table 2-16 – Sequences of the oligonucleotides used to generate the *dia* and *diaUT* fragments in this study.

Name of primer	Sequence
5569 XbaNde FRW	GGATCATCTAGAGCATATGgtggacgtgcagaacaagctcg
5569 Eco REV	GCTACGAATCCctagagcccgcggccctgttcg
5569 EcoUT REV	GCTACGAATCCgagcccgcggccctgttcg

Table 2-17 – Sequences of the oligonucleotides used to isolate the *lcy* gene in this study.

Name of primer	Sequence
LcyEgfp_Eco FRW	ctagcGAATTCcgaatccatgcgcctcacgg
LcyEgfp_Eco REV	ctagcGAATTCtgctcccaagctgaaagccg

Table 2-18 – Sequences of the oligonucleotides used to generate the Spec cassette in this study.

Name of primer	Sequence
Apratospec FRW	GTCATCAGCGGTGGAGTCGAATGTCGTGCAATACGAATGGatgagggaagcggtgatcgc
Apratospec REV	TCAGCCAATCGACTGGCGAGCGGCATCGCATTCTTCGCATttatttgccgactaccttgg

Table 2-19 – Sequences of the oligonucleotides used to extend the ProEgfpApra cassette in this study.

Name of primer	Sequence
LcyEgfp_KI FRW	ggtctacaccatgctggcccacgccagcggccggctcgacCCGGTCGCCACCGTGAGCAAGG
LcyEgfp_KI REV	aggcggccagtggccgccttttttcgtgtctgactgtccgtCATATGTGTAGGCTGGAGCTGC

Amplification of *dia* and *diaUT* used 2 similar primer sets. *dia* was amplified using 5569 XbaNde FRW and 5569 Eco REV whereas *diaUT* was amplified using 5569 XbaNde FRW and 5569 EcoUT REV.

PCR reactions were set up as 50 µl Master mixes and run under specific program conditions using the corresponding primers to their template. Each PCR reaction used specific collections of components, conditions and primers to

successfully run: the *dia* and *diaUT* PCRs (Table 2-8, 2-12 and 2-16); the *lcy* PCR (Table 2-9, 2-13 and 2-17); the Spec cassette PCR (Table 2-10, 2-14 and 2-18) and the ProEgfpApra cassette PCR (Table 2-11, 2-15 and 2-19). 3 µl aliquot of each PCR product were reserved and analysed by Agarose gel electrophoresis (run at 30 V).

2.4.2 Colony PCR

Used for diagnostic purposes through Low-Fidelity PCR.

All PCR programs were performed using a Bio-Rad T100 Thermal Cycler and GoTaq DNA Polymerase (Invitrogen). The components, programs and primer sequences for each PCR can be seen below (Table 2-20 to 2-24).

Table 2-20 – PCR Master mix components used to confirm the pUT18c-Dia, pKT25-Dia, pUT18-Dia constructs in this study.

Components	Volume (µl)
dH_2O	53.25
5 x GC buffer	30
dNTPs (1.25 mM each)	30
50% DMSO	15
$MgCl_2$ (25 mM)	15
GoTaq polymerase	0.75
50 pmole/µl 5569 XbaNde FRW	3
50 pmole/µl reverse primer	3

Table 2-21 – PCR Master mix components used to confirm the pET28-Dia construct in this study.

Components	Volume (µl)
dH_2O	53.25
5 x GC buffer	30
dNTPs (1.25 mM each)	30
50% DMSO	15
$MgCl_2$ (25 mM)	15
GoTaq polymerase	0.75
50 pmole/µl 5569 XbaNde FRW	3
50 pmole/µl reverse primer	3

Table 2-22 – PCR Master mix components used to confirm the pKNT25-Dia construct in this study.

Components	Volume (µl)
dH_2O	142
5 x GC buffer	80
dNTPs (1.25 mM each)	80
50% DMSO	40
$MgCl_2$ (25 mM)	40
GoTaq polymerase	2
50 pmole/µl 5569 XbaNde FRW	8
50 pmole/µl reverse primer	8

Table 2-23 – PCR conditions used to confirm the pUT18c-Dia, pKT25-Dia, pUT18-Dia, pKNT25-Dia and pET28-Dia constructs in this study.

Temperature (°C)	Duration	
96	2 min	
92	1 min	
58	30 sec	Repeat 29 times
72	72 sec	
72	5 min	
20	5 min	

Table 2-24 – Sequences of the oligonucleotides used to confirm the pUT18c-Dia, pKT25-Dia, pUT18-Dia, pKNT25-Dia and pET28-Dia constructs in this study.

Solution	Sequence
5569 XbaNde FRW	GGATCATCTAGAGCATATGgtggacgtgcagaacaagctcg
5569 Eco REV	GCTACGAATCCctagagcccgcggccctgttcg
5569 EcoUT REV	GCTACGAATCCgagcccgcggccctgttcg

Analysis of pUT18C-Dia, pKT25-Dia, pUT18-Dia, pKNT25-Dia and pET28-Dia used similar primer sets. pUT18C-Dia, pKT25-Dia, pUT18-Dia, pKNT25-Dia were amplified using 5569 XbaNde FRW and 5569 EcoUT REV. pET28-Dia was amplified using 5569 XbaNde FRW and 5569 Eco REV.

Each analysis PCR reaction was set up as 10 µl aliquot of their corresponding Master mix (Table 2-20 to 2-22) using specific primers complementary to the clump of transformed DNA added (Table 2-24). A clump of transformed DH5α containing the desired DNA (pUT18C-Dia, pKT25-Dia, pUT18-

Dia, pKNT25-Dia and pET28-Dia) was inoculated into each aliquot. 10 µl positive controls (using 0.2 µl pUT18 DNA as a template) were set up for each colony PCR.

Each PCR reaction was run under the specific conditions (Table 2-23). Each 10 µl aliquot of each PCR product were analysed by Agarose gel electrophoresis (run at 30 V).

2.4.3 PCR purification

47 µl of the remaining PCR was washed with 53 µl dH_2O in the PCR tube, then mixed with 50 µl phenol (in a 1.5 ml Eppendorf tube) and spun at maximum speed in a centrifuge for 5 minutes. The top layer is removed very carefully (avoid extracting any of the middle layer) and mixed with 150 µl Chloroform (in a 1.5 ml Eppendorf tube) and spun at maximum speed in a centrifuge for 1 minutes. The top layer is removed carefully again and mixed with 10 µl 3M Sodium acetate and 400 µl cold 100% Ethanol (in a 1.5 ml Eppendorf tube). The tube is placed on dry ice for 5 minutes, then in a -20°C freezer for 1 hour. The tube is spun in the cold room for 10 minutes, followed by removal of the supernatant, addition of 200 µl cold 70% Ethanol and the tube being spun at maximum speed for 5 minutes. The supernatant is removed, and the pellet is left to air dry.

2.4.4 Cosmid/plasmid DNA isolation from E. coli

The solution components used for the DNA isolation procedure can be seen below (Table 2-25).

Table 2-25 – Components for extraction solutions used in the isolation of *E. coli* DNA constructs.

Solution	Composition
Solution I	50 mM Tris/HCl pH 8
	10 mM EDTA
Solution II	200 mM NaOH
	1% SDS
Solution III	3 M Potassium acetate pH 5.5

A single colony of DH5α containing the desired plasmid DNA were inoculated and incubated overnight, in 50 ml LB containing the appropriate antibiotic at 37˚C and 250 rpm. Overnight cultures of each DH5α strain were transferred into 50 ml falcon tubes and collectively centrifuged at 4000 rpm and 4˚C for 10 minutes.

All pellets were resuspended in 40 ml ice cold solution I and centrifuged under the previous conditions. Supernatants were discarded and the pellets were resuspended in 2 ml ice cold solution I through vortexing. Upon the solution becoming homogenous, 4 ml aliquots of solution II were added to each resuspension and gently mixed via inverting the tubes. All resuspensions were incubated on ice for 4 minutes. Afterwards, 3 ml aliquots of ice-cold solution III were added to each lysate, followed by vigorous mixing. All resuspensions were incubated on ice for 10 minutes, followed by centrifugation at 4000 rpm and 4˚C for 10 minutes. The supernatants were transferred to separate 15 ml falcon tubes, with 2 ml 1:1 phenol:chloroform added afterwards. All extracts were vortexed for 30 sec and centrifuged at 4000 rpm and 4˚C for 7 minutes. All upper aqueous phases were transferred to new 50 ml falcon tubes. 12 ml aliquots of -20°C 2-propanol was added to each tube, mixed thoroughly, followed by incubation on ice for 20 minutes. Afterwards, all tubes were centrifuged at 4000 rpm and 4˚C for 15 minutes, with the supernatants discarded. Washing of pellets was performed with 500 µl -20°C 70% ethanol and centrifuged at 4000 rpm and 4˚C for 2 minutes. All supernatants were removed carefully, followed by the addition of 600 µl dH_2O to each tube. All tubes were incubated on ice for 5 minutes before resuspending the pellet in the dH_2O. The resuspensions were transferred to new 1.5 ml Eppendorf tubes. 1 µl aliquots of 30 mg/ml RNase was added to each tube, mixed and subsequently incubated for 1 hour at 37˚C. The resuspensions were collected at the bottom of the tube via a quickly spun in the centrifuge, followed by the addition of 500 µl 1:1 phenol:chloroform. All tubes were vortexed for 30 sec and centrifuged at 5000 rpm and room temperature for 5 minutes. All upper phases were moved to new 1.5 ml Eppendorf tubes and 600 µl of chloroform was added and vortexed for 30 seconds. Centrifugation of each tube at 5000 rpm at room temperature for 2 minutes occurred afterwards. The upper aqueous phases were transferred to new Eppendorf tubes, with addition of 30 µl 3M sodium acetate (pH 5.5) and mixed. Afterwards, 900 µl -20°C 100% ethanol was added and mixed by inversion of the tube, watching for precipitation, and incubated on ice for 20 minutes. The tubes were centrifuged at max speed in the cold room (4˚C) for 15 minutes and the supernatants were discarded. To each tube, 200 µl aliquots of -20°C 70% ethanol was added, tapped to mix and centrifuged at max speed in the cold room (4˚C) for 5 minutes. All supernatants were discarded and the pellets were left to dry for 10 minutes. Prior to resuspension, 200 µl dH_2O was added to each pellet and left for 5 minutes.

2.4.5 Mini plasmid DNA isolation from E. coli

The solution components used for the mini DNA isolation procedure can be seen below (Table 2-26).

Table 2-26 – Components for extraction solutions used in the mini isolation of *E. coli* DNA constructs.

Solution	Composition
Solution I	50 mM Tris/HCl pH 8
	10 mM EDTA
Solution II	200 mM NaOH
	1% SDS
Solution III	3 M Potassium acetate pH 5.5

A single colony of BW25113 containing the desired plasmid DNA were inoculated and incubated overnight, in 10 ml LB containing the appropriate antibiotics at 37˚C and 250 rpm. 1.5 ml of the overnight cultures of each BW25113 strain were transferred into 1.5 ml Eppendorf tubes and centrifuged at maximum speed for 2 minutes. The supernatant was discarded and the previous step repeated for each strain. All pellets were resuspended in 150 µl ice cold solution I through vortexing. Upon the solution becoming homogenous, 300 µl aliquots of solution II were added to each resuspension and gently mixed via inverting the tubes. All resuspensions were incubated on ice for 4 minutes. Following the 4 minutes, 225 µl aliquots of ice-cold solution III were added to each lysate, followed by vigorous mixing via inverting the tube. All resuspensions were incubated on ice for 10 minutes, followed by centrifugation at maximum speed for 5 minutes at 4°C. The supernatants were transferred to separate 1.5 ml Eppendorf tubes, with 300 µl 1:1 phenol:chloroform added afterwards. All extracts were vortexed for 10 sec and centrifuged at maximum speed for 5 minutes at room temperature. All upper phases were transferred to new 1.5 µl Eppendorf tubes and 800 µl aliquots of -20°C 2-propanol was added to each tube, mixed thoroughly, followed by incubation on ice for 10 minutes. Afterwards, all tubes were centrifuged at maximum speed for 10 minutes, with the supernatants discarded. Washing of pellets was performed with 200 µl 70% cold Ethanol and centrifuged at maximum speed for 2 minutes at 4˚C. All supernatants were removed carefully, with the pellets being left to air dry for 5 minutes and the addition of 50 µl dH_2O.

2.4.6 Agarose gel electrophoresis

The components used to make the gel, buffer and dye for the Agarose gel electrophoresis procedure can be seen below (Table 2-27).

Table 2-27 – Components of the buffer, gel and dyes used to visualise the DNA samples in the Agarose gel electrophoresis.

Component	Composition
0.7% Agarose gel	2.1 g Agarose
	300 ml 1x TAE buffer
	15 µl of 10 mg/ml Ethidium Bromide
50x TAE	2 M Tris acetate
	50 mM EDTA pH 8
	Made up in dH_2O
1x TAE	1 ml 50x TAE made up in 50ml of dH_2O
10x DNA loading dye	50 mM Tris
	50 mM EDTA
	50% Glycerol pH 7.4
	0.05% Xylene Cyanol
	0.05% Bromophenol Blue

Procedure for making components:

0.7% Agarose gel – the components were melted into the 1x TAE buffer, followed by the 10 mg/ml Ethidium Bromide being added.

10x DNA loading dye – all the components up to the 50% Glycerol (pH 7.4) were mixed thoroughly together and autoclaved. The remaining components were added afterwards.

Procedure for Agarose gel electrophoresis:

If the DNA sample volume was less than 10 µl, dH_2O was added to make the total volume 10 µl. 1x DNA loading dye was added to the desired DNA sample.

0.7% Agarose gels were set in Bio-Rad Mini-Sub and Wide Mini-sub cell GT trays. The stained samples were run on solid gels submerged in 1x TAE buffer and viewed using a Bio-Rad UV transilluminator. All samples run were compared to a λ DNA ladder digested with *EcoRI* and *HindIII* (5 µl was loaded with each gel).

2.4.7 Preparative and confirmational restriction digests

Used to isolate specific DNA fragments and prepare plasmids and DNA fragments for ligation.

All restriction digests were performed using Roche restriction enzymes and their recommended digestion buffer. The components for each preparative or confirmation restriction digest can be seen below (Table 2-28 to 2-35).

Table 2-28 – Components and their corresponding volumes used for the *dia* and *diaUT* PCR restriction digest.

Components	Volume (µl)
dH_2O	88
Buffer H	10
XbaI	1
EcoRI	1

Table 2-29 – Components and their corresponding volumes used for the *Icy* PCR pellet restriction digest.

Components	Volume (µl)
dH_2O	89
Buffer H	10
EcoRI	1

Table 2-30 – Components and their corresponding volumes used for the pUT18c-Dia preparative restriction digest.

Components	Volume (µl)
dH_2O	80
Buffer H	12
NdeI	1.5
EcoRI	1.5
pUT18c-Dia DNA	25

Table 2-31 – Components and their corresponding volumes used for the pUC18-ProEgfpApra preparative restriction digest.

Components	Volume (µl)
dH_2O	73
Buffer H	10
HindIII	1
EcoRI	1
pUC18-ProEgfpApra DNA	15

Table 2-32 – Components and their corresponding volumes used for the pET28 preparative restriction digest.

Components	Volume (µl)
dH_2O	94
Buffer H	12
NdeI	1
EcoRI	1
pET28 DNA	12

Table 2-33 – Components and their corresponding volumes used for the pK18mobsacB preparative restriction digest.

Components	Volume (µl)
dH_2O	79
Buffer H	10
EcoRI	1
pK18mobsacB DNA	10

Table 2-34 – Components and their corresponding volumes used for the pK18mobsacB-Lcy confirmational restriction digest.

Components	Volume (µl)
dH_2O	15.5
Buffer H	2
EcoRI	0.5
pK18mobsacB-Lcy DNA	2

Table 2-35 – Components and their corresponding volumes used for the pK18mobsacB-LcyProEgfpApra confirmational restriction digest.

Components	Volume (µl)
dH_2O	14.5
Buffer H	2
XbaI	0.5
pK18mobsacB-LcyProEgfpApra DNA	3

Each restriction digest was constructed using their corresponding volumes (Table 2-28 to 2-35). The *dia*, *lcy*, pK18mobsacB-Lcy, pK18mobsacB-LcyProEgfpApra and pUC18-ProEgfpApra PCR, preparative and confirmational restriction digests were incubated at 37°C overnight. The pET28 an pK18mobsacB restriction digests was incubated at 37°C for 5 hours, where upon 1µl Alkaline Phosphatase was added. The digest was incubated at 37°C for another hour.

All digests were analysed by Agarose gel electrophoresis separately. The *dia* restriction digests were run at 30 V and the pET28 restriction digest was run at 10 V overnight.

2.4.8 Isolating DNA fragments via agarose

Following the separation of the DNA fragments generated by the pUT18c-Dia preparative restriction digest, the desired DNA fragments were viewed and identified using long-wavelength UV (310 nm) through a Bio-Rad transilluminator. The desired fragments were excised with a sterile razor blade (with 80% ethanol) and purified from the Agarose gel via the Qiagen QIAquick Gel Extraction Kit (QIAGEN, 2018). Fragments were stored in sterile dH_2O at -20°C.

2.4.9 Ligation of DNA fragments

All ligations were made using 5x T4 DNA ligase and buffer (Invitrogen). The components for each ligation can be seen below (Table 2-36 to 2-39).

Table 2-36 – Components and their corresponding volumes used for the *dia* BTH construct ligation.

Components	Volume (µl)
dH_2O	8.5
5x T4 DNA ligase buffer	3
T4 DNA ligase	0.5
Desired fragment (*dia* or *diaUT*)	2
Desired BTH plasmid (pUT18c, pKT25, pUT18 or pKNT25)	1

Table 2-37 – Components and their corresponding volumes used for the pET28 *dia* ligation.

Components	Volume (µl)
dH_2O	7.5
5x T4 DNA ligase buffer	3
T4 DNA ligase	0.5
dia	3
pET28	1

Table 2-38 – Components and their corresponding volumes used for the pUC18-PromCherryApra ligation.

Components	Volume (µl)
dH_2O	6.5
5x T4 DNA ligase buffer	3
T4 DNA ligase	0.5
PromCherryApra cassette	4
pUC18	1

Table 2-39 – Components and their corresponding volumes used for the pK18mobsacB-Lcy ligation.

Components	Volume (µl)
dH_2O	7
5x T4 DNA ligase buffer	3
T4 DNA ligase	0.5
lcy restriction digest	3
pK18mobsacB restriction digest	1.5

The *dia* BTH constructs, pET28-Dia, pUC18-ProEgfpApra and pK18MobSacB-Lcy ligations were set up using different components and volumes (Table 2-36 to 2-39). Each prepared vector and fragment were mixed within the sterile dH_2O, incubated at 65°C for 2 minutes and then cooled on ice. Subsequently, the 5x T4 DNA ligase buffer and T4 DNA ligase was thoroughly mixed into the mixture. All ligations were incubated on ice at 4°C overnight.

2.4.10 Transformation of competent E. coli cells via electroporation

A single colony of the appropriate *E. coli* strain was grown overnight in 10 ml LB containing appropriate antibiotics/recombination-inducing compounds in a shaking incubator at 250 rpm at 37°C (BL21 and Rosetta) or 30°C (for BW25113). A 100 µl aliquot of each overnight culture was subcultured in 10 ml LB with the appropriate antibiotics/recombination-inducing compounds and incubated at the same conditions for 5 hours. The cells were collected by centrifuging at 4000 rpm for 10 minutes, with the supernatant being removed afterwards. The pellet was washed by resuspension in 25 ml 10% glycerol and centrifuged using the same conditions. This last 2 steps were repeated twice more. The supernatant was discarded, and the cells were resuspended in 300 µl 10% glycerol for the final volume. In individual ice-cold 0.2 cm cuvettes, 45 µl of competent cells and 1 µl of the desired DNA were combined for electroporation. Electroporation was carried using the Bio-Rad Gene Pulser 2 (set to 200 Ω, 25 µF and 2.5 kV). Afterwards, 500 µl ice-cold LB was added and mixed with the cells, followed by moving the resuspended cells to a 1.5 ml Eppendorf tube. Each tube was incubated in a shaking incubator at 30˚C for 1 hour and 30 minutes. After incubation, 100 µl of the transformed cells was streaked onto LB plates containing appropriate antibiotics and supplements, then incubated at the 37°C or 30°C (for the BW25113 pK18mobsacB-Lcy).

2.4.11 Transformation of competent E. coli cells via chemical competence

An overnight culture of *E. coli* DH5α cells were grown in 10 ml LB, with a 500 µl aliquot subcultured in 50 ml LB. The subculture was incubated at 37˚C and 250 rpm for 2 hours and 20 minutes. The 50 ml subculture was transferred to a 50 ml falcon tube and span at 4000 rpm and 4˚C for 10 minutes in a centrifuge. The supernatant was discarded following centrifugation and the pellet was resuspended with 20 ml ice-cold 10 mM NaCl. The resuspended cells were centrifuged at 4000 rpm at 4˚C for 7 minutes. The supernatant was discarded immediately afterwards. The pellet was resuspended in 20 ml 30 mM CaCl2: 10 mM RbCl, followed by incubation for 20 minutes on ice. The resuspended cells were centrifuged using the previous conditions and the supernatant discarded afterwards. The pellet was resuspended in 500 µl 30 mM CaCl2: 10 mM RbCl. For transformation, 50 µl competent cells were transferred to Eppendorf tubes containing 4 µl of the ligated DNA and incubated on ice for 30 minutes. After the incubation on ice, all Eppendorfs were heat shocked at 42˚C for 1 minute and immediately transferred back to ice. 500 µl LB were added to each transformation with all tubes being incubated on a lab shaker at 37˚C for 1 hour. 100 µl aliquots of each transformed cells were plated onto LB plates containing appropriate antibiotics and incubated at 37˚C overnight.

2.4.12 BACTH assay streaking

Transformed bacteria containing the successful combinations of the BACTH constructs were taken from the original chemical competence transformation plates and streaked onto LB plates supplemented with the appropriate antibiotics and the compounds for overexpressing and staining functioning BACTH interactions. The new streaks were incubated at 30°C for 48 hours, with images being taken after 24 and 48 hours.

2.5 Protein molecular and structural computational analysis:

2.5.1 Identification of Dia homolog occurrence, conservation, and prediction of possible domains within Dia

The identification of the extent to which Dia homologs occur within bacterial species was performed via various BLAST searches against the *S. coelicolor* sequence, with certain searches excluding specific groups of bacteria. The excluded groups include: the *Streptomyces* genus and the Actinobacteria phyla. From each comparative BLAST search, between 10-13 possible homologs (depends on the search) were selected and aligned against the primary sequence of Dia (UniProt ID: Q9ZBR0) in the NCBI Multiple Alignment Viewer 1.22.2 to identify regions of conservation across the species.

To identify possible shared domains between Dia homologs, 14 of the identified Dia homologs were aligned against the primary sequence of Dia (Uniprot ID: Q9ZBR0) in the NCBI Multiple Sequence Alignment Viewer 1.22.2 and the NCBI Conserved Domains search.

To identify possible similarities between the domains of Dia (Uniprot ID: Q9ZBR0) and DivIVA (Uniprot ID: Q9S2X4) or AtpF (Uniprot ID: Q9K4D7) of *S. coelicolor*, the primary sequence of Dia was aligned against the primary sequence of DivIVA and AtpF separately using Clustal Omega.

2.5.2 Analysis of GPR homolog occurrence

The identification of the extent to which GPR (Uniprot ID: A9CJ72) homologs occur within bacterial species was performed via various BLAST searches against the *A. tumefaciens* sequence, with certain searches excluding and including specific groups of bacteria. The excluded and included groups included: *A. tumefaciens* and the Hyphomicrobiales Order.

2.5.3 Identification and prediction of possible domains within Lcy

The identification of Lcy and its homologs within *L. aggregata* was seen in the BLAST searches against the *A. tumefaciens* GPR (Uniprot ID: A9CJ72) sequence which excluded *A. tumefaciens* but included the Hyphomicrobiales Order.

To identify possible shared domains and conserved regions between GPR (Uniprot ID: A9CJ72) and Lcy (a GPR homolog), the identified Lcy homologs were

aligned against the primary sequence of GPR in the NCBI Multiple Sequence Alignment Viewer 1.22.2 and the NCBI Conserved Domains search.

To demonstrate the similarity between the *L. aggregata* Lcy homologs of the published IAM genome and the species strain LZB033, the primary sequences of the LZB033 (Unpublished ordered locus name: FIG00742013) and IAM (UniProt ID: A0A0M6Y2Z9) homologs were aligned against each other using Clustal Omega.

2.5.4 Prediction of the physical and chemical properties of selected proteins

To predict the physical and chemical properties of the selected proteins, the primary sequences of Dia and Lcy were separately passed through the Expasy ProtParam program to predict a collection of different properties including: molecular weight, theoretical isoelectric point (pI) and the instability index (II).

2.5.5 Prediction of secondary and tertiary structures within Dia

Within PCOILS, the secondary structure was predicted with the pSIPRED system. Predictions were made through analysis of the Dia primary sequence run through with weighted and unweighted hydrophobic residues. The results from both predictions were compared to confirm any consistently predicted secondary structures.

PCOILS, alongside DeepCoil and MARCOIL, were used to predict the possibility of coiled-coil regions within the Dia secondary structure based on the primary structure. For each program, a collection of different parameters was run either predicting based on: the Dia primary structure alone and when compared to similar proteins in the nr90 database. The results from the deep learning algorithms were compared with each other to identify regular regions of possible coiled-coils.

2.5.6 Theoretical structural modelling

To analyse the theoretical tertiary structure of Dia, the AlphaFold structural predictions on UniProt for Dia, DivIVA and AtpF of *S. coelicolor* strain M145 were taken and orientated to similar planes for comparison of the structures throughout the protein and at identified aligned regions within their respective sequences.

2.6 Protein purification:

2.6.1 Cell lysis:

Overexpression of the desired protein was produced using the optimal conditions as set out in the result section. The overexpressing cells were collected with separate 50 ml falcon tubes, using an Eppendorf 5430 centrifuge (with an Eppendorf Fixed Angle QuickLock F-35-6-30 rotor) at 6200 rpm for 5 minutes. The supernatant was discarded, and the remaining cells were collected in the same tubes under the same conditions. 3 of the pellets were re-suspended in 15 ml of Binding buffer (Table 2-22) and transferred to the falcon tube with the remaining pellet, whereupon the remaining pellet was re-suspended. Cell lysis was achieved through a JENCONS Ultrasonic Processor via sonicating the cells for 5x 30 seconds on ice. After lysis, the lysate was deposited into a 50 ml falcon tube and centrifuged using an Eppendorf 5430 centrifuge (with an Eppendorf Fixed Angle QuickLock F-35-6-30 rotor) at 6200 rpm for 15 minutes, with the supernatant transferred to another 50 ml tube. The supernatant was sterile filtered using 0.2 micron pipette filters.

2.6.2 Protein purification under native conditions:

All protein purification was carried out using a Severn Biotech disposable purification column containing GE Healthcare Ni Sepharose 6 Fast Flow reagent. The buffer components for the native protein purification can be seen below (Table 2-40).

Table 2-40 – Composition of the native buffers used in protein purification in this study.

Solution	Composition
Binding buffer	50 mM Tris
	20 mM $MgCl_2$
	300 mM NaCl
	10 mM imidazole pH 8.0
Elution buffer	50 mM Tris
	20 mM $MgCl_2$
	300 mM NaCl
	300 mM imidazole pH 8.0
Cleaning buffer	50 mM Tris
	20 mM $MgCl_2$
	300 mM NaCl
	500 mM imidazole pH 8.0

The purification column was prepared for the protein sample by passing 20 ml sterile dH_2O through, followed by a 10 ml Binding buffer. To bind the protein to the column, a 200 µl supernatant aliquot (called the preload) was loaded and run through and collected from the column. The remaining supernatant was loaded into the column (10 ml at a time) and run through and collected as the Flowthrough. With the sample bound to the nickel matrix, the column was washed to remove any unwanted proteins. The column was washed with 10 ml Binding buffer twice separately which was collected each time (called wash 1 and 2). Following the washes, a series of 10x 1 ml elutions of Elution buffer were run and collected from the wash column consecutively (called elution 1 – 10). Finally, remove any remaining protein from the column, 2 separate 1 ml elutions of Cleaning buffer were run and collected in succession through the column (called elution 11 and 12 respectively). Finally, 10 ml of Cleaning buffer was run through and collected from the column (called elution 13).

2.6.3 Dialysis:

Dialysis was carried out using Spectra/Por® molecularporous membrane tubing with a molecular weight cut-off of 3500 Da. The buffer components for dialysis can be seen below (Table 2-41).

Table 2-41 – Components used for the construction of the dialysis buffer in this study.

Component	Composition
Dialysis buffer (pH 8)	50 mM Tris
	150 mM NaCl
	10 mM $MgCl_2$

The dialysis tubing was cut open and 5 cm^2 squares were cut and equilibrated in the dialysis buffer for 10 minutes. A 1 ml aliquot of the purified Dia mixture was placed into a 2 ml microcentrifuge tube with the lid removed. The O-ring from the lid was pushed past the underside of the lip to secure the membrane over the top of the microcentrifuge tube. The membrane was pulled tight across the top of the tube to generate a flat surface for dialysis. The tube was secured in a float and placed upside down in the target buffer. The buffer was gently stirred at 4°C for at least 2-hour sessions. The buffer was replaced between each session with a total of 3 separate dialysis session carried out on the same aliquot.

2.6.4 SDS-PAGE:

All SDS-PAGE gels were made and run using an ATTO AE-6450 Dual Mini Slab Kit. The components for the gel, buffer and staining/destaining dyes for the SDS-PAGE can be seen below (Table 2-42 and 2-43).

Table 2-42 – Components use to construct the gels and buffers used for SDS-PAGE in this study.

Component	Composition
12% SDS-PAGE resolving gel (in dH_2O)	5.02 ml sterile dH_2O
	6 ml 30% acrylamide
	3.75 ml 1.5 M Tris pH 8.8
	150 µl 10%SDS
	60 µl 25% APS
	9 µl TEMED
12% SDS-PAGE stacking gel (in dH_2O)	5.6 ml sterile dH_2O
	1.6 ml 30% acrylamide
	670 µl 1.5M Tris pH 8.8
	80 µl 10%SDS
	40 µl 25% APS
	8 µl TEMED
10x SDS Buffer solution (1L)	30 g Tris
	144 g Glycine
	10 g SDS
	1 L dH_2O

Procedure for making the components:

Resolving gel – the components were added to a 30 ml glass universal tube, with APS added last, mixed thoroughly and added carefully to a sealed SDS-PAGE gel glass container in 1 ml aliquots. 500 µl dH_20 is placed on top to level the gel and left to set for 30 minutes, whereupon the dH_20 was removed.

Stacking gel – the components were added to a 30 ml glass universal tube, with APS added last, mixed thoroughly and added carefully on top of the resolving gel in 1 ml aliquots. A comb is placed into the stacking gel (with no bubbles trapped underneath), left to set for 30 minutes, with the comb being removed afterwards.

Table 2-43 – Components used to construct the protein staining dyes and destaining solutions used in this study.

Solution	Composition
Colloidal Coomassie blue R250 (in dH_2O)	40% ethanol
	10% acetic acid
	0.1% R250
4x Protein loading dye (in dH_2O)	200 mM Tris pH 6.8
	20% β-mercaptoethanol
	8% SDS
	40% Glycerol
	0.4% bromophenol blue
	Made up with dH_2O
De-staining solution (in dH_2O)	40% ethanol
	10% acetic acid
	Made up with dH_2O

Procedure for SDS-PAGE:

3.5 µl of 4x Protein loading dye was added to 10.5 µl of the protein samples and heated at 95°C for 10 minutes to denature before being cooled on ice.

The SDS-PAGE gel was fixed into the vertical tank containing an inch of 10x SDS buffer (with no bubbles trapped underneath the gel). The tank is topped up with 1x SDS buffer afterwards. The protein samples were loaded into the stacking gel and run at 200 V for 45 minutes. After running the gels were stained with Colloidal Coomassie blue R250 for at least 1 hour with gentle agitation. The gels were de-stained with the De-staining solution for 1 hour before being washed and submerged in dH_2O. The gels were then viewed with white light illumination using a Bio-Rad trans-illuminator.

2.7 Mass photometry:

2.7.1 REFEYN Mass Photometry

A sample of dialysed His-Dia was sent off for analysis using the REFEYN Mass Photometry.

2.8 Cross species transformation:

2.8.1 Triparental mating

Clumps from the selected DH5α pK18mobsacB-LcyProEgfpApra colony, HB101 pRK2013 and LZB033 were mixed placed on the centre of a YTSS cellophane agar plate. The bacterial clumps were mixed thoroughly and spread evenly across the cellophane of the plate. The streaked plate was placed into a 30°C incubator and grown for 3 days.

2.9 Fluorescent microscopy:

2.9.1 Preparation of fluorescent microscopy samples

For microscopic observations of *L. aggregata* LcyProEgfpApra, the strain was grown in 10 ml liquid YTSS containing apramycin for 2 days at 30°C and 200 rpm. 1µl aliquots of the strain were grown on coverslips inserted into YTSS media at 45° angles and incubated at 30°C. At chosen timepoints, the coverslips were removed and fixed using 400µl methanol, stained using 30µl propidium iodide (PI) (1 mg/PI) and washed twice with 200µl 1xPBS. The coverslips were then mounted using 8µl 20% glycerol and viewed under a fluorescent microscope. All observations were taken with an oil immersion at a 100x objective on a Carl Zeiss Axio scope with Zen software.

Chapter 3:

Analysis of the protein structure of Dia

3.1 Introduction:

As previously mentioned in Chapter 1, our knowledge of the *S. coelicolor* genome is incomplete with only about 54% of the genes characterised into Clusters of Orthologous Groups. Within that, about only 11.7% of genes have been determined to be involved in Cellular and signalling processes and a total of 0.5% of them being involved specifically in Cell division and the cytoskeleton. Additionally, there is still about 46% of the genome uncharacterised, with 30.8% of the remaining genes with their function unidentified (Heinsch *et al*., 2019). Collectively, this suggests there are many genes still left to classified. Within these undetermined genes, one such gene: *SCO5569*, has been previously researched. *SCO5569* had been highlighted to possibly have similar domains to known *S. coelicolor* polar growth protein (a theoretical coiled-coil domain) and appears to be involved in determining the hyphal diameter during active polar growth; through disruptions of the gene resulting in drastically varying diameters throughout the hyphae. An additional documented effect of Dia is its involvement in the distance between sporulation septa of aerial hyphae (Ireland, Unpublished). With these reported functions, *SCO5569* has been dubbed '*dia*'. Alongside these cellular effects, the localisation pattern of Dia-EGFP has been shown to possess similarities to that of DivIVA-EGFP (Hutchinson, Unpublished; Ireland, Unpublished) (Figure 3-1). With these documented capacities of Dia, Dia could be directly involved in the polar growth mechanics of *S. coelicolor*. To investigate this, we plan to analyse the structural properties of Dia to explain its role within hyphal diameter and the parallels seen within its localisation pattern to DivIVA.

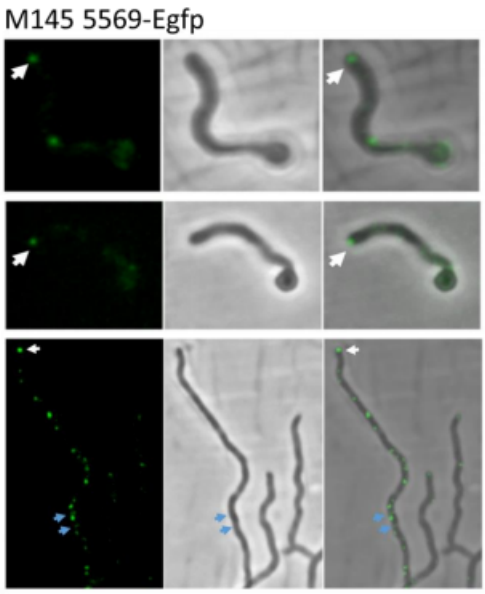


Figure 3-1. Fluorescent microscopy images of the Dia single crossover fusion (DiaEgfp) within the germination and vegetative growth of *S. coelicolor* strain M145. The conjoined DiaEgfp fusion localises at the apical tip of the germ tube and vegetative hyphae during active polar growth. Additional signals of DiaEgfp can be seen throughout the germ tube and the vegetative hyphae. Taken from (Ireland, Unpublished).

As explained in the introduction, the TIPOC system within *Streptomyces* is pivotal for the initiation and maintenance of polar growth. A vital component in this system is the DivIVA protein, which is required for the initiation of polar growth through the recruitment of cell wall synthesis machinery. The localisation of DivIVA determines where and when polar growth is initiated. DivIVA has been found to concentrate along curved hyphal walls, especially at hyphal tips, alongside localising at future branch positions along the hyphal wall before any growth was seen. In addition, during polar growth of hyphae, DivIVA appeared to develop into a large cluster at the apical tip (Hempel *et al*., 2008). Given the case that DivIVA was seen to concentrate at the site of polar growth and potentially form a higher structure, extracted complexes of DivIVA were confirmed to arrange into a homo-oligomeric complex that align with the hyphal apex but possess a weak association with the membrane (Wang *et al*., 2009). Although DivIVA complexes initiate and drives polar growth, the localisation of the complexes is dynamic due to their weak association. With this in mind, partner proteins of DivIVA could help to provide stability to the DivIVA complexes. The FilP component of the TIPOC helps provide this function. FilP was found to localise near the hyphal tip during polar growth and was redistributed upon arrest. Further analysis revealed FilP was important in shaping the structure of the DivIVA complex and which in turn helps maintain the stability of the complex (Fröjd & Flärdh, 2019). Alongside the dynamic relationship of DivIVA and FilP, the TIPOC components Scy influences the dynamic of the TIPOC through

organising the complexes of DivIVA and FilP. Scy is capable of binding to DivIVA and FilP complexes simultaneously, helping to localise the TIPOC. In addition, Scy helps to limit the number of DivIVA complexes forming near the hyphal tip and in turn limit the number of TIPOCs forming and maintain the integrity of formed TIPOCs (Holmes *et al*., 2013). Even though Scy helps to organise both DivIVA and FilP within the TIPOC, it is not vital for the localisation and assembly of both components. DivIVA has been found to stimulate the oligomerisation of FilP into its hexagon, self-assembled network. (Fuchino *et al*., 2013).

3.2 Results:

3.2.1 Analysis of Dia and its homologs

Initially, the primary sequence of Dia (UniProt ID: Q9ZBR0) was passed through the Expasy ProtParam program to determine some of the physical and chemical properties of the protein. Dia is constructed of 379 amino acids, has a predicted molecular weight of 40.97684 kDa, the theoretical isoelectric point (pI) is 4.36 and has an instability index (II) of 49.43. Since the II is greater than 40, the protein is deemed to be unstable *in vitro*. Additionally, a cysteine residue was identified as the 26th amino acid in the sequence. With knowledge of the length of Dia and the possibility it may be involved with components of the TIPOC, a universal mechanism seen across *Streptomyces* species, we were unsure if Dia homologs was found across the genus. Consequently, we reviewed if Dia had homologs across various species or was specific to *S. coelicolor* through a BLAST search of the primary sequence of Dia (UniProt ID: Q9ZBR0). The BLAST search revealed that Dia appeared to possessed homologs in a variety of different species of *Streptomyces* with a high percentage of similarity: at least 80% (Table 3-1 and 3-2). All the matches were unique homologs, due to the high E-values (with some E-values being so significantly small that the program rounded them to 0). The top 1000 matches for Dia being identified in various *Streptomyces* species. Following the top 1000 matches, multiple hits for potential Dia homologs within other genus were identified including: *Kitasatospora*, *Streptacidiphilus*, *Peterkaempfera*, *Phaecidiphilus* and *Saccharothrix* (all the hits were unique due to the high E-values, with sequence similarity between 44-52%). *Kitasatospora*, *Streptacidiphilus*, *Peterkaempfera* and *Saccharothrix* hits were all from a bacterial species within the Actinomycetota (Actinobacteria) phylum.

Table 3-1. The amino acid sequence of Dia (UniProt ID: Q9ZBR0) from *S. coelicolor* was BLAST against other bacterial genomes. The top hit matched with an identical copy of Dia and the following hits identified potential highly conserved homologs of Dia found across various *Streptomyces* species. The potential homologs possessed between 91-100% similarity to Dia.

Description	Scientific name	E value	Percentage Identity	Accession
hypothetical protein [Streptomyces]	*Streptomyces*	0.0	100.00%	WP_011030311.1
hypothetical protein [Streptomyces]	*Streptomyces*	0.0	99.47%	WP_003973425.1
cell division initiation protein [unclassified Streptomyces]	*unclassified Streptomyces*	0.0	99.21%	WP_093457067.1
cell division initiation protein [Streptomyces sp. LRa12]	*Streptomyces sp. LRa12*	0.0	99.21%	WP_136209400.1
cell division initiation protein [Streptomyces sp. E5N91]	*Streptomyces sp. E5N91*	0.0	98.42%	WP_121714210.1
cell division initiation protein [Streptomyces rubogriseus]	*Streptomyces rubogriseus*	0.0	97.89%	WP_109036097.1
cell division initiation protein [Streptomyces sp. SID5926]	*Streptomyces sp. SID5926*	0.0	97.80%	MYS74452.1
cell division initiation protein [Streptomyces]	*Streptomyces*	0.0	100.00%	WP_202493697.1
cell division initiation protein [Streptomyces rubogriseus]	*Streptomyces rubogriseus*	0.0	97.63%	NEC38046.1
cell division initiation protein [Streptomyces sp. RK76]	*Streptomyces sp. RK76*	0.0	99.46%	WP_247704091.1
cell division initiation protein [Streptomyces sp. CB09001]	*Streptomyces sp. CB09001*	0.0	95.51%	WP_115744393.1
cell division initiation protein [Streptomyces rubogriseus]	*Streptomyces rubogriseus*	0.0	97.55%	WP_203697443.1

cell division initiation protein [Streptomyces]	*Streptomyces*	0.0	91.62%	WP_153177767.1
hypothetical protein [Streptomyces sp. KO7888]	*Streptomyces sp. KO7888*	0.0	91.62%	NHI06919.1
cell division initiation protein [Streptomyces sp. NRRLWC-3753]	*Streptomyces sp. NRRLWC-3753*	0.0	92.67%	KPC80448.1

Table 3-2. The amino acid sequence of Dia (UniProt ID: Q9ZBR0) from *S. coelicolor* was BLAST against other bacterial genomes. After roughly 1000 hits with Dia, hits were identified as potential partially conserved homologs of Dia and ATP synthase subunit B alignments found across various bacterial genomes including: *Kitasatospora*, *Streptacidiphilus*, *Peterkaempfera*, *Phaecidiphilus* and *Saccharothrix*. The identified hits possessed between 44-52% similarity to Dia.

Description	Scientific name	E value	Percentage Identity	Accession
ATP synthase F0 subunit B [Kitasatospora sp. CB02891]	*Kitasatospora sp. CB02891*	4e-82	49.49%	WP_100590045.1
cell division initiation protein [Streptomyces griseocameus]	*Streptomyces griseocameus*	5e-82	52.71%	WP_207248167.1
ATP synthase F0 subunit B [Streptomyces sp. NBRC 109706]	*Streptomyces sp. NBRC 109706*	6e-82	46.17%	WP_062205868.1
hypothetical protein [Kitasatospora sp. NRRL B-11411]	*Kitasatospora sp. NRRL B-11411*	9e-82	48.40%	WP_030458838.1
cell division initiation protein [Strepacidiphilis sp. ASG 303]	*Strepacidiphilis sp. ASG 303*	1e-81	44.50%	MCD0484933.1
ATP synthase F0 subunit B [Streptomyces kaniharaensis]	*Streptomyces kaniharaensis*	1e-81	50.00%	WP_195911236.1

ATP synthase F0 subunit B [Peterkaempfera griseoplana]	*Peterkaempfera griseoplana*	1e-81	47.69%	WP_055589177.1
hypothetical protein [Streptomyces sp. 114.5]	*Streptomyces sp. 114.5*	1e-81	45.28%	WP_121174185.1
cell division initiation protein [Streptomyces bohaiensis]	*Streptomyces bohaiensis*	2e-81	46.37%	NJQ15295.1
hypothetical protein [Phaeacidiphilis oryzae]	*Phaeacidiphilis oryzae*	2e-81	47.85%	WP_037572496.1
hypothetical protein [Kitasatospora sp. CB01950]	*Kitasatospora sp. CB01950*	2e-81	49.00%	WP_073813948.1
cell division initiation protein [Kitasatospora humi]	*Kitasatospora humi*	2e-81	49.48%	WP_249800785.1
ATP synthase F0 subunit B [Streptomyces sp. 1331.2]	*Streptomyces sp. 1331.2*	6e-81	46.89%	WP_097236115.1
ATP synthase F0 subunit B [Kitasatospora cheerisanensis]	*Kitasatospora cheerisanensis*	6e-81	48.01%	WP_035866494.1
ATP synthase F0 subunit B [Saccharothix sp. ST-888]	*Saccharothix sp. ST-888*	8e-81	44.47%	WP_045301172.1
ATP synthase F0 subunit B [Kitasatospora sp. MBT63]	*Kitasatospora sp. MBT63*	1e-80	45.82%	WP_033818758.1
ATP synthase F0 subunit B [Kitasatospora phosalcinea]	*Kitasatospora phosalcinea*	1e-80	47.56%	WP_033215105.1

As potential homologs of Dia were identified in separate various species outside of the *Streptomyces* genus, we decided to investigate to what extent Dia homologs could be found within other less related species through another BLAST search against the primary sequence of Dia (UniProt ID: Q9ZBR0) excluding the *Streptomyces* genus. The hits returned potential homologs of Dia within multiple different genus of Actinobacteria including *Actinospica*, *Kitasatospora*, *Streptacidiphilus* and *Peterkaempfera* (Table 3-3 and 3-4). Most of the hits identified were of unique homologs and alignments due to the high E-values. The sequence similarity of the hits varied though, anywhere between 43-72%. Most of the newly aligned species are species of Actinobacteria and from a variety of different Families including Streptomycetaceae and Actinospicaceae. A few hits were returned for *Streptomyces* species though they were excluded. However, was a highly significant and unexpected hit with a species that's not an Actinobacteria (*Mesorhizobium*) and is from entirely different phyla (Pseudomonadota). With the possibility of a bacteria outside of the Actinobacteria phylum that utilise homologs of Dia, we decided to investigate the extent and range that these Dia homologs could be found outside the Actinobacteria phylum.

Table 3-3. The amino acid sequence of Dia (UniProt ID: Q9ZBR0) from *S. coelicolor* was BLAST against other bacterial genomes, excluding *Streptomyces* genus. Multiple different hits identified potential Dia homologs alongside ATP synthase subunit B alignments for bacterial species including: *Actinospica*, *Mesorhizobium* and *Kitasatospora*. A few hits for *Streptomyces* species appeared although the genus was excluded. The identified homologs and alignments possessed between 62-72% similarity to Dia.

Description	Scientific name	E value	Percentage Identity	Accession
vacuolar-type H+-ATPase subunit H [Streptomyces murinus]	*Streptomyces murinus*	1e-172	72.98%	MBA90556252.1
vacuolar-type H+-ATPase subunit H [Streptomyces murinus]	*Streptomyces murinus*	2e-169	72.47%	MBA90483448.1
cell division initiation protein [Actinospica acidiphila]	*Actinospica acidiphila*	3e-165	72.61%	NEC48338.1

ATP synthase F0 subunit B [Actinospica acidiphila]	*Actinospica acidiphila*	1e-164	72.35%	MBM4828382.1
cell division initiation protein [Actinospica acidiphila]	*Actinospica acidiphila*	4e-164	72.21%	NEA81882.1
cell division initiation protein [Actinospica acidiphila]	*Actinospica acidiphila*	1e-162	72.21%	NEA95533.1
hypothetical protein [Actinobacteria bacterium OV320]	*Actinobacteria bacterium OV320*	1e-160	69.44%	KPI29515.1
hypothetical protein [Actinobacteria bacterium OK006]	*Actinobacteria bacterium OK006*	2e-158	69.29%	KPI03357.1
hypothetical protein [Actinobacteria bacterium OK074]	*Actinobacteria bacterium OK074*	8e-148	67.35%	KPI12408.1
cell division initiation protein [Streptomyces cinereoruber]	*Streptomyces cinereoruber*	1e-140	64.38%	MBY88114370.1
cell division initiation protein [Mesorhizobium sp. B2-3-3]	*Mesorhizobium sp. B2-3-3*	3e-135	62.50%	TPN29827.1
cell division initiation protein [Streptomyces microflavus]	*Streptomyces microflavus*	5e-134	63.25%	NEB69738.1
cell division initiation protein [Kitasatospora albolonga]	*Kitasatospora albolonga*	8e-134	64.96%	WP084749869_.1
hypothetical protein OV450_0611 [Actinobacteria bacterium OV450]	*Actinobacteria bacterium OV450*	5e-129	62.83%	KPI01605.1
vacuolar-type H+-ATPase subunit H [Streptomyces virbiniae]	*Streptomyces virbiniae*	1e-120	62.30%	MBP2346759.1

Table 3-4. The amino acid sequence of Dia (UniProt ID: Q9ZBR0) from *S. coelicolor* was BLAST against other bacterial genomes, excluding *Streptomyces* genus. After roughly 100 hits with Dia, hits were identified as potential partially conserved homologs of Dia and ATP synthase subunit B alignments found across bacterial species including: *Kitasatospora*, *Streptacidiphilus* and *Peterkaempfera*. The identified hits possessed between 43-63% similarity to Dia.

Description	Scientific name	E value	Percentage Identity	Accession
cell division initiation protein [Kitasatospora gansuensis]	*Kitasatospora gansuensis*	6e-76	45.69%	WP_221504102.1
hypothetical protein [Streptacidiphilus rugosus]	*Streptacidiphilus rugosus*	7e-76	46.33%	WP_037607620.1
ATP synthase F0 subunit B [Kitasatospora sp. A2-31]	*Kitasatospora sp. A2-31*	7e-76	46.41%	WP_255753310.1
ATP synthase F0 subunit B [Kitasatospora sp. SID7827]	*Kitasatospora sp. SID7827*	1e-75	48.95%	WP_202524123.1
hypothetical protein [Streptacidiphilus jiangxiensis]	*Streptacidiphilus jiangxiensis*	2e-75	47.26%	WP_042454107.1
cell division initiation protein [Streptacidiphilus pinicola]	*Streptacidiphilus pinicola*	1e-74	46.92%	WP_111504744.1
hypothetical protein SAMN05414137 [Streptacidiphilus jiangxiensis]	*Streptacidiphilus jiangxiensis*	1e-74	46.65%	SEM36912.1
hypothetical protein [Streptacidiphilus albus]	*Streptacidiphilus albus*	1e-73	49.87%	WP_034090117.1
cell division initiation protein [Streptacidiphilus fuscans]	*Streptacidiphilus fuscans*	1e-73	45.63%	MBF9067326.1

cell division initiation protein [Streptacidiphilus sp. 4-A2]	*Streptacidiphilus sp. 4-A2*	2e-72	45.77%	MBC3844111.1
cell division initiation protein [Streptacidiphilus ASG 303]	*Streptacidiphilus ASG 303*	9e-72	43.17%	WP_255744691.1
cell division initiation protein [Streptacidiphilus P02-A3a]	*Streptacidiphilus P02-A3a*	4e-70	47.15%	WP_182454912.1
ATP synthase F0 subunit B [Peterkaempfera bronchialis]	*Peterkaempfera bronchialis*	8e-69	61.33%	WP_111492988.1
hypothetical protein [Streptacidiphilus jeojiense]	*Streptacidiphilus jeojiense*	9e-67	47.53%	WP_030263937.1
cell division initiation protein [Kitasatospora sp. SUK 42]	*Kitasatospora sp. SUK 42*	1e-66	63.18%	WP_254802937.1
hypothetical protein [Streptacidiphilus carbonis]	*Streptacidiphilus carbonis*	1e-65	46.12%	WP_042392355.1

To investigate this, additional BLAST searches were performed against the primary sequence of Dia (UniProt ID: Q9ZBR0) but excluding the entire Actinobacteira phylum. The hits returned potential homologs of Dia within a variety of different species including *Mesorhizobium*, *Pseudorhodobacter*, *Rubrivirga*, Acidobacteria, *Rhodococcus*, Clostridia, Elusimicrobia, Clostridiales, Firmicutes, Deltaproteobacteria, *Gordonia*, *Chloroflexi* and *Limnochordia* (Table 3-5 and 3-6). The aligned species we found a across a range of different phyla including Pseudomonadota (*Mesorhizobium* and *Pseudorhodobacter*), Rhodothermaeota (*Rubrivirga*), Acidobacteriota (Acidobacteria), Bacillota – known as Firmicutes (Clostridia, Clostridiales and *Limnochordia*), Elusimicrobiota (Elusimicrobia), Myxococcota – known as Deltaproteobacteria and Chloroflexota – known as Chloroflexi (the returned hits were of unique homologs and alignments due to the

high E-values, the sequence similarity varied between 30-73% though). A collection of hits was returned for species of Actinobacteria (*Streptomyces*, *Rhodococcus* and *Gordonia*) though they were excluded. Collectively, the BLAST searches revealed potential homologs of Dia across a plethora of different bacteria, with some species being distantly related to *S. coelicolor*.

Table 3-5. The amino acid sequence of Dia (UniProt ID: Q9ZBR0) from *S. coelicolor* was BLAST against other bacterial genomes, excluding the Actinobacteria phylum. Multiple different hits identified potential Dia homologs alongside ATP synthase subunit B alignments for bacterial species including: *Mesorhizobium*, *Pseudorhodobacter*, *Rubrivirga*, *Acidobacteria* and *SAR202 cluster*. A few hits for Actinomycetota species appeared although the genus was excluded. All the identified homologs possessed between 40-73% similarity to Dia.

Description	Scientific name	E value	Percentage Identity	Accession
vacuolar-type H+-ATPase subunit H [Streptomyces murinus]	*Streptomyces murinus*	1e-172	72.98%	MBA9056252.1
vacuolar-type H+-ATPase subunit H [Streptomyces murinus]	*Streptomyces murinus*	2e-169	72.47%	MBA9048348.1
cell division initiation protein [Streptomyces cinereoruber]	*Streptomyces cinereoruber*	1e-140	64.38%	MBY8814370.1
cell division initiation protein [Mesorhizobium sp. B2-3-3]	*Mesorhizobium sp. B2-3-3*	3e-135	62.50%	TPN29827.1
cell division initiation protein [Streptomyces microflavus]	*Streptomyces microflavus*	5e-134	63.25%	NEB69738.1
vacuolar-type H+-ATPase subunit H [Streptomyces virginiae]	*Streptomyces virginiae*	9e-123	62.30%	MBP2346759.1

cell division initiation protein [Streptomyces buecherae]	*Streptomyces buecherae*	1e-100	72.43%	MBC3992238.1
cell division initiation protein [Streptomyces buecherae]	*Streptomyces buecherae*	2e-100	72.43%	MBC3985258.1
cell division initiation protein [Streptomyces alkaliterrae]	*Streptomyces alkaliterrae*	5e-92	63.22%	MBB1254118.1
cell division initiation protein [Streptomyces alkaliterrae]	*Streptomyces alkaliterrae*	7e-92	63.22%	MBB1259457.1
cell division initiation protein [Streptomyces alkaliphilus]	*Streptomyces alkaliphilus*	3e-79	45.09%	MQS07157.1
hypothetical protein [Pseudohodobacter sp.]	*Pseudohodobacter sp.*	7e-36	53.79%	MBC7679085.1
hypothetical protein [Rubrivirga sp.]	*Rubrivirga sp.*	3e-32	46.53%	MBR24884.1
hypothetical protein DCC49_08730 [Acidobacteria bacterium]	*Acidobacteria bacterium*	3e-30	43.36%	RIK08684.1
ATP synthase subunit B/B' [Acidobacteria bacterium]	*Acidobacteria bacterium*	8e-29	42.47%	MBU6254263.1
ATP synthase subunit B/B' [SAR202 cluster bacterium]	*SAR202 cluster bacterium*	9e-27	41.50%	NQW06729.1

Table 3-6. The amino acid sequence of Dia (UniProt ID: Q9ZBR0) from *S. coelicolor* was BLAST against other bacterial genomes, excluding the Actinobacteria phylum. After roughly 30 hits with Dia, hits were identified as potential partially conserved homologs of Dia, ATP synthase subunit B and DivIVA alignments found across various bacterial genomes including: *Rhodococcus*, *Clostridia*, BMS3Abin01_01157, *Elusimicrobia*, *Clostridales*, *Acidobacteria*, *Firmicutes*, *Deltaproteobacteria*, HRbin41_00973, *Gordonia*, *Chloroflexi*, DLM65_10790 and *Limnochordia*. The identified hits and alignments possessed between 30-43% similarity to Dia.

Description	Scientific name	E value	Percentage Identity	Accession
cell division septum initiation protein [Rhodococcus opacus]	*Rhodococcus opacus*	3e-22	33.33%	MBA8960431.1
ATPase [Clostridia bacterium]	*Clostridia bacterium*	2e-21	36.24%	MBC7341733.1
hypothetical protein BMS3Abin01_01157 [bacterium BMS3Abin01]	*bacterium BMS3Abin01*	2e-21	34.46%	GBE58225.1
ATPase [Elusimicrobia bacterium]	*Elusimicrobia bacterium*	3e-21	36.11%	MCR4296034.1
ATPase [Clostridales bacterium]	*Clostridales bacterium*	3e-21	35.33%	MBS3957128.1
ATPase [Acidobacteria bacterium]	*Acidobacteria bacterium*	4e-21	35.22%	MBU6336392.1
ATPase [Firmicutes bacterium]	*Firmicutes bacterium*	6e-21	38.03%	NPV52129.1
hypothetical protein [Acidobacteria bacterium]	*Acidobacteria bacterium*	1e-20	43.61%	MBI4727913.1
ATPase [Clostridia bacterium]	*Clostridia bacterium*	1e-20	32.65%	MQY74842.1
ATPase [Deltaproteobacteria bacterium]	*Deltaproteobacteria bacterium*	1e-20	34.25%	MBA3821031.1
hypothetical protein [Acidobacteria bacterium]	*Acidobacteria bacterium*	1e-20	33.33%	MBW4079070.1
ATPase [Acidobacteria bacterium]	*Acidobacteria bacterium*	1e-20	34.93%	KAA0273641.1
hypothetical protein HRbin41_00973 [bacterium HR41]	*bacterium HR41*	2e-20	36.30%	GBD46151.1

DivIVA domain-containing protein [Acidobacteria bacterium]	*Acidobacteria bacterium*	3e-20	32.76%	RIK12837.1
DivIVA domain-containing protein [Gordonia asplenii]	*Gordonia asplenii*	4e-20	31.03%	NMN99701.1
ATPase [Chloroflexi bacterium]	*Chloroflexi bacterium*	5e-20	38.85%	MBX6771757.1
hypothetical protein DLM65_10790 [Candidatus Dormibacter sp. RRmetagenome_bin12]	*Candidatus Dormibacter sp. RRmetagenome_bin12*	2e-19	38.46%	PZR79409.1
hypothetical protein [Limnochordia bacterium]	*Limnochordia bacterium*	2e-19	36.81%	MCK9221145.1
DivIVA domain-containing protein [Rhodococcus erythropolis]	*Rhodococcus erythropolis*	3e-19	30.80%	MYV30590.1

Across all three BLAST searches, various hits against ATPases, DivIVA and ATP synthase subunit B-based proteins were also identified. As multiple hits for potential Dia homologs were found across bacteria with various degrees of similarity, we decided to explore the homology between these potential homolog hits compared to Dia.

3.2.2 Potential domains of Dia

To investigate the homology between the Dia homologs across bacteria species, 3 different collections of potential homologs within each previous BLAST search were selected and aligned separately against Dia (UniProt ID: Q9ZBR0) using the NCBI Multiple Alignment Viewer 1.22.2. For each collection, the aligned Dia homologs against Dia were from different species of bacteria depending on the BLAST search. The initial BLAST search allowed for potential Dia homologs from various species of *Streptomyces* to be aligned, whereas the BLAST search excluding Streptomyces genus allowed for the alignment of potential Dia homologs from various species of Actinobacteria. Finally, the BLAST search excluding Actinobacteria phyla, allowed for the alignment of potential Dia homologs from various bacterial species across many different phyla including Acidobacteriota and Bacillota.

Within the alignment of the 15 Dia homologs from the initial BLAST search against Dia, most aligned homologs were of similar lengths to Dia, with some homologs being either 1 amino acid longer or between 2-16 amino acids shorter (depending on the homolog). The N-terminus (approximately 20 amino acids) and C-terminus (approximately 64 amino acids) of Dia were not conserved over the *Streptomyces* homologs (Figure 3-2). The remaining middle section of Dia (approximately 295 amino acids – residues 20-315) demonstrated high levels of conservation (Figure 3-2), with the same 8 conserved regions identified. Within every *Streptomyces* Dia homolog (Accession numbers: WP_011030311.1, WP_003973425.1, WP_093457067.1, WP_136209400.1, WP_121714210.1, WP_109036097.1, MYS74452.1, WP_202493697.1, NEC38046.1, WP_247704091.1, WP_115744393.1, WP_203697443.1, WP_153177767.1, NHI06919.1, KPC80448.1) the 8 conserved regions were seen between residues: 20-231, 235, 258-261, 264, 270-285, 277-300, 302-311 and 313-315. Between these conserved regions across the 15 *Streptomyces* homologs were multiple non-conserved regions of varying sizes. The non-conserved regions throughout the Dia homologs were likely a result of multiple substitutions and deletions (Figure 3-2).

For the second BLAST search Dia homolog alignment, 13 homologs were selected from the vast range of different genus within the Actinobacteria phyla including *Kitasatospora* and *Streptacidiphilus*. Some of the aligned homologs against Dia possessed similar lengths to Dia, though only 377 amino acids of Dia were aligned. Some homologs were found to be 3 amino acids longer or between 6-15 amino acids shorter than the shortened Dia sequence, with one aligned homolog being drastically shorter by 181 amino acids due to lacking the entire C-terminus of Dia. In a similar fashion to the initial BLAST alignment, no conservation of the N-terminus (approximately 20 amino acids) and C-terminus was seen but a larger non conserved C-terminus was identified in the alignment (approximately 161 amino acids). Between residues 20-216 were 3 highly conserved sections of across the 13 homologs (Accession numbers: WP_221504102.1, WP_037607620.1, WP_042454107.1, WP_111504744.1, SEM36912.1, WP_03409117.1, MBF9067326.1, MBC3844111.1, WP_255744691.1, WP_182454912.1, WP_030263937.1, WP_2548002178.1, WP_042392355.1). The 3 conserved regions of the aligned homologs were identified between residues 20-156, 158-170 and 176-216. The first 2 identified conserved regions across the homologs possessed large segments of their sequences which were highly conserved (thick red bars). The highly conserved segments were separated by sporadic single

residue segments which were poorly conserved (thick blue bars). The highly conserved segments were of varying lengths across the conserved regions. The non-conserved regions throughout the Dia homologs were likely a result of multiple substitutions and deletions (Figure 3-3).

Finally, the Dia homolog alignment against last BLAST search included 7 homologs selected from the vast range of phyla including Acidobacteriota and Bacillota. Most of the aligned homologs against Dia possessed much shorter lengths to Dia, though only 205 amino acids of Dia were aligned. The homologs shorter than Dia were between 32-59 amino acids shorter and one aligned homolog being longer than Dia by 29 amino acids. There was no conservation of the N-terminus (approximately 35 amino acids) and C-terminus was seen. One of the homologs possessed large non conserved C-terminus within their alignments (approximately 61 amino acids). Between residues 35-144 were 2 highly conserved sections of across the 7 homologs (Accession numbers: WP_037607620.1, WP_042454107.1, WP_111504744.1, SEM36912.1, WP_03409117.1, MBF9067326.1, MBC3844111.1, WP_182454912.1, WP_030263937.1, WP_2548002178.1). The 2 conserved regions of the aligned homologs were identified between residues 35-90 and 92-144. The identified conserved regions across the homologs demonstrated repeating segments of their sequences which were highly conserved and poorly conserved. The highly and poorly conserved sequences were of varying lengths across the conserved regions. The non-conserved regions throughout the Dia homologs were likely a result of multiple substitutions and deletions (Figure 3-4).

Collectively, the conserved section (between residues 59-141) across all the homolog alignments of Dia demonstrates a shared ancestry (though the conservation of the region differs over the different alignment groups). With the conserved region identified throughout and outside of the Actinobacteria phyla, some of the functions of Dia may have been retained across this shared ancestry.

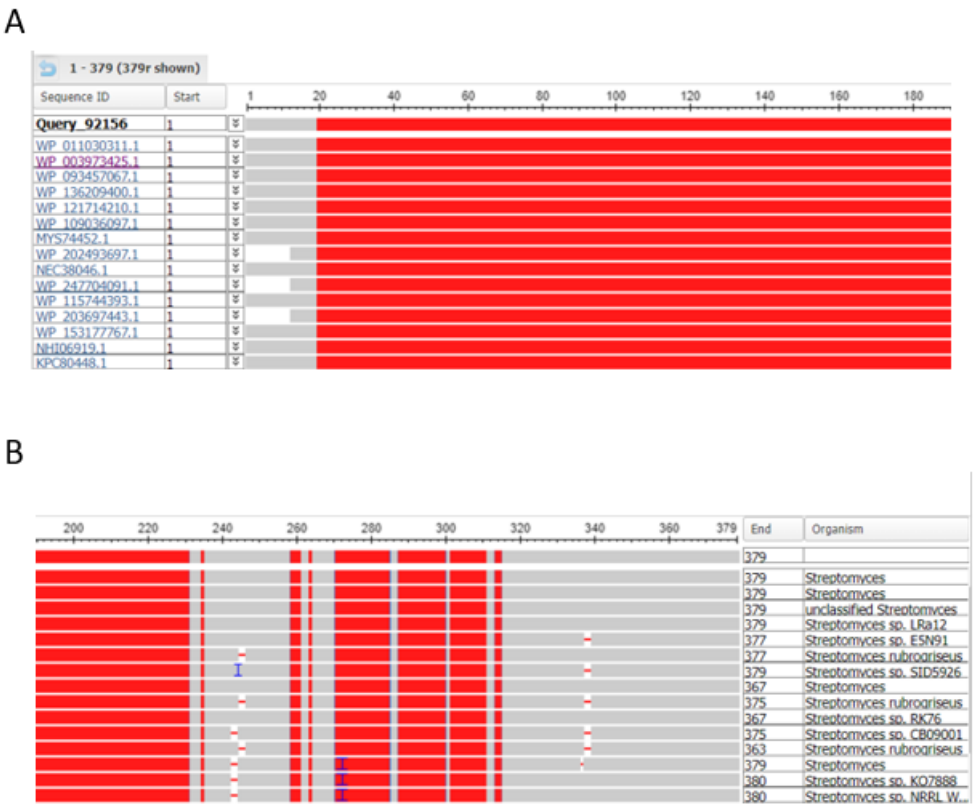


Figure 3-2. The conserved similarity between Dia (UniProt ID: Q9ZBR0) and potential homologs identified in the initial BLAST search. The homologs were aligned within a multiple alignment search using NCBI Multiple Sequence Alignment Viewer 1.22.0. (A) The alignment of Dia (query) against a collection of potential homologs between residue 1-190 approximately of Dia. (B) The continued alignment of Dia against the potential homologs between approximately residue 190-379 of Dia. The NCBI accession number and species name of each aligned homolog can be found under the sequence ID and Organism columns respectively. The labelled sequences are as such: Query_92156: The *S. coelicolor* Dia sequence from UniProt (ID: Q9ZBR0), Accession number WP_011030311.1: first *Streptomyces* homolog hit aligned, Accession number WP_003973425.1: second *Streptomyces* homolog hit aligned, Accession number WP_093457067.1: an unclassified *Streptomcyes* homolog hit aligned, Accession number WP_136209400.1: a *Streptomcyes LRa12* homolog hit aligned, Accession number WP_121714210.1: a *Streptomyces E5N91* homolog hit aligned, Accession number WP_109036097.1: first *Streptomyces rubrogriseus* homolog hit aligned, Accession number MYS74452.1: a *Streptomyces sp. SID5926* homolog hit aligned, Accession number WP_202493697.1: third *streptomyces* homolog hit aligned, Accession number NEC38046.1: second *Streptomyces rubrogriseus* homolog hit aligned, Accession number WP_247704091.1: a *Streptomyces sp. RK76* homolog hit aligned, Accession number WP_115744393.1: a *Streptomyces sp. CB9001* homolog hit aligned, Accession number WP_203697443.1: third *Streptomyces rubrogriseus* homolog hit aligned, Accession number WP_153177767.1: fourth *Streptomyces* homolog hit aligned, Accession number NHI06919.1: a *Streptomyces sp. KO7888* homolog hit aligned, Accession number KPC80448.1: a *Streptomyces sp. NRRL WC-3753* homolog hit aligned. The initial and final aligned residues of each homolog can be found under the Start and End columns respectively. All the aligned homologs possessed a similar N-terminus to Dia, between approximately residues 1-20, though the termini were not conserved – possibly due sufficient substitutions (thick grey bars). 8 conserved regions of the aligned homologs were identified between residues 20-231, 235, 258-261, 264, 270-285, 277-300, 302-311 and 313-315. The identified conserved regions across the homologs vary in length of their sequences, with all being highly conserved (thick red bars). Many of the homologs possessed gaps in their sequences when aligned against Dia (thin red lines). Non-conserved sections of the homologs are found between the conserved regions. The C-terminus of the aligned homologs from approximately residue 315-379 possessed multiple aligned regions with Dia though the termini were not conserved (possibly due sufficient substitutions and deletions).

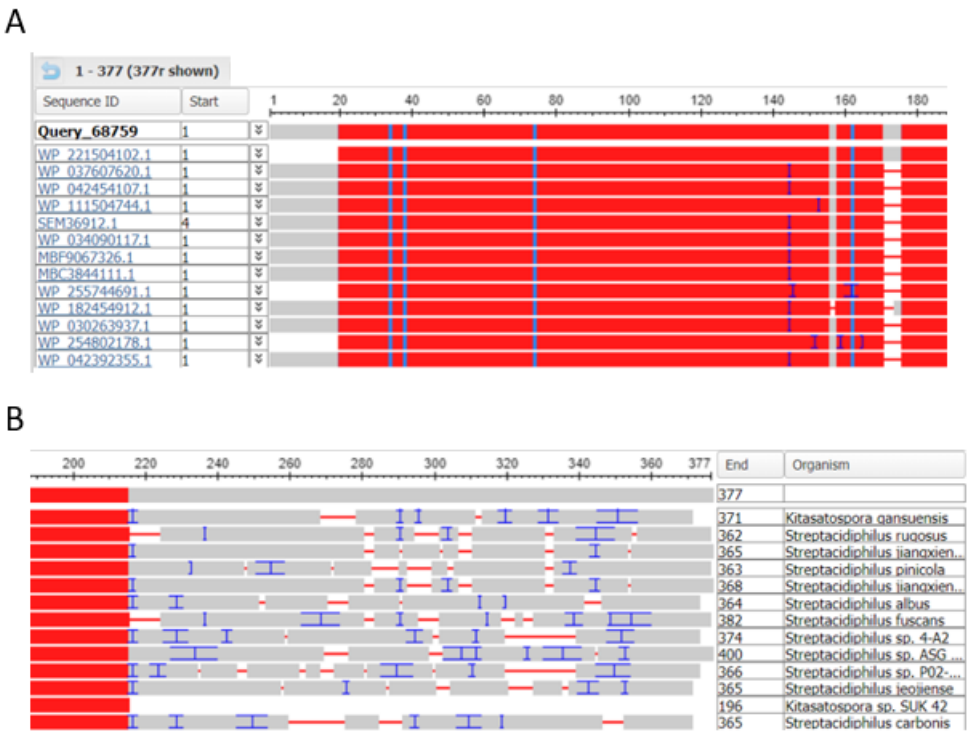


Figure 3-3. The conserved similarity between Dia (UniProt ID: Q9ZBR0) and potential homologs identified in the BLAST search excluding *Streptomyces*. The homologs were aligned within a multiple alignment search using NCBI Multiple Sequence Alignment Viewer 1.22.2. (A) The alignment of Dia (query) against a collection of potential homologs between residue 1-184 approximately of Dia. (B) The continued alignment of Dia against the potential homologs between approximately residue 184-377 of Dia. The NCBI accession number and species name of each aligned homolog can be found under the sequence ID and Organism columns respectively. The labelled sequences are as such: Query_68759: The *S. coelicolor* Dia sequence from UniProt (ID: Q9ZBR0), Accession number WP_221504102.1: a *Kitasatospora gansuensis* homolog hit aligned, Accession number WP_037607620.1: a *Streptacidiphilus rugosus* homolog hit aligned, Accession number WP_042454107.1: first *Streptacidiphilus jiangxiensis* homolog hit aligned, Accession number WP_111504744.1: a *Streptacidiphilus pinicola* homolog hit aligned, Accession number SEM36912.1: second *Streptacidiphilus jiangxiensis* homolog hit aligned, Accession number WP_03409117.1: a *Streptacidiphilus albus* homolog hit aligned, Accession number MBF9067326.1: a *Streptacidiphilus fuscans* homolog hit aligned, Accession number MBC3844111.1: a *Streptacidiphilus sp. 4-A2* homolog hit aligned, Accession number WP_255744691.1: a *Streptacidiphilus sp. ASG 303* homolog hit aligned, Accession number WP_182454912.1: a *Streptacidiphilus sp. P02-A3a* homolog hit aligned, Accession number WP_030263937.1: a *Streptacidiphilus jeojiense* homolog hit aligned, Accession number WP_2548002178.1: a *Kitasatospora sp. SUK 42* homolog hit aligned, Accession number WP_042392355.1: a *Streptacidiphilus carbonis* homolog hit aligned. The initial and final aligned residues of each homolog can be found under the Start and End columns respectively. 10 of the aligned homologs possessed a similar N-terminus to Dia, between approximately residues 1-20, though the termini were not conserved – possibly due sufficient substitutions (thick grey bars). The remaining 3 homologs lacked the non-conserved N-terminus of Dia. 3 conserved regions of the aligned homologs were identified between residues 20-156, 158-170 and 176-216. The first 2 identified conserved regions across the homologs possessed large segments of their sequences which were highly conserved (thick red bars), which were separated by several single residue segments which were poorly conserved (thick blue bars). The last conserved domain didn't contain any poorly conserved segments. The highly conserved segments were of varying lengths across the conserved regions. Many of the homologs possessed gaps in their sequences when aligned against Dia (thin red lines). Non-conserved sections of the homologs are found between the conserved regions. The C-terminus of the aligned homologs from approximately residue 216-377 possessed multiple aligned regions with Dia though the termini were not conserved (possibly due sufficient substitutions and deletions). 1 homolog lacked the non-conserved C-terminus of Dia.

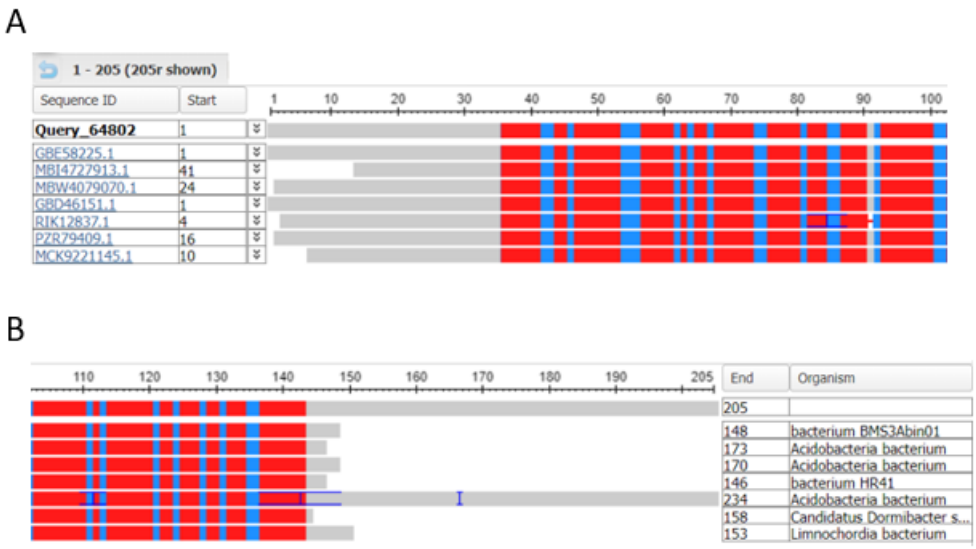
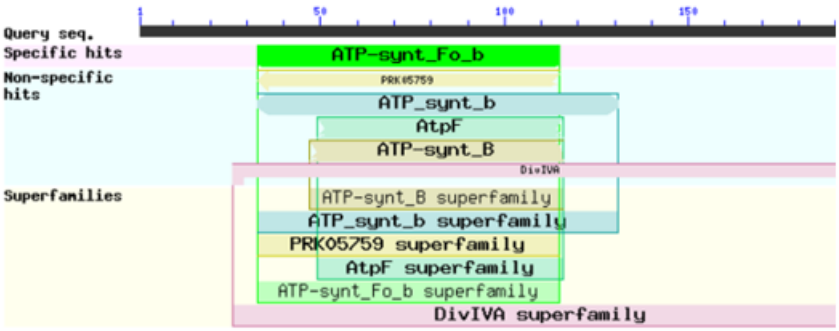


Figure 3-4. The similarity between Dia (UniProt ID: Q9ZBR0) and potential homologs identified in the BLAST search excluding the Actinobacteria phylum. The homologs were aligned within a multiple alignment search using NCBI Multiple Sequence Alignment Viewer 1.22.2. (A) The alignment of Dia (query) against a collection of potential homologs between residue 1-102 approximately of Dia (indicated by the numbered scale bar above the coloured aligned sequences). (B) The continued alignment of Dia against the potential homologs between approximately residue 102-205 of Dia. The NCBI accession number and species name of each aligned homolog can be found under the sequence ID and Organism columns respectively. The labelled sequences are as such: Query_92513: The *S. coelicolor* Dia sequence from UniProt (ID: Q9ZBR0), Accession number GBE_58225.1: a *bacterium BMS3Abin01* homolog hit aligned, Accession number MBI4727913.1: first *Acidobacteria bacterium* homolog hit aligned, Accession number MBW4079070.1: second *Acidobacteria bacterium* homolog hit aligned, Accession number GBD46151.1: a *bacterium HR41* homolog hit aligned, Accession number RIK12837.1: third *Acidobacteria bacterium* homolog hit aligned, Accession number PZR79409.1: a *Canadidatus Dormibacter* homolog hit aligned, Accession number MCK9221145.1: a *Limochordia bacterium* homolog hit aligned. The initial and final aligned residues of each homolog can be found under the Start and End columns respectively. Only 2 of the aligned homologs possessed a similar N-terminus to Dia, between approximately residues 1-35, though the termini were not conserved – possibly due sufficient substitutions (thick grey bars). The remaining 5 homologs possessed majority of this non-conserved N-terminus, though the length of each N-terminus varied (anywhere between residue 4-35 and 13-35 approximately). 2 conserved regions of the aligned homologs were identified between residues 35-90, 92-144. All the identified conserved regions across the homologs showed repeating segments of their sequences which were either highly conserved (thick red bars) or poorly conserved (thick blue bars) sequences. The highly and poorly conserved segments were of varying lengths across the conserved regions. A homolog possessed a gap in their sequence when aligned against Dia (thin red line). Non-conserved sections of the homologs are found between the conserved regions. The C-terminus of the aligned homologs from approximately residue 144-205 possessed multiple aligned regions with Dia though the termini were not conserved (possibly due sufficient substitutions and deletions).

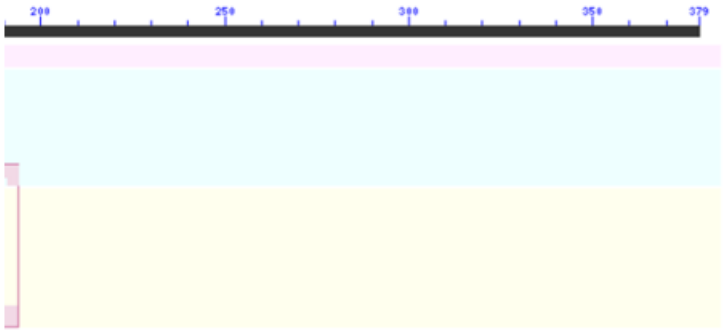
Further analysis of the highest match from the BLAST search revealed that Dia may contain 2 functional domains, a DivIVA based region between 26-194 aa and F0F1 ATP synthase subunit B region between 33-115 aa. The overlapping of these predicted domains is unexpected as both domains possess different characteristics to each other. The DivIVA-based domain is known for helping maintain the cell shape and cell wall integrity through recruitment of cell wall synthesis machinery whereas the F0F1 ATP synthase subunit B is documented for its role in the formation of the peripheral stalk which links together the catalytic complex and to act as a stator to hod the rotating catalytic complex in place (Walker & Dickson, 2006). Both proteins however are found across a wide range of different bacteria and are highly conserved in both sequence and structure. The remaining homologs also predicted the DivIVA-based and F0F1 ATP synthase subunit B domains within roughly the same position in their primary sequences.

With the universal conserved region between residues 35-144 of Dia, we which to infer any conserved function within the region by analysing the primary sequence of Dia through the NCBI Conserved Domain search. The search revealed a specific hit for an F-type ATP synthase, membrane subunit B (ATP-synt_Fo_b) domain between residues 33-115. Overlapping with this specific hit were multiple non-specific ATP synthase subunit B domains (PRK05759, ATP_synt_b, AtpF, ATP-synt_B) between the residues 33-131. Finally, a non-specific hit for a DivIVA domain overlapped with the various ATP synthase subunit B domains between residues 26-194 (Figure 3-5).

A



B



C

Name	Accession	Description	Interval	E-value
PRK05759	PRK05759	F0F1 ATP synthase subunit B; Validated	33-115	4.84e-08
ATP-synt_Fo_b	cd06503	F-type ATP synthase, membrane subunit b; Membrane subunit b is a component of the Fo complex ...	33-115	1.25e-07
ATP_synt_b	TIGR01144	ATP synthase, F0 subunit b; This model describes the F1/F0 ATP synthase b subunit in bacteria ...	33-131	1.60e-06
AtpF	COG0711	FoF1-type ATP synthase, membrane subunit b or b' [Energy production and conversion];	49-116	2.73e-05
ATP-synt_B	pfam00430	ATP synthase B/B' CF(0); Part of the CF(0) (base unit) of the ATP synthase. The base unit is ...	47-116	9.26e-04
DivIVA	COG3599	Cell division septum initiation DivIVA, interacts with FtsZ, MinD and other proteins [Cell ...	26-194	1.32e-03

Figure 3-5. The conserved theoretical domains predicted with Dia (UniProt ID: Q9ZBR0) using the NCBI Conserved Domains search. (A) The potential conserved domains identified between approximately residue 1-190 of the primary sequence of Dia. (B) The potential conserved domains identified between approximately residue 190-379 of the primary sequence of Dia. (C) NCBI Accession numbers, descriptions, aligned residues of Dia (interval) and the significance of each returned hit. The Query sequence and scale bar indicate the primary sequence of Dia in segments of 10 residues. A specific hit for an F-type ATP synthase, membrane subunit B domain (cd06503) was identified between residues 33-115. 4 non-specific hits were identified for a collection of additional ATP synthase subunit B domains (PRK05759, TIGR01144, COG0711, pfam00430) within overlapping range of residues as the specific hit (between residues 33-131). A non-specific DivIVA domain (COG3599) was identified overlapping with the various ATP synthase subunit B domains within residues 26-194. Superfamilies of the specific and non-specific hits were returned within the same ranges.

With the Dia conserved domain search suggesting the potential conserved domains of DivIVA and ATP synthase subunit B within the conserved segment of Dia (residue 35-144), we decided to explore the similarity of these potential domains to Dia. We carried out alignment of Dia (UniProt ID: Q9ZBR0) against documented homologs of DivIVA (UniProt ID: Q9S2X4) and AtpF (UniProt ID: Q9K4D7) of *S. coelicolor* using Clustal Omega.

The alignment of Dia against DivIVA revealed within the aligned residues 26-194 of Dia, partial similarity through conservation of multiple individual residues alongside the substitution of residues with strong and weak similar properties. These conserved and substituted aligned residues of Dia were within clusters of the DivIVA sequence of varying sizes. A second aligned section of Dia between residues 195-324 was indicated, with possession of a similar pattern of conservation and substitution as the initial predicted alignment of Dia with DivIVA (residues 26-194). Gaps of various sizes were interspaced with the partially conserved alignments with DivIVA (Figure 3-6).

```
Dia         MDVQNKLDEITAMVSGARAMPMSASCVVNR-------------AELLSMLEELRAELPGS        47
DivIVA      -------------------MPLTPEDVRNKQFTTVRLREGYDEDEVDAFLDEVEAELTRL        41
                             **:: . * *:              *: ::*:*:.***

Dia         LAQAQELIGDREQM---VAQARQEAD----------------RIIEGAHAERGSLIADTE        88
DivIVA      LRENEDLRAKLAAATRAAAQNQQNMRKPPEPPQDQQHQQGPPQHQQGPP-QQGGMPQQG-        99
            * : ::* ..       .** :*:                  :  :*   ::*.:  :

Dia         VARRSQAEADRILAEARQEAEEVRAEADDYVDSKLANFEV------------VLTKTLGS        136
DivIVA      MRGPGAPVPAGISGPPQQQMGGPM-GGPPQLPSGAPQLPAGPGGQGGPQGPGPMGQGPGP        158
            :   .      * .  :*:       .   : *   :: .             : :  *

Dia         VGRGREKLL---G-----------------TGPGLDENGYEDEDAPERS-----------        165
DivIVA      MGQGPGPMQGQMGPGPMGGPMGGPQGPGGPGGPGMPGQGGPGGDSAARVLSLAQQTADQA        218
            :*:*   :    *                  ***:  :*  . *:  *

Dia         --------------------------HDPETLRRDADAYVDTKLGAFEAVL---AKTLD        195
DivIVA      IAEARSEANKIVGEARSRAEGLERDARAKADALERDAQEKHRVAMGSLESARATLERKVE        278
                                  . ::*.***:    . :*::*:.     :.::

Dia         AVGRGRQKLHGRIATDDLGALADDMTTVQHSSDADY----LAGLAGLADAPASAPAPA--        249
DivIVA      DLRGFEREYRTRLK----SYLESQLRQLETQADDSLAPPRTPATASLPPSPAPSMAPAGA        334
             :   .:: : *:     . * .::  :: .:* .       . *.*  :** : ***

Dia         --------------EQQGQPQYGEQQPVAARMPAQAVPEMPSQEPVYGYASQQQPDPYAA        295
DivIVA      SAPSYGGNQSMGGGPGQSGPSYGGQQQMSPAMTQPMAPVR--------------------        374
                       *. *.** ** ::  *    .*

Dia         YQQTYDGGPDPYGYQQQGADPYAYQYDGGQQAYDAQQGYAQPPQPQQPPHAPQTHPQALD        355
DivIVA      -----PQGPSPMGQAPSPMRGFLIDEDDN------------------------------        398
                  **.* *   .    :  : *..

Dia         ETSLFDTSMISAEQLRAYEQGRGL    379
DivIVA      ------------------------    398
```

Figure 3-6. Protein sequence alignment between Dia (UniProt ID: Q9ZBR0) and DivIVA (UniProt ID: Q9S2X4) of *S. coelicolor*, aligned using Clustal Omega. (*) represents positions with a conserved residue, (:) indicates conserved amino acid groups with strongly similar properties and (.) represents conserved amino acid groups with weakly similar properties.

The alignment of Dia against AtpF demonstrated a pattern of partial conservation and substitution of residues within the predicted alignment of residues (49-116) of Dia. 2 additional alignments between Dia and AtpF were indicated between residues 1-48 and 138-227. Both residues possessed similar patterns of conservation and substitution as the predicted alignment between Dia and AtpF. Multiple gaps and varying sizes were interspaced between the partially conserved alignments AtpF (Figure 3-7).

```
Dia       ---------MDVQNK-LDEITAMVSGARAMPMSASCVVN--RAELLSMLEELRAELPGSL  48
AtpF      MSPMLQIAAEEMENPLIPPIPELVIGLIAFVIVFGFLAKKLLPNINKVLEERREAIEGGI  60
                    :::*  :  *  :* *  *: :  . :.:    :: .:*** *  : *.:

Dia       AQAQE-------LIGDREQMVAQARQEADRIIEGAHAERGSLIADTEVARRSQAEADRIL  101
AtpF      EKAEAAQTEAQSVLEQYKAQLAEARHEAARLRQEAQEQGATLIAEM--------------  106
           :*:        :: : :  :*:**:** *: : *: : .:***:

Dia       AEARQEAEEVRAEADDYVDSKLANFEVVLTKTLGSVGRGREKLLGTGPGLDENGYEDEDA  161
AtpF      ---RAEG----------------------------QRQREEIIAAGHAQIQ--------  126
             * *.                              * **:::.:* .  :

Dia       PERSHDPETLRRDADAYVDTKLGAFEAVLAKTLDAVGRGRQKLHGRIATDDLGALADDMT  221
AtpF      ADRKAAASALRQD--------VGKLATELAGKLVGESLEDHARQSRVIDRFLDELDDKAT  178
           :*.   .:**:*        :* : : ** .* . .   :  :.*:    *. * *. *

Dia       TVQHSSDADYLAGLAGLADAPASAPAPAEQQGQPQYGEQQPVAARMPAQAVPEMPSQEPV  281
AtpF      TAEATR------------------------------------------------------  184
          *.: :

Dia       YGYASQQQPDPYAAYQQTYDGGPDPYGYQQQGADPYAYQYDGGQQAYDAQQGYAQPPQPQ  341
AtpF      ------------------------------------------------------------  184


Dia       QPPHAPQTHPQALDETSLFDTSMISAEQLRAYEQGRGL        379
AtpF      ------------------------------------        184
```

Figure 3-7. Protein sequence alignment between Dia (UniProt ID: Q9ZBR0) and AtpF (UniProt ID: Q9K4D7) of *S. coelicolor*, aligned using Clustal Omega. (*) represents positions with a conserved residue, (:) indicates conserved amino acid groups with strongly similar properties and (.) represents conserved amino acid groups with weakly similar properties.

The comparison of Dia against DivIVA and AtpF separately demonstrated areas of partial conservation (Figure 3-6 and 3-7) throughout Dia against both proteins, especially within predicted conserved domains of DiviVA (residues 26-194) and AtpF (residues 49-116). Curiously, new aligned sections within Dia outside of these predicted domains were indicated for both alignments and possessed partial conservation. Additionally, both predicted conserved domains of Dia for DivIVA and

AtpF demonstrated similar levels of conservation within their overlapping regions (residues 49-116).

The current knowledge of Dia, through BLAST, UniProt and sequence alignments, has shown that Dia has a highly conserved section (residues 35-144) found within countless species across many phyla. Alongside this, Dia has been revealed to contain 2 possible overlapping domains for: DivIVA and ATP synthase subunit B. These domains are found within the conserved region and Dia possess similarity to known homologs of these domains within *S. coelicolor*. Additionally, these hits for proteins of these domains were found across all BLAST searches. To understand these possible domains further, we decided to continue investigating the possible structural capacities of Dia of *S. coelicolor* through analysing its secondary and tertiary structure.

3.2.3 Potential secondary and tertiary structure of Dia

To predict the secondary structures Dia may contain, as well as help confirm if the predicted coiled-coils are left-handed, the primary sequence of Dia (UniProt ID: Q9ZBR0) was passed through the PCOILS program using select parameters. These parameters were: the use of the RSCB Protein Data Bank matrix (to compare structures of similar proteins) and the weighting of the hydrophobic residues within the program (to allow for the detection of any false positives). Since the weighting of the residues required two different programs to be run for comparison of a false positive, 2 PCOILS program were run with either the hydrophobic residue weighting on or off. The PCOILS programs both predicted the same secondary structure for Dia: a majority of α-helices throughout the sequence, with a high probability between 0-110 aa and 150-205 aa. Additionally, an β-sheet was predicted to likely form around residue 120 (Figure 3-8).

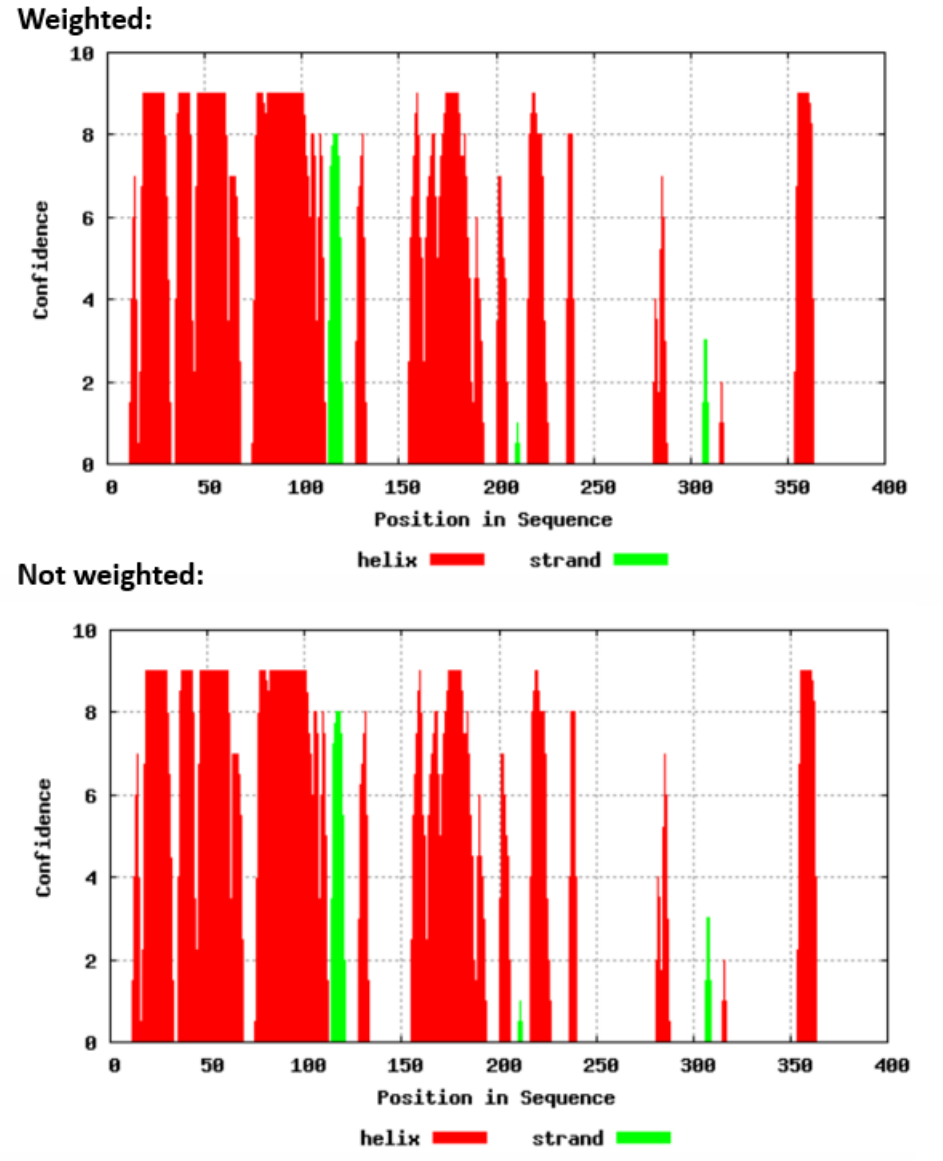


Figure 3-8. PSIRED secondary structure predictions of Dia (UniProt ID: Q9ZBR0), based on the primary sequence alone, using the PCOILS analytic algorithm. Dia is predicted to contain large segments of α-helices between 0-110 aa and 150-205 aa with small segments of α-helixes scattered throughout the remaining sequence. Dia is likely to contain a β-sheet at about 120 aa. The predicted secondary structure was identical for Dia whether the searches were weighted or not.

With Dia mostly predicted to form α-helices, which are useful in forming coiled-coils in the tertiary structures, we investigated the predicted Dia tertiary structure from UniProt. The UniProt data predicted 2 coiled-coil regions between residues 48-68 and 90-121 which are within the predicted ranges for the domains of Dia (35-144). Additionally, 2 disordered regions between 141-173 aa and 242-356 aa were also predicted (Figure 3-9).

Figure 3-9. The initial structure revealed within Dia (UniProt ID: Q9ZBR0) from the UniProt database. 2 coiled-coil regions (black bars) are predicted within 48-68 aa and 90-121 aa. 2 disordered regions are predicted afterwards between 141-173 aa and 242-356 aa.

Alongside the secondary structure predictions, the PCOILS coiled-coil predictions for Dia predicted the occurrence of coiled-coils between roughly 30-60 aa and 75-110 aa within the 28 residue screening window for both programs. However, the weighted program predicted at least a 60% probability for 30-60 aa coiled-coil and about a 20% probability for the 75-110 aa coiled-coil. The non-weighted program predicted a 90% chance for the coiled-coils. Due to the difference in probability between the programs being greater that 30%, the predicted coiled-coils are likely to be false positives (Figure 3-10).

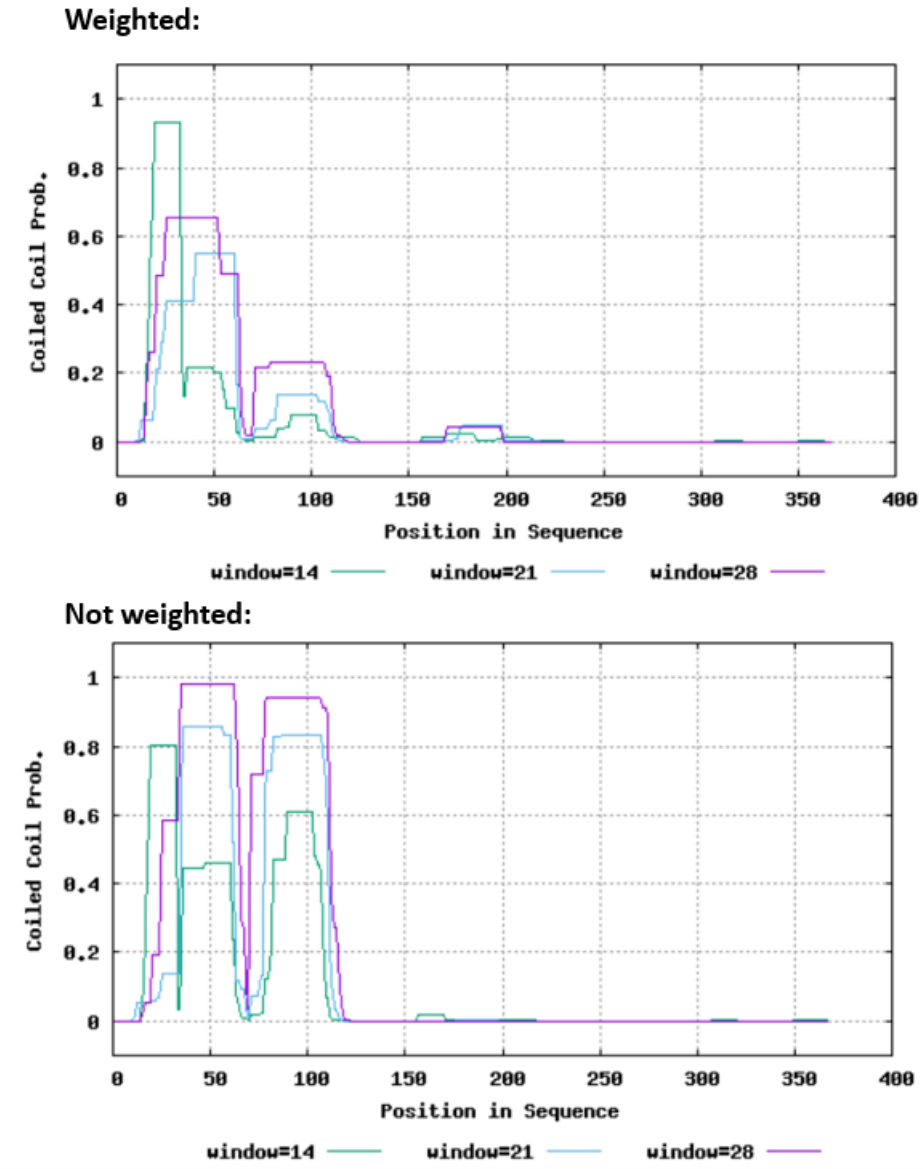


Figure 3-10. Tertiary structure predictions of Dia (UniProt ID: Q9ZBR0), based on the primary sequence alone, using the PCOILS analytic algorithm. The weighted prediction in the 28 residue screening window suggested a low probability of coiled-coil regions between roughly 30-50 aa and 75-110 aa. The not weighted prediction in the 28 residue screening window suggests a high probability of coiled-coil regions between roughly 30-60 aa and 75-110 aa.

Since the PCOILS program coiled-coil prediction for left-handed coiled-coils disagreed with the prior knowledge of the tertiary structure of Dia, the MARCOIL and the DeepCoil algorithms were used to help determine whether Dia contained any right-hand coiled-coils. Firstly, the primary sequence of Dia (UniProt ID: Q9ZBR0) was passed through the MARCOIL algorithm, in a similar manor to the PCOILS program. The primary sequence was inputted into the program and run using the parameters: 9FAM matrix (useful for first time screening due to the lack of structural

data) and the transitional probability of the coiled-coil formation in the Markov chain model. 2 programs were run both using the 9FAM matrix and either having the transitional probability to high or low. The programs both predicted a large, coiled-coil region between 30-105 aa with at least 90% probability. The low transitional probability program predicted a uniform coiled-coil region whereas the high transitional probability program prediction had a decreased probability (of 90%) for the coiled-coil region at about reside 65 (Figure 3-11).

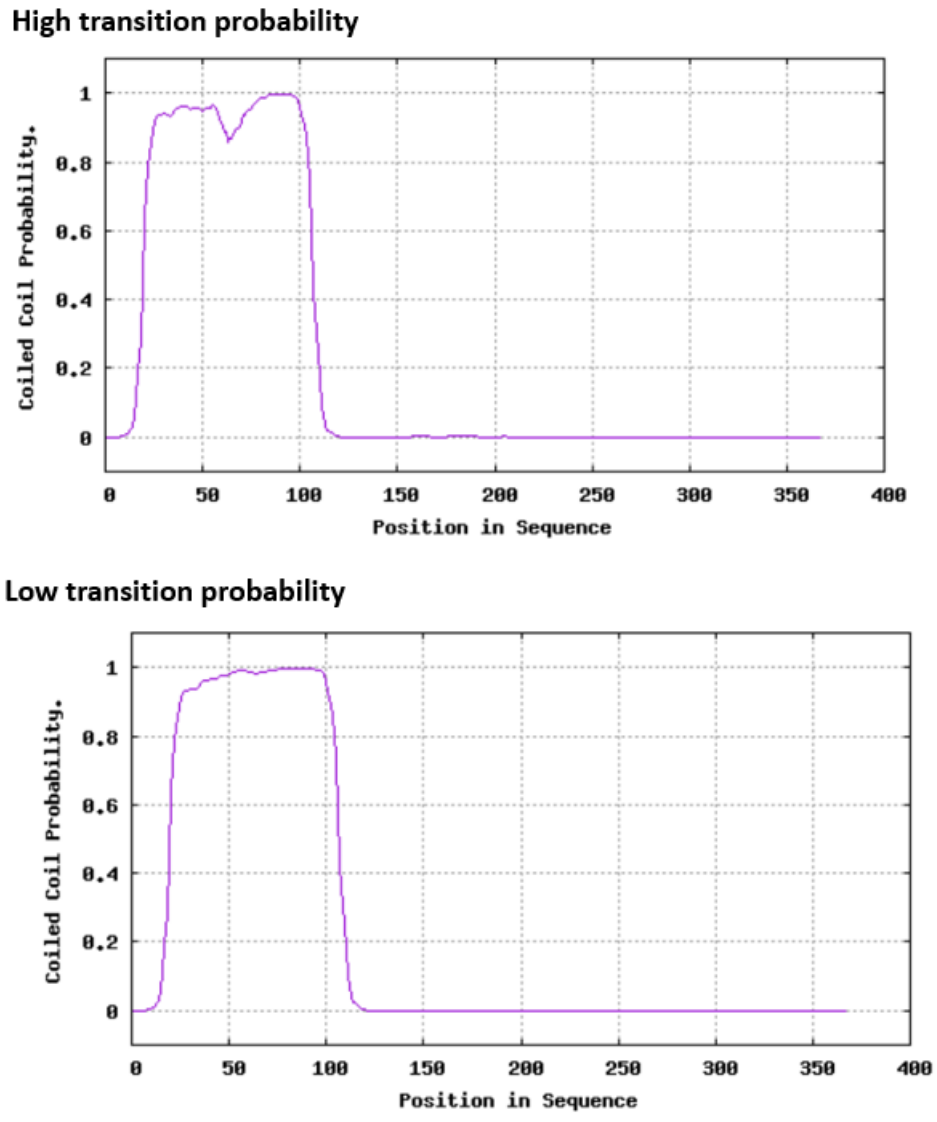


Figure 3-11. Tertiary structure predictions of Dia (UniProt ID: Q9ZBR0) using the MARCOIL analytic algorithm. The high transitional probability prediction suggests a coiled-coil region throughout residues 30-105 with at least 90% probability. The low transitional probability prediction suggests the same coiled-coil region throughout residues 30-105 with almost 100% probability.

Finally, the primary sequence of Dia (UniProt ID: Q9ZBR0) was analysed through the DeepCoil algorithm to help confirm the results from the MARCOIL algorithm. The sequence was analysed under the parameter: Use_input_sequence (the primary sequence is compared using the algorithm's default settings). The program predicted the occurrence of coiled-coils between 20-100 aa. Within this range, there are 2 likely clusters of coiled-coils between roughly 20-65 aa and 65-100 aa with at least 50% probability. Both clusters have regions with considerably higher probability (at least 80%) being roughly between 30-50 aa and 90-95 aa (Figure 3-12).

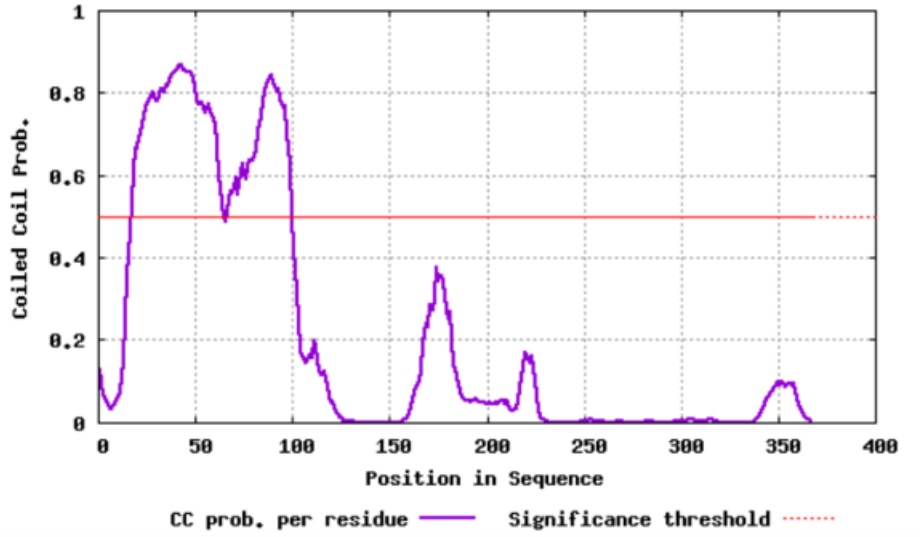


Figure 3-12. Tertiary structure predictions of Dia (UniProt ID: Q9ZBR0) using the DeepCoil analytic algorithm. Throughout residues 20-100, a coiled-coil region was predicted with a probability of 50% and above. Residues 30-50 and 90-95 had a highest probability (of 80% and above) to form a coiled-coil region.

Since the analysis of the secondary and tertiary structure of Dia helped to confirm the likelihood of the predicted coiled-coil regions within the possible domains, we decided to investigate possible 3D models of these regions and Dia as a whole. The theoretical tertiary structure of Dia (UniProt ID: Q9ZBR0), DiviVA (UniProt ID: Q9S2X4) and AtpF (UniProt ID: Q9K4D7) for *S. coelicolor* strain M145 had already been generated and modelled using the AlphaFold program. The models of Dia, DivIVA and AtpF were orientated to similar planes so the aligned regions in each protein could be compared with each other. 180° horizontally inverted versions of each protein model were also captured for analysis of other sections of the tertiary structure of the proteins. The AlphaFold models are coloured according to the predicted Local Distance Difference Test (pLDDT) score which is a

confidence metric for the per-residue LDDT-Ca score for the analysed protein. LDDT-Ca measures the percentage of correctly predicted interatomic distances and together the two metrics allow for a confidence metric for predicting locally correct structures and correct domains. The predicted protein tertiary structure is coloured according to the pLDDT score of each residue to highlight the confidence of the predicted structures throughout. The colouring ranges from: very high confidence as a dark blue colour (pLDDT > 90), high confidence as a light blue colour (pLDDT > 70 but < 90), low confidence as a yellow colour (pLDDT > 50 but < 70) and very low confidence as an orange colour (pLDDT < 50) (Figure 3-13).

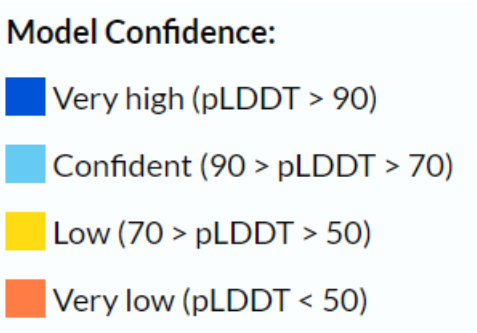


Figure 3-13. Key of the pLDDT confidence values for the AlphaFold models. Very high confidence in the modelled structural segment is identified as dark blue (confidence score higher than 90). High confidence in the modelled structural segment is identified as light blue (confidence score between 70-90). Low confidence in the modelled structural segment is identified as yellow (confidence score between 50-70). Very low confidence in the modelled structural segment is identified as orange (confidence score less than 50).

Analysis of the 3 protein models (Dia, DivIVA and AtpF) revealed a range of different coiled-coil structures within each protein. Dia AlphaFold model predicted 4 predicted coiled-coils (varying between high to very confidence values) between the residue 3-15, 30-85, 88-144 and 168-206. 2 more coiled-coil regions (with low confidence values) were seen between residues 211-239 and 367-375. Residues 239-367 of the Dia model are mostly unstructured, with 2 very low confidence structures predicted between residues between 246-255 and 335-341. No discernible shape could be seen in these possible structures. All the predicted and possible structures were connected by unfolded segments of varying size and confidence (Figure 3-14).

The DivIVA model identified 3 predicted coiled-coil regions between residue 2-15, 24-67 and 201-310. All 3 predicted structures had high to very high confidence values. Between 126-138, 172-183, 315-324 and 372-389 were several possible structures with no noticeable shape to them (all the possible structures possessed

low to very confidence). All the predicted and possible structures were connected by unfolded segments of varying size and confidence (Figure 3-15).

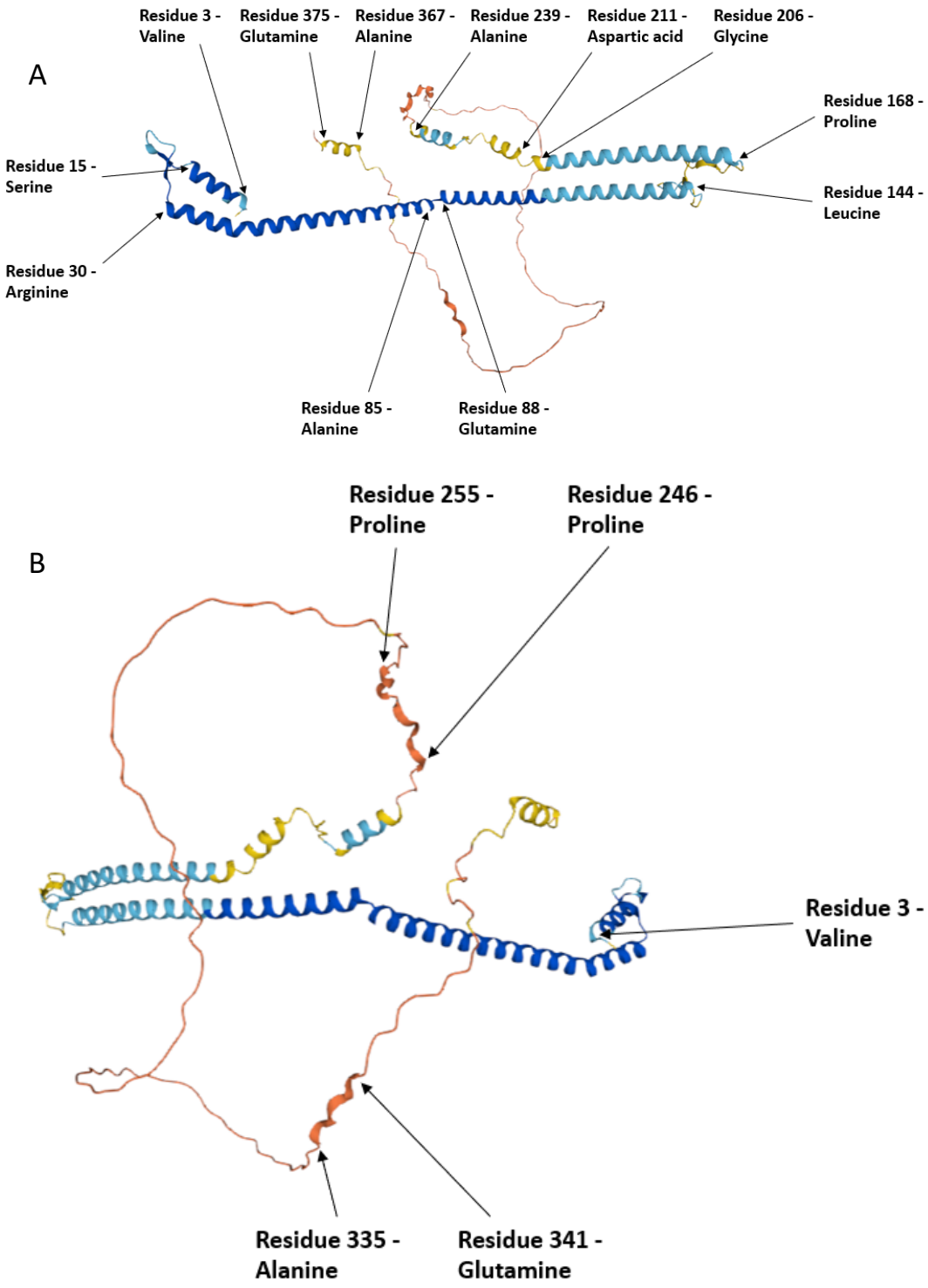


Figure 3-14. Modelled tertiary structure of Dia (UniProt ID: Q9ZBR0) from *S. coelicolor* through the AlphaFold software. (A) An orientation of the Dia model with the predicted coiled-coils in the foreground. (B) An orientation of the Dia model with the disordered regions in the foreground (the image has been roughly inverted 180° horizontally). 4 predicted coiled-coils were modelled between residue 3-15, 30-85, 88-144 and 168-206 with confidence scores (pLDDT) ranging between 70-100. 2 predicted coiled-coils were modelled between residue 211-239 and 367-375 with confidence scores ranging between 50-70. The region between residue 239-367 was modelled to form a disordered region with a confidence score of <50. Within these regions are predicted structures between 246-255 and 335-341.

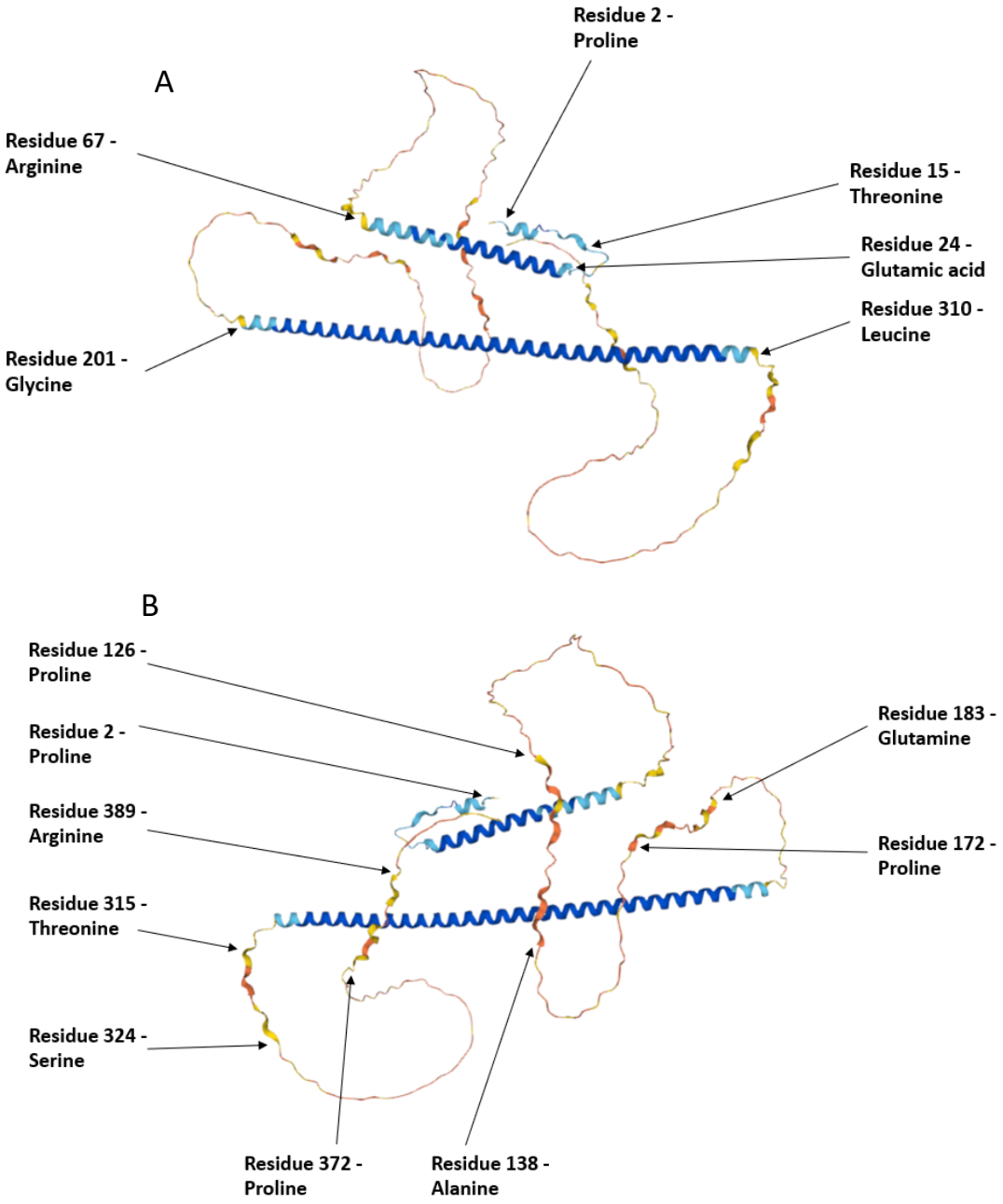


Figure 3-15. Modelled tertiary structure of DivIVA (UniProt ID: Q9S2X4) from *S. coelicolor* using the AlphaFold software. (A) An orientation of the DivIVA model with the predicted coiled-coils in the foreground. (B) An orientation of the DivIVA model with the disordered regions in the foreground (the image has been roughly inverted 180° horizontally). 3 predicted coiled-coils were modelled between residue 2-15, 24-67 and 201-310 with confidence scores (pLDDT) ranging between 70-100. The regions between residue 67-201 and 310-398 were modelled to form disordered regions mostly with a confidence score of <50. Within these regions are predicted structures between 126-138, 172-183, 315-324 and 372-389 with a confidence score of 50-70.

Finally, the AtpF model demonstrated 1 long predicted coiled-coil structure between residue 20-182 (with mostly very high confidence values). Residues 1-20 indicated an unfolded region (with low to very low confidence values). All the predicted structures were connected by unfolded segments of varying size and confidence (Figure 3-16).

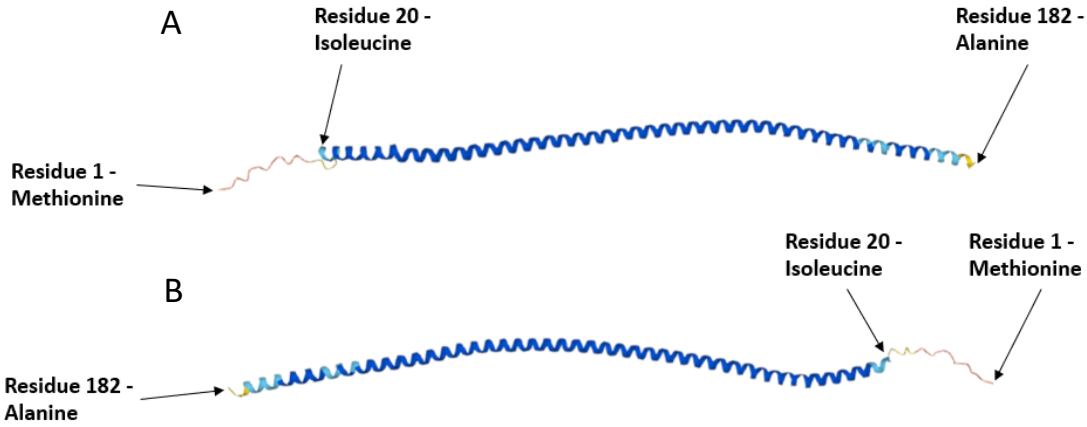


Figure 3-16. Modelled tertiary structure of AtpF (UniProt ID: Q9K4D7) from *S. coelicolor* using the AlphaFold software. (A) An orientation of the AtpF model with the predicted coiled-coils in the foreground. (B) An orientation of the AtpF model with the disordered regions in the foreground (the image has been roughly inverted 180° horizontally). A predicted coiled-coil region was modelled between residue 20-182 with a high confidence score (pLDDT) ranging between 70-100. The region between residue 1-20 was modelled to form a disordered region with a confidence score (pLDDT) of <50.

Collectively, all three protein models predicted confident coiled-coil structures which overlap across the same range of residues of each protein (roughly between residues 20-144). This overlapping range is from the coiled-coil structures of Dia (residues 30-85, 88-144 and 168-206), DivIVA (residues 24-67) and AtpF (residues 20-182). Additionally, confident coiled-coil structures of Dia (residues 3-15 and 168-206) and DiviVA (residues 2-15 and 201-310) overlapped in 2 separate ranges of residues. The regions of each protein with low and very low pLDDT scores can be unstructured in isolation.

3.3 Conclusions:

In this chapter we decided to investigate the range of bacteria that utilise Dia homologs and shed light on: the possible domains, the secondary and tertiary structure of Dia through computational analysis. A variety of different software were utilised to inspect each aspect of the primary sequence of Dia, through programs integrated in the MPI Bioinformatics server and advanced structural prediction software.

In the first part of this chapter, we decided to investigate the extent to which Dia homologs occur within bacterial species and the similarity shared between them. Additionally, we analysed the possible domains of Dia and compared their similarity to known homologs of the protein within *S. coelicolor*. We conducted 3 separate BLAST searches identifying that Dia homologs highly prevalent within species of Actinobacteria and across other phyla including Acidobacteria and Bacillota. Alignment of some of these homologs from each BLAST search, separately against Dia, revealed high levels of conservation across all phyla (though the amount varied). The Actinobacteria species, especially in the Streptomycetaceae Family, possessed the highest levels of conservation to Dia (between residues 20-216). A universal conserved region between Dia homologs across all bacterial species was seen between residues 35-144. This demonstrates that Dia may be found across a variety of different bacterial species. Moreover, the functions of Dia may be conserved through a shared ancestry and play a more universal role within bacterial species. Further analysis of the universal conserved region of Dia (residues 35-144) by a NCBI Conserved Domain search revealed multiple hits for ATP synthase subunit B domains and a DivIVA domain overlapping within the conserved region (between residues 33-131). These possible domains were reaffirmed through alignment of known proteins of these domains in *S. coelicolor* (DivIVA and AtpF) through Clustal Omega. The alignments aligned conserved sections of both proteins against Dia within the part of the predicted overlapping region of Dia (between residues 49-116). Additionally, these predicted domains were also inferred in periodic BLAST hits across all three searches. This demonstrates 2 functional but overlapping domains of Dia which are mostly contained within the universal conserved region. This suggests a possible conserved function throughout the identified homologs which may play a role in a variety of bacteria outside of the *Streptomyces* genus.

Within the second half of this chapter, we began investigating the secondary and tertiary structure of Dia through various computational analysis. In addition, we

used AlphaFold models of the tertiary structure of Dia, DivIVA and AtpF (a ATP synthase subunit B homolog) to compare any structural similarity they share. Firstly, the computational prediction of the secondary structure of Dia through the PSIRED section of the PCOILS algorithm predicted a large majority of α-helices between residues 0-110 and residues 150-205, with a predicted β-sheet predicted around residue 120. Within these same regions, the suggested tertiary structure for Dia on UniProt is of coiled-coil structures (between residues 48-68 and 90-121) In addition, the domains of Dia (DivIVA and ATP synthase subunit B) confirmed within residues 49-110 region are known to form coiled-coil structures. Both DivIVA and ATP synthase subunit B are well documented proteins consisting of large, coiled regions (Wang *et al*., 2009; Deckers-Hebestreit *et al*., 2000). The PCOILS program then speculated that the coiled-coil domains of Dia within the residues 0-110 region were likely to be a false positive which contradicted current knowledge of Dia within the UniProt database. The PCOILS program is designed to identify left-handed coiled-coils which could explain the highlighting of the coiled-coils as a false positive. Due to this contradiction, the possible coiled-coils of Dia were then analysed with the MARCOIL and DeepCoil algorithms to discern if the coiled-coils within residues 0-110 were likely to form. Both algorithms determined that coiled-coils were likely to form within the range of residues 30-100, with the DeepCoil algorithm predicting 2 separate coiled-coil regions at residues 30-50 and 90-95. The predictions for coiled-coils formation within Dia from the computational analysis overlapped with the UniProt entry predictions (residues 48-68 and 90-121) as well. Finally, having consistent predictions of the tertiary structure of Dia, we decided to reaffirm our tertiary structure predictions through analysing AlphaFold models of Dia, DivIVA and AtpF of *S. coelicolor* and to determine any of the remaining tertiary structure of Dia. The three protein models predicted coiled-coil structures overlapping between residues 20-144 (all with high confidence). Moreover, 2 coiled-coil structures of Dia and DivIVA overlapped between residues 3-15 and 168-206 (with high confidence values). The Dia model also predicted 2 more low confidence coiled-coil regions were seen between residues 211-239 and 367-375. Between residues 239-367, most the sequence was unstructured and contained 2 very low confidence structures between residues 246-255 and 335-341. This low pLDDT section of Dia can be unstructured in single monomers and may requiring other components to fold into a higher structure. Comparison of the AlphaFold models reaffirmed the possible coiled-coil structure within the conserved region of Dia (residues 35-144) and the possible conserved domains of DivIVA and ATP synthase subunit B (residues 49-116).

3.4 Future directions:

Although the predictions of the secondary and tertiary structure of Dia and its potential domains were mostly consistent with previous data and predictions; further experimental analysis of Dia is required. A suitable method for helping confirm the secondary structure of Dia would be through Circular dichroism spectroscopy to confirm the secondary structure and the folding and binding properties of Dia. Additionally, X-ray crystallography could be used to generate an accurate model of the tertiary structure of Dia to help confirm any of the predicted sections. Finally, to help confirm the potential domains of Dia, comparing the X-ray crystallography of Dia against the experimental X-ray crystallography data of DivIVA and AtpF should help confirm the predicted similarities from the computational analysis. Moreover, generating specific mutants that remove the sections of Dia that contain the possible domains and monitoring the effects within *S. coelicolor* may help to infer the functions they play within the bacteria.

Chapter 4:

Interactions Between Dia, the TIPOC and the ParAB system

4.1 Introduction:

Previously in the Kelemen lab Dia was characterised as a protein that controls the diameter of hyphae as they grow within *S. coelicolor*. In Chapter 3, we predicted possible domains within the primary structure of Dia that were capable of forming coiled-coil oligomers, which were identified to contain multiple ATP synthase subunit B domains. Intriguingly other TIPOC components, such as DivIVA, Scy and FilP are all coiled-coil proteins. By highlighting the structural properties of Dia, we are starting to understand how the unique hyphal regulatory function of Dia may occu. Since DivIVA monomers were known to self-assemble into a high order assembly whilst interacting with the TIPOC partner proteins, Scy and FilP; we hypothesized that Dia may also be able to interact with the TIPOC proteins, and we wished to investigate these potential interactions.

DivIVA is highly conserved protein found throughout a wide variety of rod-shaped bacteria usually linked to the bacterial cell division or driving the growth of the bacteria. In *B. subtilis*, DivIVA possesses a dynamic relationship where DivIVA is constantly shifting from the poles of the cell to the new division septum (Bach, Albrecht & Bramkamp, 2014). Additionally, the loss of the rod shape increases the DivIVA dynamics for quick relocalisation from the poles to new septa (Giacomelli *et al*., 2022). The role of DivIVA in *B. subtilis* is controlling cell division, by blocking FtsZ polymerisation via positioning the FtsZ inhibitor, MinD to the poles and at division sites, to block re-initiation of division at these sites. On the other hand, within Actinobacteria such as *S. coelicolor* or *S. venezuelae*, DivIVA localises around and along concave cell walls where upon filament-like multimeric complexes naturally form within the nearby cytoplasm. Although DivIVA complexes naturally accumulates near the cell wall curvature, the complexes themselves can only weakly associate with the membrane and as a result the assemblies constantly assemble and disassemble in a dynamic fashion. The establishment of the multimeric DivIVA complexes at the hyphal tip appears to stabilise the complex and allow for the initiation of cell wall synthesis (Hempel *et al*., 2008). A 22 amino acid coiled-coil domain at the beginning of DivIVA is critical for the establishment of the multimeric complexes (Wang *et al*., 2009). This, consequently, in species of *Streptomyces* allows for necessary coordination of the important cellular function, polar growth. These complexes, alongside other partner proteins (Scy and FilP)

allow for the formation of large multiprotein structure to regulate polar growth. This large structure has been called the Tip Organising Centre (TIPOC), though an alternative name used for multiprotein complex is the polarisome. Additionally, the interactions between DivIVA with TIPOC partner proteins can also help to regulate the partner proteins and their high order assemblies alongside regulating DivIVA. DivIVA-based polarisomes have been found to help FilP to localise at the hyphal tip and generate a gradient for the self-assembly of FilP filaments. In turn though, FilP appears to be involved in regulating the structure of the DivIVA-based polarisomes and delimit the amount of stable polarisomes that can form (Fröjd & Flärdh, 2019). Furthermore, the DivIVA-based polarisomes have been shown to stimulate the oligomerisation of FilP into larger stress-bearing filaments at the hyphal tip (Fuchino *et al*., 2013). The other TIPOC partner protein: Scy was capable regulating the quantity of stable DivIVA-based polarisomes. Scy appears to limit the amount of polarisomes through organising the dispersed DivIVA polarisomes into limited foci called polarity centres (Holmes *et al*., 2013).

Alongside DivIVA and its partner proteins, which influence polar growth within Actinobacteria, the TIPOC proteins play an important role in other important cellular processes, such as chromosome organisation. ParA and ParB were vital for the regulation and coordination of the chromosome. ParA forms helical filaments from the hyphal tip which extends back to the chromosome whereupon it binds to ParB to initiate the formation of ParB complexes (Jakimowicz *et al*., 2007). With the activation of ParB, a complex was formed around a several kilobase long chromosomal region which, with ParA, regulates septation in hyphae (Donczew *et al*., 2016). Although interactions between DivIVA and ParA or ParB have not been identified *in vivo* within *Streptomyces*, the interactions have been identified *in vivo* within other bacterial species. One such species, the actinomycete *Mycobacterium smegmatis*, DivIVA was found to directly compete with the nucleoid to bind to ParA, with the DivIVA-ParA interaction being vital for modulating the change between cell elongation to chromosome segregation (Pióro *et al*., 2018). Alternatively in a synthetic *E. coli* system, DivIVA was shown to directly interact with the ParB-*parS* chromosome centromere and tether the chromosome to the pole of the cell. This was demonstrated for a range of different Actinobacterial proteins including DivIVA and ParB from *M. tuberculosis* and *S. coelicolor* (Donovan *et al*., 2012). Intriguingly, another TIPOC component, Scy was shown to directly interact with ParA (Ditkowski *et al*, 2013) and control the switch from active growth to cell division. Moreover, recent work in the Kelemen lab has established a direct link between Scy and SepF,

the latter protein being a key component of the divisome, which also includes FtsZ to initiate septum formation (Cassettari, Unpublished)

Based on the fact that, the *dia* knockout mutant has altered hyphal diameter suggests that Dia could be a novel component of the TIPOC. Therefore, we aimed to test interactions between Dia and the established TIPOC components, DivIVA, Scy and FilP to help determine if Dia is a novel TIPOC component and consequently how Dia interacts within the *Streptomyces* polar growth mechanism. Moreover, we aimed to test interactions between Dia and other proteins that were previously shown to interact with the TIPOC. These included ParA, ParB and a novel ParA homologue, ParH, which has been investigated in Dr Kelemen's lab. ParH has been shown to directly interact with both Scy and ParB (Gillespie, Unpublished; Alanazi, Unpublished).

In this Chapter we present a screening for interacting partners for Dia using the Bacterial adenylate Cyclase-Based Two Hybrid (BACTH) assay (Karimova, Ullmann & Ladant, 2000), testing for protein-protein interactions *in vivo* in the heterologous host *E. coli*.

4.2 Results:

4.2.1 The BACTH system

The BACTH assay utilises adenylate cyclase within a heterologous *E. coli* host to test the potential protein-protein interaction (Karimova, Ullmann & Ladant, 2000). The adenylate cyclase used within this system, from *Bordetella pertussis* (*B. pertussis*), was composed of two non-functional domains (T18 and T25) which must interact with each other correctly to generate functional adenylate cyclase (Figure 4.1). Adenylate cyclase was responsible for the production of cyclic AMP (cAMP) from AMP. cAMP was capable of binding to the catabolite activator protein (CAP) which the BACTH system uses to regulate the reporter genes. By fusing the two domains to separate proteins of interest and co-expressing them, the proteins of interest capable of interacting and allowing for T18 and T25 domains to form a functional adenylate cyclase that will activate the expression of *lacZ* through the generation of cAMP and formation of the cAMP-CAP complex (Figure 4-1). The reporter gene for the BACTH system was the *lacZ* gene, which was responsible for the generation of β-galactosidase. When produced, β-galactosidase was able to hydrolise X-gal to release an insoluble blue dye to act as a visual marker within *E. coli* colonies for any successful protein-protein interactions.

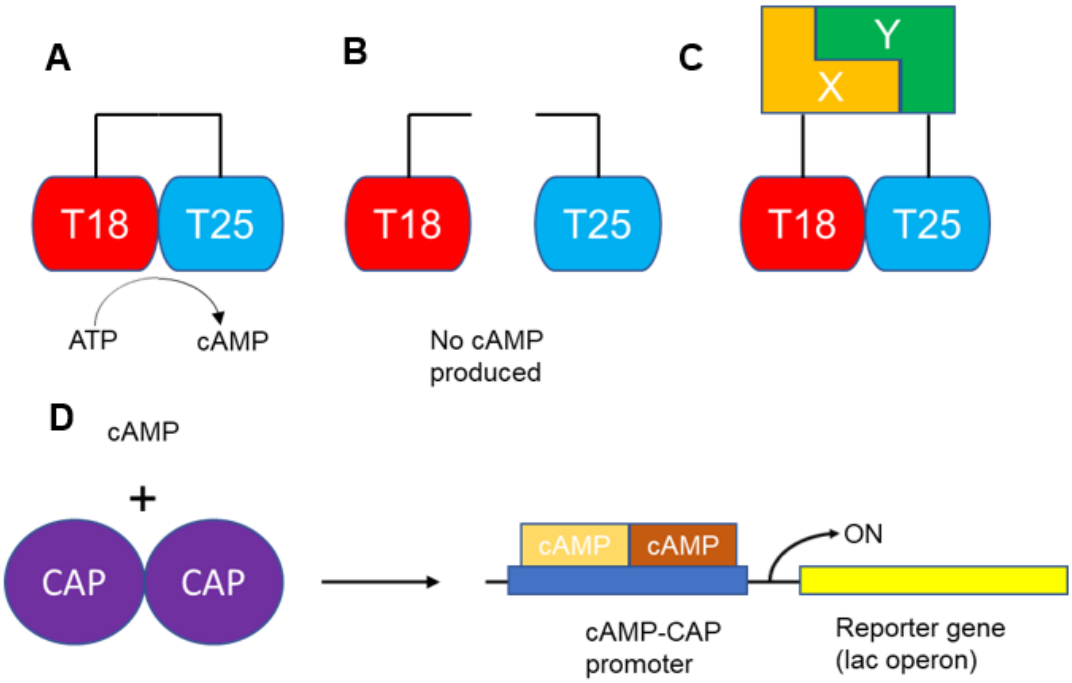


Figure 4-1. The fundamentals of the BACTH assay. (A) Adenylate cyclase from *B. pertussis* consists of two domains, T18 and T25, which allows the production of cAMP from ATP. (B) When the two domains were separated, a functional adenylate cyclase can't form, and the cells can no longer produce cAMP. (C) Translational fusions of the T18 and T25 domains to proteins that interact leads to the reconstitution of adenylate cyclase and production of cAMP. (D) Production of cAMP was monitored using *lacZ*, as a reporter gene which in the presence of X-gal will allow blue/white screening .

To generate the fused proteins, four different vectors were used: pUT18c, pUT18, pKT25 and pKNT25 (BACTH vectors, Figure 4-2). Each vector contains either the T18 or T25 domain with a Multiple cloning site (MCS) upstream or downstream of said domain. This allows for the generation of the T18 or T25 domains at either the C-terminus or the N-terminus of the protein to be tested. adenylate. This allowed us to test several combinations for each pairwise interaction. The T18 encoding vector set were: pUT18c (T18 tag was at the C-terminus of the protein) and pUT18 (T18 tag was at the N-terminus of the protein); and the T25 encoding vector sets were: pKT25 (T25 tag was at the C-terminus of the protein) and pKNT25 (the T25 tag was at the N-terminus of the protein). Furthermore, the T18 and T25 vector sets carry a different resistance marker and origin of replication allowing for selection and co-transformation of different combinations of vectors in the adenylate cyclase deficient *E. coli* BTH101. In our design, the genes of interests were inserted into the MCS via the use of *XbaI* and

EcoRI to generate a functional, in-frame fusion to the desired adenylate cyclase domain.

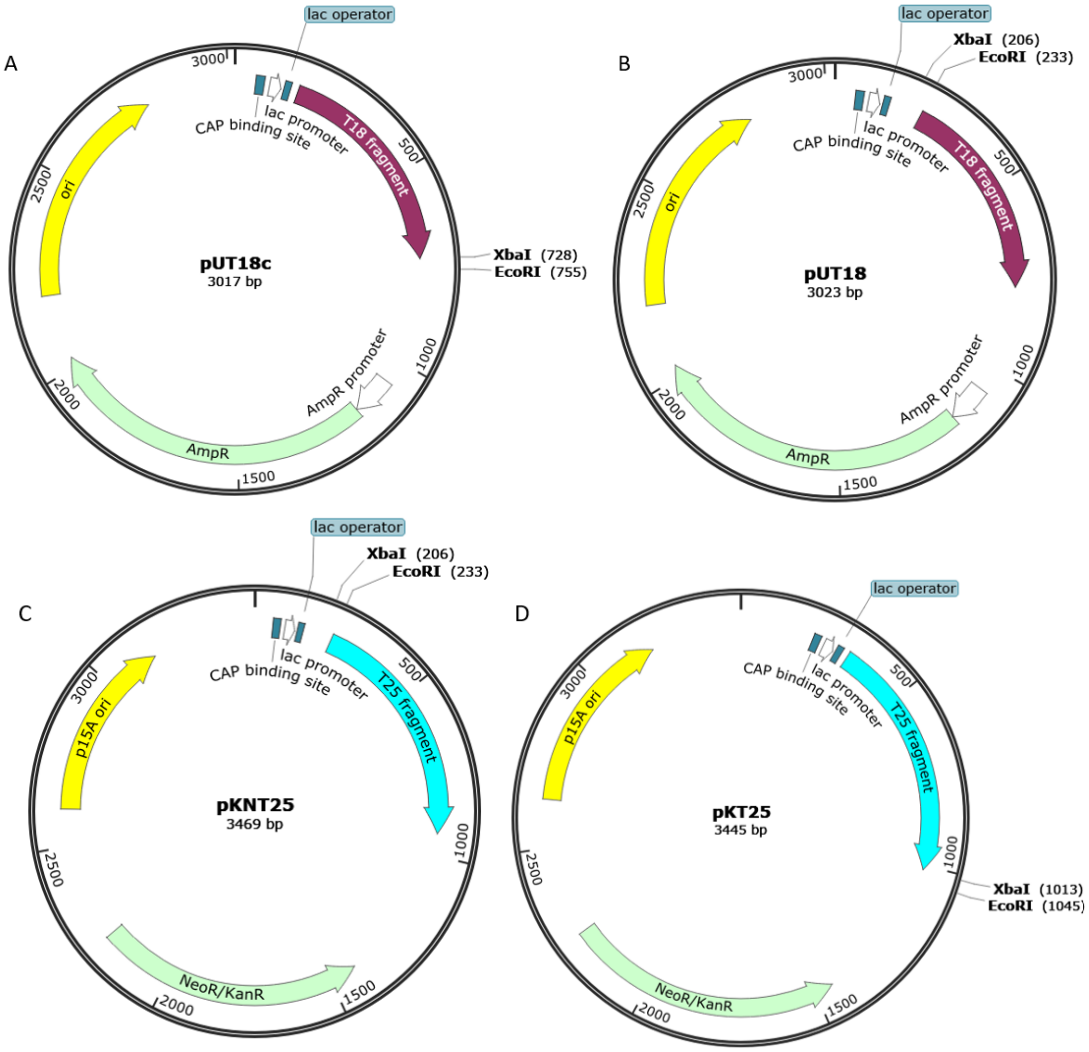


Figure 4-2. The vector maps used to produce plasmid constructs used in BACTH assays; each plasmid (pUT18c, pUT18, pKT25 and pKNT25) contains a domain of adenylate cyclase from *B. pertussis*. (A) The BACTH vector pUT18c contains the T18 domain which proceeds the MCS containing cloning restriction sites *XbaI* and *EcoRI*. The T18 domain was driven by the *lac* promoter. It also contains an origin of replication from pUC18 and an Ampicillin resistance gene. (B) The BACTH vector pUT18 contains the gene for the T18 domain which precedes the MCS containing cloning restriction sites *XbaI* and *EcoRI*. The gene for the T18 domain was driven by the *lac* promoter. It also contains an origin of replication from pUC18 and an ampicillin resistance gene. (C) The BACTH vector pKT25 contains the T25 domain which proceeds the MCS containing cloning restriction sites *XbaI* and *EcoRI*. The T25 domain was driven by the *lac* promoter. It also contains a kanamycin resistance gene and the *E. coli* p15A origin of replication which was distinct and compatible with the origin of replication in pUT18c allowing co-transformation. (D) The BACTH vector pKNT25 contains the the gene for theT25 domain which precedes the MCS containing cloning restriction sites *XbaI* and *EcoRI*. The gene for the T25 domain was driven by the *lac* promoter. It also contains a kanamycin resistance gene and the *E. coli* p15A origin of replication which was distinct and compatible with the origin of replication in pUT18 allowing co-transformation. Vector maps were generated using SnapGene.

4.2.2 Analysis of possible self-interactions between dia oligomers and the interactions between Dia and components of the TIPOC

To determine if *dia* monomers were capable of self-assembly, the BACTH assay was used with *dia* being cloned into 4 different BACTH vectors (pUT18c, pUT18, pKT25 and pKNT25). The vectors expressed Dia fused to either the C-terminus or N-terminus of the adenylate cyclase domains: T18 and T25. This allowed for fusions which test interactions with Dia in all possible orientations. To create functional fusions with the adenylate cyclase domains, 2 different variants of *dia* were used dependent on the terminus: a Wildtype copy (*dia*) and a variant lacking a stop codon (called *diaUT*) (Figure 4-4). The *dia* variant was fused to the C-terminus containing vectors (pUT18c and pKT25) and the *diaUT* variant was fused to the N-terminus containing vectors (pUT18 and pKNT25). Collectively, this resulted in 4 different constructs being created: pUT18c-Dia, pUT18-Dia, pKT25-Dia and pKNT25-Dia (Figure 4-3). To generate the *dia* constructs, the 2 variants of *dia* were amplified using 2 similar sets of primers and the 7A1 cosmid as a template. *dia* was amplified with the primers '5569 XbaNde FRW' and '5569 Eco REV', whereas *diaUT* was amplified with the primers '5569 XbaNde FRW' and '5569 EcoUT REV'. The primers allowed for the *dia* variants to be restricted and ligated into their respective vectors via *XbaI* and *EcoRI*. Successful cloning of the *dia* variants was confirmed using Colony PCR, followed by sequencing of each recombinant construct to confirm they were free of mutations (Figure 4-4).

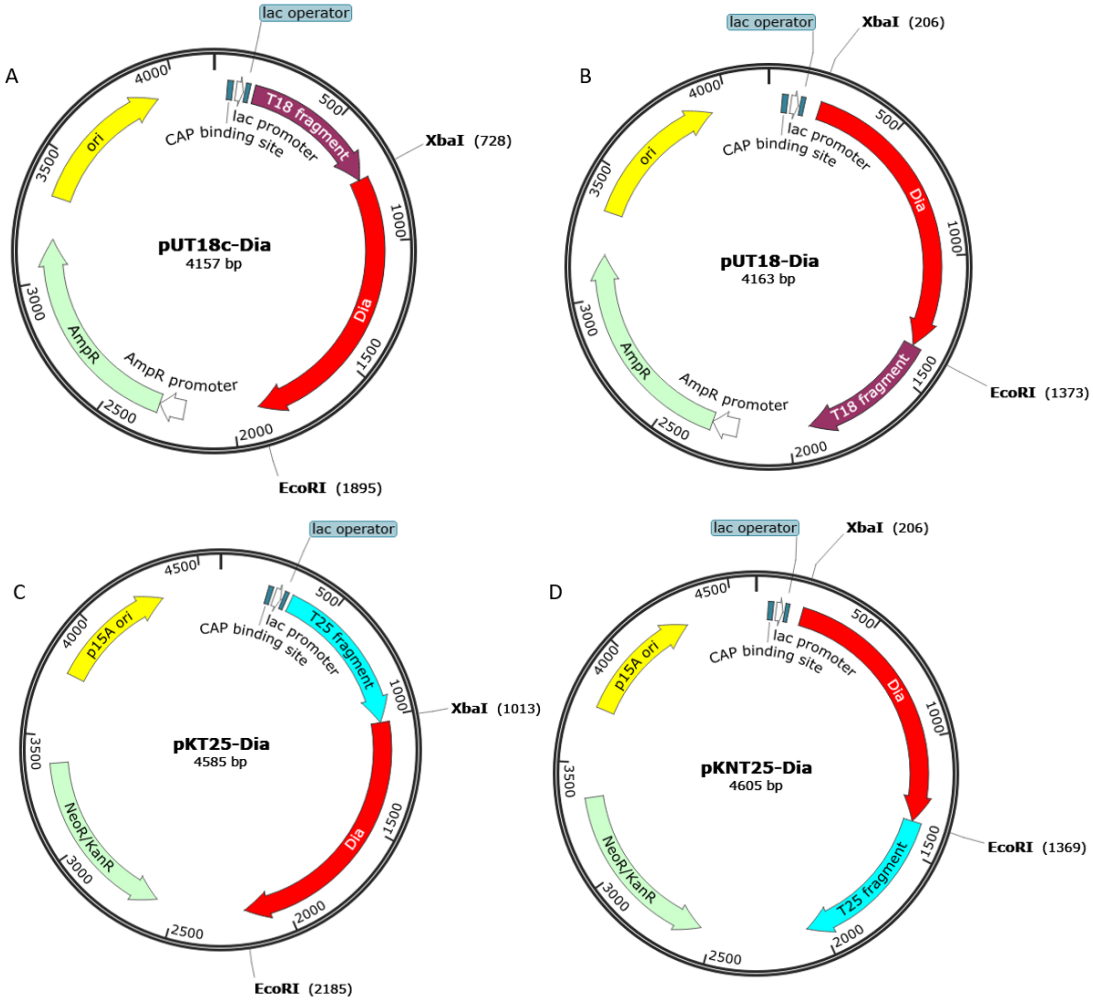


Figure 4-3. The BACTH Dia plasmid maps of pUT18c-Dia, pUT18-Dia, pKT25-Dia and pKNT25-Dia. (A) pUT18c-Dia was an altered copy of pUT18c with the *dia* gene cloned using the restriction sites *XbaI* and *EcoRI*. (B) pUT18-Dia was an altered copy of pUT18 with the *diaUT* gene cloned using the restriction sites *XbaI* and *EcoRI*. (C) pKT25-Dia was an altered copy of pKT25 with the *dia* gene cloned using the restriction sites *XbaI* and *EcoRI*. (D) pKNT25-Dia was an altered copy of pKNT25 with the *diaUT* gene cloned using the restriction sites *XbaI* and *EcoRI*. The *dia* gene was amplified from the 7A1 cosmid using the primers 5569 XbaNde FRW and 5569 Eco REV. The *diaUT* gene was amplified from the 7A1 cosmid using the primers 5569 XbaNde FRW and 5569 EcoUT REV.

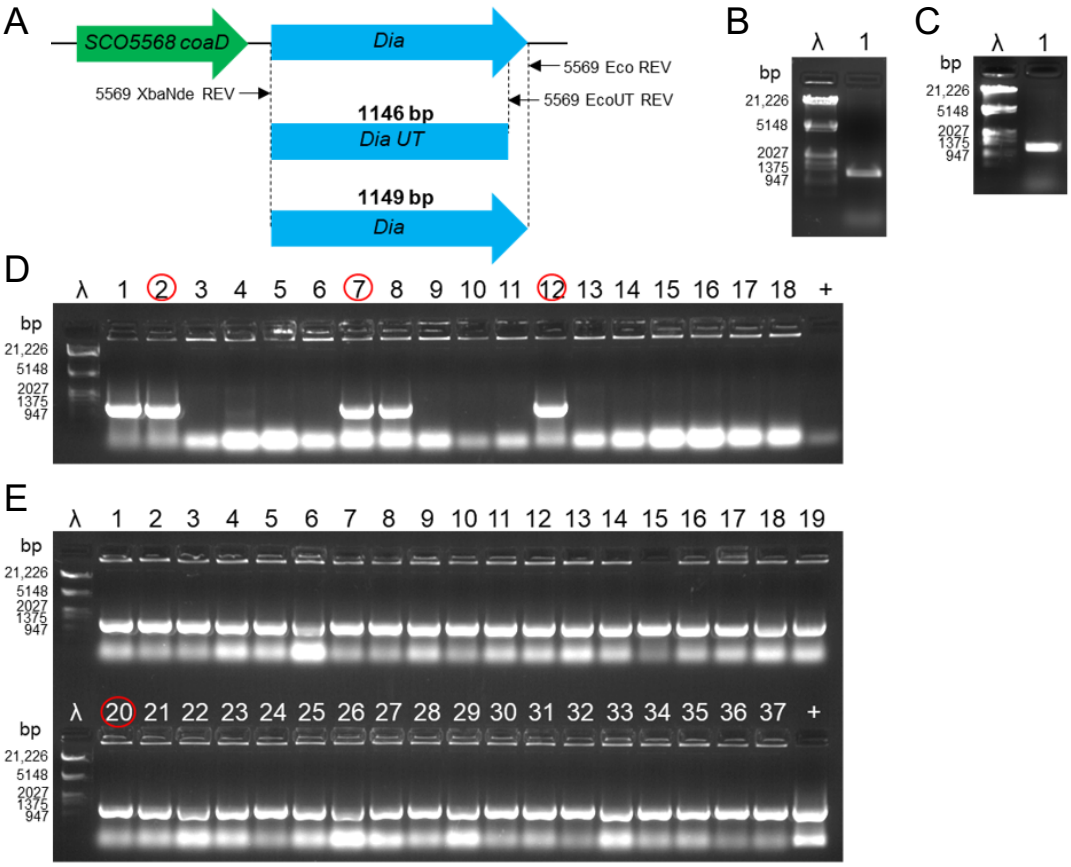


Figure 4-4. Generation of pUT18c-Dia, pUT18-Dia, pKT25-Dia and pKNT25-Dia. (A) The predicted PCR amplifications of the *diaUT* and *dia* fragments using the 7A1 template and two primer pairs to generate the different fragments: 5569 XbaNde FRW and 5569 EcoUT REV (to generate *diaUT*); 5569XbaNde FRW 5569 Eco REV (to generate *dia*). (B) Confirmation of the predicted PCR amplified *diaUT* fragment via 0.7% Agarose gel electrophoresis. Lane λ: Lambda DNA *XbaI* + *EcoRI* ladder. Lane 1: *diaUT* PCR. (C) Confirmation of the predicted PCR amplified *dia* fragment via 0.7% Agarose gel electrophoresis. Lane λ: Lambda DNA *XbaI* + *EcoRI* ladder. Lane 1: *dia* PCR. (D) Confirmation of the successful DH5α pUT18c-Dia, DH5α pKT25-Dia and DH5α pUT18-Dia transformants using 5569 XbaNde FRW and 5569 EcoUT REV primers via colony PCR. The colony PCR was analysed via 0.7% Agarose gel electrophoresis. Lane λ: Lambda DNA *XbaI* + *EcoRI* ladder. + contains the *dia* PCR product (positive control). Lane 1-4: pUT18c-Dia colony PCRs; Lane 5-9: pKT25-Dia colony PCRs; Lane 10-13: pUT18-Dia colony PCRs; Lane 14-18: pKNT25-Dia colony PCRs. The pUT18c-Dia, pKT25-Dia and pUT18-Dia colony PCRs identified 5 positive colonies (2, 2 and 1 positive colonies respectively). No positive colonies were identified for the pKNT25-Dia colony PCRs. (E) Confirmation of the successful DH5α DH5α pKNT25-Dia transformants using 5569 XbaNde FRW and 5569 EcoUT REV primers via colony PCR. The colony PCR was analysed via 0.7% Agarose gel electrophoresis. Lane λ: Lambda DNA *XbaI* + *EcoRI* ladder. + contains the *dia* PCR product (positive control). Lane 1-37: pKNT25-Dia colony PCRs. The pKNT25-Dia colony PCRs confirmed 37 positive colonies. pUT18c-Dia #2, pKT25-Dia #7, pUT18-Dia #12 and pKNT25-Dia #20 were selected for use in the BTH combinations (red circle).

To test the possible interaction, the different combinations of pUT18c-Dia/pUT18-Dia and pKT25-Dia/pKNT25-Dia (8 in total) were co-transformed into *E. coli* BTH101; which were deficient in adenylate cyclase. pUT18c/pUT18 were produced with a high copy number whereas pKT25/pKNT25 were produced with a low copy number, which results a higher expression of fusions from the pUT18c/pUT18 constructs. This difference in expression levels may affect the interaction of the fusions, in addition to the different positions of the tags. Afterwards, 3 colonies from each transformation alongside 3 colonies of each control (positive and negative) were selected and streaked onto LB agar plates containing the X-gal and incubated at 30°C for 48 hours to allow the colours of blue/white screening to be as distinct as possible (Figure 4-5). All combinations testing Dia-Dia interaction turned blue with a comparable intensity to the positive control. This demonstrates that Dia was capable of self-interaction and might be able to form higher order assemblies. Additionally, the expression levels had no effect on the interactions between Dia monomers.

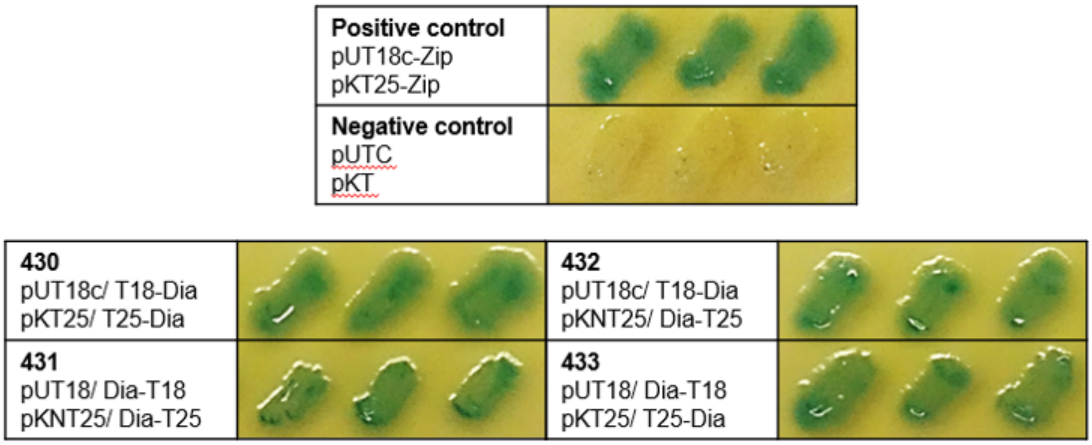


Figure 4-5. BACTH assay testing the interactions between different orientations of *dia*, alongside the controls. Blue BTH colonies indicated a positive interaction between the proteins of the tested genes, with colourless colonies indicating no interaction. The constructs tested and the orientations of their fused adenylate cyclase subunit were detailed next to their corresponding result. 3 colonies of each co-transformation were streaked onto LB agar plates containing ampicillin, kanamycin, IPTG and X-gal. All streaks were analysed after 48 hours of growth at 30°C after inducing expression of the BACTH system.

After inspecting the self-interactions between Dia monomers, we decided to investigate the potential interactions between Dia and components of the TIPOC. As previously demonstrated, Dia-EGFP has been found to localise in a similar pattern to DivIVA-EGFP throughout the hyphae of *S. coelicolor* (Hutchinson, Unpublished) and the rudimentary analysis of the translated Dia sequence indicated the potential for a coiled-coil domain within the first 100 aa of the sequence. This indicates that Dia appears to localise in a pattern similar to a key component of the TIPOC and possess a coiled-coil domain, which are found within the TIPOC components: DivIVA, Scy and FilP. Due to these highlighted similarities of Dia to the TIPOC components, we decided to test for possible interactions between Dia and DivIVA, Scy and FilP.

To test these possible interactions, the BACTH assay was utilised. The constructs used in the BACTH assay for *divIVA, scy* and *filP* were generated previously in the Kelemen lab. To test the possible interactions, the different combinations of Dia constructs (Figure 4-3) were co-transformed with previously generated BACTH constructs for each gene of interest: *divIVA*, *scy*, or *filP* (Figure 4-6 a and b) into *E. coli* BTH101. There were 4 combinations generated for the Scy-Dia and DivIVA-DIA interactions and 8 combinations for the FilP-Dia interaction. Afterwards, 3 colonies from co-transformation alongside 3 colonies of each control (positive and negative) were selected and streaked onto LB agar plates containing the appropriate compounds and incubated at 30°C for 48 hours (Figure 4-7).

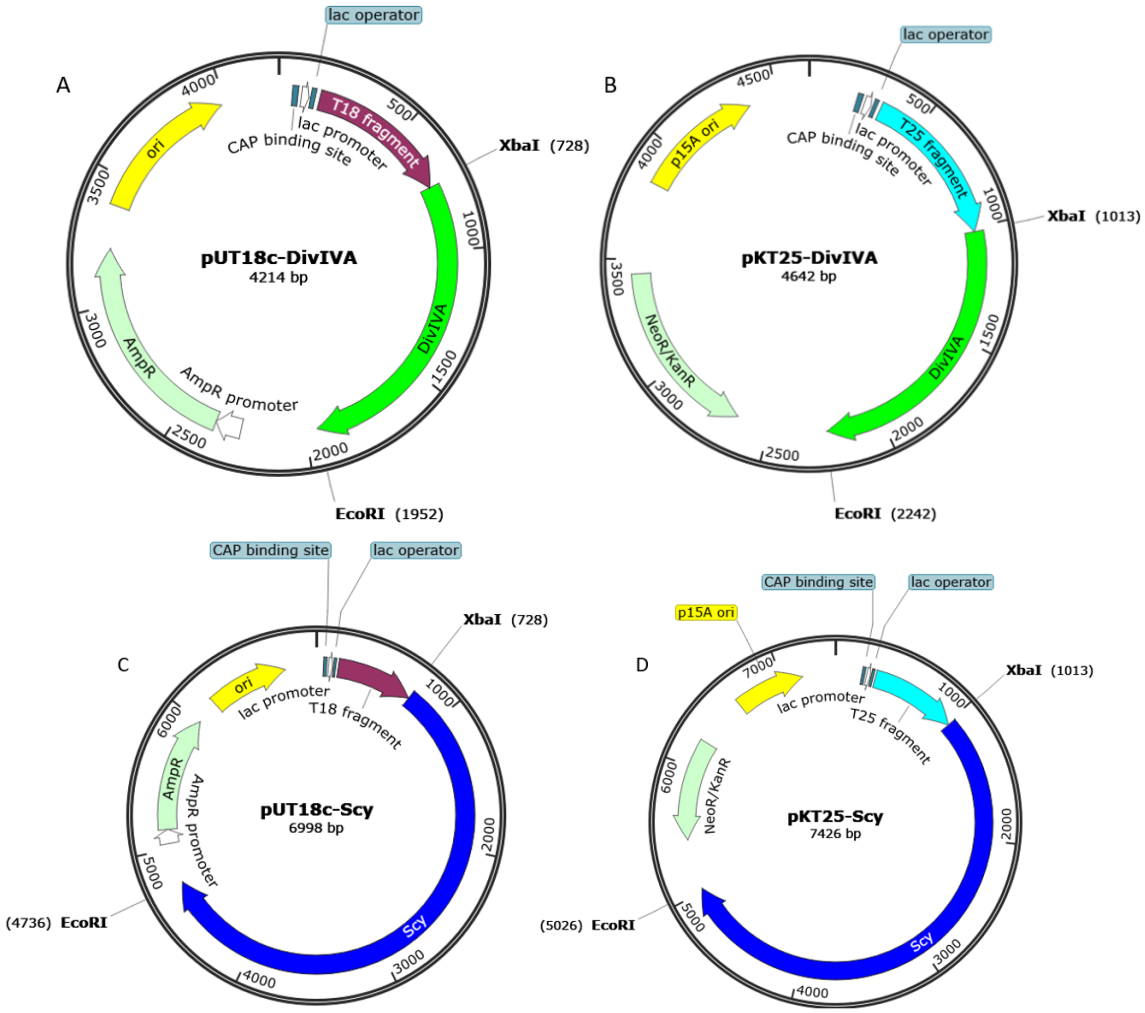


Figure 4-6a. The BACTH plasmid maps carrying TIPOC components for DivIVA (pUT18c-DivIVA and pKT25-DivIVA) and Scy (pUT18c-Scy and pKT25-Scy). (A) pUT18c-DivIVA carries the *divIVA* gene cloned using the restriction sites *XbaI* and *EcoRI*. (B) pKT25-DivIVA carries the *divIVA* gene cloned using the restriction sites *XbaI* and *EcoRI*. These constructs produce DivIVA fused to the C-terminus of the T18 and T25 domains. (C) pUT18c-Scy carries the *scy* gene cloned using the restriction sites *XbaI* and *EcoRI*. (D) pKT25-Scy carries the *scy* gene cloned using the restriction sites *XbaI* and *EcoRI*. The DivIVA and Scy constructs produce their respective genes fused to the C-terminus of the T18 and T25 domains.

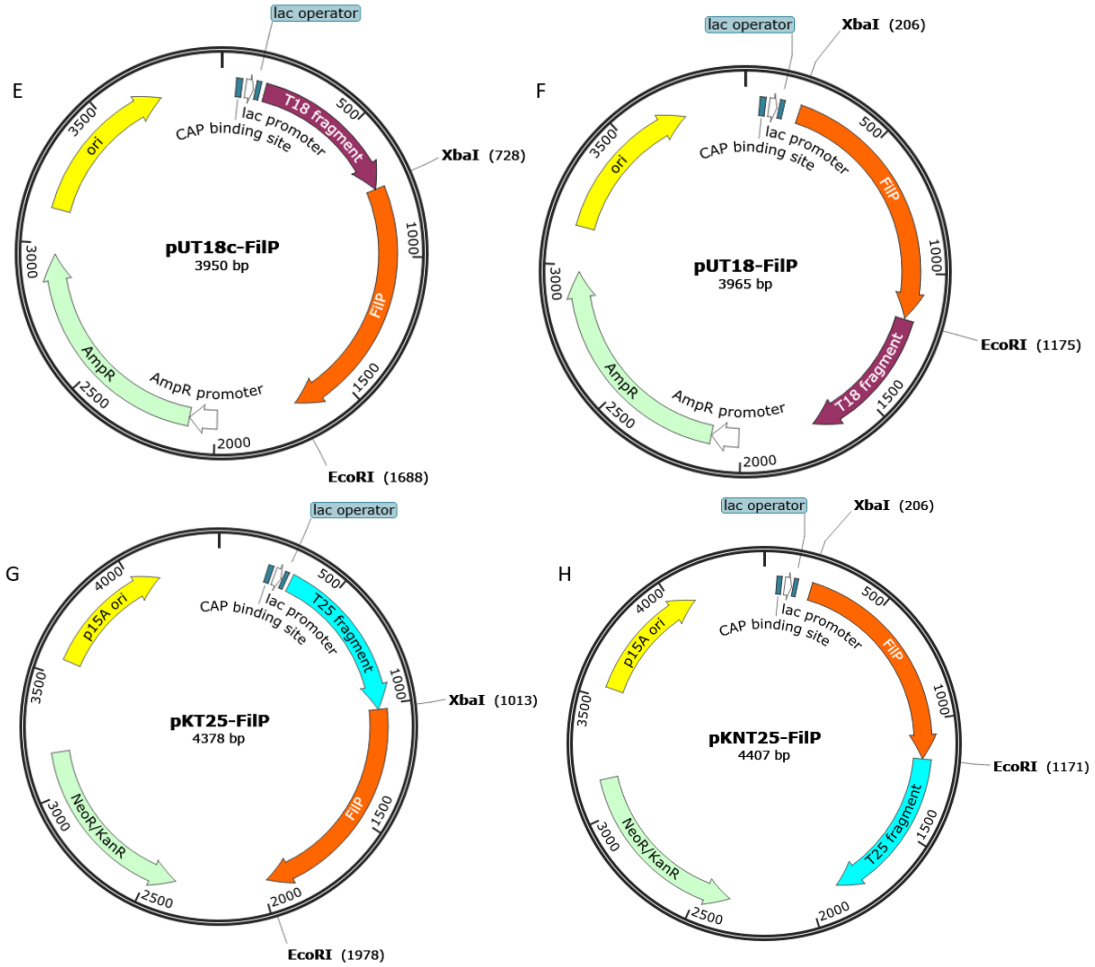


Figure 4-6b. The BACTH plasmid maps carrying TIPOC components for FilP (pUT18c-FilP, pUT18-FilP, pKT25-FilP and pKNT25-FilP). (E) pUT18c-FilP carries the *filP* gene cloned using the restriction sites *XbaI* and *EcoRI*. (F) pUT18-FilP carries the *filP* gene cloned using the restriction sites *XbaI* and *EcoRI*. (G) pKT25-FilP carries the *filP* gene cloned using the restriction sites *XbaI* and *EcoRI*. (H) pKNT25-FilP carries the *filP* gene cloned using the restriction sites *XbaI* and *EcoRI*. The pUT18c-FilP and pKT25-FilP constructs produce their respective genes fused to the C-terminus of the T18 and T25 domains. The remaining FilP constructs produce FilP fused to the N-terminus of the T18 and T25 domains.

For the combinations testing DivIVA-Dia interactions, all combinations generated a weak; interspersed blue signal when compared to the negative control though it wasn't as intense as the positive control. This appears to demonstrate that Dia might be able to interact with DivIVA but further, independent tests are needed to be carried out to confirm this interaction. On the other hand, the interactions between Dia and Scy generated a blue signal as intense as the positive control. This suggests that Scy and Dia were capable of strongly interacting despite variation in expression in the different constructs. The fact that strong interaction was detected independent of where the tag was (N- or C-terminus) might suggest that the interaction was part of a complex higher order assembly, where both the N and C-termini of proteins were closely located.

The combinations testing FilP-Dia interactions resulted in a collection of different signals. There were two orientations/combinations where a clear positive signal was detected suggesting that FilP was capable of interacting with Dia. The expression levels and the orientation of the tags in the different combinations may have affected the interaction between Dia and FilP. Interestingly, clear positive signals were only seen in the combinations (442 and 444) where the T18 and T25 tags were both at the N-terminus of Dia and FilP, suggesting that the N-terminus of both proteins may be required for direct interaction.

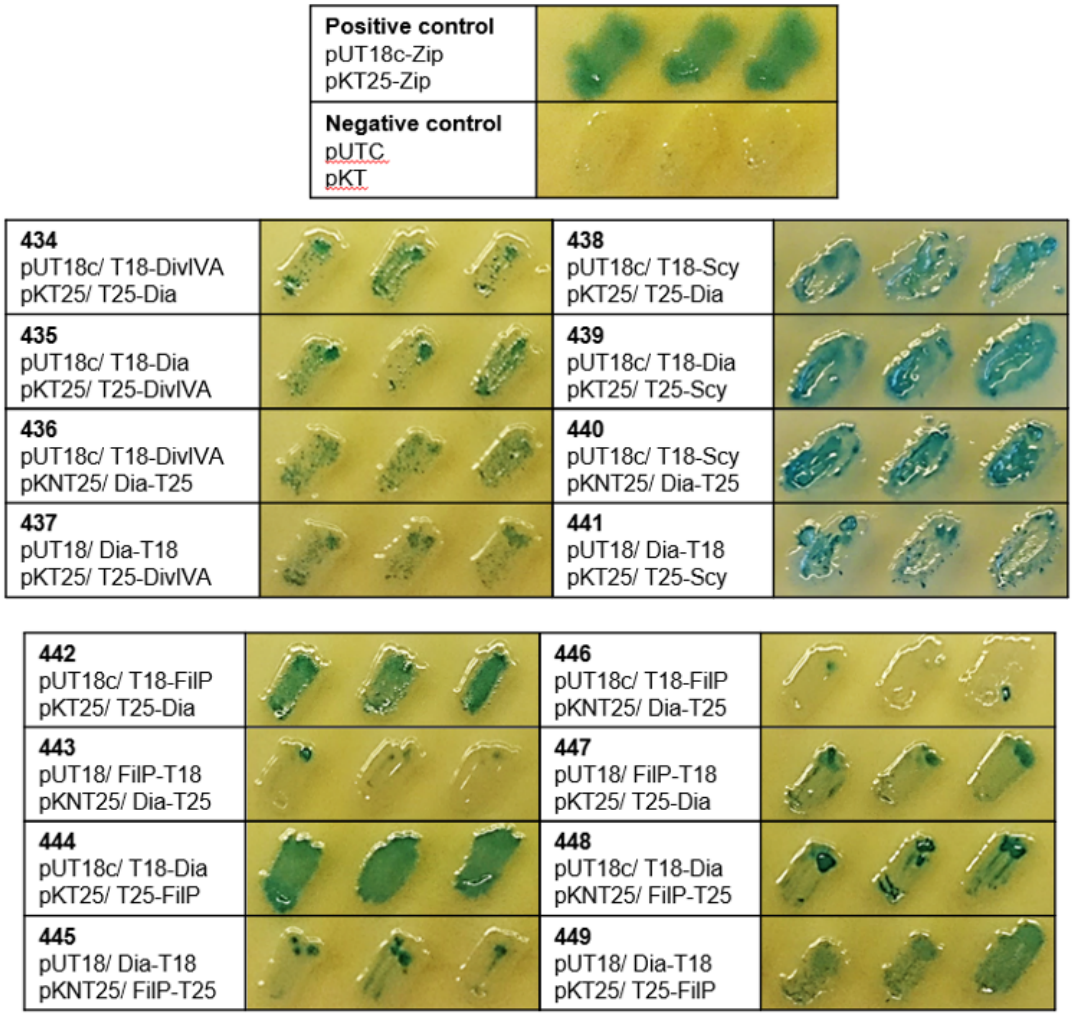


Figure 4-7. BACTH assay testing the interactions between *dia* and *divIVA*, *scy* or *filP*, alongside the controls. Blue BTH colonies indicated a positive interaction between the proteins of the tested genes, with colourless colonies indicating no interaction. The constructs tested and the orientations of their fused adenylate cyclase subunit were detailed next to their corresponding result. 3 colonies of each co-transformation were streaked onto LB agar plates containing ampicillin, kanamycin, IPTG and X-gal. All streaks were analysed after 48 hours of growth at 30°C after inducing expression of the BACTH system.

Collectively, this set of tested interactions has shown Dia was capable of interacting with a range of TIPOC components and suggests that Dia might be a novel component of the TIPOC, playing an active role in growth of *S. coelicolor*.

4.2.3 Interactions between Dia and components of the ParAB system

After determining the Dia can interact directly with the TIPOC, we decided to investigate the potential interactions between Dia and components of the ParAB system. As previously demonstrated, Scy and ParA have been found to not only directly interact with each other, but also maintain a dynamic interplay between each other resulting in the regulation of Scy higher order structures and ParA polymerisation. Furthermore, in related Actinobacteria, it has been shown that DivIVA can directly interact with ParB with in *C. glutamicum*, with ParB responsible for mobilising DivIVA. Due to the capacities that DivIVA and Scy possess to interact with the ParAB system in Actinobacteria, we began testing the possible interactions between Dia and ParA, ParB and ParH (a homologue of ParA).

The constructs used in the BACTH assay for *parA* and *parB* were previously generated in the Kelemen lab. To test the possible interactions, the different combinations of Dia-ParA and Dia-ParB were tested (Figure 4-8a, 4-8b, 4-9).

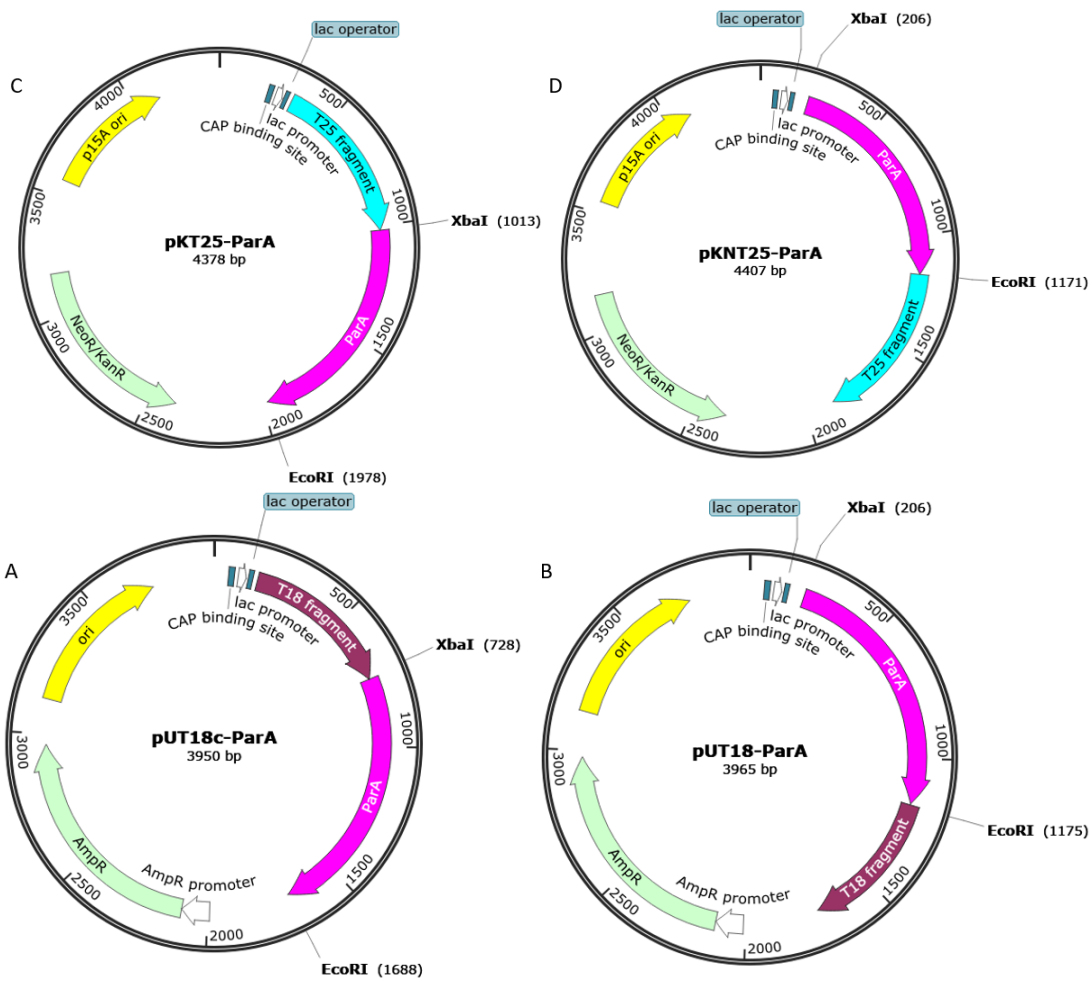


Figure 4-8a. The BACTH plasmid maps carrying ParAB components for ParA (pUT18c-ParA, pUT18-ParA, pKT25-ParA and pKNT25-ParA). (A) pUT18c-ParA carries the *parA* gene cloned using the restriction sites *XbaI* and *EcoRI*. (B) pUT18-ParA carries the *parA* gene cloned using the restriction sites *XbaI* and *EcoRI*. (C) pKT25-ParA carries the *parA* gene cloned using the restriction sites *XbaI* and *EcoRI*. (D) pKNT25-ParA carries the *parA* gene cloned using the restriction sites *XbaI* and *EcoRI*. The pUT18c-ParA and pKT25-ParA constructs produce their respective genes fused to the C-terminus of the T18 and T25 domains. The remaining ParA constructs produce ParA fused to the N-terminus of the T18 and T25 domains.

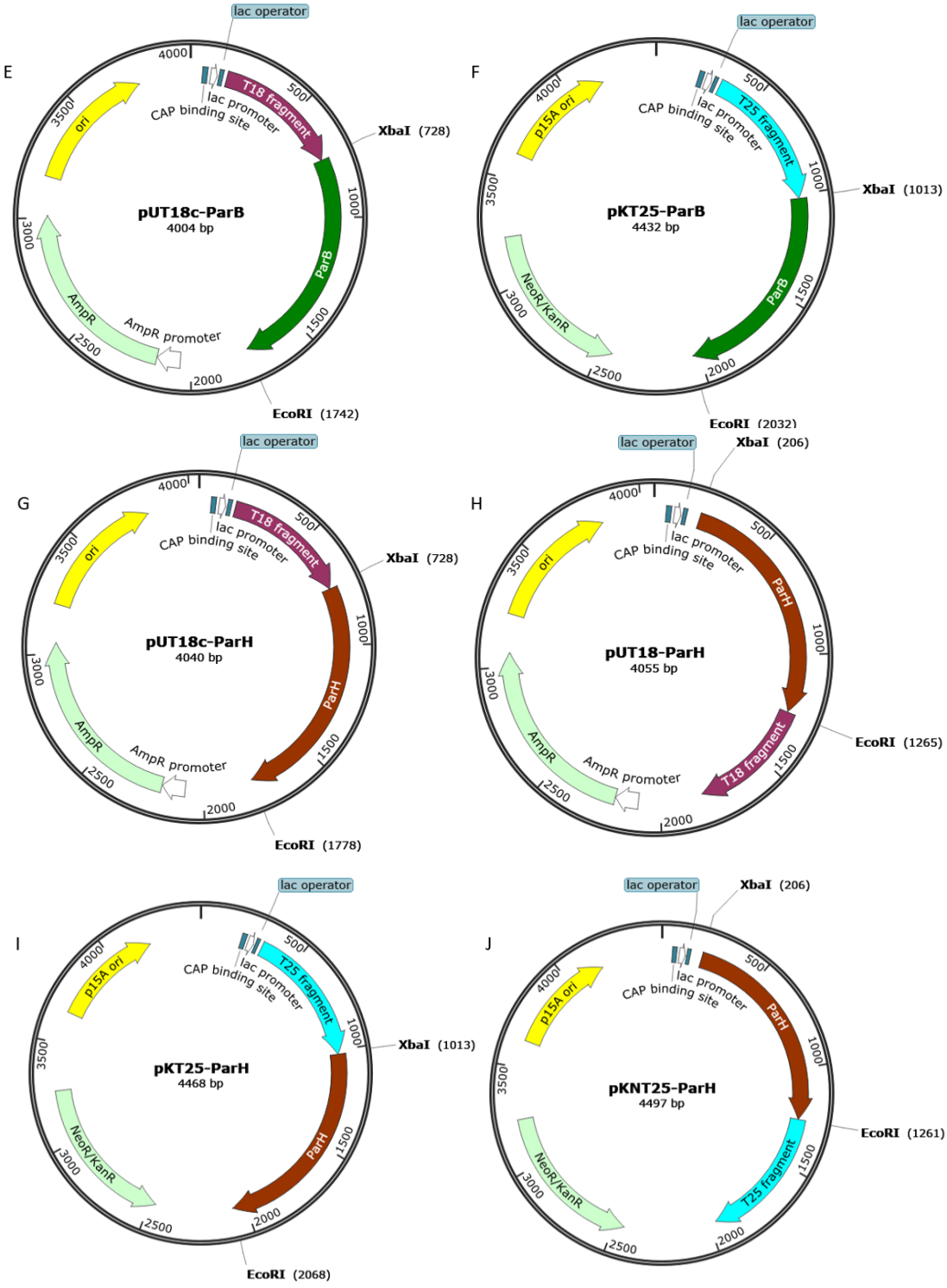


Figure 4-8b. The BACTH plasmid maps carrying ParAB components for ParB (pUT18c-ParB and pKT25-ParB) and ParH (pUT18c-ParH, pUT18-ParH, pKT25-ParH and pKNT25-ParH). (E) pUT18c-ParB carries the *parB* gene cloned using the restriction sites *XbaI* and *EcoRI*. (F) pKT25-*ParB* carries the *parB* gene cloned using the restriction sites *XbaI* and *EcoRI*. (G) pUT18c-ParH carries the *parH* gene cloned using the restriction sites *XbaI* and *EcoRI*. (H) pUT18-ParH carries the *parH* gene cloned using the restriction sites *XbaI* and *EcoRI*. (I) pKT25-ParH carries the *parH* gene cloned using the restriction sites *XbaI* and *EcoRI*. (J) pKNT25-ParH carries the *parH* gene cloned using the restriction sites *XbaI* and *EcoRI*. The ParB, pUT18c-ParH and pKT25-ParH constructs produce their respective genes fused to the C-terminus of the T18 and T25 domains. The remaining ParH constructs produce ParH fused to the N-terminus of the T18 and T25 domains.

The interactions between Dia and ParA generated no blue signal when compared to the positive control and was identical to the negative control. This demonstrates that ParA and Dia do not directly interact under these conditions. Contrary to ParA, when testing Dia-ParB interactions, all combinations generated some blue signals when compared to the negative control though most of the signals weren't as intense as the positive control. In the strongest combination, the T18 and T25 tags were at the C-terminus of Dia and ParB, which might indicate that Dia and ParB interact at their C-terminal ends.

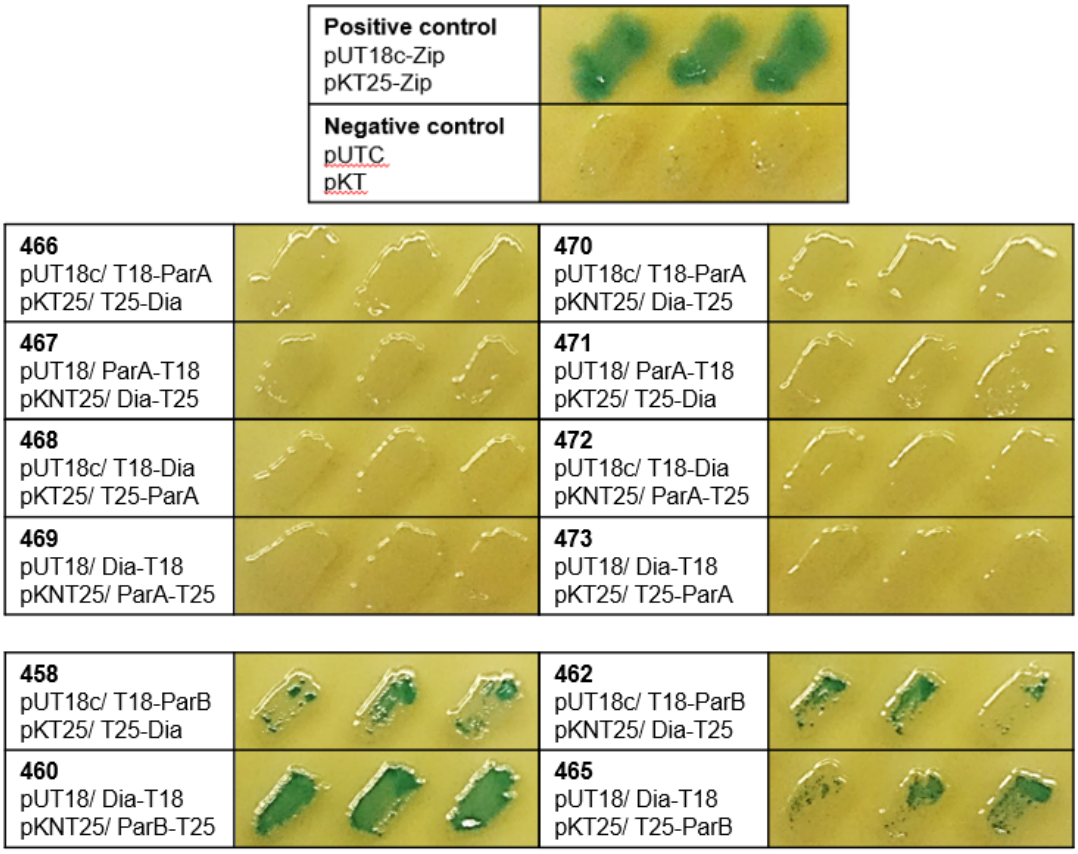


Figure 4-9. BACTH assay testing the interactions between *dia* and *parA* or *parB*, alongside the controls. Blue BTH colonies indicated a positive interaction between the proteins of the tested genes, with colourless colonies indicating no interaction. The constructs tested and the orientations of their fused adenylate cyclase subunit were detailed next to their corresponding result. 3 colonies of each co-transformation were streaked onto LB agar plates containing ampicillin, kanamycin, IPTG and X-gal. All streaks were analysed after 48 hours of growth at 30°C after inducing expression of the BACTH system.

In addition to testing the ParA-ParB system, we also tested the ParA homologue, ParH for potential interaction with Dia. Previously, ParH has been shown to interact with Scy and ParB directly (Gillespie, Unpublished). To test these the Dia-ParH interaction, we used constructs for *parH* that were previously generated in the Kelemen lab (Figure 4-8b and 4-10).

The 8 combinations testing Dia-ParH interactions resulted in a collection of different signals. An unambiguous clear positive signal was found when the tags T18 and T25 were at the N-terminus of Dia and ParH, although strong positive signal was also observed when ParH was tagged at its C-terminus and Dia was tagged at its N-terminus. This suggests the Dia and ParH interact, but we cannot make predictions about the way in which they interact. The Dia-ParH interaction was unexpected, as the homologous ParA did not show any interaction with Dia.

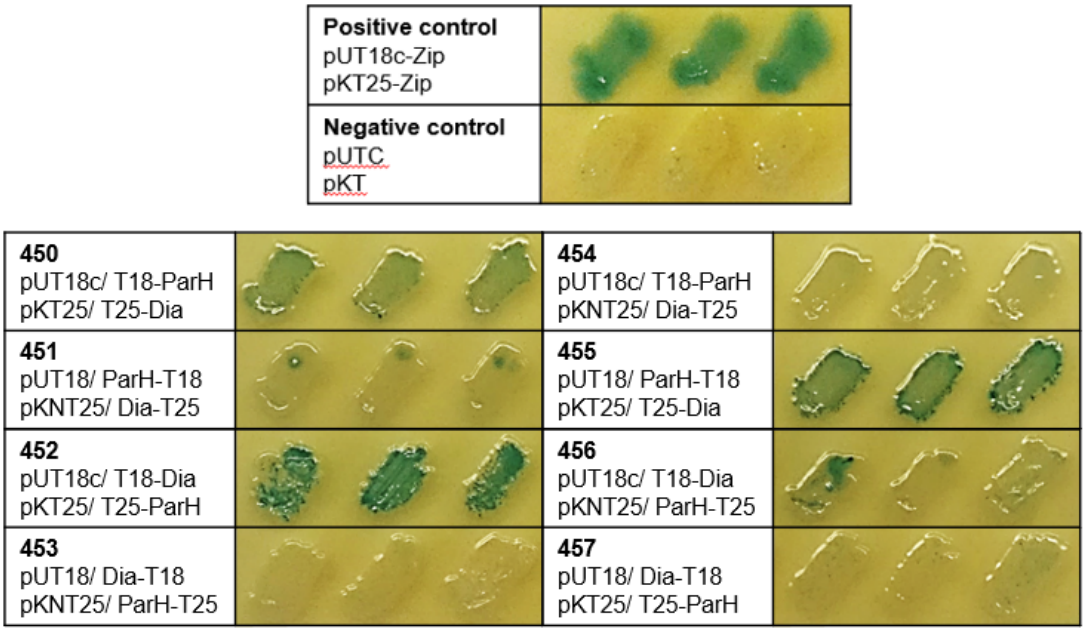


Figure 4-10. BACTH assay testing the interactions between *dia* and *parH*, alongside the controls. Blue BTH colonies indicated a positive interaction between the proteins of the tested genes, with colourless colonies indicating no interaction. The constructs tested and the orientations of their fused adenylate cyclase subunit were detailed next to their corresponding result. 3 colonies of each co-transformation were streaked onto LB agar plates containing ampicillin, kanamycin, IPTG and X-gal. All streaks were analysed after 48 hours of growth at 30°C after inducing expression of the BACTH system.

4.3 Conclusions:

In this chapter we sought to identify potential partner proteins for Dia through the use of the BACTH assay. We started by testing for self-interaction of Dia and found that Dia was capable of self-interaction. This was not surprising as, all coiled-coil proteins generate dimers, tetramers or other higher order assemblies.

With the known localisation pattern of Dia similar to other TIPOC components and possessing a coiled-coil domain, we next pursued detailing which possible interactions Dia possessed with components of the TIPOC. We investigated the 3 selected TIPOC components: Scy, FilP and DivIVA. The most convincing interaction was observed when Dia-Scy interaction was tested. Further possible interaction between Dia and FilP were detected in some of the BACTH combinations. However, in the BACTH system Dia and DivIVA did not show clear interaction (Figure 4-11). These interactions indicate that Dia is capable of interacting with the TIPOC components: Scy and FilP and may be a new novel component of the system. This could be the case with the observations of *S. coelicolor* Δ*dia* strain which demonstrated irregular hyphal diameter of various amounts through the hyphae after polar growth. In light of these findings, Dia appears to possess the ability to interact with key components of the TIPOC directly, though the necessary molecular interactions of how this occurs was yet to be confirmed. This raise the question to what extent does the interactions of Dia with the TIPOC influence the mechanisms of apical growth and whether Dia could be considered a novel component of the TIPOC. Since Dia has been confirmed to directly interact with key components of the TIPOC, we explored the idea of Dia interacting with documented components of the chromosome segregation ParAB system: being ParA and ParB. Previously, Scy, the TIPOC component was shown to interact with ParA (Ditkowski *et al*., 2013). Dia showed no interaction with ParA when tested using the BACTH system. Interestingly, Dia was capable of interacting with ParB directly (Figure 4-11). Considering this and the Dia-Scy interaction, Dia could be responsible of maintaining the interplay between the TIPOC and the ParAB system through a different mechanism to the interplay between Scy and ParA, playing an active role between growth and chromosome segregation of *S. coelicolor*. Previous research has showed that ParB influences the growth of hyphae through deletions of the gene. ParB disruptions have the reduced growth rate but increased extension time in *S. venezuelae* which has been suggested to be a potential checkpoint mechanism (Donczew *et al*., 2016). The highlighted interactions between

Dia and components of the TIPOC (DivIVA, Scy and FilP) and Dia and ParB provides evidence that Dia could be a component of this potential mechanism.

Finally, following on from the previous research conducted on ParH by Michael Gillespie, we tested whether Dia could interact with ParH, a homologue of ParA. Structurally, ParH and ParA were expected to be very similar, that being ParH has conserved the important motifs and residues needed for the key ParA functions: DNA binding and ATP hydrolysis (Gillespie, Unpublished). Surprisingly, whilst Dia did not show any interaction with ParA, we observed interaction between Dia and ParH in certain combinations (Figure 4-11).

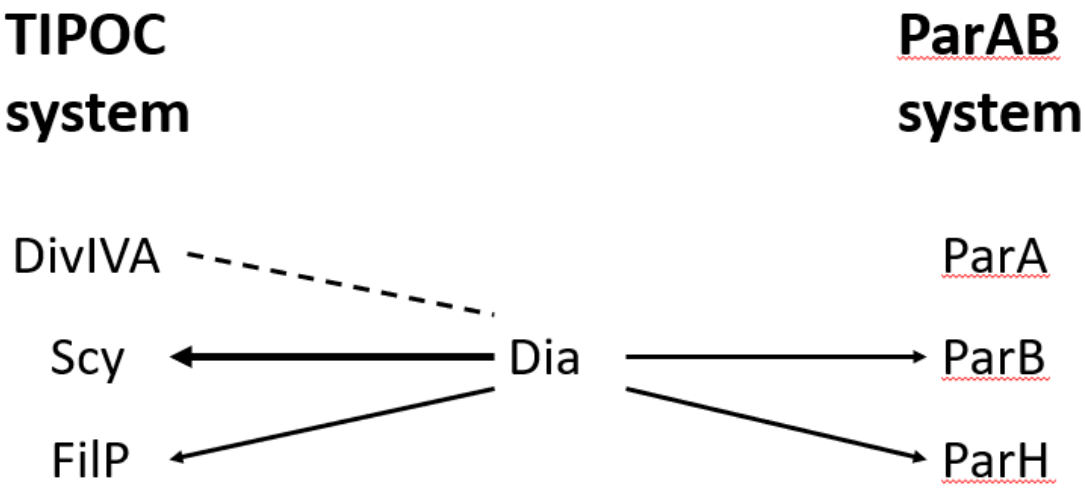


Figure 4-11. The diagram summarising the BACTH assay interactions between Dia and the TIPOC and ParAB components. The tested TIPOC components (DivIVA, Scy and FilP) can be seen in the left column and the tested ParAB components (ParA, ParB and ParH) can be seen in the right column. Multiple direct interactions are indicated by bold, solid arrows and certain direct interactions with solid arrows. Possible interactions are indicated by dashed lines and no interaction is indicated by no arrow. Dia is capable of directly interacting with Scy, certain combinations of FilP and no clear interactions were seen between DivIVA. Dia was capable of directly interacting with ParB and certain combinations of ParH. No interactions were seen between Dia and ParA.

4.4 Future directions:

Although we demonstrated the capacity for Dia to interact with various partner proteins using the BACTH system, these potential interactions must be also confirmed using an independent method. Furthermore, for DivIVA, Scy and ParB we didn't have all the possible constructs available for testing, those that expressed the T18 or T25 tags at the C-terminus of the appropriate proteins. These constructs can be generated and tested using the BACTH assay to determine if the interaction with these partner proteins in affected by the orientation of the tags. Finally, to help confirm and provide more detail of these interactions, *in vitro* biochemical assays, such as the pelleting assays of all the tested interactions could be undertaken and analysed on SDS-PAGE gels. Any changes in the position of the Dia between the pellet or supernatant columns on the SDS-PAGE gel, when one of the interacting proteins was added, would not only confirm the direct interaction but provide insight into how the interaction affects the polymerisation or depolymerisation of Dia.

Chapter 5:

Oligomerisation of Dia

5.1 Introduction:

The previous data chapters have confirmed and revealed different aspects of Dia. Through the computational analysis of Chapter 3, Dia has been confirmed to possess a coiled-coil domain which likely forms into a similar coiled-coil tertiary structure. Within this coiled-coil domain, multiple aligned domains for ATP synthase subunit B were highlighted. Alongside this, results from Chapter 4 unveiled that Dia had the capacity to interact with many partner proteins including interaction between Dia monomers. These aspects of Dia we have identified start help to explain the unique function of Dia to help provide consistent and stable hyphal diameter and what which proteins Dia interacts with as a possible component of the TIPOC and ParAB systems. The many partner proteins of Dia (Scy, ParB and to a lesser extent, DivIVA and FilP) have been shown to interact with DivIVA within *S. coelicolor*. DivIVA and FilP of *S. coelicolor* have been shown to form large self-assembling higher order structures due to coiled-coil domains. A large coiled-coil domain of DivIVA monomers was identified to be pivotal in the formation of oligomers (Wang *et al*., 2009). The FilP higher order assemblies formed long cables along the hyphae due to an apical gradient established via interactions with DivIVA (Fuchino *et al*., 2013; Alcock, Unpublished). Further analysis of FilP oligomerisation identified the formation of the higher order assemblies were due to the C-terminus of the coil 2 domain (Alcock, Unpublished). With Dia being capable of self-interaction and possessing a coiled-coil domains, among having many partner proteins, it is possible Dia could form a higher order assembly built from its own monomers or even the other partner proteins to facilitate its known function of regulation of hyphal diameter. Through investigating the quaternary structure of Dia, we will be able to understand any structures Dia may form and begin integrate these mechanics to expand the current models we have for polar growth and chromosome segregation within the *Streptomyces* genus.

5.2 Results:

To characterise the Dia protein we decided to purify the desired proteins of *S. coelicolor* using a well-established vector system and a heterogenous *E. coli* host. The pET28a plasmid (Figure 5-1) was selected for the amplification and overexpression of Dia due to the simple overexpression system and the medium copy number of the plasmid (around 20-60 copies per host cell). The gene of interest (*dia*) was inserted into the MCS via the *EcoRI* and *NdeI* restriction sites. The insertion would express a protein with an N-terminal His tag, following the thrombin cleavage site, to Dia and the kanamycin resistance gene allows for selection of the construct. The insertion of Dia should allow the ATG of *dia* and the ATG of the *NdeI* restriction site to overlap. This in turn accurately places the beginning of the gene in frame with the N-terminal His tag, consequently placing control of the expression under the T7 promoter.

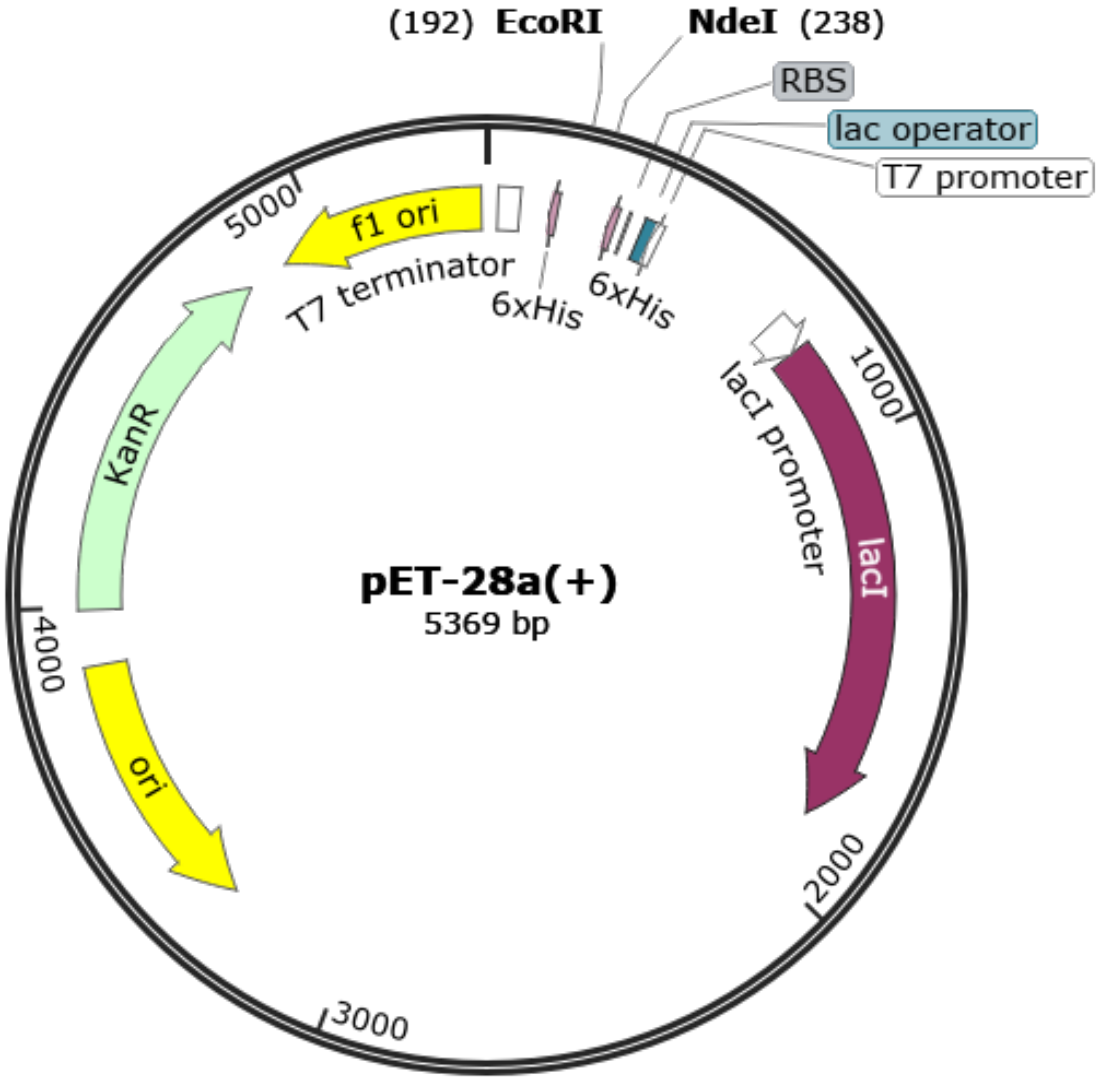


Figure 5-1. Plasmid map of pET28a used for the overexpression of His-Dia. The pET28a vector contains a 6x His-tag which is proceeded the MCS containing cloning restriction sites *EcoRI* and *NdeI*, for the fusion of the target protein to the 6x His-tag. The 6x His-tag can be induced via a *lac* promoter proceeding the tag. It also contains a Kanamycin resistance gene for selection. Vector maps were generated using SnapGene.

To generate the components for the overexpression construct, the *dia* fragment was digested from pUT18c-Dia (Figure 4-3), using the restriction enzymes *EcoRI* and *NdeI*, and isolated via an Agarose gel (Figure 5-2). The pET28a plasmid was also prepared for ligation with *EcoRI* and *NdeI*, with both digestions confirmed via Agarose gel electrophoresis (Figure 5-2). The prepared *dia* fragment was ligated into the *EcoRI* and *NdeI* restriction sites of the prepared pET28a plasmid and transformed into DH5α for isolation of the recombinant plasmid (pET28a-Dia) via plasmid DNA isolation (Figure 5-2 and 5-3).

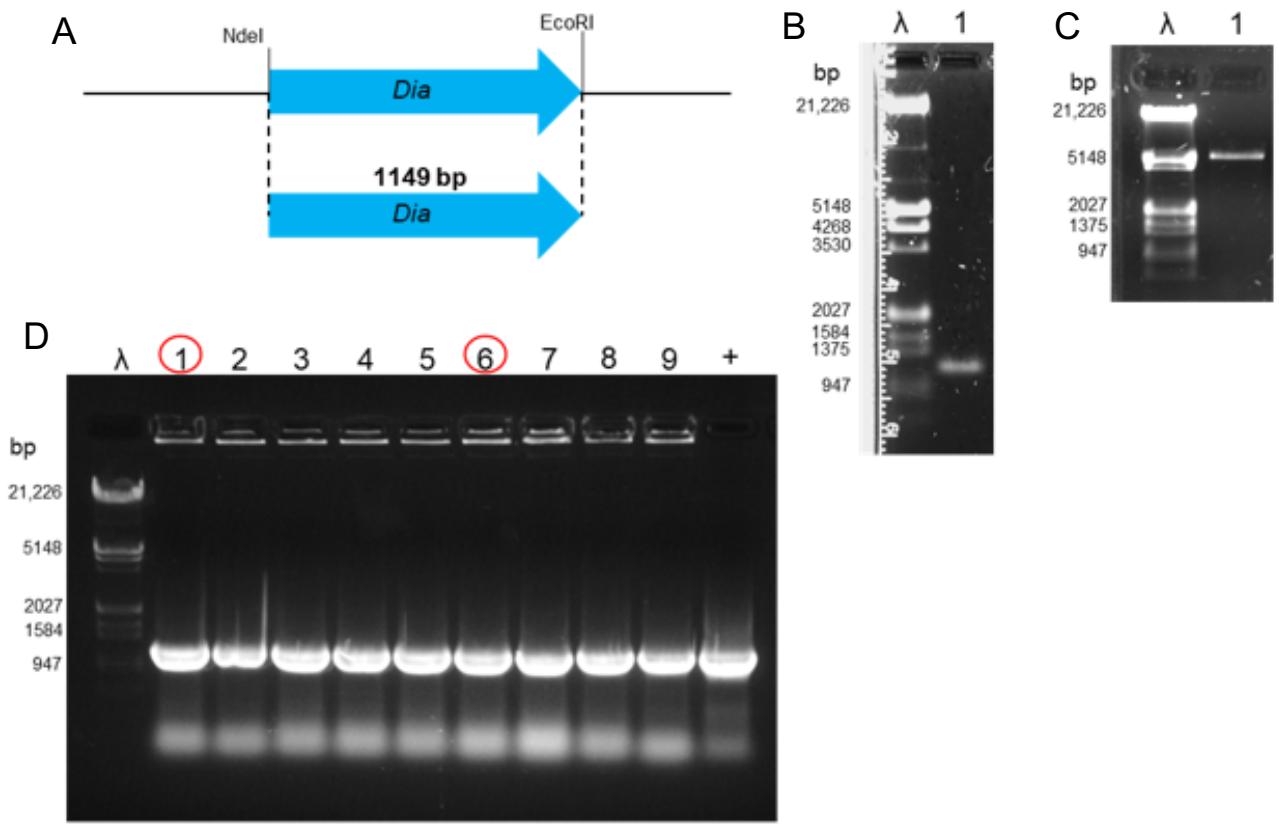
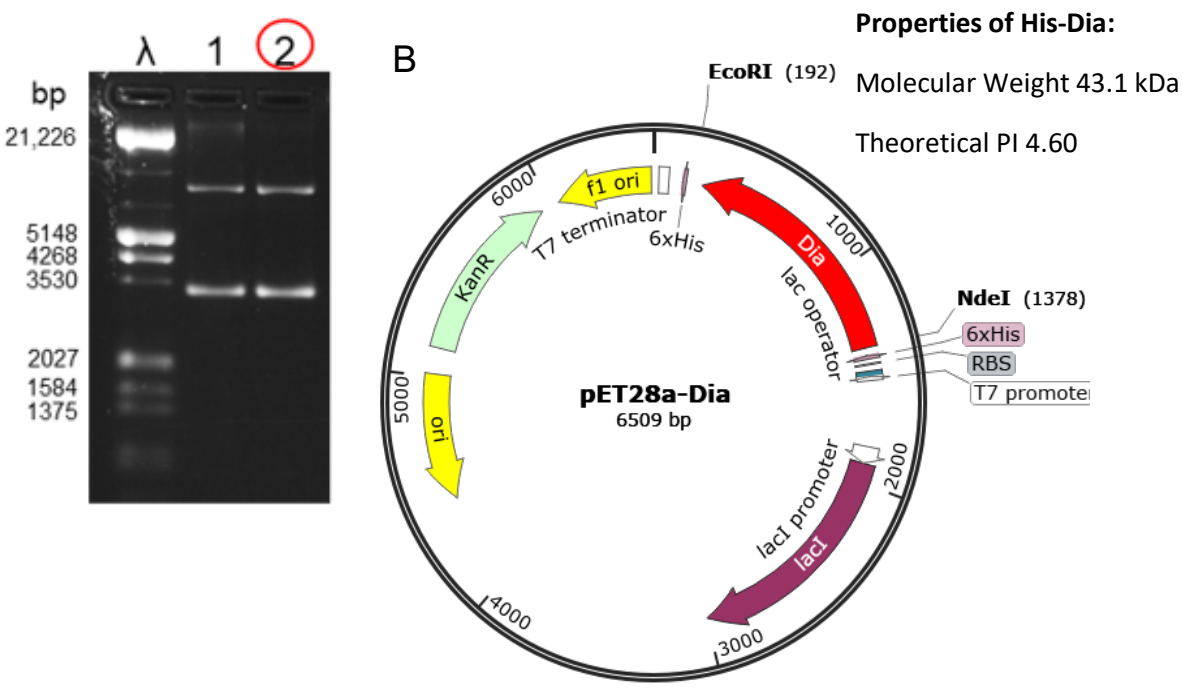


Figure 5-2. Generation of pET28a-Dia constructs. (A) The predicted restriction digest of *dia* fragment using the pUT18c-Dia template with the restriction enzymes: *NdeI* and *EcoRI*. (B) Confirmation of the *NdeI-EcoRI* restricted *dia* fragment via 0.7% Agarose gel electrophoresis. Lane λ: Lambda DNA *XbaI* + *EcoRI* ladder. Lane 1: restricted *dia* fragment. (C) Confirmation of the *NdeI-EcoRI* restricted pET28a cosmid vector via 0.7% Agarose gel electrophoresis. Lane λ: Lambda DNA *XbaI* + *EcoRI* ladder. Lane 1: restricted pET28a vector. (D) Confirmation of successful DH5α pET28a-Dia transformants using 5569 XbaNde FRW and 5569 Eco REV primers via colony PCR. The colony PCR was analysed via 0.7% Agarose gel electrophoresis. Lane λ: Lambda DNA *XbaI* + *EcoRI* ladder. + contains the pET28a PCR product (control). Lanes 1-9: pET28a-Dia colony PCRs. The pET28a-Dia colony PCRs identified positive colonies. pET28a -Dia #1 and pET28a-Dia #6 were selected for use if future overexpression trails (red circle).



MGSSHHHHHHSSGLVPRGSHMDVQNKLDEITAMVSGARAMPMSASCVVNRAELLSMLEELRAELPGS
LAQAQELIGDREQMVAQARQEADRIIEGAHAERGSLIADTEVARRSQAEADRILAEARQEAEEVRAEA
DDYVDSKLANFEVVLTKTLGSVGRGREKLLGTGPGLDENGYEDEDAPERSHDPETLRRDADAYVDTKL
GAFEAVLAKTLDAVGRGRQKLHGRIATDDLGALADDMTTVQHSSDADYLAGLAGLADAPASAPAPAEQ
QGQPQYGEQQPVAARMPAQAVPEMPSQEPVYGYASQQQPDPYAAYQQTYDGGPDPYGYQQQGADPYAY
QYDGGQQAYDAQQGYAQPPQPQQPPHAPQTHPQALDETSLFDTSMISAEQLRAYEQGRGL

Figure 5-3. Large scale plasmid preparations and plasmid map of pET28a-Dia. (A) Previously selected pET28a-Dia constructs (see Figure 5-2) were analysed via 0.7% Agarose gel electrophoresis. Lane λ: Lambda DNA digested with *EcoRI* and *HindIII*. Lane 1: pET28a-Dia #1 and Lane 2: pET28a-Dia #6. (B) pET28a-Dia is an altered copy of pET28a with the *dia* gene cloned using the restriction sites *EcoRI* and *NdeI*. Induction of the fused protein, through the use of IPTG, should produce His-Dia with the calculated properties of a molecular weight of 43.1 kDa and a theoretical PI of 4.60. The primary sequence for His-Dia can be seen below the plasmid map, with the His-tag is labelled in green including the individual bold His residues.

From the recombinant constructs we selected two plasmids: pET28a-Dia #1 and pET28a-Dia #6. Both plasmids were transformed into the 2 suitable *E. coli* hosts. BL21 pLysS and Rosetta via electroporation, resulting in 4 different combinations being tested (BL21 pET28a-Dia #1, BL21 pET28a-Dia #6, Rosetta pET28a-Dia #1 and Rosetta pET28a-Dia #6). We then performed overexpression trial to test whether BL21 or Rosetta were better host for Dia overexpression and to test whether His-Dia could indeed be overproduced using these plasmids. All 4 strains were inoculated into separate 10ml liquid LB, containing the appropriate antibiotics and grown in a shaking incubator at 200 rpm for 3 hours. 1 ml of the uninduced pET28a-Dia #1 cultures was collected into separate 1.5ml Eppendorf

tube and cells were frozen overnight. 5 µl of IPTG (0.2mM) was added to each culture and left to grow for 18 hours at 250 rpm at 37oC. 1 ml of the induced cultures was collected and the samples (6 in total: 2 uninduced and 4 induced) were resuspended in 300 µl Tris based buffer (20 mM Tris pH 8, 10 mM $MgCl_2$). The resuspended pellets were lysed via sonication on ice and spun down. The supernatant was transferred to a separate 1.5ml Eppendorf, and the pellets were resuspended in 100 µl of the same buffer. The resuspended pellets and supernatant for each combination were denatured and analysed using 12% SDS-PAGE. The overexpression levels were much higher in the Rosetta strains than in BL21 pLysS strains (Figure 5-4 and 5-5). Both constructs pET28a-Dia #1 and pET28a-Dia #6 produced good levels of His-Dia expression in Rosetta, and importantly, the majority of His-Dia was present in the soluble fraction as opposed to the pellet fraction. This trial experiment confirmed that using the construct generated we could overproduce His-Dia in *E. coli* and the protein is in the soluble fraction, which suggests that we will be able to purify it using non-denaturing methods. For the large scale preparations we chose the *E. coli* Rosetta pET28a-Dia #6 strain.

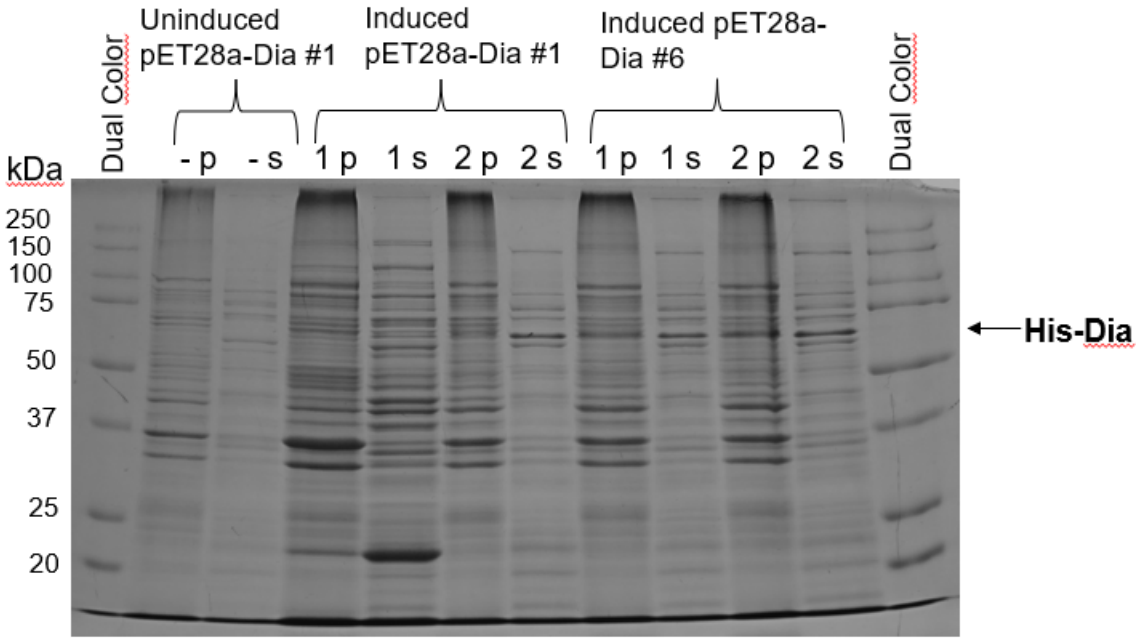


Figure 5-4. His-Dia trail overexpression within BL21 pET28a-Dia analysed through SDS-PAGE. The supernatant and pellet fractions from the *E. coli* BL21 His-Dia overexpression trail were analysed within 12% SDS-PAGE. Pellet fractions were loaded into the p labelled columns and all supernatant fractions were loaded into the s labelled columns. The samples were loaded in the following order: Uninduced pET28a-Dia #1 (-), Induced pET28a-Dia #1 (isolated colonies 1 and 2) and Induced pET28a-Dia #6 (isolated colonies 1 and 2). All dual color lanes contained 5µl BioRad dual colour protein ladder. The molecular weight for each marker protein shown in kDa. The gel was visualised with Coomassie Blue staining.

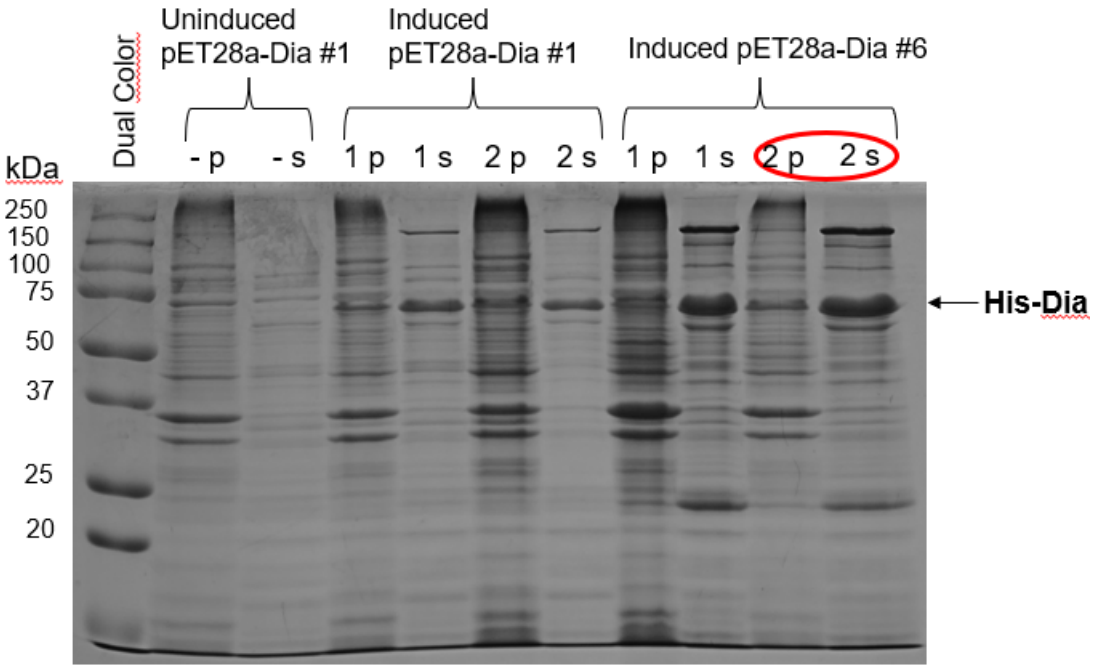


Figure 5-5. His-Dia trail overexpression within Rosetta pET28a-Dia analysed through SDS-PAGE. The supernatant and pellet fractions from the *E. coli* Rosetta His-Dia overexpression trail were analysed within 12% SDS-PAGE. Pellet fractions were loaded into the p labelled columns and all supernatant fractions were loaded into the s labelled columns. The samples were loaded in the following order: Uninduced pET28a-Dia #1 (-), Induced pET28a-Dia #1 (isolated colonies 1 and 2) and Induced pET28a-Dia #6 (isolated colonies 1 and 2). All dual color lanes contained 5µl BioRad dual colour protein ladder. The molecular weight for each marker protein shown in kDa. The gel was visualised with Coomassie Blue staining. The selected construct (Induced pET28a-Dia #6 isolated colony 2) for large scale purification is highlighted in red circle.

With the suitable *E. coli* host and construct for overexpression selected, the transformed host was inoculated into 4 separate 100 ml liquid LB, containing the appropriate antibiotics, and placed in a shaking incubator at 200 rpm and the desired temperature for 3 hours. Then 50 µl of 1M IPTG was added to each culture, with each culture grown at 250 rpm for 18 hours at 37°C. The overexpressing cells were collected, resuspended in the Tris Binding buffer (50 mM Tris, 20 mM $MgCl_2$, 300 mM NaCl, 10 mM imidazole pH 8) and lysed on ice using sonication, centrifuged to generate a clear supernatant that was also filtered using a 0.2 micron filter. The filtered cell lysate was loaded onto immobilised nickel affinity chromatography column in separate 10 ml aliquots until all of the cell lysate had been run through. The column was washed and eluted separately with appropriate amounts of the Binding, Elution and Cleaning buffers and collected individually as separate fraction respectively. This resulted in several groups of fractions: Preload, Flowthrough, Wash 1-2, Elution 1-10 and Elution 11-13.

All samples including pre-load, flowthrough, wash and elution samples were denatured at 95°C analysed on a 12%SDS-PAGE (Figure 5-6). The samples were loaded in the same order they were collected during the large scale purification. The SDS-PAGE gel was distorted at the elution samples due to the sheer amount of protein in these samples. During metal affinity chromatography, when cell extracts are loaded onto the column, all His-tagged proteins should be retained on the column generating a flowthrough sample that lacks the His-tagged protein. During His-Dia purification there was copious amount of His-Dia detected in the flowthrough samples, which suggests that we reached the capacity of the column, and we have overloaded the column. For optimal purification, the affinity chromatography could have been repeated applying less cell extracts onto the column.

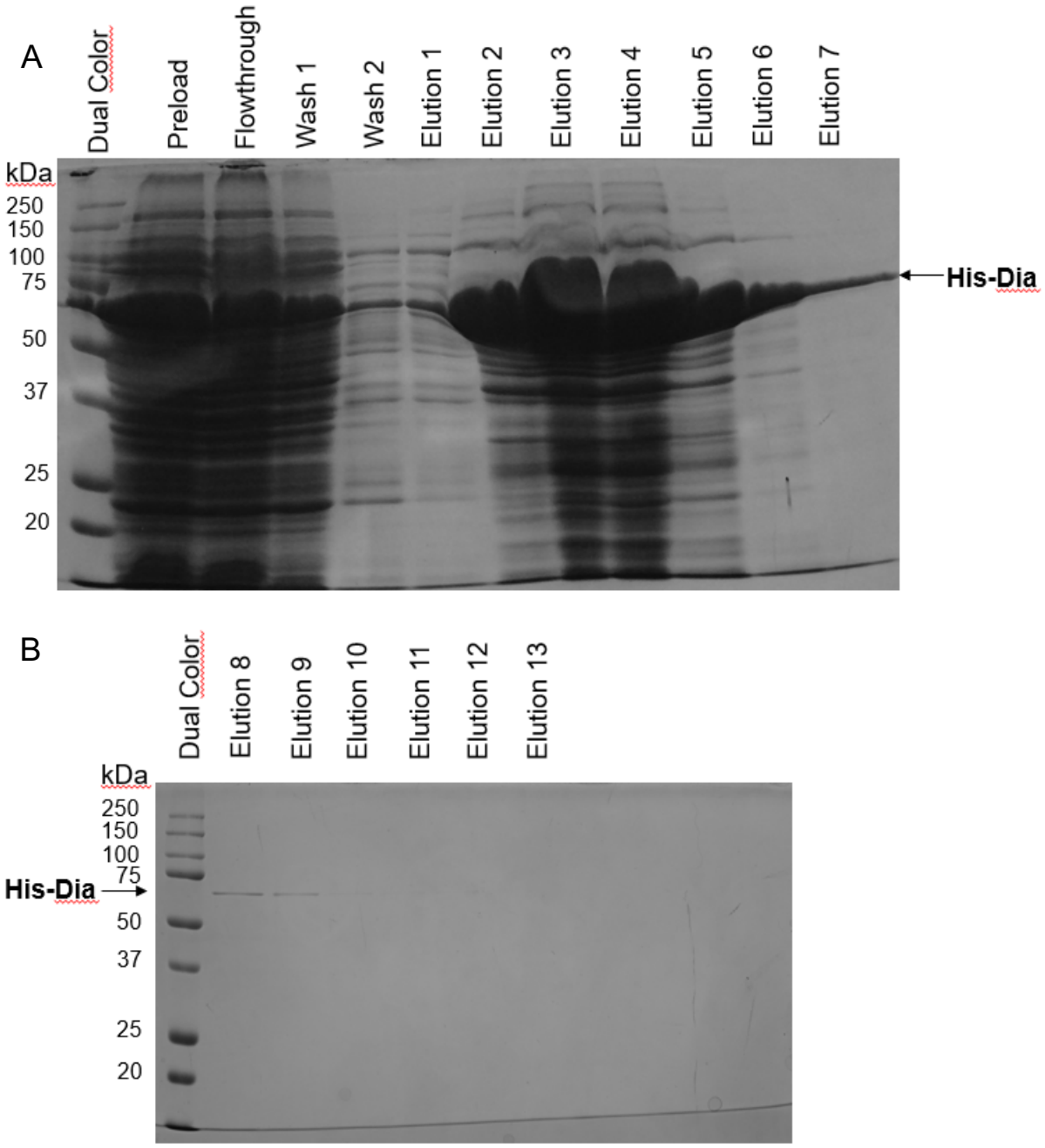


Figure 5-6. Large scale purification of His-Dia analysed through SDS-PAGE. Fractions of the large scale purification of pET28a-Dia #6 His-Dia were analysed on 12% SDS-PAGE. The samples loaded on to gel A were: Preload, Flowthrough, Wash 1-2 and Elution fractions 1-7. The samples loaded on to gel A were Elution fractions 8-13. All dual color lanes contained 5µl BioRad dual colour protein ladder. The molecular weight for each marker protein shown in kDa. The gel was visualised with Coomassie Blue staining.

To generate a clearer analysis of the samples, we repeated the SDS-PAGE using 1/10th of the amount that was loaded previously (Figure 5-7). Intriguingly in all the SDS-PAGE analysis, His-Dia was running at the apparent molecular weight of approximately 65 kDa. This is larger than the predicted weight of a His-Dia monomer, 43.1 kDa. This might mean that Dia can form a dimer, in fact Dia does

possess a Cysteine (Cys) amino acid within the potential coiled-coil structure that could be oxidised into a disulfide bond during dimerisation. We did use freshly prepared loading dye containing β-mercaptoethanol, to test whether in the presence of the reducing agent we could detect Dia monomers. However, in all of our SDS-PAGE attempts Dia did run at the apparent molecular weight of ~65 kDa. This is unusual and might suggest that the disulfide bond might be very much buried in the coiled-coil structure, perhaps inside the hydrophobic streak, hence our inability to generate Dia monomers even in the presence of β-mercaptoethanol as a reducing agent.

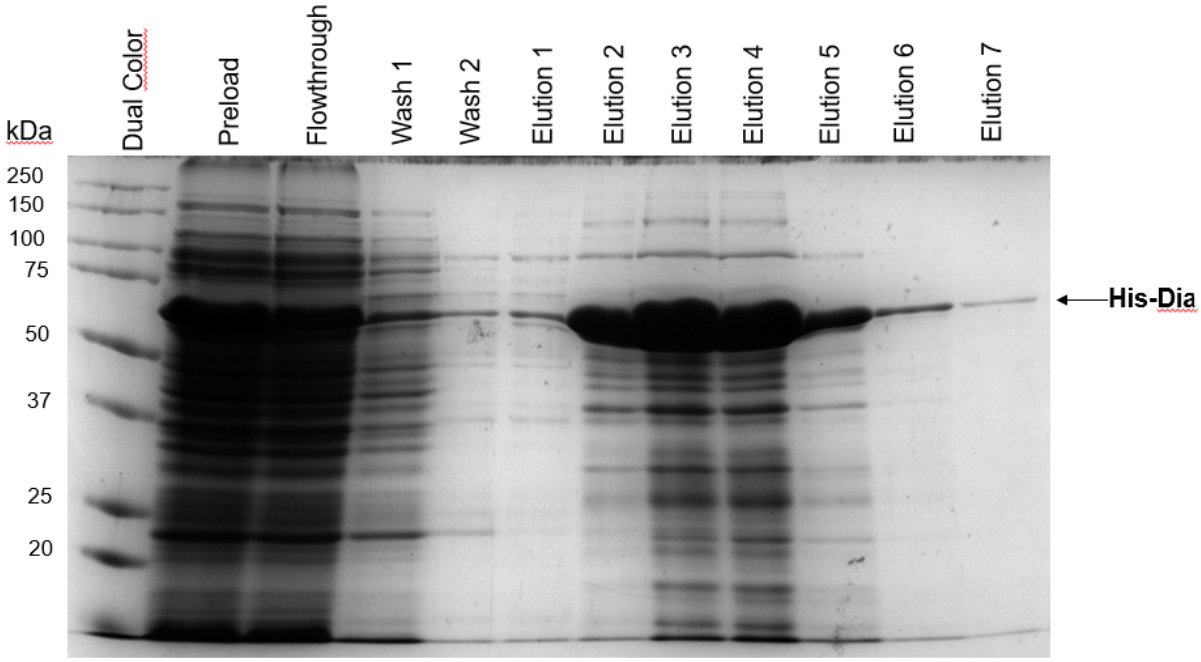


Figure 5-7. Diluted His-Dia analysed through SDS-PAGE. Fractions of the large scale purification of diluted pET28a-Dia #6 His-Dia were analysed on 12% SDS-PAGE. The samples loaded on to gel A were: Preload, Flowthrough, Wash 1-2 and Elution fractions 1-7. All dual color lanes contained 5µl BioRad dual colour protein ladder. The molecular weight for each marker protein shown in kDa. The gel was visualised with Coomassie Blue staining.

In order to have a more accurate molecular weight determination for His-Dia, we used the REFEYN Mass Photometry system that measures the scattering of light, and the subsequent interference signal, by the molecule in solution to determine the molecular mass. Precise image analysis of the interference signal allows for direct correlation of the molecules' measured mass. A 1 ml sample of the remaining cell lysate was dialysed against a Tris based buffer (50mM Tris, 150mM NaCl, 10mM $MgCl_2$, pH 8) and sent off for analysis, with the data converted into a histogram (Figure 5-8). The measured molecular weight for His-Dia was 77 kDa with

SD=17.5, which then support the hypothesis that that the band on the SDS-PAGE gels is a dimer of His-Dia.

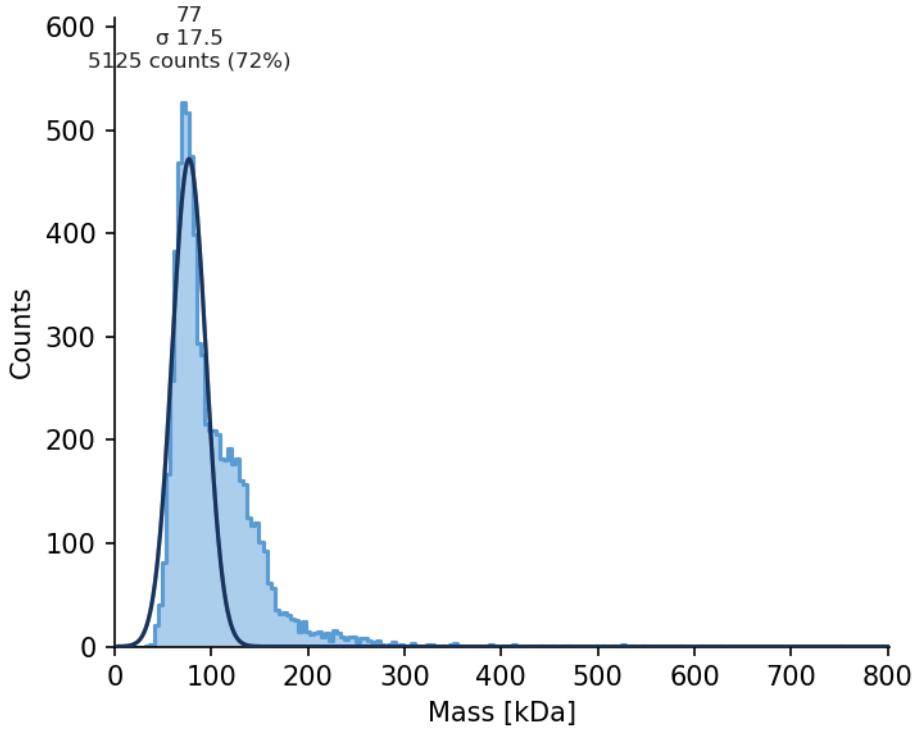


Figure 5-8. The mass measurement histogram of the dialysed His-Dia solution measured via REFEYN Mass Photometry. The predicted mass of His-Dia was 77 kDa with a standard deviation (σ) of 17.5 at the peak of the Gaussian curve.

5.3 Conclusions:

In this chapter we wished to overexpress and purify the Dia protein for future biochemical analysis.

We have generated a construct using pET28a as a vector for protein overexpression in *E. coli*. We have tested two different host strains, *E. coli* BL21 and Rosetta and we found that Dia overexpression was much more elevated in the Rosetta strain. This is not unexpected, as the Rosetta strain was developed to express genes that require tRNAs with codons that are otherwise rare in *E. coli*, including the GC rich, AGG or CCC, codons.

We have optimised the overexpression and successfully purified His-Dia in a large scale. Intriguingly, we found that the purified protein did not run at the expected molecular weight of 43.1 kDa, instead it showed an apparent molecular weight of ~65 kDa in SDS-PAGE. As proteins sometimes do not run according to their molecular weight, such as acidic proteins are known for their aberrant mobility

in SDS-PAGE, we had used a novel technology of Mass Photometry to gather information about the molecular weight of His-Dia. The molecular weight of His-Dia was estimated to be 77 kDa (SD=17.5) which makes it likely that His-Dia forms a dimer. Interestingly this Dia dimer could not be separated into its monomer during SDS-PAGE. The bonds of the possible complex were unaffected when exposed to the denaturing conditions (SDS, heat) and β-mercaptoethanol. The Cys residue in Dia might stabilise the Dia dimers, but the potential disulfide bond within Dia dimers might be located within a hydrophobic core of the coiled-coil.

5.4 Future directions:

Although we managed to purify His-Dia, the results appear to suggest that His-Dia was extracted as a dimer. If a complex was formed, the bonds within were unaffected through denaturing conditions and the used β-mercaptoethanol to break the potential disulfide bond. As a result, we suggest altering the denaturing conditions for the possible His-Dia complex by using different denaturing compounds, such as Dithiothreitol or Tris(2-carboxylethyl)phosphine hydrochloride, to see if they separate into individual components. Lastly, running the possible His-Dia complex through NMR or Mass spectroscopy would be useful for helping determine the molecular weight of His-Dia.

Chapter 6:

Analysis of GPR homologs and the generation and localisation of LcyEgfp within *L. aggregata*

6.1 Introduction:

The previous data chapters have expanded our knowledge of the DivIVA-based polar growth mechanism in Gram-positive bacteria known as the TIPOC through a possible new component, Dia for controlling hyphal diameter. Dia may be a new component of the TIPOC due to the possession of a coiled-coil domain and the variety of different partner proteins found within the TIPOC. However, is not the only known system for polar growth with the recent discovery of the GPR (Growth Pole Ring) protein in Gram-negative bacteria which typically lack any DivIVA homologues. The GPR protein was discovered in *A. tumefaciens* where 6 GPR monomers appear to assemble at a single pole of the cell throughout its life cycle, possibly as a hexameric ring (Zupan *et al*., 2019). GPR were also seen localising at the midcell prior to formation of the division septum and at the growing poles of newly formed cells (Zupan *et al*., 2019). Deletion of the C-terminus and human apolipoprotein A-IV coiled-coil domains resulted in defects in cell morphology and demonstrated the role of GPR in peptidoglycan synthesis (Zupan *et al*., 2021). In addition, homologs of GPR have been identified in a variety of different Rhizobiales (Zupan *et al*., 2019), with one such homolog (RgsE) confirmed in *Sinorhizobium meliloti*, with deletion also resulting in cell morphological defects coupled with reduced growth and division (Krol *et al*., 2020). Since homologs of this system has been shown across abroad range of different Gram-negative bacteria, we planned to investigate this occurrence further to expand our knowledge of polar growth systems which utilise alternatives to DivIVA.

6.2 Results:

6.2.1 Analysis of GPR homologs

As the GPR homologs have been found in *A. tumefaciens* and *S. meliloti* which are both Gram-negative bacteria from the Rhizobiales Family (sometimes called Rhizobiaceae Family), we searched for other GPR homologs across other species through a BLAST search against the GPR primary sequence (UniProt ID: A9CJ72) with *A. tumefaciens* being excluded. The search revealed that the top 500 matches aligned with homologs from a variety of different species from the Rhizobiales Family. These matches include mostly *Agrobacterium*, *Rhizobium* and *Sinorhizobium* species, with all of the identified matches being unique homologs due to possessing high E-values (some E-values were so significantly small that the program rounded them to 0). A few other species from the Rhizobiales Family were identified, such as *Shinella* and *Neorhizobium*, though with far fewer individual hits. For the *Agrobacterium* alignments, most of the alignments possessed high percentages of similarity of at least 80% or higher (Table 6-1). The *Rhizobium* alignments possessed a variety of different percentages of similarity ranging from between 40 – 95% (Table 6-1 and 6-2) and the *Sinorhizobium* alignments were possessed a similarity percentage between 40 – 45% roughly (Table 6-2). Following the top 500 matches, there are still hits for various species in the Rhizobiales Family but some of the aligned hits were for species from entirely different Families. These distantly related species include *Mesorhizobium*, *Pseudaminobacter* and *Aminobacter* species and share their connection to *A. tumefaciens* through the Hyphomicrobiales Order (Table 6-3). These alignments are still unique homologs due to their high E-values, though the matches had roughly 30% similarity to GPR.

Table 6-1. The amino acid sequence of GPR (Uniprot ID: A9CJ72) from *A. tumefaciens* was BLAST against other *bacterial* genomes and excluded *A. tumefaciens*. The top 500 hits identified GPR homologs from a variety of different species in the Rhizobiales Family. The identified homologs near the top of the 500 hits possessed at least 90% similarity to GPR.

Description	Scientific name	E value	Percentage Identity	Accession
hypothetical protein [Agrobacterium tumefaciens complex]	*Agrobacterium tumefaciens complex*	0.0	100.00%	WP_010971556.1
hypothetical protein [unclassified Agrobacterium]	*unclassified Agrobacterium*	0.0	99.53%	WP_006312583.1
hypothetical protein [Agrobacterium fabrum]	*Agrobacterium fabrum*	0.0	99.53%	WP_092770376.1
hypothetical protein [Agrobacterium fabrum]	*Agrobacterium fabrum*	0.0	99.53%	WP_121651540.1
hypothetical protein [Agrobacterium fabrum]	*Agrobacterium fabrum*	0.0	99.57%	WP_174020161.1
apolipoprotein A1/A4/E domain-containing protein [Agrobacterium fabrum]	*Agrobacterium fabrum*	0.0	99.53%	SDB19294.1
lipoprotein [Agrobacterium fabrum]	*Agrobacterium fabrum*	0.0	99.53%	AYM57051.1
hypothetical protein [Agrobacterium fabrum]	*Agrobacterium fabrum*	0.0	99.48%	WP_080812008.1
hypothetical protein SRABI05_00537 [Agrobacterium fabrum]	*Agrobacterium fabrum*	0.0	99.48%	CAH0150677.1
hypothetical protein [Agrobacterium fabrum]	*Agrobacterium fabrum*	0.0	99.43%	WP_121690965.1

lipoprotein [Agrobacterium fabrum]	*Agrobacterium fabrum*	0.0	99.43%	AYM62144.1
hypothetical protein [Agrobacterium saltinitolerans]	*Agrobacterium saltinitolerans*	0.0	96.69%	WP_173988981.1
hypothetical protein [Agrobacterium saltinitolerans]	*Agrobacterium saltinitolerans*	0.0	96.45%	WP_142912153.1
hypothetical protein [Agrobacterium sp. LAD9]	*Agrobacterium sp. LAD9*	0.0	96.51%	WP_107675494.1
hypothetical protein [Hyphomicrobiales bacterium]	*Hyphomicrobiales bacterium*	0.0	96.22%	MBA4777572.1
apolipoprotein A1/A4/E domain-containing protein [unclassified Agrobacterium]	*unclassified Agrobacterium*	0.0	96.46%	WP_020012392.1
hypothetical protein [Agrobacterium saltinitolerans]	*Agrobacterium saltinitolerans*	0.0	96.33%	WP_217004674.1
hypothetical protein [Agrobacterium saltinitolerans]	*Agrobacterium saltinitolerans*	0.0	96.33%	WP_137410230.1
hypothetical protein [Agrobacterium saltinitolerans]	*Agrobacterium saltinitolerans*	0.0	96.26%	WP_077982396.1
hypothetical protein [Agrobacterium]	*Agrobacterium*	0.0	96.46%	WP_060726168.1
hypothetical protein [Agrobacterium deltaense]	*Agrobacterium deltaense*	0.0	96.08%	WP_080817135.1
hypothetical protein [Agrobacterium]	*Agrobacterium*	0.0	96.04%	WP_080790835.1

Table 6-2. The amino acid sequence of GPR (Uniprot ID: A9CJ72) from *A. tumefaciens* was BLAST against other *bacterial* genomes and excluded *A. tumefaciens*. The top 500 hits identified GPR homologs from a variety of different species in the Rhizobiales Family. The identified homologs near the bottom of the top 500 hits possessed at least 40% similarity to GPR.

Description	Scientific name	E value	Percentage Identity	Accession
kinesin [Sinorhizobium sp. BJ1]	*Sinorhizobium sp. BJ1*	0.0	42.66%	WP_097520628.1
hypothetical protein [Neorhizobium sp. NCHU2750]	*Neorhizobium sp. NCHU2750*	0.0	45.47%	WP_119940087.1
hypothetical protein [Rhizobium leguminosarum]	*Rhizobium leguminosarum*	0.0	41.73%	WP_130772308.1
hypothetical protein [Rhizobium leguminosarum]	*Rhizobium leguminosarum*	0.0	41.68%	WP_130776185.1
membrane protein [Rhizobium sp. CF142]	*Rhizobium sp. CF142*	0.0	42.18%	WP_007826829.1
kinesin [Sinorhizobium meliloti]	*Sinorhizobium meliloti*	0.0	43.61%	WP_127642506.1
kinesin [Rhizobium sp. Root482]	*Rhizobium sp. Root482*	0.0	44.05%	WP_056335304.1
kinesin [Sinorhizobium meliloti]	*Sinorhizobium meliloti*	0.0	43.61%	WP_248448830.1
kinesin [Sinorhizobium meliloti]	*Sinorhizobium meliloti*	0.0	43.66%	WP_003532967.1
kinesin [Sinorhizobium meliloti]	*Sinorhizobium meliloti*	0.0	43.61%	WP_028003655.1
kinesin [Sinorhizobium meliloti]	*Sinorhizobium meliloti*	0.0	43.61%	WP_027994052.1
kinesin [Sinorhizobium meliloti]	*Sinorhizobium meliloti*	0.0	43.57%	WP_102903537.1
kinesin [Sinorhizobium meliloti]	*Sinorhizobium meliloti*	0.0	43.57%	WP_127540085.1
kinesin [Sinorhizobium meliloti]	*Sinorhizobium meliloti*	0.0	43.57%	WP_15828553.1
kinesin [Sinorhizobium meliloti]	*Sinorhizobium meliloti*	0.0	43.61%	WP_127517740.1

kinesin [Sinorhizobium meliloti]	*Sinorhizobium meliloti*	0.0	43.57%	WP_015007629.1
putative membrane protein [Rhizobium freirei PRF 81]	*Rhizobium freirei PRF 81*	0.0	43.94%	ENN89216.1
kinesin [Sinorhizobium meliloti]	*Sinorhizobium meliloti*	0.0	43.61%	WP_127615601.1
kinesin [Sinorhizobium meliloti]	*Sinorhizobium meliloti*	0.0	43.57%	WP_127642987.1
kinesin [Sinorhizobium meliloti]	*Sinorhizobium meliloti*	0.0	43.57%	WP_127641858.1
hypothetical protein [Rhizobium sp. WYJ-E13]	*Rhizobium sp. WYJ-E13*	0.0	42.34%	WP_216758284.1
kinesin [Sinorhizobium meliloti]	*Sinorhizobium meliloti*	0.0	43.61%	WP_122102741.1
kinesin [Sinorhizobium meliloti]	*Sinorhizobium meliloti*	0.0	43.52%	WP_028011718.1
kinesin [Sinorhizobium meliloti]	*Sinorhizobium meliloti*	0.0	43.40%	WP_066870287.1
kinesin [Sinorhizobium meliloti]	*Sinorhizobium meliloti*	0.0	43.52%	WP_252962200.1

Table 6-3. The amino acid sequence of GPR (Uniprot ID: A9CJ72) from *A. tumefaciens* was BLAST against other *bacterial* genomes and excluded *A. tumefaciens*. The hits after following the top 500 hits identified GPR homologs from a variety of different species in various Families. All the identified homologs possessed at roughly 30% similarity to GPR.

Description	Scientific name	E value	Percentage Identity	Accession
hypothetical protein [Rhizobium sp. C1]	*Rhizobium sp. C1*	0.0	34.50%	WP_230836141.1
hypothetical protein [Rhizobium oryzae]	*Rhizobium oryzae*	0.0	30.85%	WP_075626440.1
kinesin [Aminobacter sp. AP02]	*Aminobacter sp. AP02*	0.0	31.04%	WP_109576131.1
kinesin [Pseudaminobacter salicylatoxans]	*Pseudaminobacter salicylatoxans*	0.0	30.35%	WP_109611866.1
kinesin [Aminobacter anthylidis]	*Aminobacter anthylidis*	0.0	30.64%	WP_214386438.1
kinesin [Mesorhizobium sp.]	*Mesorhizobium sp.*	0.0	29.84%	MCO5163466.1
kinesin [Aminobacter sp. SR38]	*Aminobacter sp. SR38*	0.0	30.42%	WP_192316795.1
kinesin [Aminobacter NyZ550]	*Aminobacter NyZ550*	0.0	30.43%	WP_263007297.1
hypothetical protein [Hoeflea olei]	*Hoeflea olei*	0.0	36.43%	WP_066178506.1
kinesin [Aminobacter carboxidus]	*Aminobacter carboxidus*	0.0	30.42%	WP_192568032.1
kinesin [Mesorhizobium ciceri]	*Mesorhizobium cicero*	0.0	30.10%	WP_029352944.1
kinesin [Mesorhizobium sp.]	*Mesorhizobium sp.*	0.0	29.54%	RWO35173.1
hypothetical protein COA37_17520 [Hoeflea sp.]	*Hoeflea sp.*	0.0	32.89%	PHR19715.1
kinesin [Aminobacter aganoensis]	*Aminobacter aganoensis*	0.0	30.82%	WP_184697454.1
hypothetical protein [Hoeflea sp.]	*Hoeflea sp.*	0.0	34.65%	MBC7283043.1
kinesin [Mesorhizobium sp.]	*Mesorhizobium sp.*	0.0	29.44%	RWO02131.1

hypothetical protein [Hoeflea halophila]	*Hoeflea halophila*	0.0	32.62%	WP_097104370.1
hypothetical protein [Constotaella salsifontis]	*Constotaella salsifontis*	0.0	28.28%	WP_078707475.1
kinesin [Aminobacter DSM_101952]	*Aminobacter DSM_101952*	0.0	30.73%	WP_055977195.1
kinesin [Aminobacter lissarensis]	*Aminobacter lissarensis*	0.0	30.12%	WP_184767566.1
kinesin [Mesorhizobium sp.]	*Mesorhizobium sp.*	0.0	29.78%	RWI69326.1
kinesin [Mesorhizobium sp.]	*Mesorhizobium sp.*	0.0	30.37%	TIS69436.1
kinesin [Mesorhizobium sp.]	*Mesorhizobium sp.*	0.0	29.62%	TIQ22327.1
kinesin [Mesorhizobium sp.]	*Mesorhizobium sp.*	0.0	30.32%	RWE63449.1
kinesin [Aminobacter]	*Aminobacter*	0.0	30.03%	WP_067962528.1

As we identified new species with homologs of GPR outside of the Rhizobiales Family, we decided to investigate this further through a separate BLAST search against GPR (UniProt ID: A9CJ72) excluding *A. tumefaciens* and including species from the Hyphomicrobiales Order. The hits returned were then filtered to exclude any species within the Rhizobiaceae Family. The results from this search identified homologs from multiple different species including *Mesorhizobium*, *Bradyrhizobium*, *Pseudohoeflea*, and *Roseibium* (Table 6-4 and 6-5). Majority of the returned hits were unique homologs of GPR due to their high E-values, though the percentage similarity varied across the hits; most hits had a percentage similarity of roughly 30% or above. Most of the identified species were found across a range of different Families including Nitrobacteraceae, Phyllobateriaceae and Stappiaceae.

Table 6-4. The amino acid sequence of GPR (Uniprot ID: A9CJ72) from *A. tumefaciens* was BLAST against other *bacterial* genomes and excluded *A. tumefaciens* and included species from the Hyphomicrobiales Order. The list of hits was filtered to exclude any species from the Rhizobiaceae Family. Multiple different hits identified GPR homologs from a variety of different species in various Families. Majority of the identified homologs possessed at roughly 30% similarity to GPR.

Description	Scientific name	E value	Percentage Identity	Accession
hypothetical protein [Hyphomicrobiales bacterium]	*Hyphomicrobiales bacterium*	0.0	96.22%	MBA4777572.1
hypothetical protein C241_07106 [Bradyrhizobium lupini HPC(L)]	*Bradyrhizobium lupini HPC(L)*	0.0	94.16%	EKJ96551.1
kinesin [Mesorhizobium loti]	*Mesorhizobium loti*	0.0	43.61%	PST25130.1
kinesin [Mesorhizobium plurifarium]	*Mesorhizobium plurifarium*	0.0	43.53%	PST21802.1
kinesin [Bradyrhizobium sp. BRP14]	*Bradyrhizobium sp. BRP14*	0.0	42.94%	MCA1369893.1
hypothetical protein [Hyphomicrobiales bacterium]	*Hyphomicrobiales bacterium*	0.0	34.85%	MBG19624.1
hypothetical protein [Hoeflea sp. WL0058]	*Hoeflea sp. WL0058*	0.0	33.70%	WP_220227129.1
hypothetical protein [Martonella mediterranea]	*Martonella mediterranea*	0.0	30.85%	WP_132310275.1
hypothetical protein [Martonella endophytica]	*Martonella endophytica*	0.0	32.50%	WP_045682490.1
hypothetical protein [Hyphomicrobiales bacterium]	*Hyphomicrobiales bacterium*	0.0	45.79%	MBA4784453.1
hypothetical protein [Martonella alba]	*Martonella alba*	0.0	31.79%	WP_181409845.1

hypothetical protein [Hyphomicrobiales bacterium]	*Hyphomicrobiales bacterium*	0.0	34.55%	MBN9033798.1
hypothetical protein [Martonella alba]	*Martonella alba*	0.0	31.79%	TPW29776.1
hypothetical protein [unclassified Martonella]	*unclassified Martonella*	0.0	31.93%	WP_180897497.1
hypothetical protein [Pseudohoeflea sp. DP4N28-3]	*Pseudohoeflea sp. DP4N28-3*	0.0	33.30%	WP_219201995.1
hypothetical protein [Martonella radicus]	*Martonella radicus*	0.0	30.52%	WP_183484914.1
hypothetical protein [Martonella sp. AD-3]	*Martonella sp. AD-3*	0.0	30.35%	WP_024709408.1
hypothetical protein [Martonella mediterranea]	*Martonella mediterranea*	0.0	31.81%	WP_230718793.1
hypothetical protein [Martonella lutilitoris]	*Martonella lutilitoris*	0.0	30.77%	WP_200333518.1
hypothetical protein [Martonella limonli]	*Martonella limonli*	0.0	30.76%	WP_174801903.1
hypothetical protein [Martonella sp.]	*Martonella sp.*	0.0	31.46%	MAU22788.1
hypothetical protein [Hyphomicrobiales bacterium]	*Hyphomicrobiales bacterium*	0.0	39.65%	MBA4798710.1
hypothetical protein [Martonella lutilitoris]	*Martonella lutilitoris*	0.0	30.75%	WP_138749599.1

Table 6-5. The amino acid sequence of GPR (Uniprot ID: A9CJ72) from *A. tumefaciens* was BLAST against other *bacterial* genomes and excluded *A. tumefaciens* and included species from the Hyphomicrobiales Order. The list of hits was filtered to exclude any species from the Rhizobiaceae Family. Multiple different hits identified GPR homologs from a variety of different species in various Families. All the identified homologs possessed at roughly 30% similarity to GPR.

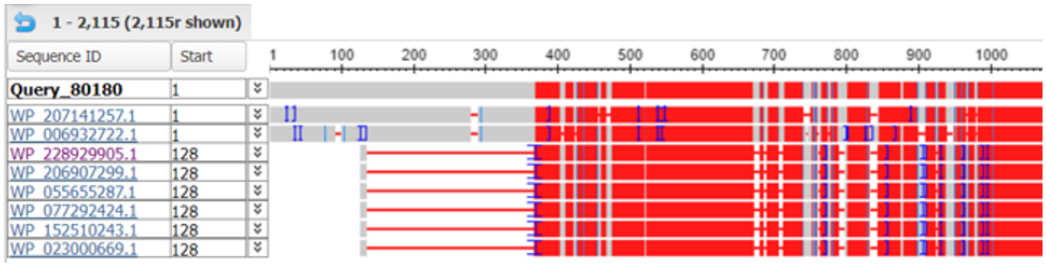
Description	Scientific name	E value	Percentage Identity	Accession
antitoxin [Rosebium aggregatum]	*Rosebium aggregatum*	2e-116	26.20%	WP_228929905.1
kinesin [Brucella intermedia]	*Brucella intermedia*	2e-116	29.73%	WP_253124229.1
apolipoprotein A1/A4/E family protein [Stappia sp. TSB10P1A]	*Stappia sp. TSB10P1A*	3e-116	27.11%	WP_186386156.1
hypothetical protein [Pseudovibrio sp. Ad46]	*Pseudovibrio sp. Ad46*	3e-116	25.28%	WP_063307951.1
hypothetical protein [Pseudovibrio sp. Ad5]	*Pseudovibrio sp. Ad5*	3e-116	25.28%	WP_063292796.1
antitoxin [Rosebium aggregatum]	*Rosebium aggregatum*	3e-116	26.20%	WP_206907299.1
antitoxin [Rosebium aggregatum]	*Rosebium aggregatum*	4e-116	26.20%	WP_055655287.1
antitoxin [Rosebium aggregatum]	*Rosebium aggregatum*	4e-116	26.20%	WP_077292424.1
antitoxin [Rosebium aggregatum]	*Rosebium aggregatum*	4e-116	26.20%	WP_152510243.1
antitoxin [Rosebium aggregatum]	*Rosebium aggregatum*	4e-116	26.20%	WP_023000669.1
kinesin [Aquamicrobium sp.]	*Aquamicrobium sp.*	4e-116	29.49%	MBX9461513.1
kinesin [Mesorhizobium sp. BH1-1-5]	*Mesorhizobium sp. BH1-1-5*	4e-116	27.68%	WP_224668286.1
kinesin [unclassified Mesorhizobium]	*unclassified Mesorhizobium*	5e-116	27.96%	WP_127860177.1
Apolipoprotein A1/A4/E family	*Labrenzia sp. THAF35*	5e-116	26.31%	QFT68236.1

protein [Labrenzia sp. THAF35]				
kinesin [unclassified Mesorhizobium]	*unclassified Mesorhizobium*	5e-116	27.45%	WP_142865853.1
hypothetical protein N6A79_06165 [Bartonella sp. HY761]	*Bartonella sp. HY761*	5e-116	23.79%	UXN07568.1
kinesin [Ochrobactrum sp. C6C9]	*Ochrobactrum sp. C6C9*	6e-116	29.00%	MCI0999804.1
kinesin [Mesorhizobium sp.]	*Mesorhizobium sp.*	7e-116	27.86%	TIP03727.1
kinesin [Mesorhizobium sp. B4-1-3]	*Mesorhizobium sp. B4-1-3*	7e-116	27.53%	WP_140750362.1
hypothetical protein [Nitratireductor aquibiodomus]	*Nitratireductor aquibiodomus*	8e-116	29.61%	WP_065815191.1
apolipoprotein A1/A4/E family protein [Stappia sp. P2PMeth1]	*Stappia sp. P2PMeth1*	1e-115	27.11%	WP_186400116.1
kinesin [Brucella thiophenivorans]	*Brucella thiophenivorans*	1e-115	28.34%	WP_094507162.1
kinesin [Mesorhizobium albiziae]	*Mesorhizobium albiziae*	1e-115	27.24%	WP_149760117.1
antitoxin [Rosebium sp.]	*Rosebium sp.*	1e-115	26.08%	MBO6860122.1
kinesin [[Ochrobactrum] telegrylli]	*[Ochrobactrum] telegrylli*	2e-115	28.81%	WP_140025183.1

One species stood out being *Roseibium aggregatum* (also known as *Labrenzia aggregata*) (Table 6-5). This species stood out due to its recent reclassification and its preference to live in an aquatic environment. *Stappia aggregata* was formally reclassified to *L. aggregata* due to differences highlighted by 16S rRNA gene sequences, chemotaxonomic analysis and distinct biochemical and physiological differences (Biebl *et al*., 2007). Following the recent reclassification, *L. aggregata* has been identified for the role it plays in marine environments. *L. aggregata* is vital for the production of dimethylsulfoniopropionate (DMSP) in photic zones of marine environments (Curson *et al*., 2017). We chose to further analyse the identified homologs for *L. aggregata* by aligning the homologs against GPR using the NCBI Multiple Sequence Alignment Viewer 1.22.2.

Alignment of 8 of the GPR homologs from *L. aggregata* against GPR from *A. tumefaciens* revealed that a variety of different regions have been conserved across the homologs. The *A. tumefaciens* GPR protein was between 187-192 amino acids longer than the aligned homologs of *L. aggregata* (depends on the homolog). The N-terminus (approximately 370 amino acids) and C-terminus (approximately 565 amino acids) of GPR were not conserved by the *L. aggregata* homologs, through either substitutions or deletions. However, the middle 1180 remaining amino acids (residues 370-1550) demonstrated high levels of conservation (Figure 6-1). Across all 8 *L. aggregata* homologs (Accession numbers: WP_207141257.1, WP_006932722.1, WP_228929905.1, WP_206907299.1, WP_055655287.1, WP_077292424.1 WP_152510243.1, WP_023000669.1) the same conserved regions were identified. Approximately 42 individual conserved regions were seen, with these highly conserved regions varying highly in length. Between these conserved regions across the 8 *L. aggregata* homologs were multiple non-conserved substitutions and deletions separating them (Figure 6-1). These conserved regions throughout the middle section of the 8 *L. aggregata* homologs indicate their shared ancestry with GPR and suggests the function of GPR's middle section may have been conserved as well.

A



B

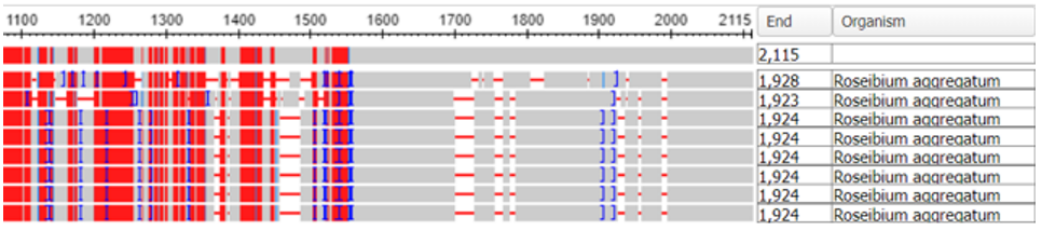
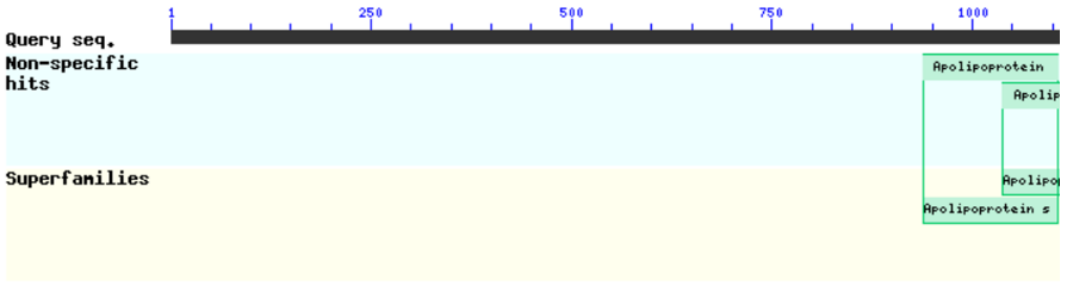


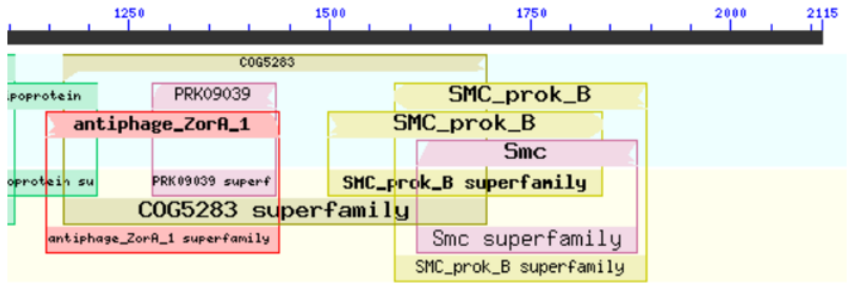
Figure 6-1. The similarity between GPR of *A. tumefaciens* and homologs identified in *L. aggregata* within a multiple alignment search using NCBI Multiple Sequence Alignment Viewer 1.22.2. (A) The alignment of GPR of *A. tumefaciens* (query) and the 8 identified *L. aggregata* homologs (with highly conserved regions highlighted in red, no conserved indicated in grey and red lines indicating gaps between aligned sequences. The alignment of GPR and the *L. aggregata* homologs from residue 1-1070 approximately of GPR (indicated by the numbered scale bar above the coloured aligned sequences). The Sequence ID column labels the aligned sequences from the second refined BLAST search. The labelled sequences are as such: Query_80180: The *A. tumefaciens* GPR sequence from UniProt (ID: A9CJ72), Accession number WP_207141257.1: first *L. aggregata* homolog hit aligned, Accession number WP_006932722.1: second *L. aggregata* homolog hit aligned, Accession number WP_228929905.1: third *L. aggregata* homolog hit aligned, Accession number WP_206907299.1: fourth *L. aggregata* homolog hit aligned, Accession number WP_055655287.1: fifth *L. aggregata* homolog hit aligned, Accession number WP_077292424.1: sixth *L. aggregata* homolog hit aligned, Accession number WP_152510243.1: seventh *L. aggregata* homolog hit aligned, Accession number WP_023000669.1: eight *L. aggregata* homolog hit aligned. The Start column indicates the first residue aligned from their respective sequence. (B) The continued alignment of GPR and the *L. aggregata* homologs from approximately residue 1070-2115 of GPR. The End column indicates the last residue aligned from their respective sequence. The Organism column indicates the organism the aligned sequence originated from. Only 2 of the aligned homologs possessed a similar N-terminus to GPR though the termini were not conserved (possibly due sufficient substitutions and deletions) between roughly residues 1-370. The remaining 6 homologs possessed a small section of this non-conserved N-terminus from residue 128-140 approximately. Between approximately residue 370-1550, multiple highly conserved regions of varying lengths were found across all 8 homologs (approximately 42 conserved regions). The C-terminus of the aligned homologs from approximately residue 1550-2115 possessed multiple aligned regions with GPR though the termini were not conserved (possibly due sufficient substitutions and deletions).

Next, to indicate if the function of GPR's middle section was conserved within the 8 *L. aggregata* homologs, the primary sequence of GPR (Uniprot ID: A9CJ72) and one of the *L. aggregata* homolog were run through the NCBI Conserved Domain Search using a Standard display. Only one *L. aggregata* homolog was run through the Conserved Domain search as the same highly conserved regions were identified across all 8 *L. aggregata* homologs so any conserved domains found within the highly conserved regions should be found across all 8 *L. aggregata* homologs. The aligned *L. aggregata* homolog (Accession number: WP_055655287.1) was used as its primary sequence had been published on UniProt (Uniprot ID: A0A0M6Y2Z9). The GPR domain search revealed a collection of different potential conserved domains across the protein, though all of the returned hits were non-specific. Within the N terminus (the first 940 amino acids) and C-terminus (last 321 amino acids) of GPR no conserved domains were identified, though within the middle section of GPR of 954 amino acids (between residues 940-1894) multiple overlapping non-specific domains were returned (Figure 6-2). Within these domains, 2 overlapping Apolipoprotein A1/A4/E domains (pfam01442) were highlighted between residues 940-1209. Following the Apolipoprotein domains, between residues 1146-1436, 2 overlapping domains: an anti-phage defense ZorAB system protein (NF033914) and a peptidoglycan-binding protein (PRK09039) were identified. Next 2 Chromosome segregation protein, SMC domains (TIGR01268) and a Chromosome segregation ATPase domain (COG1196) were identified between residues 1497-1894. All 3 of these domains overlap with each other. Finally, a Phage-related tail protein domain (COG5283) was also identified between residues 1169-1695 and overlapped with every previously mentioned domain except for one of the Apolipoprotein domains (Figure 6-2).

A



B



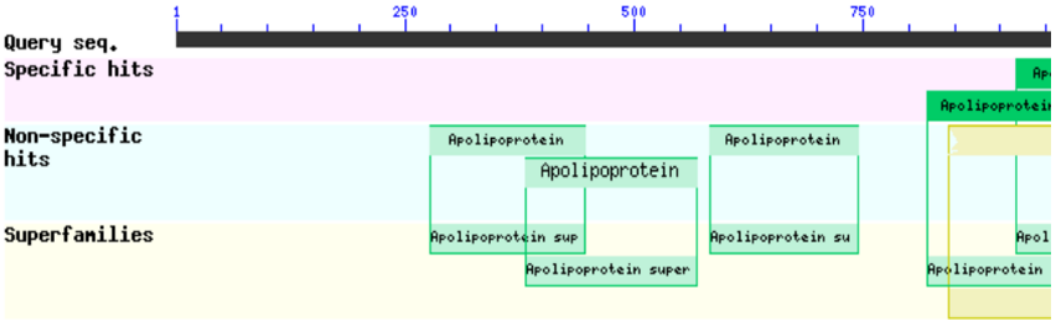
C

Name	Accession	Description	Interval	E-value
COG5283	COG5283	Phage-related tail protein [Mobilome: prophages, transposons];	1169-1695	9.57e-06
SMC_prok_B	TIGR02168	chromosome segregation protein SMC, common bacterial type; SMC (structural maintenance of ...	1581-1894	1.97e-03
Apolipoprotein	pfam01442	Apolipoprotein A1/A4/E domain; These proteins contain several 22 residue repeats which form a ...	940-1107	2.30e-03
PRK09039	PRK09039	peptidoglycan -binding protein;	1279-1432	5.50e-03
SMC_prok_B	TIGR02168	chromosome segregation protein SMC, common bacterial type; SMC (structural maintenance of ...	1497-1840	5.98e-03
Smc	COG1196	Chromosome segregation ATPase [Cell cycle control, cell division, chromosome partitioning];	1608-1883	7.65e-03
antiphage_ZorA_1	NF033914	anti-phage defense ZorAB system protein ZorA; Proteins of this subfamily are putative H+ ...	1146-1436	8.30e-03
Apolipoprotein	pfam01442	Apolipoprotein A1/A4/E domain; These proteins contain several 22 residue repeats which form a ...	1037-1209	8.70e-03

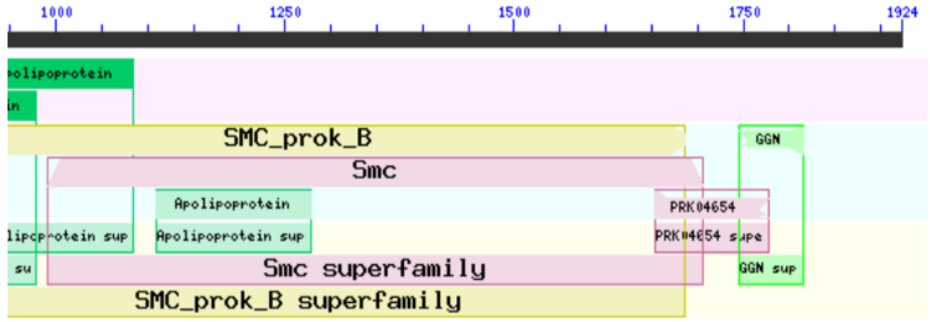
Figure 6-2. The conserved theoretical domains predicted with GPR (UniProt ID: A9CJ72) using the NCBI Conserved Domains search. (A) The potential conserved domains identified between approximately residue 1-1110 of the primary sequence of the GPR. (B) The potential conserved domains identified between approximately residue 1110-2115 of the primary sequence of the GPR. (C) NCBI Accession numbers, descriptions, aligned residues of the GPR (interval) and the significance of each returned hit. The Query sequence and scale bar indicate the primary sequence of the GPR in segments of 50 residues. 2 overlapping non-specific hits for Apolipoprotein A1/A4/E domains using the same Accession number (pfam01442) were identified between residues 940-1107 and 1037-1209 respectively. 2 non-specific hits for Chromosome segregation protein, SMC domains with the same Accession numbers (TIGR01268) were identified between residues 1581-1894 and 1497-1840 respectively. 1 non-specific hit was identified for a Phage-related tail protein domain (COG5283) between residues 1169-1695. 1 non-specific hit was identified for an anti-phage defense ZorAB system protein, ZorA domain (NF033914) between residues 1146-1436. 1 non-specific hit was identified for a peptidoglycan-binding protein domain (PRK09039) between residues 1279-1432. 1 non-specific hit was identified for a Chromosome segregation ATPase domain (COG1196) between residues 1603-1883. A non-specific Apolipoprotein domain (pfam01442), the non-specific Phage-related tail protein domain (COG5283) and the non-specific anti-phage defense ZorA protein domains (NF033914) are shown to overlap with each other. A non-specific peptidoglycan-binding protein domain (PRK09039), the non-specific Phage-related tail protein domain (COG5283) and the non-specific anti-phage defense ZorA protein domains (NF033914) are shown to overlap with each other. The non-specific Chromosome segregation protein and ATPase domains (TIGR02168 and COG1196 respectively) and the non-specific Phage-related tail protein domain (COG5283) are shown to overlap with each other. Superfamilies of the specific and non-specific hits were returned within the same ranges.

The *L. aggregata* homolog domain search returned a similar collection of domains to the GPR search, though over a larger section of the protein. The small N terminus (the first 278 amino acids) and C-terminus (last 109 amino acids) of the GPR homolog highlighted no conserved domains and the middle section of the GPR homolog of 1528 amino acids (between residues 278-1815) multiple overlapping specific and non-specific domains were returned (Figure 6-3). In a similar fashion to the GPR search, 6 Apolipoprotein A1/A4/E domains (pfam01442), both specific and non-specific domains were returned. 2 specific overlapping Apolipoprotein domains were identified between residues 820-1085. The other 4 non-specific Apolipoprotein domains were identified between residues 820-744 (contained 3 non-specific domains, 2 of them overlapping) and 1110-1279. Within these domains, 2 overlapping Apolipoprotein A1/A4/E domains (pfam01442) were highlighted between residues 940-1209. Overlapping with the specific Apolipoprotein domains, 2 overlapped domains: a Chromosome segregation protein, SMC domain (TIGR02168) and a Chromosome segregation ATPase domain (COG1196) between residues 844-1705 were displayed. Finally, a sec-independence translocase domain (PRK04654) and a Gametogenetin, GGN domain (pfam15685) were identified between residues 1653-1777 and 1745-1815 respectively. The sec-independence translocase domain overlapped with both the Chromosome segregation protein and ATPase domains and the Gametogenetin domain (Figure 6-3).

A



B



C

Name	Accession	Description	Interval	E-value
Apolipoprotein	pfam01442	Apolipoprotein A1/A4/E domain; These proteins contain several 22 residue repeats which form a ...	917-1085	1.09e-07
Apolipoprotein	pfam01442	Apolipoprotein A1/A4/E domain; These proteins contain several 22 residue repeats which form a ...	820-979	1.56e-07
SMC_prok_B	TIGR02168	chromosome segregation protein SMC, common bacterial type; SMC (structural maintenance of ...	844-1686	9.97e-07
Apolipoprotein	pfam01442	Apolipoprotein A1/A4/E domain; These proteins contain several 22 residue repeats which form a ...	278-447	4.21e-06
Smc	COG1196	Chromosome segregation ATPase [Cell cycle control, cell division, chromosome partitioning];	990-1705	1.02e-05
GGN	pfam15685	Gametogenetin; GGN is a family of proteins largely found in mammals. It reacts with POG in the ...	1745-1815	1.16e-03
Apolipoprotein	pfam01442	Apolipoprotein A1/A4/E domain; These proteins contain several 22 residue repeats which form a ...	381-568	1.29e-03
Apolipoprotein	pfam01442	Apolipoprotein A1/A4/E domain; These proteins contain several 22 residue repeats which form a ...	583-744	1.66e-03
PRK04654	PRK04654	sec-independent translocase; Provisional	1653-1777	1.95e-03
Apolipoprotein	pfam01442	Apolipoprotein A1/A4/E domain; These proteins contain several 22 residue repeats which form a ...	1110-1279	3.83e-03

Figure 6-3. The conserved theoretical domains predicted with IAM 12612 strain GPR homolog (UniProt ID: A0A0M6Y2Z9) using the NCBI Conserved Domains search. (A) The potential conserved domains identified between approximately residue 1-950 of the primary sequence of the GPR homolog. (B) The potential conserved domains identified between approximately residue 950-1924 of the primary sequence of the GPR homolog. (C) NCBI Accession numbers, descriptions, aligned residues of the GPR homolog (interval) and the significance of each returned hit. The Query sequence and scale bar indicate the primary sequence of the GPR homolog in segments of 50 residues. 2 overlapping specific hits for Apolipoprotein A1/A4/E domains using the same Accession number (pfam01442) were identified between residues 820-979 and 917-1085 respectively. 4 non-specific hits were identified as a collection of additional Apolipoprotein domains with the same Accession number (pfam01442) across 4 different intervals: residues 278-447, 381-568, 538-744 and 1110-1279 respectively. 1 non-specific hit was identified for a Chromosome segregation protein, SMC domain (TIGR02168) between residues 844-1686. 1 non-specific hit was identified for a Chromosome segregation ATPase domain (COG1196) between residues 990-1705. 1 non-specific hit was identified for a Gametogenetin, GGN domain (pfam15685) between residues 1745-1815. 1 non-specific hit was identified for a sec-independence translocase domain (PRK04654) between residues 1653-1777. The 2 specific and a non-specific Apolipoprotein domain (pfam01442) and the non-specific Chromosome segregation protein and ATPase domains (TIGR02168 and COG1196 respectively) are shown to overlap with each other. 2 of the non-specific Apolipoprotein domains (pfam01442) are shown to overlap with each other. The non-specific Chromosome segregation protein and ATPase domains (TIGR02168 and COG1196 respectively) and the non-specific sec-independence translocase domain (PRK04654) are shown to overlap with each other. The non-specific sec-independence translocase domain (PRK04654) and the non-specific sec-independence translocase domain (pfam15685) are shown to overlap with each other. Superfamilies of the specific and non-specific hits were returned within the same ranges.

The GPR domain search revealed that between residues 370-1695 contained 2 Apolipoprotein domains (pfam01442) alongside a collection of other domains: an anti-phage defense ZorAB system protein (NF033914), a peptidoglycan-binding protein (PRK09039) and a Phage-related tail protein domain (COG5283) were virtually all confined within the conserved middle section of GPR (residues 370-1550) indicated by the alignment of GPR against 8 *L. aggregata* homologs (Figure 6-1). Moreover, between residues 1497-1894, were 3 overlapping domains: 2 Chromosome segregation protein, SMC domains (TIGR01268) and a Chromosome segregation ATPase domain (COG1196) were contained within the conserved middle section and non-conserved C-terminus (residues 1550-2115) of GPR (Figure 6-1).

Furthermore, the *L. aggregata* homolog domain search also returned many of the same hits. 6 individual hits for Apolipoprotein domain (pfam01442), a Chromosome segregation protein, SMC domain (TIGR02168) and a Chromosome segregation ATPase domain (COG1196) were identified between residues 278-1705 (the middle section of the *L. aggregata* homolog). Additionally, the deletion of Apolipoprotein domains and the C-terminus of GPR have been shown to play a role in the synthesis of peptidoglycan through cell morphology defects (Zupan *et al*., 2021). As many of the same hits were returned within the middle section of the *L. aggregata* homolog, which has been shown to be a highly conserved region between the homolog and GPR (Figure 6-1), this suggests the function of the Apolipoprotein (pfam01442), the Chromosome segregation protein (TIGR02168) and Chromosome segregation ATPase (COG1196) domains of GPR have been conserved within the *L. aggregata* homologs.

With the GPR homologs in *L. aggregata* possessing the many of the conserved domains as GPR which have shown to be vital to coordinating polar growth, we choose to investigate the localisation pattern of the *L. aggregata* GPR homolog to confirm if the *L. aggregata* homolog is involved in polar growth. We were able to acquire the *L. aggregata* strain LZB033 however the sequenced genome is yet to be submitted. We identified a GPR homolog from *L. aggregata* strain LZB033 (Gene annotation: FIG00742013) which showed remarkable similarity (2 single base substitutions identified by Clustal Omega) to the GPR homolog (UniProt ID: A0A0M6Y2Z9) from the published *L. aggregata* IAM 12614 genome on UniProt (Figure 6-4 and 6-5). Due to the high similarity of the LZB033 homolog to the IAM 12614 homolog, we continued investigating the function of the LZB033 homolog and designated the gene *lcy* (*labrenzia* cytoskeletal protein).

```
FIG00742013     MANPTKAKDPAEAALSAVEEALKLDFGGPDTTETTSAAEASPAQAEPARRSQGETRGQSQ   60
A0A0M6Y2Z9      MANPTKAKDPAEAALSAVEEALKLDFGGPDTTETTSAAEASPAQAEPARRSQGETRGQSQ   60
                ************************************************************

FIG00742013     REEQRQSRRRGRGGRPPAANDDRRNIGNLIYSLQRRPSSAPFWGALALSALWAALGSSLF   120
A0A0M6Y2Z9      REEQRQSRRRGRGGRPPAANDDRRNIGNLIYSLQRRPSSAPFWGALALSALWAALGSSLF   120
                ************************************************************

FIG00742013     MTAFPDKVGSLTDPQTLLSSPEMILAVVGIVVPIIFFFVMAMMIWRAQEMRIVARGMTEV   180
A0A0M6Y2Z9      MTAFPDKVGSLTDPQTLLSSPEMILAVVGIVVPIIFFFVMAMMIWRAQEMRIVARGMTEV   180
                ************************************************************

FIG00742013     ALRLAEPEDMAKESILSVGQAIRREVAAMGDGIERAIARASELEVLVHNEVSSLERSYND   240
A0A0M6Y2Z9      ALRLAEPEDMAKESILSVGQAIRREVAAMGDGIERAIARASELEVLVHNEVSSLERSYND   240
                ************************************************************

FIG00742013     NELKIRALIDELISQRESIVMNAERVRETIAGAHESFASQLSSTSGELGSTVDHATQRMI   300
A0A0M6Y2Z9      NELKIRALIDELISQRESIVMNAERVRETIAGAHESFASQLSSTSGELGSTVDHATQRMI   300
                ************************************************************

FIG00742013     DAVNSRVEELTSTVDSRIESLGATLNASGNELVDSLTVRAEDYVARLSSTGNDLVDSLAE   360
A0A0M6Y2Z9      DAVNSRVEELTSTVDSRIESLGATLNASGNELVDSLTVRAEDYVARLSSTGNDLVDSLAE   360
                ************************************************************

FIG00742013     TGSTMWETLTARGNEVNERFADTANTFVETLSARSTTINETLATSSNSVINTLSEKADEF   420
A0A0M6Y2Z9      TGSTMWETLTARGNEVNERFADTANTFVESLSARSTTINETLATSSNSVINTLSEKADEF   420
                ****************************:*******************************

FIG00742013     RLTLETTGNNVGEVITARGEEINANLSLTSGRLIDTITSRTEELVATVDNRVTTLDESLA   480
A0A0M6Y2Z9      RLTLETTGNNVGEVITARGEEINANLSLTSGRLIDTITSRTEELVATVDNRVTTLDESLA   480
                ************************************************************

FIG00742013     ETGNRVVESISEKGQAVTDTISIRGAEIVETLSSRSTEVAEILRGTGESIVVDLSLRGGE   540
A0A0M6Y2Z9      ETGNRVVESISEKGQAVTDTISIRGAEIVETLSSRSTEVAEILRGTGESIVVDLSLRGGE   540
                ************************************************************

FIG00742013     IASKLDETAGSLTQTISVRGGELAEKLDNISERIYTAISINGSELDERLAARSNEMATIL   600
A0A0M6Y2Z9      IASKLDETAGSLTQTISVRGGELAEKLDNISERIYTAISINGSELDERLAARSNEMATIL   600
                ************************************************************

FIG00742013     EQQTIGFRETLESVSGEFSASLGDQTGALTQKLAETGTQLAELIGSRGDRVAGDINQISG   660
A0A0M6Y2Z9      EQQTIGFRETLESVSGEFSASLGDQTGALTQKLAETGTQLAELIGSRGDRVAGDINQISG   660
                ************************************************************

FIG00742013     KIAETLEVRGQALQDGLSSRLTELETIVTDRGGQLIDAFDTKTLLLSEALDSRLSTLDTT   720
A0A0M6Y2Z9      KIAETLEVRGQALQDGLSSRLTELETIVTDRGGQLIDAFDTKTLLLSEALDSRLSTLDTT   720
                ************************************************************

FIG00742013     FATRIDSMDASLGDRIATMDASLDQRYSAIDATLTDRIAAMDATLEQRYTAIDTTLTDRI   780
A0A0M6Y2Z9      FATRIDSMDASLGDRIATMDASLDQRYSAIDATLTDRIAAMDATLEQRYTAIDTTLTDRI   780
                ************************************************************

FIG00742013     ATMDATLEQRYTAIDTTLTDRIATMDSSLEHRFTTMDAGLGERIGTMDASLEQRYALMDA   840
A0A0M6Y2Z9      ATMDATLEQRYTAIDTTLTDRIATMDSSLEHRFTTMDAGLGERIGTMDASLEQRYALMDA   840
                ************************************************************

FIG00742013     TLTDRISTLDNSLEHRLSSMDSSLSERISTMDSSLGSHINTLDLTLDQRTAAFEATLEAR   900
A0A0M6Y2Z9      TLTDRISTLDNSLEHRLSSMDSSLSERISTMDSSLGSHINTLDLTLDQRTAAFEATLEAR   900
                ************************************************************

FIG00742013     TQILSDAIENRTTGISDALEEKTRNITDVLADRSDAITLQLGQRIEAAGNTLAERAEEIG   960
A0A0M6Y2Z9      TQILSDAIENRTTGISDALEEKTRNITDVLADRSDAITLQLGQRIEAAGNTLAERAEEIG   960
                ************************************************************

FIG00742013     QSLSQRVDTATSAMADKAELISASMTSGTDRIDETLDARARQISETLISRTKEIAKAFVD   1020
A0A0M6Y2Z9      QSLSQRVDTATSAMADKAELISASMTSGTDRIDETLDARARQISETLISRTKEIAKAFVD   1020
                ************************************************************

FIG00742013     GQDEMTSALDNRLAEAGSVLGKQSEQLTESLSERIAEINVSLGAKVFEVAETLDSRATQL   1080
A0A0M6Y2Z9      GQDEMTSALDNRLAEAGSVLGKQSEQLTESLSERIAEINVSLGAKVFEVAETLDSRATQL   1080
                ************************************************************
```

Figure 6-4. Protein sequence alignment between the LZB033 strain GPR homolog (Gene annotation: FIG00742013) and IAM 12612 strain GPR homolog (UniProt ID: A0A0M6Y2Z9) primary sequences of *L. aggregata*, aligned using Clustal Omega. (*) represents positions with a conserved residue, (:) indicates conserved amino acid groups with strongly similar properties and (.) represents conserved amino acid groups with weakly similar properties. Residues that differ between the aligned homologs are highlighted.

```
FIG00742013     ETLLNERLESISGTLTGESERARDILTAVISEAGATLSTESTRLRDVVQEAVTAAVQTLS   1140
A0A0M6Y2Z9      ETLLNERLESISGTLTGESERARDILTAVISEAGATLSTESTRLRDVVQEAVTAAVQTLS   1140
                ************************************************************

FIG00742013     DERNRTVEIVDSALSQATEKLSGEGDRMRQIVLGSVGEARGVLVGESQKAADAVTSAMSQ   1200
A0A0M6Y2Z9      DERNRTVEIVDSALSQATEKLSGEGDRMRQIVLGSVGEARGVLVGESQKAADAVTSAMSQ   1200
                ************************************************************

FIG00742013     ATGAMSGESGKIRELVLSAVGDAARAMAAESEKARTLYAGTLAEFSGSLTGESDKVRNEL   1260
A0A0M6Y2Z9      ATGAMSGESGKIRELVLSAVGDAARAMAAESEKARTLYAGTLAEFSGSLTGESDKVRNEL   1260
                ************************************************************

FIG00742013     SGLIAEISGNLSAESEQARMTLAKTLEEIRGQMSSEAGMVRARVNSAVTEAADLLVGRGN   1320
A0A0M6Y2Z9      SGLIAEISGNLSAESEQARMTLAKTLEEIRGQMSSEAGMVRARVNSAVTEAADLLVGRGN   1320
                ************************************************************

FIG00742013     EVANELLDKATALNEAFGARSGELAKIVGTDGNELISAIELRANDLTGRLSEVHTAILEA   1380
A0A0M6Y2Z9      EVANELLDKATALNEAFGARSGELAKIVGTDGNELISAIELRANDLTGRLSEVHTAILEA   1380
                ************************************************************

FIG00742013     ITVKGRDVTDTFAHTGLDATRSLVEAGDRIVASINERSEQATTLLTDTKRRLEADVTEIL   1440
A0A0M6Y2Z9      ITVKGRDVTDTFAHTGLDATRSLVEAGDRIVASINERSEQATTLLTDTKRRLEADVTEIL   1440
                ************************************************************

FIG00742013     NKIETSNSNLQGIVATAGENLNEVEGNLARRAGEFRSAVDRAVSETNSTTALITEQVANL   1500
A0A0M6Y2Z9      NKIETSNSNLQGIVATAGENLNEVEGNLARRAGEFRSAVDRAVSETNSTTALITEQVANL   1500
                ************************************************************

FIG00742013     RDVTDTTLADIKNLTGRFGDQSEELTKAARHLEDTNRSVESRVSERRTAIEDVADTLLAK   1560
A0A0M6Y2Z9      RDVTDTTLADIKNLTGRFGDQSEELTKAARHLEDTNRSVESRVSERRTAIEDVADTLLAK   1560
                ************************************************************

FIG00742013     TEAVDTLMRSFTKTLSETLETADDKARDAANMLSAAAEAASKQVSEQFESMRLTAGMEGQ   1620
A0A0M6Y2Z9      TEAVDTLMRSFTKTLSETLETADDKARDAANMLSAAAEAASKQVSEQFESMRLTAGMEGQ   1620
                ************************************************************


FIG00742013     KAREAIRSAQDDIIAEMTRTVTEASDRFNDAAARMRDVARDVHRELEATRSELKQGVLNL   1680
A0A0M6Y2Z9      KAREAIRSAQDDIIAEMTRTVTEASDRFNDAAARMRDVARDVHRELEATRSELKQGVLNL   1680
                ************************************************************

FIG00742013     PEEAEESSAALRKVVNEQIRALTELSEIVAKQSNALDISRPQAQAAVASAAPAPAPQQQA   1740
A0A0M6Y2Z9      PEEAEESSAALRKVVNEQIRALTELSEIVAKQSNALDISRPQAQAAVASAAPAPAPQQQA   1740
                ************************************************************

FIG00742013     PVNYAPAAPARQQSAPASEPRPAPVTTRRQQPAPAETTSSGGKGWVADLLRRASRDDDAG   1800
A0A0M6Y2Z9      PVNYAPAAPARQQSAPASEPRPAPVTTRRQQPAPAETTSSGGKGWVADLLRRASRDDDAS   1800
                ************************************************************.

FIG00742013     SDTDMSRTPLHTVESLNSLSVDIARAIDHETFVDLWNRYRNGERHVFTRRLYTLQGQQTF   1860
A0A0M6Y2Z9      SDTDMSRTPLHTVESLNSLSVDIARAIDHETFVDLWNRYRNGERHVFTRRLYTLQGQQTF   1860
                ************************************************************

FIG00742013     DEIRQKYSRDPEFRTAVERYVADFEQLLAQVSRNDRDNMLGQTYLTSDTGKVYTMLAHAS   1920
A0A0M6Y2Z9      DEIRQKYSRDPEFRTAVERYVADFEQLLAQVSRNDRDNMLGQTYLTSDTGKVYTMLAHAS   1920
                ************************************************************

FIG00742013     GRLD   1924
A0A0M6Y2Z9      GRLD   1924
                ****
```

Figure 6-5. Continued protein sequence alignment between the LZB033 strain GPR homolog (Gene annotation: FIG00742013) and IAM 12612 strain GPR homolog (UniProt ID: A0A0M6Y2Z9) of *L. aggregata*, aligned using Clustal Omega. (*) represents positions with a conserved residue, (:) indicates conserved amino acid groups with strongly similar properties and (.) represents conserved amino acid groups with weakly similar properties. Residues that differ between the aligned homologs are highlighted.

To investigate this potential homolog of GPR, the primary sequence of Lcy was passed through the Expasy ProtParam program to determine some of the physical and chemical properties of the protein. Lcy is constructed of 1924 amino acids, has a predicted molecular weight of 207.51621 kDa, a theoretical pI of 4.66 and an II of 37.80. Since the II is less than 40, the protein is deemed to be stable *in vitro*. Further analysis of the potential GPR homolog was conducted through the fusion of fluorescent tags to the C-terminus of the gene to monitor how and where Lcy localises to infer any similarities to GPR in function. A collection of fluorescent tags was generated, containing a range of differences in the linkers, fluorescent tags and resistance markers. A suitable fluorescent tag was selected from the collection and fused to the end of Lcy. The fusion was introduced into the *L. aggregata* LZB033 genome to monitor Lcy localisation.

6.2.2 Generation of fluorescent tags

In order to knock-in different fluorescent tags we use *E. coli* BW25113, which possess the pIJ790 plasmid. The pIJ790 plasmid contains the λ-RED recombinase, which is inducible with arabinose. Expression of λ-RED recombinase allows for the recombination of a complementary short, double-stranded DNA donor segment at a targeted location. To knock-in fluorescence genes we need to screen for successful recombination and this is achieved by using resistance markers that are also part of the knock-in cassettes. Previous cassettes use *egfp* and apramycin resistance genes. These cassettes are included in plasmids generated previously in the Kelemen lab which were: pUC18-ProEgfpApra, pUC18-GlyEgfpApra and pUC18-GlymCherryApra, where *egfp* and *mCherry* represent green and red fluorescent proteins, Apra - apramycin resistance and Pro and Gly refers to a four amino acid long linker upstream of the fluorescent proteins. The Kelemen lab had also generated the components for a pUC18-PromCherryApra construct. The pUC18-PromCherryApra construct was generated using the ligation of a digested pUC18-Apra plasmid and PromCherry cassette. All the constructs were transformed into *E. coli* DH5α for identification and purification of functional constructs. With successful confirmation of the Apra-based fluorescent tags, 4 new fluorescent tags were generated though recombination of a complementary donor DNA segment (containing a Spec cassette) the use of the λ-RED recombinase system on each Apra-based tag. This was to generate Spec-variants of each Apra-based fluorescent tag. The donor DNA segments were complementary due to being flanked by 2 extensions (a 46 bp extension before the start of Spec gene, and a 6 bp extension

following the end of the gene) on both ends which are homologous to the desired region within the Apra gene (the first 42 bp ad last 36 bp of the Apra gene). The complementary donor DNA segment will be generated by specific primers through PCR. The Apratospec forward primer and Apratospec reverse primer were used to produce a complementary Spec cassette using a Spec cassette generated by a previous researcher from Dr. Kelemen's lab. With this in mind, the complementary Spec/Strep resistance cassette to the Apra resistance marker is transformed into the *E. coli* BW25113 strains containing the Apra-based tags. Induced expression of the λ-RED recombinase swaps the Apra resistance marker within each construct for the Spec/Strep resistance cassette, resulting in the generation of Spec variants of each Apra-based tag. Collectively, this should result in 8 distinct fluorescent tags being generated.

As stated before, the generated pUC18-ProEgfpApra and pUC18-GlyEgfpApra constructs were transformed into DH5α and isolated using plasmid DNA isolation. Both constructs were confirmed via Agarose gel electrophoresis (Figure 6-6). To generate the Spec variant of the *egfp* tags, the Apra-based constructs were electroporated separately into BW25113 and both strains' λ-RED recombinase was induced via Arabinose. The complementary Spec cassette was electroporated into the induced BW25113 strains, with successful recombination selected for on SOB containing streptomycin. Since the successful Spec-based constructs have been selected for, suitable colonies were selected for the Spec-based constructs and isolated via mini plasmid DNA isolation. The mini plasmid DNA isolations, alongside the Apra-based constructs, were separately transformed into DH5α for isolation via plasmid DNA isolation. All isolations were confirmed via Agarose gel electrophoresis (Figure 6-7).

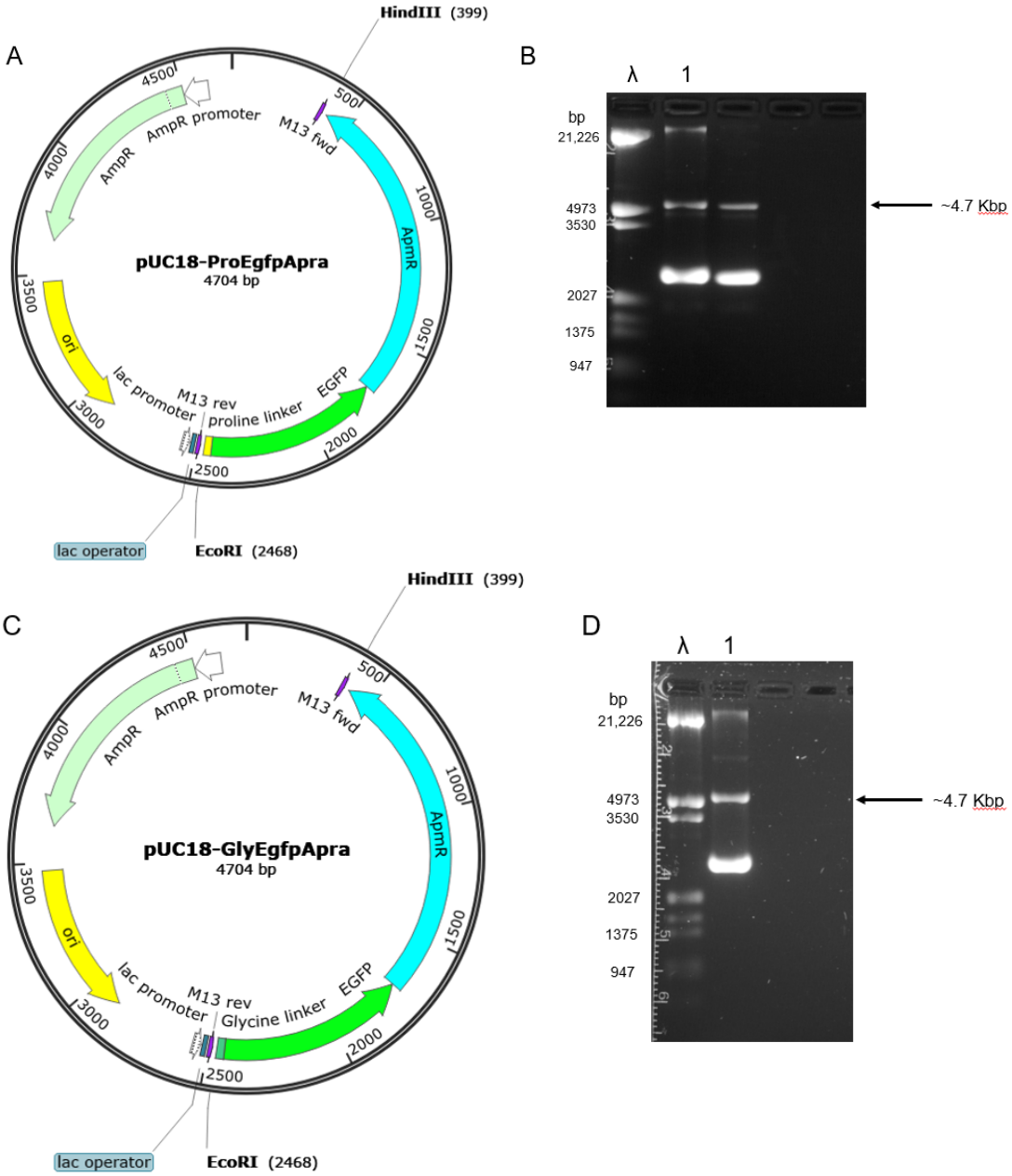


Figure 6-6. Vector maps and confirmations of the egfpApra-based constructs. (A) The vector map of the pUC18-ProEgfpApra construct. The ProEgfpApra cassette is situated between the *EcoRI* and *HindIII* restriction sites and possess a pMB1 replication origin. (B) Confirmation of the isolated ProEgfpApra construct on 0.7% Agarose gel. The loaded lanes are as such: Lane λ: Lambda DNA *EcoRI-HindIII* ladder, Lane 1: pUC18-ProEgfpApra. (C) The vector map of the pUC18-GlyEgfpApra construct. The GlyEgfpApra cassette is situated between the *EcoRI* and *HindIII* restriction sites and possess a pMB1 replication origin. (D) Confirmation of the isolated GlyEgfpApra construct on 0.7% Agarose gel. The loaded lanes are as such: Lane λ: Lambda DNA *EcoRI-HindIII* ladder, Lane 1: pUC18-GlyEgfpApra.

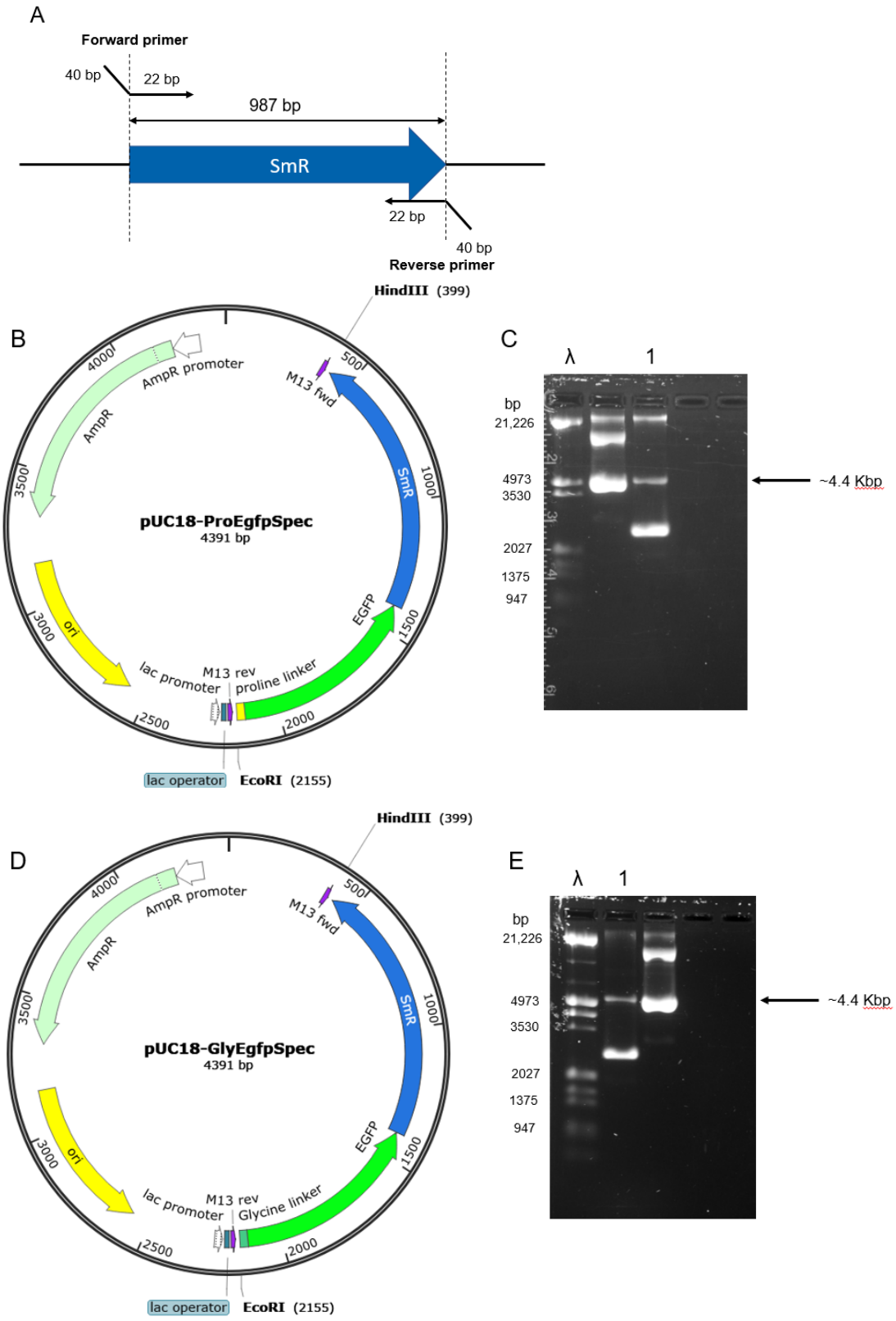


Figure 6-7. Vector maps and confirmations of the egfpSpec-based constructs. (A) A diagram of the generation of the complementary Spec cassette via PCR using primers that anneal 40 bp before and after the template Spec cassette. (B) The vector map of the pUC18-ProEgfpSpec construct. The ProEgfpSpec cassette is situated between the *EcoRI* and *HindIII* restriction sites and possess a pMB1 replication origin. (C) Confirmation of the isolated ProEgfpSpec construct on 0.7% Agarose gel. The loaded lanes are as such: Lane λ: Lambda DNA *EcoRI-HindIII* ladder, Lane 1: pUC18-ProEgfpSpec. (D) The vector map of the pUC18-GlyEgfpApra construct. The GlyEgfpSpec cassette is situated between the *EcoRI* and *HindIII* restriction sites and possess a pMB1 replication origin. (E) Confirmation of the isolated GlyEgfpSpec construct on 0.7% Agarose gel. The loaded lanes are as such: Lane λ: Lambda DNA *EcoRI-HindIII* ladder, Lane 1: pUC18-GlyEgfpSpec.

In the same fashion as the *egfp*-based constructs, the mCherryApra-based constructs (pUC18-PromCherryApra and pUC18-GlymCherryApra) were transformed into DH5α and isolated via plasmid DNA isolation. Both constructs were confirmed via Agarose gel electrophoresis (Figure 6-8). To generate the Spec

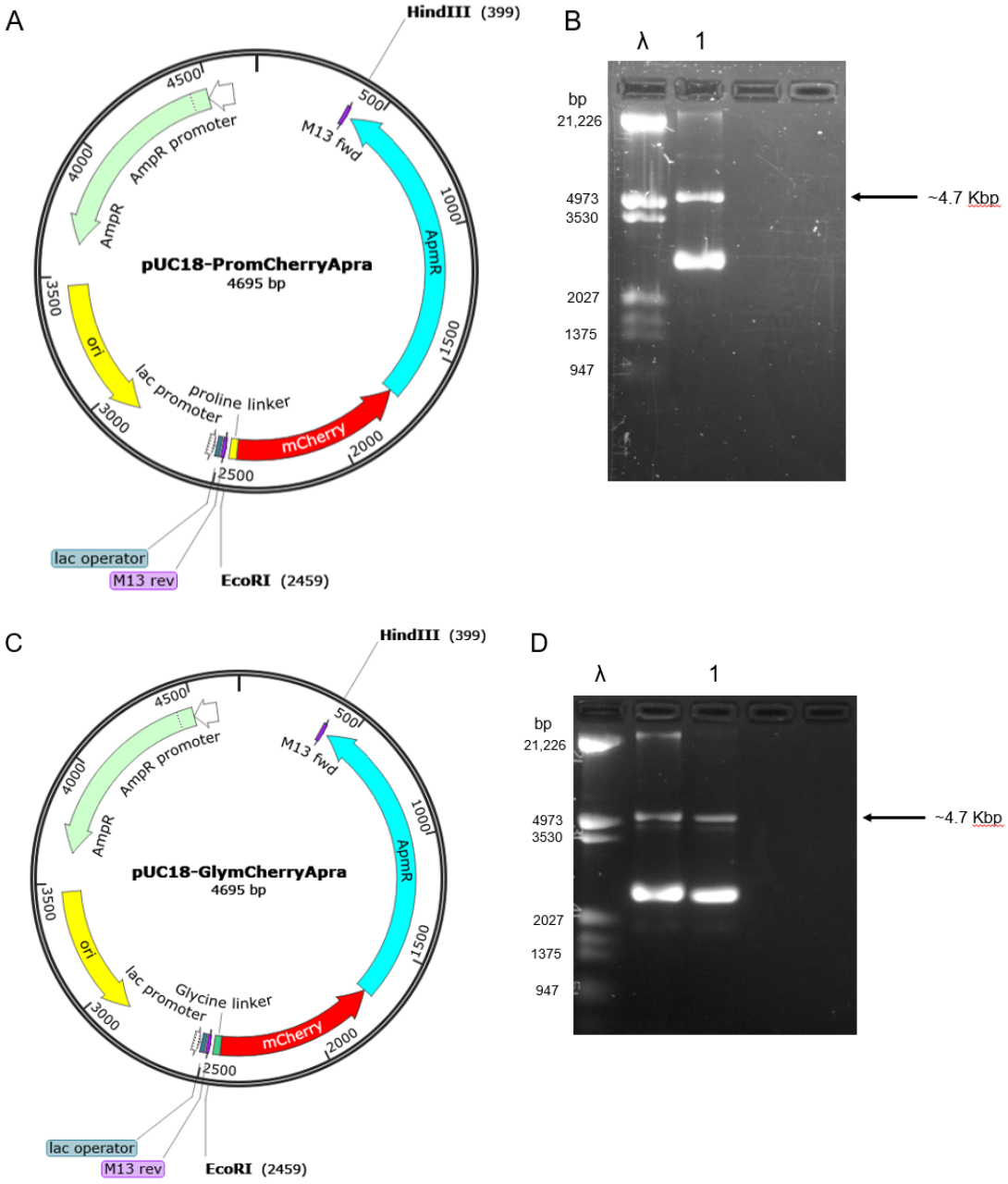


Figure 6-8. Vector maps and confirmations of the mCherryApra-based constructs. (A) The vector map of the pUC18-PromCherryApra construct. The PromCherryApra cassette is situated between the *EcoRI* and *HindIII* restriction sites and possess a pMB1 replication origin. (B) Confirmation of the isolated PromCherryApra construct on 0.7% Agarose gel. The loaded lanes are as such: Lane λ: Lambda DNA *EcoRI-HindIII* ladder, Lane 1: pUC18-PromCherryApra. (C) The vector map of the pUC18-GlymCherryApra construct. The GlymCherryApra cassette is situated between the *EcoRI* and *HindIII* restriction sites and possess a pMB1 replication origin. (D) Confirmation of the isolated GlymCherryApra construct on 0.7% Agarose gel. The loaded lanes are as such: Lane λ: Lambda DNA *EcoRI-HindIII* ladder, Lane 1: pUC18-GlymCherryApra.

variant of the *mCherry* tags, the same techniques for the *egfp*-based Spec constructs were employed. Successful *mCherry*-based Spec recombinations were selected for on SOB containing streptomycin, with suitable colonies being selected and the *mCherry*-based Spec constructs being isolated via mini plasmid DNA isolation. The mini plasmid DNA isolations, alongside the Apra-based constructs, were separately transformed into DH5α for isolation via plasmid DNA isolation. The mCherrySpec isolations were confirmed via Agarose gel electrophoresis (Figure 6-9).

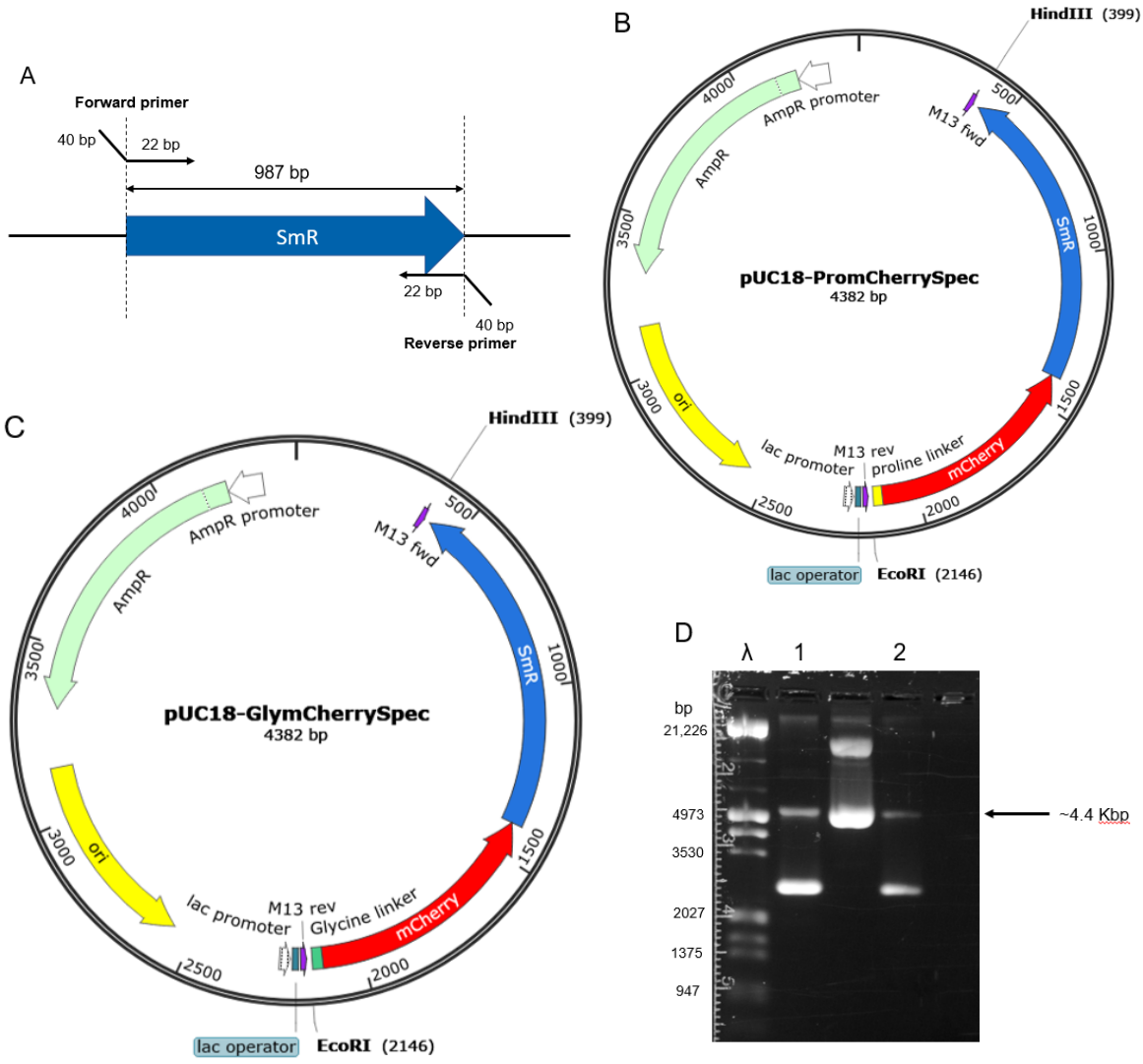


Figure 6-9. Vector maps and confirmations of the mCherrySpec-based constructs. (A) A diagram of the generation of the complementary Spec cassette via PCR using primers that anneal 40 bp before and after the template Spec cassette. (B) The vector map of the pUC18-PromCherrySpec construct. The PromCherrySpec cassette is situated between the *EcoRI* and *HindIII* restriction sites and possess a pMB1 replication origin. (C) The vector map of the pUC18-GlymCherrySpec construct. The GlymCherrySpec cassette is situated between the *EcoRI* and *HindIII* restriction sites and possess a pMB1 replication origin. (D) Confirmation of the isolated PromCherrySpec and GlymCherrySpec constructs on 0.7% Agarose gel. The loaded lanes are as such: Lane λ: Lambda DNA *EcoRI-HindIII* ladder, Lane 1: pUC18-PromCherrySpec, Lane 2: pUC18-GlymCherrySpec.

With all 8 of the fluorescent tag constructs confirmed and isolated, the pUC18-ProEgfpApra construct was selected for fusion with Lcy due to the purity of the DNA isolation.

6.2.3 Generation and monitoring of LcyEgfp fusion

With the most suitable tag being selected from the 8 fluorescent tags generated, the ProEgfpApra tag was prepared for fusion with the *lcy* gene. To construct the *lcy* fusion, we first generate a Wildtype copy of the *lcy* gene from *L. aggregata* LZB033 via PCR using specific forward and reverse primers. With generation of the *lcy* gene, the pK18MobSacB plasmid and isolated *lcy* gene was digested with *EcoRI*, ligated together and transformed into BW25113 for insertion of the selected fluorescent tag. The selected fluorescent tag was digested with *EcoRI* and *HindIII* to isolate of the fluorescent tag and resistance marker (the fluorescent tag cassette) from the pUC18 vector, via an Agarose gel. The isolated fluorescent tag cassette was extended through PCR to possess complementary 40 bp extensions to allow for insertion of the complementary fluorescent tag cassette after *lcy* fragment through the λ-RED recombinase system in BW25113. The isolated extended fluorescent tag cassette was transformed into induced BW25113 pK18mobsacB-Lcy for the insertion of the fluorescent tag cassette to generate a fusion with Lcy. The Lcy fusion will then be transformed into DH5α for use as a donor strain in Triparental mating to transfer the construct to *L. aggregata*. The helper strain *E. coli* HB101 pRK2013 will be utilised to facilitate the Triparental mating as pRK2013 is a self-transmissible plasmid which is transferred to the donor strain. Within the donor strain, pRK2013 can interact with the pK18mobsacB-Lcy fusion construct (which isn't self-transmissible) and transfer the construct into *L. aggregata* through conjugation. Within *L. aggregata*, homologous recombination between the genome and Lcy fusion can occur wherein the Lcy fusion construct can be switched with the Wildtype *lcy* gene in the chromosome due to both complementary 40 bp extensions flanking the construct. However, due to the unpredictable nature of homologous recombination, the entire pK18mobsacB vector can be inserted if only a single recombination event occurs. For this reason, the pK18mobsacB plasmid (Figure 6-10) was selected for the vector to transfer the Lcy fusion to *L. aggregata* as the sacB gene prevents growth on sucrose supplemented media. As a result, allows for the selection of the pK18mobsacB plasmid, alongside the resistance marker for the Lcy fusion, to confirm transformation with a double recombination event (double crossover).

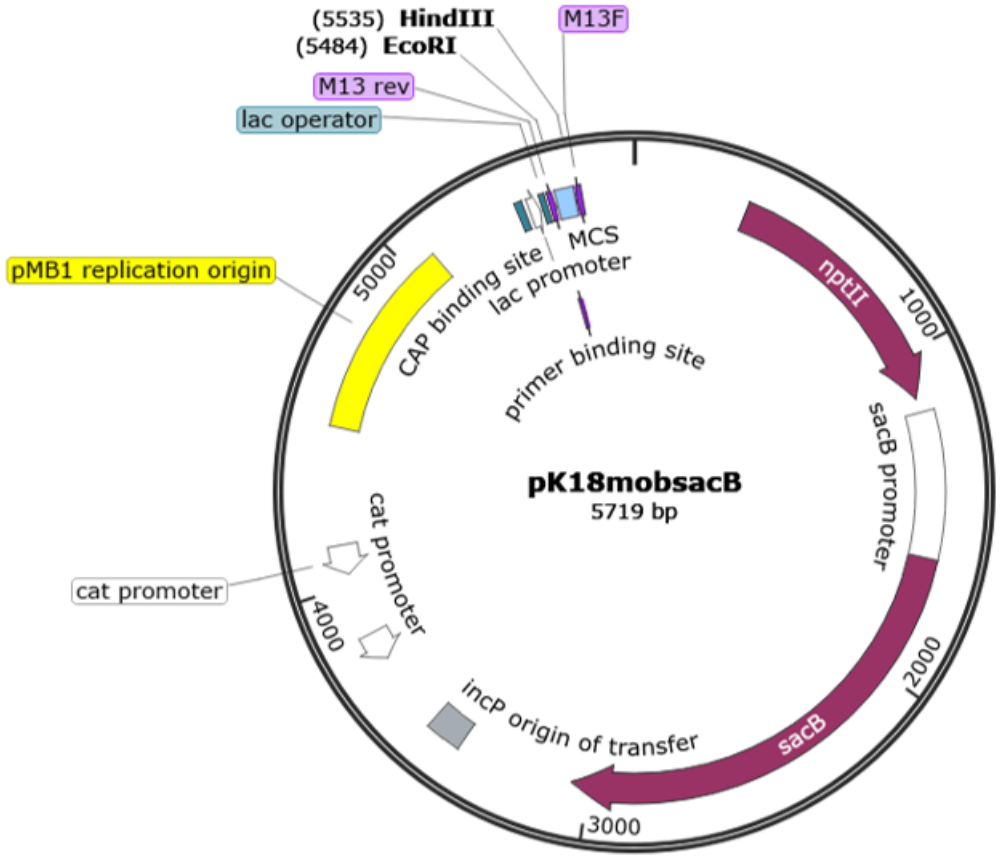


Figure 6-10. The pK18mobsacB vector used for the transportation of the Lcy fusion into *L. aggregata*. pK18mobsacB contains a MCS which contains restriction sites for *EcoRI* and *HindIII*. Following the MCS is the sacB gene for selection of pK18mobsacB vector (single crossover events) through Sucrose inhibition of growth. A pMB1 replication origin is also carried within the plasmid.

As stated previously, a copy of the *lcy* gene from *L. aggregata* was generated from *L. aggregata* chromosomal DNA via PCR using the LcyEgfpEco FRW forward primer and the LcyEgfpEco REV reverse primer. The isolated *lcy* gene will be purified via PCR purification and digested with *EcoRI*, alongside the pK18mobsacB plasmid. The pK18mobsacB digestion was gel extracted before ligation. Both digestions will be ligated together (creating pK18mobsacB-Lcy) and transformed into DH5α for isolation of the construct through plasmid DNA isolation. pK18mobsacB-Lcy was digested with *EcoRI* for confirmation (Figure 6-11).

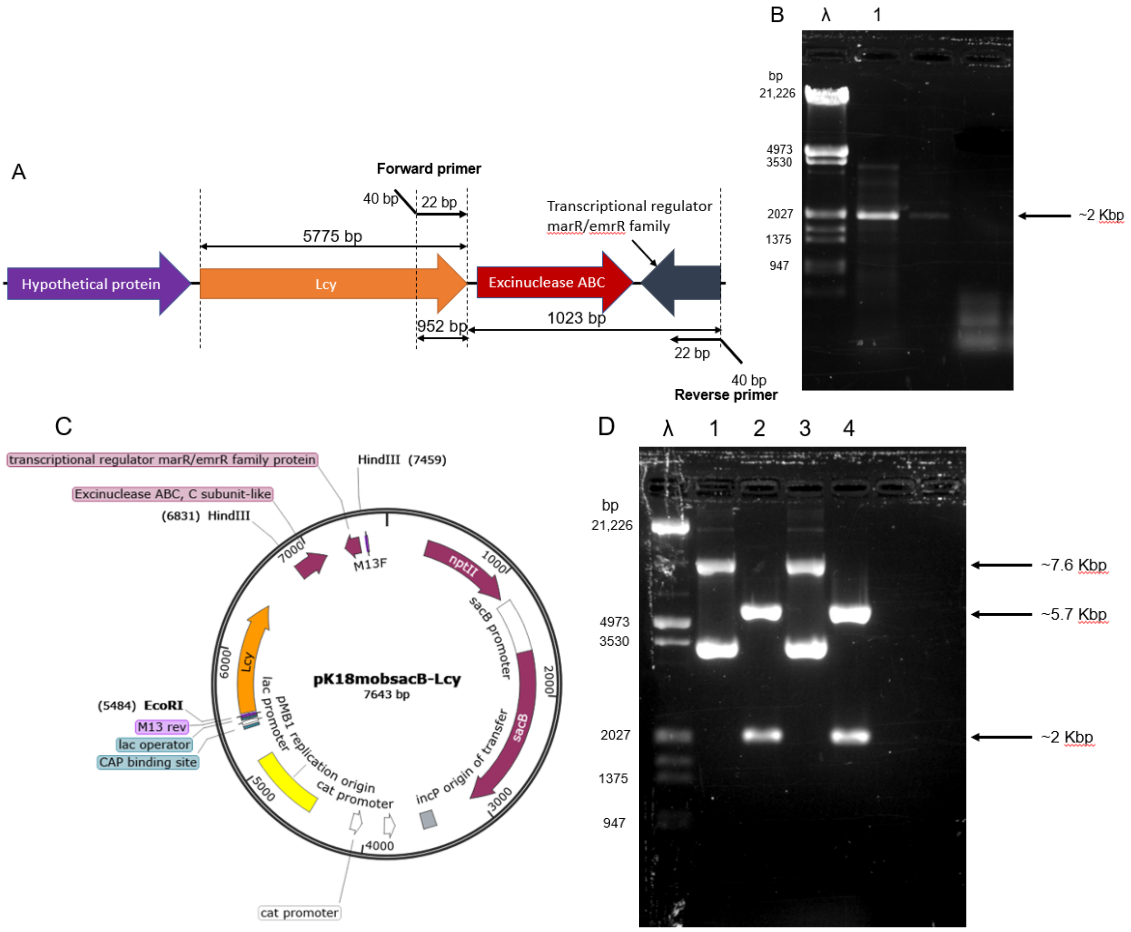


Figure 6-11. Diagram of the generation of the *lcy* fragment from *L. aggregata* chromosome, with vector maps and confirmation of the pK18mobsacB-Lcy construct. (A) A diagram of isolating the *lcy* fragment from the *L. aggregata* chromosome via PCR using primers that anneal 40 bp before and after the gene. (B) Confirmation of the isolated *lcy* fragment on 0.7% Agarose gel. The loaded lanes are as such: Lane λ: Lambda DNA *EcoRI-HindIII* ladder, Lane 1: *lcy* fragment. (C) The vector map of the pK18mobsacB-Lcy construct. The *lcy* gene is situated between the *EcoRI* and *HindIII* restriction sites and possess a pMB1 replication origin. (D) Confirmation of the isolated and digested pK18mobsacB-Lcy constructs on 0.7% Agarose gel. The constructs were digested with *EcoRI* and *HindIII*. The loaded lanes are as such: Lane λ: Lambda DNA *EcoRI-HindIII* ladder, Lane 1: pK18mobsacB-Lcy #1, Lane 2: digested pK18mobsacB-Lcy #1, Lane 3: pK18mobsacB-Lcy #2, Lane 4: digested pK18mobsacB-Lcy #2.

The isolated pK18mobsacB-Lcy construct was then transformed into BW25113 via electroporation for recombination with the ProEgfpApra cassette. In addition, the ProEgfpApra cassette, from pUC18-ProEgfpApra, was digested with *EcoRI* and *HindIII* and gel extracted to isolate the cassette. The cassette was then extended via PCR using the LcyEgfpKI FRW forward primer and the LcyEgfpKI REV reverse primer to generate the 40 bp complementary extensions. The complementary ProEgfpApra cassette was transformed into induced BW25113 pK18mobsacB-Lcy via electroporation for recombination and selected for on LB containing apramycin. The successfully generated pK18mobsacB-LcyProEgfpApra

constructs were isolated and extracted via mini plasmid DNA isolation and subsequently transformed into DH5α for Triparental mating and plasmid DNA isolation. The isolated pK18mobsacB-LcyProEgfpApra construct was digested with *XbaI* and confirmed via Agarose gel electrophoresis (Figure 6-12). DH5α

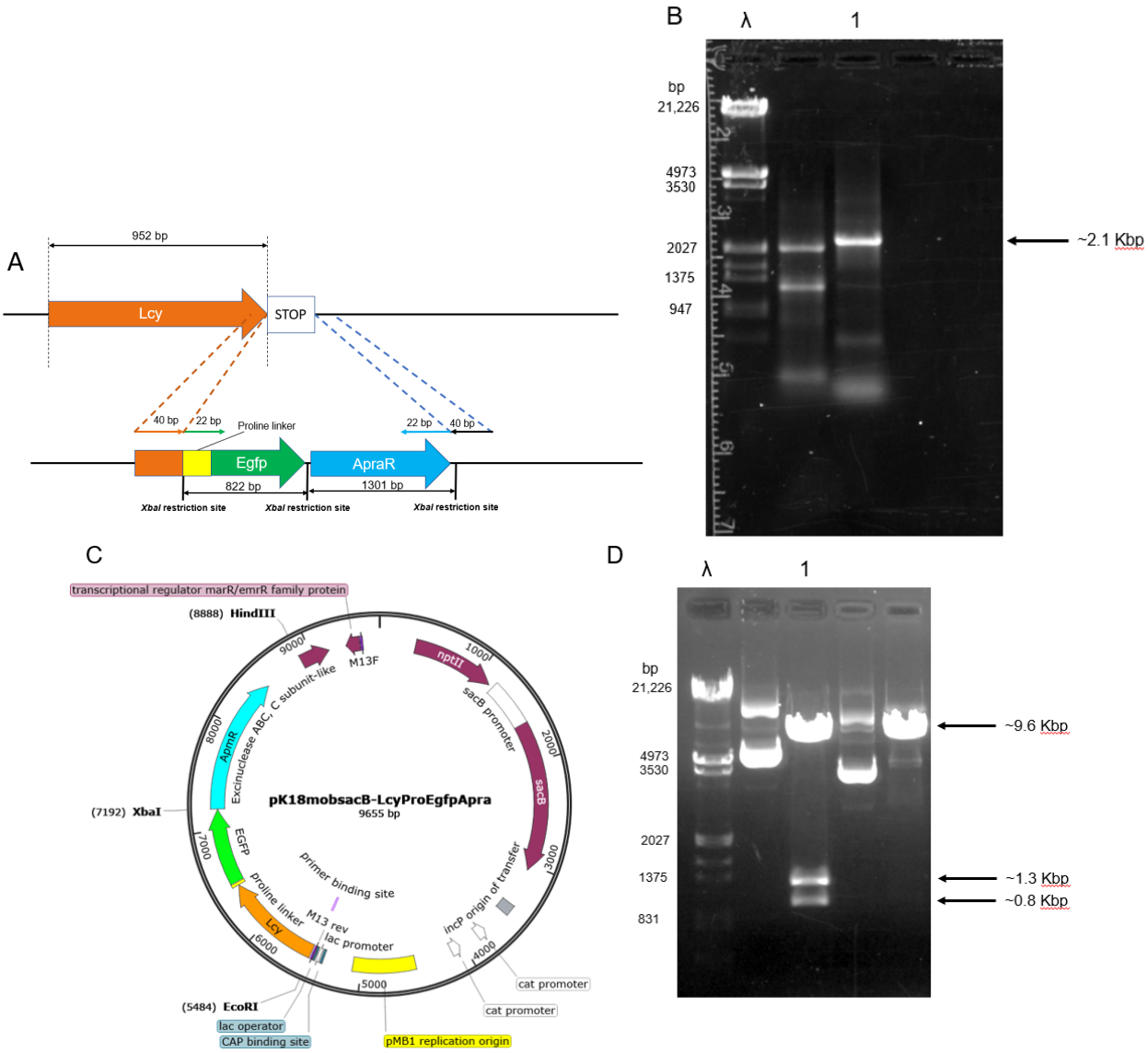


Figure 6-12. Diagram of the insertion of the complementary ProEgfpApra cassette into pK18mobsacB-Lcy, with vector maps and confirmation of the pK18mobsacB-LcyProEgfpApra construct. (A) A diagram of the insertion of the complementary ProEgfpApra cassette via homologous recombination due to the complementary 40 bp flanking extensions. The expected sizes of the components when digested are shown. (B) Confirmation of the extended complementary ProEgfpApra cassette on 0.7% Agarose gel. The loaded lanes are as such: Lane λ: Lambda DNA *EcoRI-HindIII* ladder, Lane 1: ProEgfpApra. (C) The vector map of the pK18mobsacB-LcyProEgfpApra construct. The LcyProEgfpApra gene is situated between the *EcoRI* and *HindIII* restriction sites and possess a pMB1 replication origin. (D) Confirmation of the isolated and digested pK18mobsacB-LcyProEgfpApra construct on 0.7% Agarose gel. The construct was digested with *XbaI*. The loaded lanes are as such: Lane λ: Lambda DNA *EcoRI-HindIII* ladder, Lane 1: pK18mobsacB-LcyProEgfpApra, Lane 2: digested pK18mobsacB-LcyProEgfpApra.

pK18mobsacB-LcyProEgfpApra was copied to *L. aggregata* via triparental mating with HB101 pRK2013. The DH5α pK18mobsacB-LcyProEgfpApra, HB101 pRK2013 and *L. aggregata* were mixed together for Triparental mating.

The transformed *L. aggregata* LcyProEgfpApra (a double crossover event occurred) was selected for on YTSS + 10% Sucrose agar plates containing apramycin, with successful colonies being selected and grown for microscopy. The prepared samples were grown over 3 days on YTSS, followed by being prepared for microscopy. Within the microscopy images of LcyEgfp, we were able to localise Lcy within growing *L. aggregata* at one of the poles. Only a single cluster of Lcy appeared within the rod-shaped bacteria at a high intensity at the observed pole. The high intensity of LcyEgfp was seen at poles of cells at different stages of growth and at division septum between daughter cells (Figure 6-13).

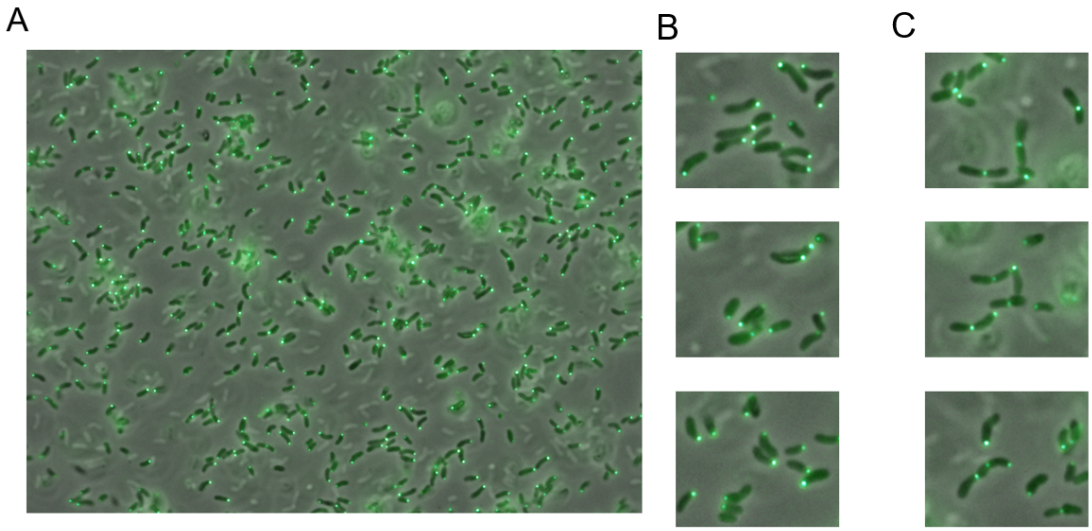


Figure 6-13. Fluorescent microscopy images of the LcyProEgfpApra within growing and dividing cells of *L. aggregata*. *L. aggregata* LcyProEgfpApra was plated on YTSS before being prepared for microscopy. (A) Localisation of LcyProEgfpApra in *L. aggregata* at various stages of its cell cycle. (B) Enhanced sections of image A to highlight the localisation pattern of LcyEgfp during cell growth of *L. aggregata*. Intense signals from LcyEgfp clustered to a single pole in *L. aggregata*. The intensity of the signal remains the same within a variety of different sized cells. (C) Enhanced sections of image A to highlight the localisation pattern of LcyEgfp during division of *L. aggregata*. Alongside the previously seen localisation pattern of LcyEgfp during growth, bright signals of LcyEgfp can be seen accumulated at the division septum between daughter cells of varying size. Some daughter cells have LcyEgfp localised at the division septum and at the opposite pole.

6.3 Conclusions:

In this chapter, we wished to understand the scope GPR homologs are found within Gram-negative bacteria and identify and analyse any retained function of these homologs within polar growth, via computational analysis and cellular localisation.

Many GPR homologs were identified across different Gram-negative species through various BLAST searches with the GPR homologs of *L. aggregata* being selected for further analysis. Alignment of the *L. aggregata* homologs against GPR revealed the same high conservation of most of the middle section of GPR (between residues 278-1815) across the homologs as multiple separated conserved regions of varying lengths. The conserved sections were separated by many small substitutions and deletions across the *L. aggregata* homologs. Further analysis of conserved domains of GPR and one of the *L. aggregata* homologs identified that the *L. aggregata* homologs should possess the same domains as GPR (Apolipoprotein A1/A4/E, Chromosome segregation protein and Chromosome segregation ATPase) within the conserved region.

To confirm if the known functions of GPR were retained within *L. aggregata* homologs via cellular localisation, a copy of a *L. aggregata* homolog would be generated and fused to a suitable fluorescent tag. The acquired strain of *L. aggregata* (LZB033) genome hadn't been published however the strain possessed an almost identical *L. aggregata* homolog to the analysed published homolog. The LZB033 homolog was designated Lcy and a collection of distinct fluorescent tags were generated and analysed to determine the most suitable tags to fuse with Lcy for monitoring Lcy during cell growth of *L. aggregata*. We were successful in generating a collection of 8 unique fluorescent tags with a different combination of tag, linker and resistance marker. In addition, the purest fluorescent tag (pUC18-ProEgfpApra) was selected for localising Lcy and successful fusion was generated and monitored.

Through the synthesised fluorescent tags, we were capable of localising Lcy within growing *L. aggregata*. Lcy appears to localise at a single pole of the rod-shaped bacteria at a high intensity. This suggests that an abundance of Lcy accumulates at a pole of the bacteria potentially for use in polar growth similar to GPR. Additionally, there were no other signals of LcyEgfp seen throughout the remainder of the cell which supports the idea of Lcy only localising to the poles of the cell. Finally, the high intensity of LcyEgfp was seen throughout the growth of the

cells and signals were seen at single poles of adjoined cells during cell division. The adjoined dividing cells were of different sizes with the majority of observed cases having the smaller cell localising LcyEgfp at the division septum. Larger cell was consistently seen with the LcyEgfp cluster localising at the opposite pole to the division septum. Together, this continues to highlight the potential use of Lcy within unipolar growth as well as in determination or formation of the division septum. Collectively, this appears to highlight a similar localisation pattern to GPR throughout the life cycle of *L. aggregata*, though the specific higher order assembly used by Lcy at the pole could not be determined.

6.4 Future directions:

Given that we managed to infer Lcy being actively involved in polar growth of *L. aggregata* via computational analysis and localisation at the pole, this correlation needs further confirmation. Since Lcy was investigated based on the similarity to GPR from *A. tumefaciens*, Lcy should be purified for X-ray crystallography and TEM imaging to analyse the potential structurally similarity to GPR. Additionally, if Lcy assembles into a higher assembly similar to GPR, Lcy could be tested in a BACTH assay against other differently orientated monomers to test for direct interaction between monomers, followed by overexpression and purification of assembled Lcy for analysis by NMR or Mass spectroscopy to confirm the assembly of multiple monomers. Finally, a knockout of *lcy* should be generated to confirm the function of Lcy within *L. aggregata* and the polar growth mechanism through cellular defects. The *lcy* null strain should be also transformed with an inducible copy of GPR to see if expression of GPR can fix the *L. aggregata* cellular defects by GPR acting as a substitute for Lcy, helping to confirm any similarities in function between the two proteins.

Chapter 7:

Discussion

7.1 Introduction:

Across the various species of rod-shaped bacteria, different mechanisms to drive polar growth have evolved. These systems are vital for maintaining the growth of the bacteria throughout their life cycle which is necessary for production of offspring later in the cell cycle. Unsurprisingly each component of these systems plays a key role though the current understanding and complexity of each known polar growth mechanism varies between bacteria. For some of the most complex systems (the TIPOC in *S. coelicolor*), a detailed and through understanding of the different components has been established. The TIPOC itself is constructed through 3 major proteins (DivIVA, FilP and Scy) each performing a specific role. The first major component of TIPOC (DivIVA) has been shown to determine the location of polar growth at the pole of the hyphae due its natural localisation to negatively curved cell membranes (Flärdh, 2003). DivIVA can then self-assemble into a stable, tetramer complex at the negatively curved membranes to help determine the location of the TIPOC. This in turn allows for the accumulation of the other important TIPOC proteins to the site. Following DivIVA, both FilP and Scy are recruited separately to the hyphal pole. FilP directly interacts with DivIVA to trigger the self-assembly of FilP filaments from the apical tip (Fuchino *et al*., 2013). These filaments provide structural support through reducing osmotic stress for newly growing hyphae (Fuchino *et al*., 2016). Finally, Scy directly interacts with both DivIVA and FilP and appears to be responsible for coordinating the localisation of the TIPOC as a whole. This is demonstrated by overexpression of Scy resulting in multiple TIPOCs forming and hyperbranching occurring (Holmes *et al*., 2013). However, for some species such as *A. tumefaciens*, the establishment of components in its polar growth mechanism have only been identified recently. *A. tumefaciens* utilises a polar growth mechanism called GPR (Growth Pole Ring) where 6 of the GPR proteins localise at the growth pole of *A. tumefaciens* in a hexameric ring during polar growth (Zupan *et al*., 2019), with deletion of GPR resulting in abnormal cell morphology and demonstrating the role of GPR in organising cell wall synthesis machinery (Zupan *et al*., 2021). GPR and DivIVA of *S. coelicolor* share the same function by helping determine where polar growth occurs. This shows how unique polar growth mechanisms have evolved independently across different rod-shaped bacteria.

Though the difference in known components and their functions within the polar growth mechanisms of *S. coelicolor* and *A. tumefaciens* demonstrate the vast difference in understanding of polar growth mechanisms across species. With this in mind, we decided to investigate and expand on the knowledge of both polar growth mechanisms in this study. We plan to investigate the structural properties, functions and possible partner proteins of Dia, a potential new component for the TIPOC, which may help regulate the structure of the multi-protein complex. Simultaneously, we plan to search for and confirm new polar growth mechanisms in Gram-negative rod-shaped bacteria through identifying potential homologues of the *A. tumefaciens* GPR system and monitoring their localisation pattern with fluorescence.

7.2 Does Dia possess a role within the TIPOC system?

From the previous research of *dia* knockouts by Daisy Ireland, we know that Dia has an ability to influence the growth of both aerial and vegetative hyphae and has a possible coiled-coil structure, similar to other *S. coelicolor* polar growth proteins (Ireland, Unpublished). In the vegetative hyphae, the diameter can be drastically different at various points, with the diameter being increased by almost 25% in Δ*dia* mutants (from a usual 0.85 μm to 1.06 μm). In addition, these changes in diameter are not consistent within the hyphae with irregular fluctuations between the increased diameter and the wild-type diameter (Ireland, Unpublished). Since the diameter of unaltered hyphae is consistently stable and determined when cell wall synthesis occurs, the removal of Dia must influence the TIPOC mechanics at the hyphal tip. Alongside these effects in diameter, the Dia knockout aerial hyphae also showed vastly different distances between each sporulation septa, resulting in the formation of compartments that were either smaller than standard compartments or up to twice the size of compartments found within wild-type *S. coelicolor* (usual distance between sporulation septa is about 2 μm, some Δ*dia* null mutants had sporulation septa almost 4 μm between). Within these larger compartments, multiple chromosomes were found within (Ireland, Unpublished). With the size of forming spores determined by precise regulation of the position of FtsZ rings and the ParAB system, Dia may influence these systems. Not only that but the effects on hyphae growth reduced growth within *S. coelicolor* as a whole and resulted in reduced colony sizes at the same time for Wild-type *S. coelicolor*. Although the colony sizes were only marginally reduced, the spore pigment was noticeably reduced within the colony when compared to Wild-type *S. coelicolor* grown for the same amount of time (Ireland, Unpublished). With these effects in mind, the knockouts of Dia have shown

that Dia can influence the structure of the hyphae determined by polar growth and the positioning of sporulation septa as they form in aerial hyphae. These 2 mechanics of *S. coelicolor* are regulated by two main systems: the TIPOC and the ParAB system which Dia seems to possibly influence or be a part of as novel components.

To start determining if Dia is involved with the systems that drive *S. coelicolor's* life cycle, Daisy also generated a conjoined DiaEgfp construct (a single crossover event) and monitored where the construct localised. The adjoined construct accumulated behind the hyphal tip and at various points further down the hyphae (Ireland, Unpublished). In previous research I conducted, we generated a double crossover version of DiaEgfp and monitored the localisation pattern. In conjunction with Daisy's research, we identified the same localisation pattern of Dia within growing and developed vegetative hyphae (Hutchinson, Unpublished). We also noticed that the DiaEgfp localisation pattern mirrored the DivIVAEgfp localisation pattern seen in previous research. With reconfirmation of the localisation pattern of Dia and a correlation with known component of the TIPOC (DivIVA), we decided to investigate the structure and function of Dia further.

To determine the range of bacteria that utilise Dia outside of *S. coelicolor*, we ran 3 separate BLAST searches which identified homologs of Dia within multiple species of Actinobacteria and a several species of different phyla including Acidobacteria and Bacillota. We further investigated the similarity of these identified homologs through aligning them in the NCBI Multiple Alignment Viewer tool. Across all analysed Dia homologs, there was a universally conserved region between residues 35-144 which possessed alternating sections of high and low conservation. Within the species more closely related to *S. coelicolor* (from the Actinobacteria phyla), a larger, highly conserved region was seen between residues 20-216. These results demonstrate a shared ancestry between Dia homologs which are contained within a variety of different bacterial species. Following this, we decided to further analysis possible functions of Dia by passing the primary sequence through the NCBI Conserved Domain search tool. Multiple hits were returned for a variety of ATP synthase subunit B domains and a DivIVA domain all overlapping within the conserved region of Dia (between residues 33-131), inferring a conserved and universal function. To affirm these possible domains, known proteins of these possible domains in *S. coelicolor* (DivIVA and AtpF) were aligned against Dia in Clustal Omega. Both alignments aligned conserved sections with Dia between residues 49-116 (within the predicted range for the possible domains and the

conserved region of Dia across all homologs). This demonstrates that not only Dia is abundantly found within species of Actinobacteria and other phyla, but also the shared conserved region may possess a shared function between the homologs.

With multiple Dia homologs possessing a possible conserved function (of DivIVA and ATP synthase subunit B) across phyla, we decided to investigate the secondary structure and tertiary structure of Dia which would allow for these possible domains. We ran the primary sequence of Dia through several computational programs to predict these structures, these programs being: PSIRED (to predict the secondary structure), PCOILS (top predict the likelihood to form left-handed coiled-coils), MARCOIL and DeepCoil programs (to predict the formation of any coiled-coils). Firstly, we started with the PSIRED section of the PCOILS program which predicted the formation of α-helices between residues 0-110 and 150-205. A β-sheet also was predicted around residue 120. The predicted α-helices overlap with ATP synthase subunit B domains which was unsurprising as this predicted domain for Dia is known to be formed from α-helices. 75% of the subunit b of F-ATPases is composed of α-helices (Deckers-Hebestreit *et al*., 2000). The PCOILS section of the program anticipated the formation of coiled-coil structures between residues 0-110, but this was indicated to be a false positive. This contradicted current knowledge on the UniProt database which predicted coiled-coils between residues 48-68 and 90-121 (overlapping with the predicted α-helices between residues 0-110). To determine if coiled-coils did form within Dia, the primary sequence of Dia was passed through the MARCOIL and DeepCoil program to verify the formation of any coiled-coils. Both programs returned predicted coiled-coils between the residues 30-100, though DeepCoil predicted 2 separate coiled between residues 30-50 and 90-95. These predictions overlapped with and reaffirmed the original predicted structure outlined in the UniProt entry (coiled-coils between residues 48-68 and 90-121). This helped confirm with the previous predictions of the secondary structure and possible domains of Dia as α-helices are known for their capacity to form larger, coiled-coil structures. Two or more α-helices are known to be arranged into larger helix structures called coiled-coils (Lupas, 1996). Finally, with consistent predictions of certain sections of the tertiary structure of Dia, we decided to model the tertiary structures of Dia, DivIVA and AtpF from *S. coelicolor* to uphold the current predictions and investigate further tertiary structures within the remainder of Dia. All three protein models predicted coiled-coil structures overlapping between residues 20-144. Furthermore, Dia and DivIVA had 2 coiled-coil structures overlapped between residues 3-15 and 168-206. All the overlapping

structures possessed high confidence values. Additionally, there was a low confidence section of the Dia model between residues 239 -367 which was unstructured which could be due to that this section of Dia requires another component to form a higher structure. All the accurate predicted coiled-coil structures within Dia model overlapped with the previously predicted α-helices and coiled-coil structures (between residues 0-110 and 30-100 respectively), the conserved region of Dia (residues 33-144) and the predicted conserved domains of DivIVA and ATP synthase subunit B (residues 49-116). This helped to affirm the previous predictions and may suggest possible conserved functions for the region. A possible function could be a retained capacity of ATP synthase subunit B. ATP synthase subunit B is distinguished for its ability to resist the rotation of the ATP synthase catalytic domain. The peripheral stalk of most F-ATPases acts as a stator to help hold the catalytical complex in place. Within *E. coli* and other bacteria, the peripheral stalk is constructed out of 2 copies subunit b (Walker & Dickson, 2006). If this were the case, it would allow for a region of the protein which provides stability against movement. If Dia does interact with the TIPOC, this in turn would allow for Dia to resist moving with any new forming complexes and hold them in a more fixed position. This would help explain the instability in hyphal diameter seen when Dia is knocked out in *S. coelicolor*.

With the possible connection between Dia and the TIPOC, we chose to determine if any direct interactions occur between Dia and specific partner proteins of the TIPOC through the BACTH assay. Since all the known key components of the TIPOC interact with each other to form the multiprotein complex, we tested Dia against all key component: DivIVA, FilP and Scy. Additionally, due to each component capable of self-interaction into larger structures, we tested Dia for self-interaction as well. The Dia self-interactions conclusively indicated that Dia monomers can interact with each other and in a variety of different orientations. The chosen partner protein interactions yielded a mixed collection of interactions. The DivIVA interaction signals generated were weak and interspersed when compared to the controls and were inconclusive. This may mean that Dia and DivIVA can interact, but further testing is required. The FilP interactions however resulted in some conclusive signals alongside a variety of weak interspersed signals. The conclusive interactions appear to occur when the N-termini of both Dia and FilP are interacting. Definitively, these signals do highlight that Dia can interact with FilP directly, although the weaker signals could be limited to specific orientations or variations in the conditions they were expressed. On the other hand, the Scy

interactions resulted in conclusive signals in all the tested orientations, allowing for interactions between both N- and C- termini of both proteins. Collectively, these results show that Dia can interact with 2 of the major components of the TIPOC directly and possibly through the formation of a larger structure. Both Scy and FilP are important components of the TIPOC playing important roles in the multiprotein complex. Scy has been shown to be important in regulating polar growth and coordinating the other components of the TIPOC. Scy has been found to directly interact with both DivIVA and FilP in BATCH and pelleting assays and Scy null mutants have been shown to reduce polar growth. Scy overexpression resulted in hyperbranching of the hyphae and altered the localisation of DivIVA (Holmes *et al*., 2013). FilP, on the other hand, has been shown to be linked to the polar growth mechanics whilst providing structural support. FilP forms long, branching networks along the hyphae, away from the hyphal tip, which are capable of self-assembly in the presence of DivIVA (Fuchino *et al*., 2013). FilP null mutants were seen to have reduced cell rigidity when under Atomic force microscopy (Bagchi *et al*., 2008). With the inferred structural properties of Dia and the known effects on polar growth in Dia null mutants, Dia may influence the mechanics of the TIPOC by providing structural stability to the different components and the larger structures they form. This in turn may help to localise the forming structures of the system at the hyphal tip and help maintain a stable TIPOC complex with Scy through facilitating interactions with FilP and allowing consistent growth of the hyphae. Consequently, this would allow for a consistent fixed diameter of the hyphae to occur during polar growth. To help determine this possible mechanism, each interaction between Dia and the tested TIPOC components should be reaffirmed via *in vitro* pelleting assay and visualised by transmission electron microscopy (TEM). This should help to provide further understanding into the dynamics of each Dia-TIPOC interaction both between the monomers and complexes of each interaction. In addition, mutants of Dia for the identified domains should be generated and tested in BACTH assays against other Dia mutations and the TIPOC components. The signals generated should help provide insight into the roles the different domains play in determining the capacities of Dia to self-assemble and to interact with the TIPOC components. Finally, if any of the generated Dia mutants alter its capacities, these mutants should be fluorescently tagged and transformed into the *S. coelicolor* genome for co-localisation with fluorescent tags versions of each partner protein from the TIPOC (DivIVA, FilP and Scy). These localisation patterns should help reaffirm any seen dynamics from the TEM and may provide evidence of the possible mechanism *in vivo*. Ultimately though, since Dia can interact with all three major components and influences the

TIPOC mechanics, this strongly suggests that Dia may be a new novel component in the TIPOC.

With the establishment of the capacity Dia can interact with certain components of the TIPOC and the known interactions components of the TIPOC (Scy and DivIVA) have with the ParAB system, we decided to investigate if Dia possessed this capability. We tested this through interacting Dia with the known ParAB partner proteins (ParA and ParB) and a ParA homologue (ParH) within the BACTH assay. The ParA interactions uniformly resulted in no interactions. On the other hand, ParB interacted with Dia directly but in the same manner as FilP, the generated signals were a mix of conclusive and weak interspersed signals. Since the same BACTH vectors were used, the mixed signals are unsurprising. The conclusive signal occurred when the N-termini of Dia and ParB where interacting. This was unexpected as Dia can interact with the TIPOC, possibly as a novel component. Additionally, the TIPOC components: DivIVA and Scy are known to interact ParA and ParB independently. In *M. smegmartis*, DivIVA and ParA has been found to directly interact within a heterogenous *E. coli* host. The interactions were confirmed through a collection of ParA mutants tested against DivIVA in BACTH assays, especially the ParAT3A mutation which prevented DivIVA binding to ParA (Pióro *et al*., 2018). The DivIVA and ParB-*parS* complex of *C. glutamicum* has been shown to directly interact within BACTH assays and *in vivo* in *C. glutamicum* (Donovan *et al*., 2012). On the other hand, Scy and ParA Have been shown to interact with each other directly through BACTH assays and *in vivo* co-localisation studies (Ditkowski *et al*., 2013). Finally, we found Dia was capable of interacting with an unpublished but researched ParA homologue (ParH) from Dr. Kelemen's lab. As seen with ParB, a similar proportion of mixed signals were seen but direct interaction was confirmed. Conclusive signals were seen for Dia and ParH interacting between both proteins' N-termini and the N-termini of Dia interacting with the C-terminus of ParH. This was unexpected as the ParA did not interact with Dia. Our current knowledge of ParH has revealed it to possess a similar capacity to ParA, is known to interact with Scy and has been suggested to form a unique complex for septa positioning. ParH has been shown to directly interact with Scy and ParB independently (Gillespie, Unpublished; Alanazi, Unpublished). ParH does not require ATP for these interactions or polymerisation and has been proposed to form a complex with ParB at the chromosome, after chromosome segregation, to help coordinate the placement of septa through affecting FtsZ polymerisation (Gillespie, Unpublished). The Dia-ParB and Dia-ParH interactions help to confirm

the known localisation of Dia seen within developed and growing hyphae and may help to shed light on the generation of spores with irregular sizes and number of chromosomes. If the same possible mechanism between Dia and the TIPOC applies to the interactions between Dia and the ParAB system, then Dia complexes may bind to the ParB-ParH complexes to help maintain their stability when bound to the chromosome after chromosome segregation. Assuming the mechanism between ParB-ParH complexes and FtsZ polymerisation suggested by Gillespie is correct, then the proposed maintained stability by Dia would allow for the consistent formation of the ParB-ParH complexes and in turn the regular signalling of FtsZ polymerisation and correct positioning of septa in sporulation. In addition, the large capacity for interactions between Dia and components of the TIPOC and ParAB system may help to coordinate interactions between both systems during active polar growth. As the same mechanism for interactions between Dia and the TIPOC have been suggested for the Dia-ParB and Dia-ParH interactions, confirmation of the possible mechanism and further insight should also be possible through the same experimental analysis as suggested for the Dia-TIPOC interactions. Although, additional co-localisation studies with fluorescently tagged FtsZ should be conducted to confirm the possible influence on FtsZ polymerisation.

Lastly, we decided to investigate the self-assembly capacity of Dia through immobilised metal affinity chromatography: Dia was fused to a 6xHis tag, overexpressed and purified through a suitable *E. coli* host. His-Dia was extracted and ran at approximately ~65 kDa which was larger than the expected His-Dia monomer (43.1 kDa). The extracted His-Dia monomers may be forming a dimer which is possible due to the identified Cysteine (Cys) residue, within a potential coiled-coil, which could form a disulfide bridge between monomers. The extracted His-Dia was ran with fresh loading dye containing β-mercaptoethanol to reduce any potential disulphide bridges, however, all attempts resulted in His-Dia running at ~65 kDa. This suggests that the Cys residue may be buried within a hydrophobic core of the coiled-coil structure. The extracted His-Dia was also resistant to standard denaturing conditions (SDS and heat) suggesting the potential dimer is highly stable. However, His-Dia has an unusually high theoretical pI for a protein (4.60) and appears to be an acidic protein, the protein may be running at an unusual weight on the gel. To determine if this is the case or Dia is forming a dimer, a dialysed sample of His-Dia was analysed by Mass Photometry to accurately determine the molecular weight. The Mass Photometry of His-Dia estimated the weight to be 77 kDa with a SD of 17.5. This allows for the predicted weight of the

His-Dia dimer (86.2 kDa). To investigate separating the His-Dia dimer, other denaturing compounds such as Dithiothreitol or Tris(2-carboxylethyl)phosphine hydrochloride maybe be able to separate the extracted monomers for analysis via SDS-PAGE. Additionally, the His-Dia dimer should be analysed by NMR or Mass spectroscopy to confirm the molecular weight of the subunits.

7.3 Do DivIVA-lacking rod-shaped bacteria use alternative proteins for polar growth?

As we have previously mentioned, the polar growth mechanisms for species of rod-shaped bacteria can differ vastly in both understanding and complexity. A common factor of these systems across species is the ability of one of the mechanism's components to naturally accumulate near the negative-curved pole of the rod-shaped bacteria to start and localise the polar growth mechanism to the desired pole, though different mechanisms use a different component for this important role. For many of the Gram-positive bacteria, different systems of differing complexity are used that utilise a homolog of DivIVA as a crucial component that underpins the entire system, with the species *S. coelicolor* possessing one of the best well-documented polar growth mechanisms. Within *S. coelicolor*, a complex multi-protein system called the TIPOC (Tip Organising Centre) is used to organise and direct polar growth. DivIVA, one of the major components, localises to the negative-curved pole to allow drive polar growth through cell wall synthesis (Flärdh, 2003). The localised DivIVA monomers are assembled into stable tetramer structure (Wang *et al*., 2009) which is regulated through phosphorylation/dephosphorylation via AfsK or SppA respectively (Hempel *et al*., 2012; Passot, *et al*., 2022). Collectively, this demonstrates the complex regulation of DivIVA within *S. coelicolor* to drive polar growth at the pole of the hyphae. However, for many rod-shaped Gram-negative bacteria, very little is understood of their polar growth mechanisms as most Gram-negative bacteria lack any homolog of DivIVA. Though an alternative polar growth mechanism which uses an entirely different component to DivIVA has been identified. This system is known as the GPR (Growth Pole Ring) system and is found within *A. tumefaciens*. The GPR system utilises a hexameric ring structure, approximately 200 nm in diameter, which localises near the growth pole of the cell (Zupan *et al*., 2019). Other homologs of the GPR system have been found within other Alphaproteobacteria such as *S. meliloti* and have been shown directly influence polar growth of the cell (Krol *et al*., 2020). With the knowledge of a different polar growth mechanism in Gram-negative bacteria, we investigated for

polar growth mechanisms in other Gram-negative species which shared homology to GPR system.

To search for homologs of the GPR mechanism, BLAST searches were conducted against other bacterial genomes. Many closely related GPR homologs were found within species of the Rhizobiaceae Family which was unsurprising as *A. tumefaciens* and *S. meliloti* are contained within this Soil-dwelling family. Surprisingly though, we found a multitude of other homologs across other families including the Stappiaceae Family which are only distantly related to GPR through the Hyphomicrobiales Order. With this being the case, the highlighted homologs found within other families contained varying amounts of sequence identity to GPR, with most possessing 30% or less. Even with such a low conservation seen of the homologs, the hits returned had exceptionally high E-values demonstrating the significance of each hit. Since the significance is unlikely to be from the sequence identity, it is most likely that these homologs share a similar structure to GPR instead. From these structurally significant homologs, a collection of 8 hits for *L. aggregata* (also sometimes called *R. aggregatum*) interested us due to the exceptionally low sequence similarity (below 30%) and the uniqueness within the search and reclassification of this species. *L. aggregata* has recently been reclassified in 2007 due to recent analysis of various aspects including its genome and cell wall structure (Biebl *et al*., 2007). *L. aggregata* has also been recently identified as an important marine bacterium in the production of DMSP (Curson *et al*., 2017). With multiple significant homologs of GPR being identified in a bacterium with a vastly different environment and potentially different life cycle, we chose to investigate these homologs. To further understand the similarity of the *L. aggregata* homologs to GPR, we aligned the 8 *L. aggregata* homologs against GPR in the NCBI Multiple Sequence Alignment Viewer 1.22.2. We found across all 8 homologs between residues 370-1550, approximately 42 individual highly conserved regions separated by either substitutions or deletions. With the highly conservation regions found across all homologs, the *A. tumefaciens* GPR and one of the *L. aggregata* homolog sequences were individually passed through the NCBI Conserved Domain Search to predict for any conserved function. Within the conserved middle section of GPR and the *L. aggregata* homolog were multiple Apolipoprotein A1/A4/E, Chromosome segregation protein and Chromosome segregation ATPase domains were identified. Deletions of Human apolipoprotein A-IV-4 like domain in GPR revealed defects in cell morphology (Zupan *et al*., 2021). This indicates the conservation of some of the function of GPR in the *L. aggregata* homologs.

The *L. aggregata* species we managed to acquire called LZB033 has not had its genome published yet but the GPR homolog we identified and isolated from LZB033 (FIG00742013) is remarkably similar to the GPR homolog (A0A0M6Y2Z9) from the published *L. aggregata* IAM 12614 genome on UniProt (only possessed 2 single base substiutions), so we continued to investigate the function of the LZB033 homolog. Assuming the *L. aggregata* LZB033 GPR homolog should be found within the cytoplasm of the cell and is likely to provide some form of cytoskeletal function for polar growth, we designated the gene lcy (*labrenzia* cytoskeletal protein). To determine if Lcy possess a similar function to GPR, we fluorescently tagged the homolog and transformed into the *L. aggregata* genome. The fusion replaced the wildtype version in the genome (through a double crossover event) and the construct was monitored and localised. We observed that LcyEgfp localised at a high intensity at one of the poles. In conjunction, the signal was seen throughout different sized cells and in dividing cells which indicates Lcy may be utilised across the life cycle of *L. aggregata* cells. In addition, in some dividing *L. aggregata* cells, LcyEgfp signals were seen at different poles during the division event. For larger (parent) cells, a strong signal was seen at the pole opposite to the division septum whereas smaller (daughter) cells were seen with the LcyEgfp signals at the division septum. Lcy may be localised to the growing pole of *L. aggregata* throughout its life cycle and being recruited for a role near the division septum. With Lcy sharing conserved homology to GPR and potential important domains for polar growth, Lcy localisation in *L. aggregata* in a unipolar fashion and being recruited to the division septum appears to mirror that of GPR. GPR-GPR was found localise at the growth pole of growing cells throughout their life cycles and form at the midcell prior to cell division *in vivo* (Zupan *et al*., 2019). The ΔApolipoprotein A-IV-4 domain GPR-GPR was found to cause severe defects in the cell morphology though was found to localise to the cell poles. Complementation with GPR deficient cells, returned the cells to a rod-shaped morphology, though the cells were abnormally long (Zupan *et al*., 2021). This indicates that the features of the GPR mechanism may be more conserved and widespread (in Gram-negative bacteria) than previously thought and other rod-shaped bacteria within the Hyphomicrobiales Order may utilise variants of it. To further confirm the function of Lcy, null mutants should be generated and monitored for any defects in the polar growth of *L. aggregata* throughout its life cycle. In addition, to investigate the similarity of function and shape between GPR and Lcy, the *lcy* null strains transformed to carry a plasmid with an inducible copy of GPR to test if the expression of GPR is able to fix any defects within the cell by being a substitute to Lcy. To confirm any structural similarities, TEM imaging and X-

ray crystallography of Lcy and GPR and comparison of the 3D structures would help to confirm any structural similarities between them. Finally, to test if Lcy can assembles into a higher order assembly similar to GPR, a BACTH assay testing interaction between differently orientated Lcy monomers would help confirm any direct interaction between Lcy monomers, coupled with the analysis of purified Lcy assemblies by NMR or Mass spectroscopy would help to confirm if and how higher order structure of Lcy assembles and how this assembly differs to GPR.

References

Adler, H. I., Fisher, W. D., Cohen, A., & Hardigree, A. A. (1967). MINIATURE escherichia coli CELLS DEFICIENT IN DNA. *Proceedings of the National Sciences of the United States of America, 57*(2), 321-326.

Alanazi, M. (Unpublished). *PhD thesis*, University of East Anglia.

Alcock, E. (Unpublished). Remarkable in vitro Assemblies of a Bacterial Cytoskeletal Protein, FilP and its Derivaties. *PhD Thesis*, University of East Anglia, Norwich.

Anderson-Furgeson, J. C., Zupan, J. R., Grangeon, R., & Zambryski, P. C. (2016). Loss of PodJ in Agrobacterium tumefaciens Leads to Ectopic Polar Growth, Branching, and Reduced Cell Division. *Journal of Bacteriology, 198*(13), 1883-1891.

Aussel, L., Barre, F. X., Aroyo, M., Stasiak, A., Stasiak, A. Z., & Sherratt, D. (2002). FtsK Is a DNA motor protein that activates chromosome dimer resolution by switching the catalytic state of the XerC and XerD recombinases. *Cell, 108*(2), 195-205.

Bach, J. N., Albrecht, N., & Bramkamp, M. (2014). Imaging DivIVA dynamics using photo-convertible and activatable fluorophores in Bacillus subtilis. *Frontiers in Microbiology*. doi:10.3389/fmicb.2014.00059

Bagchi, S., Tomenius, H., Belova, L. M., & Ausmees, N. (2008). Intermediate filament-like proteins in bacteria and a cytoskeletal function in Streptomyces. *Molecular Microbiology, 70*(4), 1037-1050.

Barka, E. A., Vasta, P., Sanchez, L., Gaveau-Vaillant, N., Jacquard, C., Meier-Kolthoff, J. P., . . . van Wezel, G. P. (2015). Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiology and molecular biology reviews (MMBR), 80*(1), 1-43.

Bentley, S. D., Chater, K. F., Cerdeño-Tárraga, A. M., Challis, G. L., Thompson, N. R., James, K. D., . . . Hopwood, D. A. (2002). Complete genome sequence of the model actinomycete Streptomyces coelicolor A3(2). *Nature, 417*(6885), 141-147.

Bi, E. F., & Lutkenhaus, J. (1991). FtsZ ring structure associated with division in Escherichia coli. *Nature, 254*(6349), 161-164.

Biebl, H., Pukall, R., Lünsdorf, H., Schulz, S., Allgaier, M., Tindall, B. J., & Wagner-Döbler, I. (2007). Description of Labrenzia alexandrii gen. nov., sp. nov., a novel alphaproteobacterium containing bacteriochlorophyll a, and a proposal for reclassification of Stappia aggregata as Labrenzia aggregata comb. nov., of Stappia marina as Labrenzia marina comb. *International Journal of systematic and evolutionary microbiology, 57*(Pt 5), 1095-1107.

Bigot, S., Corre, J., Louarn, J. M., Cornet F, & Barre, F. X. (2004). FtsK activities in Xer recombination, DNA mobilization and cell division involve overlapping and separate domains of the protein. *Molecular Microbiology, 54*(4), 876-886.

Bowman, G. R., Comolli, L. R., Zhu, J., Eckart, M., Koenig, M., Downing, K. H., . . . Shapiro, L. (2008). A polymeric protein anchors the chromosomal origin/ParB complex at a bacterial cell pole. *Cell, 134*(6), 945-955.

Breier, A. M., & Grossman, A. D. (2007). Whole-genome analysis of the chromosome partitioning and sporulation protein Spo0J (ParB) reveals spreading and origin-distal sites on the Bacillus subtilis chromosome. *Molecular Microbiology, 64*(3), 703-718.

Brown, P. J., de Pedro, M. A., Kylesa, D. T., Van der Henst, C., Kim, J., De Bolle, X., . . . Brun, Y. V. (2012). Polar growth in the Alphaproteobacterial order Rhizobiales. *Proceedings of the National Academy of Sciences of the United States of America, 109*(5), 1691-1701.

Buddelmeijer, N., & Beckwith, J. (2004). A complex of the Escherichia coli cell division proteins FtsL, FtsB and FtsQ forms independently of its localization to the septal region. *Molecular Microbiology, 52*(5), 1315-1327.

Cameron, T. A., Zupan, J. R., & Zambryski, P. C. (2015). The essential features and modes of bacterial polar growth. *Trends in Microbiology, 23*(6), 347-353.

Carballido-López, R., Formstone, A., Li, Y., Ehrlich, S. D., Noirot, P., & Errington, J. (2006). Actin homolog MreBH governs cell morphogenesis by localization of the cell wall hydrolase LytE. *Development Cell, 11*(3), 399-409.

Cassettari, G. (Unpublished). SepF: A link between cell division and polar growth in StreptomycescoelicolorGemma. *PhD thesis*, University of East Anglia, Norwich.

Cha, J. H., & Stewart, G. C. (1997). The divIVA minicell locus of Bacillus subtilis. *Journal of Bacteriology, 179*(5), 1671-83.

Champoux, J. J. (2001). DNA topoisomerases: structure, function, and mechanism. *Annual review of Biochemistry, 70*, 369-213.

Curson, A. R., Liu, J., Bermejo Martínez, A., Green, R. T., Chan, Y., Carrión, O., . . . Todd, J. D. (2017). Dimethylsulfoniopropionate biosynthesis in marine bacteria and identification of the key gene in this process. *Nature Microbiology, 2*(17009).

Dai, K., Xu, Y., & Lutkenhaus, J. (1993). Cloning and characterization of ftsN, an essential cell division gene in Escherichia coli isolated as a multicopy suppressor of ftsA12(Ts). *Journal of Bacteriology, 175*(12), 3790-3797.

Dajkovic, A., Lan, G., Sun, S. X., Wirtz, D., & Lutkenhaus, J. (2008). MinC spatially controls bacterial cytokinesis by antagonizing the scaffolding function of FtsZ. *Current biology, 18*(4), 235-244.

Daniel, R. A., & Errington, J. (2003). Control of cell morphogenesis in bacteria: two distinct ways to make a rod-shaped cell. *Cell, 113*(6), 767-776.

Davis, M. A., Martin, K. A., & Austin, S. J. (1992). Biochemical activities of the parA partition protein of the P1 plasmid. *Molecular microbiology, 6*(9), 1141-1147.

de Boer, P. A., Crossley, R. E., & Rothfield, L. I. (1992). Roles of MinC and MinD in the site-specific septation block mediated by the MinCDE system of Escherichia coli. *Journal of Bacteriology, 174*(1), 63-70.

Deckers-Hebestreit, G., Greie, J., Stalz, W., & Altendorf, K. (2000). The ATP synthase of Escherichia coli: structure and function of F(0) subunits. *Biochimica et Biophysicia Acta, 1458*(2-3), 364-373.

Defeu Soufo, H. J., & Graumann, P. L. (2006). Dynamic localization and interaction with other Bacillus subtilis actin-like proteins are important for the function of MreB. *Molecular Microbiology, 62*(5), 1340-1356.

Di Ventura, B., Knecht, B., Andreas, H., Godinez, W. J., Fritsche, M., Rohr, K., . . . Sourjik, V. (2013). Chromosome segregation by the Escherichia coli Min system. *Molceular systems biology, 9*(1), 686.

Ditkowski, B., Holmes, N., Rydzak, J., Donczew, M., Bezulska, M., Ginda, K., . . . Jakimowicz, D. (2013). Dynamic interplay of ParA with the polarity protein, Scy, coordinates the growth with chromosome segregation in Streptomyces coelicolor. *Open biology, 3*(3), 130006.

Ditkowski, B., Troć, P., Ginda, K., Magdalena, D., Chater, K. F., Zakrzewska-Czerwińska, J., & Jakimowicz, D. (2010). The actinobacterial signature protein ParJ (SCO1662) regulates ParA polymerization and affects chromosome segregation and cell division during Streptomyces sporulation. *Molecular Microbiology, 78*(6), 1403-1415.

Doi, M., Wachi, M., Ishino, F., Tomioka, S., Ito, M., Sakagami, Y., . . . Matsuhashi, M. (1988). Determinations of the DNA sequence of the mreB gene and of the gene products of the mre region that function in formation of the rod shape of Escherichia coli cells. *Journal of Bacteriology, 170*(10), 4619-4624.

Domínguez-Cuevas, P., Porcelli, I., Daniel, R. A., & Errington, J. (2013). Differentiated roles for MreB-actin isologues and autolytic enzymes in Bacillus subtilis morphogenesis. *Molecular Microbiology, 89*(6), 1084-1098.

Domínguez-Escobar, J., Chastanet, A., Crevenna, A. H., Fromion, V., Wedlich-Söldner, R., & Carballido-López, R. (2011). Processive movement of MreB-associated cell wall biosynthetic complexes in bacteria. *Science (New York), 333*(6039), 225-228.

Donczew, M., Mackiewicz, P., Wróbel, A., Flärdh, K., Zakrzewska-Czerwińska, J., & Jakimowicz, D. (2016). ParA and ParB coordinate chromosome segregation with cell elongation and division during Streptomyces sporulation. *Open Biology, 6*(4), 150263.

Donovan, C., Sieger, B., Krämer, R., & Bramkamp, M. (2012). A synthetic Escherichia coli system identifies a conserved origin tethering factor in Actinobacteria. *Molecular Microbiology, 84*(1), 105-116.

Ebersbach, G., & Gerdes, K. (2001). The double par locus of virulence factor pB171: DNA segregation is correlated with oscillation of ParA. *Proceedings of the National Sciences of the United States of America, 98*(26), 15078-15083.

Ebersbach, G., Briegel, A., Jensen, G. J., & Jacobs-Wagner, C. (2008). A self-associating protein critical for chromosome attachment, division, and polar organization in caulobacter. *Cell, 134*(6), 956-968.

Ebersbach, G., Ringgaard, S., Møller-Jensen, J., Wang, Q., Sherratt, D. J., & Gerdes, K. (2006). Regular cellular distribution of plasmids by oscillating and filament-forming ParA ATPase of plasmid pB171. *Molecular Microbiology, 61*(6), 1428-1442.

Egan, A. J., Cleverly, R. M., Peters, K., Lewis, R. J., & Vollmer, W. (2017). Regulation of bacterial cell wall growth. *The Federation of Biochemical Societies Journal, 284*(6), 851-867.

Erickson, H. P. (1998). Atomic structures of tubulin and FtsZ. *Trends in Cell Biology, 8*(4), 133-137.

Erickson, H. P. (2007). Evolution of the cytoskeleton. *Bioessays, 29*(7), 668-677.

Errington, J. (2015). Bacterial morphogenesis and the enigmatic MreB helix. *Nature Reviews (Microbiology), 13*(4), 241-248.

Eswaramoorthy, P., Erb, M. L., Gergory, J. A., Silverman, J., Pogliano, K., Pogliano, J., & Ramamurthi, K. S. (2011). Cellular architecture mediates DivIVA ultrastructure and regulates min activity in Bacillus subtilis. *mBio, 2*(6), e00257-11.

Figurski, D. H., & Helinski, D. R. (1979). Replication of an origin-containing derivative of plasmid RK2 dependent on a plasmid function provided in trans. *Proceedings of the National Academy of Sciences of the United Sates of America, 76*(4), 1648-1652.

Flärdh, K. (2003). Essential role of DivIVA in polar growth and morphogenesis in Streptomyces coelicolor A3(2). *Molecular Microbiology, 49*(6), 1523-36.

Flärdh, K. (2003). Growth polarity and cell division in Streptomyces. *Current opinion in Microbiology, 6*(6), 564-571.

Flärdh, K., & Buttner, M. J. (2009). Streptomyces morphogenetics: dissecting differentiation in a filamentous bacterium. *Nautre reviews. Microbiology, 7*(1), 36-49.

Flärdh, K., Leibovitz, E., Buttner, M. J., & Chater, K. F. (2000). Generation of a non-sporulating strain of Streptomyces coelicolor A3(2) by the manipulation of a developmentally controlled ftsZ promoter. *Molecular Microbiology, 38*(4), 737-749.

Flärdh, K., Richards, D. M., Hempel, A. M., Howard, M., & Buttner, M. J. (2012). Regulation of apical growth and hyphal branching in Streptomyces. *Current opinion of Microbiology, 15*(6), 737-743.

Fröjd, M. J., & Flärdh, K. (2019). Apical assemblies of intermediate filament-like protein FilP are highly dynamic and affect polar growth determinant DivIVA in Streptomyces venezuelae. *Molecular Microbiology, 112*(1), 47-61.

Fu, W., Lin, J., & Cen, P. (2007). 5-Aminolevulinate production with recombinant Escherichia coli using a rare codon optimizer host strain. *Applied Microbiology and Biotechnology, 75*(4), 777-782.

Fuchino, K., Bagchi, S., Cantlay, S., Sandblad, L., Wu, D., Bergman, J., . . . Ausmees, N. (2013). Dynamic gradients of an intermediate filament-like cytoskeleton are recruited by a polarity landmark during apical growth. *Proceedings of the National Academy of Sciences of the United States of America, 110*(27), 1889-1897.

Fuchino, K., Flärdh, K., Dyson, P., & Ausmees, N. (2016). Cell-Biological Studies of Osmotic Shock Response in Streptomyces spp. *Journal of Bacteriology, 199*(1), e00465-16.

Garner, E. C., Bernard, R., Wang, W., Zhuang, X., Rudner, D. Z., & Mitchison, T. (2011). Coupled, circumferential motions of the cell wall synthesis machinery and MreB filaments in B. subtilis. *Science (New York), 333*(6039), 222-225.

Gerdes, K., Howard, M., & Szardenings, F. (2010). Pushing and pulling in prokaryotic DNA segregation. *Cell, 141*(6), 927-942.

Gerding, M. A., Lui, B., Bendezú, F. O., Hale, C. A., Bernhardt, T. G., & de Boer, P. A. (2009). Self-enhanced accumulation of FtsN at Division Sites and Roles for Other Proteins with a SPOR domain (DamX, DedD, and RlpA) in Escherichia coli cell constriction. *Journal of Bacteriology, 191*(24), 7383-7401.

Giacomelli, G., Feddersen, H., Peng, F., Martins, G. B., Grafemeyer, M., Meyer, F., . . . Bramkamp, M. (2022). Subcellular Dynamics of a Conserved Bacterial Polar Scaffold Protein. *Genes (Basel), 13*(2), 278.

Gillespie, M. D. (Unpublished). 'ParH: A Novel Regulator of Septum Site Placement in Streptomyces coelicolor'. *PhD thesis*, University of East Anglia, Norwich.

Goffin, C., & Ghuysen, J. M. (1998). Multimodular penicillin-binding proteins: an enigmatic family of orthologs and paralogs. *Microbiology and molecular biology reviews, 62*(4), 1079-1093.

González, J. M., Whitman, W. B., Hodson, R. E., & Moran, M. A. (1996). Identifying numerically abundant culturable bacteria from complex communities: an example from a lignin enrichment culture. *Applied and Environmental Microbiology, 62*(12), 4433-4440.

Grangeon, R., Zupan, J. R., Anderson-Furgeson, J., & Zambryski, P. C. (2015). PopZ identifies the new pole, and PodJ identifies the old pole during polar growth in Agrobacterium tumefaciens. *Proceedings of the National Academy of Sciences of the United States of America, 112*(37), 11666-11671.

Grant, S. G., Jessee, J., Bloom, F. R., & Hanahan, D. (1990). Differential plasmid rescue from transgenic mouse DNAs into Escherichia coli methylation-restriction mutants. *Proceedings of the National Academy of Sciences, 87*(12), 4645-4649.

Grantcharova, N., Lustig, U., & Flärdh, K. (2005). Dynamics of FtsZ assembly during sporulation in Streptomyces coelicolor A3(2). *Journal of Bacteriology, 187*(9), 3227-3337.

Grenier, F., Matteau, D., Baby, V., & Rodrigue, S. (2014). Complete Genome Sequence of Escherichia coli BW2511. *Genome annoucements, 2*(5), e01038-14.

Gruber, S., & Errington, J. (2009). Recruitment of condensin to replication origin regions by ParB/SpoOJ promotes chromosome segregation in B. subtilis. *Cell, 137*(4), 685-696.

Gündoğdu, M. E., Kawai, Y., Pavlendova, N., Ogasawara, N., Errington, J., Scheffers, D. J., & Hamoen, L. W. (2011). Large ring polymers align FtsZ polymers for normal septum formation. *EMBO Journal, 30*(3), 617-627.

Haeusser, D. P., & Margolin, W. (2016). Splitsville: structural and functional insights into the dynamic bacterial Z ring. *Nature Reviews. Microbiology, 14*(5), 305-319.

Hale, C. A., & de Boer, P. A. (1999). Recruitment of ZipA to the septal ring of Escherichia coli is dependent on FtsZ and independent of FtsA. *Journal of Bacteriology, 181*(1), 167-176.

Hamoen, L. W., Meile, J. C., de Jong, W., Noirot, P., & Errington, J. (2006). SepF, a novel FtsZ-interacting protein required for a late step in cell division. *Molecular Microbiology, 59*(3), 989-999.

Harry, E., Monahan, L., & Thompson, L. (2006). Bacterial cell division: the mechanism and its precison. *International review of Cytology, 253*, 27-94.

Heinsch, S. C., Hsu, S.-Y., Otto-Hanson, L., Kinkel, L., & Smanski, M. J. (2019). Complete genome sequences of Streptomyces spp. isolated from disease-suppressive soils. *BMC Genomics, 20*(1).

Hempel, A. M., Cantlay, S., Molle, V., Wang, S. B., Naldrett, M. J., Parker, J. L., . . . Flärdh, K. (2012). The Ser/Thr protein kinase AfsK regulates polar growth and hyphal branching in the filamentous bacteria Streptomyces. *Poceedings of the National Academy of Sciences of the United States of America, 109*(35), E2371-E2379.

Hempel, A. M., Wang, S. B., Letek, M., Gil, J. A., & Flärdh, K. (2008). Assemblies of DivIVA mark sites for hyphal branching and can establish new zones of cell wall growth in Streptomyces coelicolor. *Journal of Bacteriology, 190*(22), 7579-83.

Hester, C. M., & Luthkenhaus, J. (2007). Soj (ParA) DNA binding is mediated by conserved arginines and is essential for plasmid segregation. *Proceedings of the National Academy of Sciences of the United States of America, 104*(51), 20326-20331.

Holmes, N. A., Walshaw, J., Leggett, R. M., Thibessard, A., Dalton, K. A., Gillespie, M. D., . . . Kelemen, G. H. (2013). Coiled-coil protein Scy is a key component of a multiprotein assembly controlling polarized growth in Streptomyces. *Proceedings of the National Academy of Sciences of the United States of America, 110*(5), 397-406.

Hu, Z., & Lutkenhaus, J. (2001). Topological regulation of cell division in E. coli. spatiotemporal oscillation of MinD requires stimulation of its ATPase by MinE and phospholipid. *Molecular Cell, 7*(6), 1337-1343.

Hu, Z., Gogol, E. P., & Lutkenhaus, J. (2002). Dynamic assembly of MinD on phospholipid vesicles regulated by ATP and MinE. *Proceedings of the National Academy of Sciences of the United States of America, 99*(10), 6761-6766.

Huang, K. H., Durand-Heredia, J., & Janakiraman, A. (2013). FtsZ ring stability: of bundles, tubules, crosslinks, and curves. *Journal of Bacteriology, 195*(9), 1859-1868.

Hutchings, M. I., Hong, H. J., & Buttner, M. J. (2006). The vancomycin resistance VanRS two-component signal transduction system of Streptomyces coelicolor. *Molecular microbiology, 59*(3), 923-935.

Hutchinson, A. (Unpublished). 'Cellular localization of coiled-coil proteins in Streptomyces'. *BSc dissertation*, University of East Anglia, Norwich.

Ireland, D. (Unpublished). 'Characterisation of a novel coiled-coil protein Sco5569 in Streptomyces'. *BSc dissertation*, University of East Anglia.

Ireton, K., Gunther, N. W., & Grossman, A. D. (1994). spo0J is required for normal chromosome segregation as well as the initiation of sporulation in Bacillus subtilis. *Journal of Bacteriology, 176*(17), 5320-5329.

Ishikawa, S., Kawai, Y., Hiramatsu, K., & Kuwano, M. (2006). A new FtsZ-interacting protein, YlmF, complements the activity of FtsA during progression of cell division in Bacillus subtilis. *Moelcular Microbiology, 60*(6), 1364-1380.

Jakimowicz, D., Chater, K., & Zakrzewska-Czerwi´nska, J. (2002). The ParB protein of Streptomyces coelicolor A3(2) recognizes a cluster of parS sequences within the origin-proximal region of the linear chromosome. *Molecular Microbiology, 45*(5), 1365-1377.

Jakimowicz, D., Gust, B., Zakrzewska-Czerwinska, J., & Chater, K. F. (2005). Developmental-stage-specific assembly of ParB complexes in Streptomyces coelicolor hyphae. *Journal of Bacteriology, 187*(10), 3572-3580.

Jakimowicz, D., Zydek, P., Kois, A., Zakrzewska-Czerwińska, J., & Chater, K. F. (2007). Alignment of multiple chromosomes along helical ParA scaffolding in sporulating Streptomyces hyphae. *Molecular Microbiology, 65*(3), 625-641.

Jiménez, M., Martos, A., Vicente, M., & Rivas, G. (2011). Reconstitution and organization of Escherichia coli proto-ring elements (FtsZ and FtsA) inside giant unilamellar vesicles obtained from bacterial inner membranes. *Journal of Biological Chemistry, 286*(13), 11236-11241.

Jones, L. J., Carballido-López, R., & Errington, J. (2001). Control of cell shape in bacteria: helical, actin-like filaments in Bacillus subtilis. *Cell, 104*(6), 913-922.

Jurásek, M., Flärdh, K., & Vácha, R. (2020). Effect of membrane composition on DivIVA-membrane interaction. *Biochimica et biophysica acta (Biomembranes), 1862*(8), 183144.

Jyothikumar, V., Tilley, E. J., Wali, R., & Herron, P. R. (2008). Time-lapse microscopy of Streptomyces coelicolor growth and sporulation. *Applied and environmental microbiology, 74*(21), 6774-6781.

Kang, C. M., Nyayapathy, S., Lee, J. Y., Suh, J. W., & Husson, R. N. (2008). Wag31, a homologue of the cell division protein DivIVA, regulates growth, morphology and polar cell wall synthesis in mycobacteria. *Micriobiology (Reading), 154*(Pt 3), 725-735.

Karimova, G., Ullmann, A., & Ladant, D. (2000). A bacterial two-hybrid system that exploits a cAMP signaling cascade in Escherichia coli. *Methods in Enzymology, 328*, 59-73.

Karoui, M. E., & Errington, J. (2001). Isolation and characterization of topological specificity mutants of minD in Bacillus subtilis. *Molecular Microbiology, 42*(5), 1211-1221.

Kawai, Y., & Ogasawara, N. (2006). Bacillus subtilis EzrA and FtsL synergistically regulate FtsZ ring dynamics during cell division. *Microbiology (Reading), 152*(Pt 4), 1129-1141.

Keenan, T., Dowle, A., Bates, R., & Smith, M. C. (2019). Characterization of the Streptomyces coelicolor Glycoproteome Reveals Glycoproteins Important for Cell Wall Biogenesis. *mBio, 10*(3), e01092-19.

Keijser, B. J., Noens, E. E., Kraal, B., Koerten, H. K., & van Wezel, G. P. (2003). The Streptomyces coelicolor ssgB gene is required for early stages of sporulation. *FEMS Microbiology Letters, 225*(1), 59-67.

Kelemen, G. H., & Buttner, M. J. (1998). Initiation of aerial mycelium formation in Streptomyces. *Current opinion in Microbiology, 1*(6), 656-662.

Kelemen, G. H., Brian, P., Flärdh, K., Chamberlin, L., Chater, K. F., & Buttner, M. J. (1998). Developmental regulation of transcription of whiE, a locus specifying the polyketide spore pigment in Streptomyces coelicolor A3 (2). *Jounal of Bacteriology, 190*(9), 2515-2521.

Kieser, T., Bibb, M., Buttner, M. J., Chater, K., & Hopwood, D. A. (2000). *Practical Streptomyces Genetics.* Norwich: John Innes Foundation.

Kim, H.-J., Calcutt, M. J., Schmidt, F. J., & Chater, K. F. (2000). Partitioning of the linear chromosome during sporulation of Streptomyces coelicolor A3(2) involves an oriC-linked parAB locus. *Journal of Bacteriology, 182*(5), 1313-1320.

Kleinschnitz, E. M., Heichlinger, A., Schirner, K., Winkler, J., Latus, A., Maldener, I., . . . Muth, G. (2011). Proteins encoded by the mre gene cluster in Streptomyces coelicolor A3(2) cooperate in spore wall synthesis. *Molecular microbiology, 79*(5), 1367-1379.

Kois-Ostrowska, A., Strzałka, A., Lipietta, N., Tilley, E., Zakrzewska-Czerwińska, J., Herron, P., & Jakimowicz, D. (2016). Unique Function of the Bacterial Chromosome Segregation Machinery in Apically Growing Streptomyces - Targeting the Chromosome to New Hyphal Tubes and its Anchorage at the Tips. *PLoS Genetics, 12*(12), e1006488.

Koonin, E. V. (1993). A superfamily of ATPases with diverse functions containing either classical or deviant ATP-binding motif. *Journal of Molecular Biology, 229*(4), 1165-1174.

Krol, E., Yau, H. C., Lechner, M., Schäper, S., Bange, G., Vollmer, W., & Becker, A. (2020). Tol-Pal System and Rgs Proteins Interact to Promote Unipolar Growth and Cell Division in Sinorhizobium meliloti. *mBio, 11*(3), e00306-20.

Kruse, T., Bork-Jensen, J., & Gerdes, K. (2005). The morphogenetic MreBCD proteins of Escherichia coli form an essential membrane-bound complex. *Molecular Microbiology, 55*(1), 78-89.

Kwak, J., Dharmatilake, A. J., Jiang, H., & Kendrick, K. E. (2001). Differential regulation of ftsZ transcription during septation of Streptomyces griseus. *Journal of Bacteriology, 183*(17), 5092-5101.

Lackner, L. L., Raskin, D. M., & do Boer, P. A. (2003). ATP-dependent interactions between Escherichia coli Min proteins and the phospholipid membrane in vitro. *Journal of Bacteriology, 185*(3), 725-749.

Lam, H., Schofield, W. B., & Jacobs-Wagner, C. (2006). A landmark protein essential for establishing and perpetuating the polarity of a bacterial cell. *Cell, 124*(5), 1011-1023.

Lenz, P., & Søgaard-Andersen, L. (2011). Temporal and spatial oscillations in bacteria. *Nature reviews (Microbiology), 9*(8), 565-577.

Leonard, T. A., Butler, P. J., & Löwe, J. (2005). Bacterial chromosome segregation: structure and DNA binding of the Soj dimer--a conserved biological switch. *EMBO journal, 24*(2), 270-282.

Letek, M., Ordóñez, E., Vaquera, J., Margolin, W., Flärdh, K., Mateos, L. M., & Gil, J. A. (2008). DivIVA is required for polar growth in the MreB-lacking rod-shaped actinomycete Corynebacterium glutamicum. *Journal of Bacteriology, 190*(9), 3283-92.

Levin, P. A., & Losick, R. (1994). Characterization of a cell division gene from Bacillus subtilis that is required for vegetative and sporulation septum formation. *Journal of Bacteriology, 176*(5), 1451-1459.

Levin, P. A., Margolis, P. S., Setlow, P., Losick, R., & Sun, D. (1992). Identification of Bacillus subtilis genes for septum placement and shape determination. *Journal of Bacteriology, 174*(21), 6717-6728.

Li, P., Zhang, H., Zhao, G. P., & Zhao, W. (2020). Deacetylation enhances ParB-DNA interactions affecting chromosome segregation in Streptomyces coelicolor. *Nucleic acids research, 48*(9), 4902-4914.

Lin, D. C., & Grossman, A. D. (1998). Identification and characterization of a bacterial chromosome partitioning site. *Cell, 92*(5), 675-685.

Lin, Y. S., Kieser, H. M., Hopwood, D. A., & Chen, C. W. (1993). The chromosomal DNA of Streptomyces lividans 66 is linear. *Molecular Microbiology, 10*(5), 923-933.

Livny, J., Yamaichi, Y., & Waldor, M. K. (2007). Distribution of centromere-like parS sites in bacteria: insights from comparative genomics. *Journal of Bacteriology, 189*(23), 8693-8703.

Locey, K. J., & Lennon, J. T. (2016). Scaling laws predict global microbial diversity. *Proceedings of the National Academy for Sciences for the United States of America, 113*(21), 5970-5975.

Louca, S., Mazel, F., Doebeli, M., & Parfrey, L. (2019). A census-based estimate of Earth's bacterial and archaeal diversity. *PLOS biology, 17*(2), e3000106.

Löwe, J., & Amos, L. A. (1998). Crystal structure of the bacterial cell-division protein FtsZ. *Nature, 391*(6663), 203-206.

Lupas, A. (1996). Coiled coils: new structures and new functions. *Trends in Biochemical Sciences, 21*(10), 375-382.

Lutkenhaus, J. (2012). The ParA/MinD family puts things in their place. *Trends in Microbiology, 20*(9), 411-418.

Margolin, W. (2000). Organelle division: Self-assembling GTPase caught in the middle. *Current Biology, 10*(9), 328-330.

Marston, A. L., & Errington, J. (1999). Selection of the midcell division site in Bacillus subtilis through MinD-dependent polar localization and activation of MinC. *olecular Microbiology, 33*(1), 84-96.

Marston, A. L., Thomaides, H. B., Edwards, D. H., Sharpe, M. E., & Errington, J. (1998). Polar localization of the MinD protein of Bacillus subtilis and its role in selection of the mid-cell division site. *Genes & Development, 12*(21), 3419-30.

Mason, J. M., & Arndt, K. M. (2004). Coiled coil domains: stability, specificity, and biological implications. *Chembiochem, 5*(2), 170-176.

McCormick, J. R., Su, E. P., Driks, A., & Losick, R. (1994). Growth and viability of Streptomyces coelicolor mutant for the cell division gene ftsZ. *Molecular Microbiology, 14*(2), 243-254.

Meléndez, A. B., Menikpurage, I. P., & Mera, P. E. (2019). Chromosome Dynamics in Bacteria: Triggering Replication at the Opposite Location and Segregation in the Opposite Direction. *mBio, 10*(4), e01002-19.

Mohammadi, T., van Dam, V., Sijbrandi, R., Vernet, T., Zapun, A., Bouhss, A., . . . Breukink, E. (2011). Identification of FtsW as a transporter of lipid-linked cell wall precursors across the membrane. *EMBO Journal, 30*(8), 1425-1432.

Mukherjee, P., Sureka, K., Datta, P., Hossain, T., Barik, S., Das, K. P., . . . Basu, J. (2009). Novel role of Wag31 in protection of mycobacteria under oxidative stress. *Molecular Microbiology, 73*(1), 103-119.

Murray, H., & Errington, J. (2008). Dynamic control of the DNA replication initiation protein DnaA by Soj/ParA. *Cell, 135*(1), 74-84.

Nogales, E., Downing, K. H., Amos, L. A., & Lowe, J. (1998). Tubulin and FtsZ form a distinct family of GTPases. *Nature Structural Biology, 5*(6), 451-458.

Ohashi, T., Hale, C. A., de Boer, P. A., & Erickson, H. P. (2002). Structural evidence that the P/Q domain of ZipA is an unstructured, flexible tether between the membrane and the C-terminal FtsZ-binding domain. *Journal of Bacteriology, 184*(15), 4313-4315.

Oliva, M. A., Halbedel, S., Freund, S. M., Dutow, P., Leonard, T. A., Veprintsev, D. B., . . . Löwe, J. (2010). Features critical for membrane binding revealed by DivIVA crystal structure. *EMBO Journal, 29*(12), 1988-2001.

Osawa, M., Anderson, D. E., & Erickson, H. P. (2008). Reconstitution of contractile FtsZ rings in liposomes. *Science, 320*(5877), 792-194.

Park, K. T., Wu, W., Battaile, K. P., Lovell, S., Holyoak, T., & Lutkenhaus, J. (2011). The Min oscillator uses MinD-dependent conformational changes in MinE to spatially regulate cytokinesis. *Cell, 146*(3), 396-407.

Park, K. T., Wu, W., Lovell, S., & Lutkenhaus, J. (2012). Mechanism of the asymmetric activation of the MinD ATPase by MinE. *Molecular Microbiology, 85*(2), 271-281.

Parte, A. C., Carbasse, J. S., Meier-Kolthoff, J. P., Reimer, L. C., & Göker, M. (2020). List of prokaryotic names with standing in nomenclature (LPSN) moves to the DSMZ. *International Journal of Systematic and Evolutionary Microbiology, 70*(11), 5607-5612.

Passot, F. M., Cantlay, S., & Flärdh, K. (2022). Protein phosphatase SppA regulates apical growth and dephosphorylates cell polarity determinant DivIVA in Streptomyces coelicolor. *Molecular Microbiology, 117*(2), 411-428.

Patrick, J. E., & Kearns, D. B. (2008). MinJ (YvjD) is a topological determinant of cell division in Bacillus subtilis. *Moelcular Microbiology, 70*(5), 1166-79.

Phadnis, S. H., & Das, H. K. (1987). Use of the plasmid pRK 2013 as a vehicle for transposition inAzotobacter vinelandii. *Journal of Biosciences, 12*(2), 131-135.

Pichoff, S., & Lutkenhaus, J. (2002). Unique and overlapping roles for ZipA and FtsA in septal ring assembly in Escherichia coli. *EMBO Journal, 21*(4), 685-693.

Pichoff, S., Shen, B., Sullivan, B., & Lutkenhaus, J. (2012). FtsA mutants impaired for self-interaction bypass ZipA suggesting a model in which FtsA's self-interaction competes with its ability to recruit downstream division proteins. *Moelcular Microbiology, 83*(1), 151-167.

Pióro, M., Małecki, T., Portas, M., Magierowska, I., Trojanowski, D., Sherratt, D., . . . Jakimowicz, D. (2018). Competition between DivIVA and the nucleoid for ParA binding promotes segrosome separation and modulates mycobacterial cell elongation. *Molecular Microbiology, 111*(1), 204-220.

Ptacin, J. L., Gahlmann, A., Bowman, G. R., Perez, A. M., von Deizmann, L., Eckart, M. R., . . . Shapiro, L. (2014). Bacterial scaffold directs pole-specific centromere segregation. *Proceedings of the National Academy of Sciences of the United States of America, 111*(19), E2046-E2055.

Ptacin, J. L., Lee, S. F., Garner, E. C., Toro, E., Eckart, M., Comolli, L. R., . . . Shapiro, L. (2010). A spindle-like apparatus guides bacterial chromosome segregation. *Nature cell biology, 12*(8), 791-798.

QIAGEN. (2018). *QIAGEN QIAquick Gel Extraction Kit*. Retrieved April 20, 2022, from https://www.qiagen.com/us/products/discovery-and-translational-research/dna-rna-purification/dna-purification/dna-clean-up/qiaquick-gel-extraction-kit/

Ramamurthi, K. S., & Losick, R. (2009). Negative membrane curvature as a cue for subcellular localization of a bacterial protein. *Proceedings of the National Academy of Sciences of the United States of America, 106*(32), 13541-13555.

Redenbach, M., Kieser, H. M., Denapaite, D., Eichner, A., Cullum, J., Kinashi, H., & Hopwood, D. A. (1996). A set of ordered cosmids and a detailed genetic and physical map for the 8 Mb Streptomyces coelicolor A3(2) chromosome. *Molecular Microbiology, 21*(1), 77-96.

Richards, D. M., Hempel, A. M., Flärdh, K., Buttner, M. J., & Howard, M. (2012). Mechanistic basis of branch-site selection in filamentous bacteria. *PLoS computational biology, 8*(3), e1002423.

Ringgaard, S., Ebersbach, G., Borch, J., & Gerdes, K. (2007). Regulatory cross-talk in the double par locus of plasmid pB171. *Journal of Biological chemistry, 282*(5), 3134-3145.

Ringgaard, S., van Zon, J., Howard, M., & Gerdes, K. (2009). Movement and equipositioning of plasmids by ParA filament disassembly. *Proceedings of the National Academy of Sciences of the United States of America, 106*(46), 19369-19374.

Rowlett, V. W., & Margolin, W. (2013). The bacterial Min system. *Current Biology, 23*(13), R553-R556.

Saalbach, G., Hempel, A. M., Vigouroux, M., Flärdh, K., Buttner, M. J., & Naldrett, M. J. (2013). Determination of phosphorylation sites in the DivIVA cytoskeletal protein of Streptomyces coelicolor by targeted LC-MS/MS. *Journal of proteosome research, 12*(9), 4187-4192.

Salje, J., van den Ent, F., de Boer, P., & Löwe, J. (2011). Direct membrane binding by bacterial actin MreB. *Molecular Cell, 43*(3), 478-87.

Schäfer, A., Tauch, A., Jäger, W., Kalinowski, J., Thierbach, G., & Pühler, A. (1994). Small mobilizable multi-purpose cloning vectors derived from the Escherichia coli plasmids pK18 and pK19: selection of defined deletions in the chromosome of Corynebacterium glutamicum. *Gene, 145*(1), 69-73.

Scholefield, G., Whiting, R., Errington, J., & Murray, H. (2011). Spo0J regulates the oligomeric state of Soj to trigger its switch from an activator to an inhibitor of DNA replication initiation. *Molecular Microbiology, 79*(4), 1089-1100.

Sharp, M. D., & Pogliano, K. (2003). The membrane domain of SpoIIIE is required for membrane fusion during Bacillus subtilis sporulation. *Journal of Bacteriology, 185*(6), 2005-2008.

Sharpe, M. E., & Errington, J. (1996). The Bacillus subtilis soj-spo0J locus is required for a centromere-like function involved in prespore chromosome partitioning. *Molecular Microbiology, 21*(3), 501-509.

Shen, B., & Lutkenhaus, J. (2009). The conserved C-terminal tail of FtsZ is required for the septal localization and division inhibitory activity of MinC(C)/MinD. *Molecular Microbiology, 72*(2), 410-424.

Soppa, J., Kobayashi, K., Noirot-Gros, M. F., Oesterhelt, D., Ehrlich, S. D., Dervyn, E., . . . Moriya, S. (2002). Discovery of two novel families of proteins that are proposed to interact with prokaryotic SMC proteins, and characterization of the Bacillus subtilis family members ScpA and ScpB. *Molecular Microbiology, 45*(1), 59-71.

Stogios, P. J., & Savchenko, A. (2020). Molecular mechanisms of vancomycin resistance. *Protein Science, 29*(3), 654-669.

Sullivan, N. L., Marquis, K. A., & Rudner, D. Z. (2009). Recruitment of SMC by ParB-parS organizes the origin region and promotes efficient chromosome segregation. *Cell, 137*(4), 697-707.

Swulius, M. T., Chen, S., Jane Ding, H., Li, Z., Briegel, A., Pilhofer, M., . . . Jensen, G. J. (2011). Long helical filaments are not seen encircling cells in electron cryotomograms of

rod-shaped bacteria. *Biochemical and Biopphysical Research Communications, 407*(4), 650-655.

Szafan, M., Skut, P., Ditkowski, B., Ginda, K., Chandra, G., Zakrzewska-Czerwińska, J., & Jakimowicz, D. (2013). Topoisomerase I (TopA) is recruited to ParB complexes and is required for proper chromosome organization during Streptomyces coelicolor sporulation. *Journal of Bacteriology, 195*(19), 4445-4455.

Szwedziak, P., Wang, Q., Freund, S. M., & Löwe, J. (2012). FtsA forms actin-like protofilaments. *EMBO Journal, 31*(10), 2249-2260.

van den Ent, F., Vinkenvleugel, T. M., Ind, A., West, P., Veprintsev, D., Nanninga, N., . . . Löwe, J. (2008). Structural and mutational analysis of the cell division protein FtsQ. *Molecular Microbiology, 68*(1), 110-123.

van den Ent, F., Izoré, T., Bharat, T. A., Johnson, C. M., & Löwe, J. (2014). Bacterial actin MreB forms antiparallel double filaments. *eLife, 3*(e02634).

van Wezel, G. P., van der Meulen, J., Kawamoto, S., Luiten, R. G., Koerten, H. K., & Kraal, B. (2000). ssgA is essential for sporulation of Streptomyces coelicolor A3(2) and affects hyphal development by stimulating septum formation. *Journal of Bacteriology, 182*(20), 5653-5662.

Varley, A. W., & Stewart, G. C. (1992). The divIVB region of the Bacillus subtilis chromosome encodes homologs of Escherichia coli septum placement (minCD) and cell shape (mreBCD) determinants. *Journal of Bacteriology, 174*(21), 6729-6742.

Wachi, M., Doi, M., Okada, Y., & Matsuhashi, M. (1989). New mre genes mreC and mreD, responsible for formation of the rod shape of Escherichia coli cells. *Journal of Bacteriology, 171*(12), 6511-6516.

Wachi, M., Doi, M., Tamaki, S., Park, W., Nakajima-Iijima, S., & Matsuhashi, M. (1987). Mutant isolation and molecular cloning of mre genes, which determine cell shape, sensitivity to mecillinam, and amount of penicillin-binding proteins in Escherichia coli. *Journal of Bacteriology, 169*(11), 4935-9440.

Walker, J. E., & Dickson, V. K. (2006). The peripheral stalk of the mitochondrial ATP synthase. *Biochimica et Biophysica acta, 1757*(5-6), 286-296.

Walshaw, J., Gillespie, M. D., & Kelemen, G. H. (2010). A novel coiled-coil repeat variant in a class of bacterial cytoskeletal proteins. *Journal of Structural Biology, 170*(2), 202-215.

Walter, S., Wellmann, E., & Schrempf, H. (1998). The cell wall-anchored Streptomyces reticuli avicel-binding protein (AbpS) and its gene. *Journal of Bacteriology, 180*(7), 1647-1654.

Wang, S. B., Cantlay, S., Nordberg, N., Letek, M., Gil, J. A., & Flärdh, K. (2009). Domains involved in the in vivo function and oligomerization of apical growth determinant DivIVA in Streptomyces coelicolor. *FEMS microbiology letters, 297*(1), 101-109.

Watanabe, E., Inamoto, S., Lee, M. H., Kim, S. U., Ogua, T., Mori, H., . . . Nagai, K. (1989). Purification and characterization of the sopB gene product which is responsible for

stable maintenance of mini-F plasmid. *Molecular & general Genetics (MGG), 218*(3), 431-436.

Weiss, D. S. (2004). Bacterial cell division and the septal ring. *Molecular Microbiology, 54*(3), 588-597.

Willemse, J., Borst, J. W., de Waal, E., Bisseling, T., & van Wezel, G. P. (2011). Positive control of cell division: FtsZ is recruited by SsgB during sporulation of Streptomyces. *Genes & Development, 25*(1), 89-99.

Xu, Z., Masuda, Y., Itoh, H., Ushijima, N., Shiratori, Y., & Senoo, K. (2019). Geomonas oryzae gen. nov., sp. nov., Geomonas edaphica sp. nov., Geomonas ferrireducens sp. nov., Geomonas terrae sp. nov., Four Ferric-Reducing Bacteria Isolated From Paddy Soil, and Reclassification of Three Species of the Genus Geobacter as Members of. *Frontiers in Microbiology, 10*, 2201.

Yamaichi, Y., & Niki, H. (2004). migS, a cis-acting site that affects bipolar positioning of oriC on the Escherichia coli chromosome. *EMBO Journal, 23*(1), 221-233.

Zupan, J. R., Grangeon, R., Robalino-Espinosa, J. S., Garnica, N., & Zambryski, P. (2019). GROWTH POLE RING protein forms a 200-nm-diameter ring structure essential for polar growth and rod shape in Agrobacterium tumefaciens. *Proceedings of the National Academy of Sciences of the United States of America, 116*(22), 10962-10967.

Zupan, J., Guo, Z., Biddle, T., & Zambryski, P. (2021). Agrobacterium tumefaciens Growth Pole Ring Protein: C Terminus and Internal Apolipoprotein Homologous Domains Are Essential for Function and Subcellular Localization. *mBio, 12*(3), e00764-21.