Molecular mechanisms and dynamics of the segregation of plasmid pB171 from an enteropathogenic strain of *Escherichia coli*

Iman Alnaqshabandy

PhD

University of York

Biology

January 2020

Abstract

The stable inheritance of bacterial low copy number plasmids is mediated by dedicated active segregation systems. The *Escherichia coli* B171 enteropathogenic strain specifies virulence factors that cause diarrhea and are encoded by the low copy number plasmid pB171. The plasmid contains two partition cassettes. The system under investigation is the *par2* module, which consists of three elements: two genes, *parA* encoding an ATPase and *parB* encoding a centromere-binding protein, and two partition or centromere sites, *parC1* and *parC2,* which harbour direct repeats that are recognised by the ParB protein. An initial and crucial event in plasmid segregation is the binding of the centromere-binding protein site-specifically to the centromere site to assemble a higher-order nucleoprotein complex structure.

In this work, *in vivo* experiments have shown that the *parC2* centromere is fundamental for plasmid stability. ParB is a dimer in solution. The crystal structure of the C-terminal DNA-binding domain of ParB was solved at 1.93 Å resolution and revealed that the protein folds into a dimeric ribbon-helix-helix motif similar to that of the Arc/MetJ transcriptional repressors superfamily. The solution nuclear magnetic resonance showed that the ParB N-terminal domain is unstructured and highly dynamic. The secondary structure prediction showed that the central region of ParB N-terminal domain forms a α-helix structure. ParB binds *parC1* and *parC2* with high affinity and cooperativity. Nuclear magnetic resonance experiments identified two ParB regions that recognise DNA: the C-terminus and the tip of the N-terminal tail. Replacement of residues in the tip of ParB N-terminal tail provided insights into the role played by these residues in plasmid stability. Interestingly, chemical cross-linking coupled with mass spectrometry revealed that the centromere site and ATP enhance ParA-ParB complex formation and showed that the tip of ParB N-terminal tail interacts with the partner ParA. Based on these findings, a model of nucleoprotein complex assembly is proposed.

Table of Contents

List of Tables

List of Figures

Acknowledgments

Firstly, I would like to express my sincere gratitude and appreciation to my first supervisor Dr Daniela Barillà, for the great opportunity she has given me to work on such an amazing project. She has supported me in every possible way, lending her time and giving me valuable feedback. Secondly, I would like to thank my second supervisor, Dr Michael Plevin, for his suggestions and insightful comments. He has always encouraged me and helped me gain confidence.

Many thanks to my TAP members, Professor James Moir and Dr Nathalie Signoret for their invaluable advice and feedback on my research and for always being so supportive of my work.

Special thanks to the past and current members of the Barillà group and to the L1 lab, particularly, I would like to say a very big thank you to my fellow lab mates, Dr Azhar Kabli, Dr Cecilia Pennica and John Armstrong for sharing this journey with me and for the stimulating discussions. I would also like to thank all the staff members in the Department who have been very helpful and supportive throughout.

My thanks also go out to the support I received from Dr Alex Heyam for the experimental setup and data processing, and Dr Clément Dégut for the structural analysis. I would also like to thank Dr Chris Taylor for the kind help with mass spectrometry experiments. In addition, I would like to thank Dr Andrew Leech for his help with SEC-MALLS experimental setup.

Furthermore, I gratefully acknowledge the funding received from the Higher Committee for Education Development in Iraq (HCED) to financially support me, my PhD research and my family during this truly life-changing experience.

Finally, my biggest strength and motivation, my family. My parents, Hussein and Ahlam… You have always believed in me and have always encouraged me. My siblings, for helping in whatever way they could during this challenging period. My husband, Ravak, who has been by my side throughout this PhD, living every single minute of it, and without whom, I would not have had the courage to embark on this journey in the first place. My adorable Payan and little Peleen for being such a wonderful joy in my life and making it possible for me to complete what I have started. This thesis would not have been possible if it was not for your constant love, endless support and prayers. This thesis is dedicated to you.

Author's declaration

I declare that this thesis is a presentation of original work and I am the sole author. Work concerning mass spectrometry was performed by Dr Chris Taylor. NMR experimental set up and data processing was performed by Dr Alex Heyam. Structural analysis of the protein was performed by Dr Clément Dégut. This work has not previously been presented for an award at this, or any other, University. All sources are acknowledged as References.

Chapter 1

Introduction

1.1 Plasmids

Plasmids are extra-chromosomal circular pieces of DNA that can replicate independently from the chromosome. They are mainly found in bacteria as well as in some archaea and eukaryotes (Wang, 2017). Although plasmids are non-essential for the survival of host cells (Clowes, 1972), they may encode a wide variety of genetic elements that allow their host to survive better in a challenging environment or to compete better with other microorganisms that are present in the same ecological niche (Million-Weaver and Camps, 2014; Wang, 2017). Plasmids can be involved in many essential activities in bacteria including metabolism, virulence, resistance to antibiotics and evolution (Wang, 2017). One of the key features of plasmids that represents a medical threat for both humans and animals is their ability to retain and transmit genes for resistance to drugs and pathogenicity (Actis, *et al*., 1999).

Plasmids are typically present in multiple copies, and the number of copies is usually constant under a vast variety of growth conditions. The copy number can vary from one to two copies per cell for large and low-copy-number plasmids, like P1 and F1 plasmids, or as many as more than 100 copies per cell for ColE1 plasmids. The size of plasmids can also vary, ranging from a few kilobases to more than 100 kbp (Actis, *et al*., 1999; Million-Weaver and Camps, 2014). Because of their small size, plasmids have been widely used in many areas including scientific research, industrial fermentation, in addition to agricultural and medical applications (Wang, 2017). Moreover, plasmids are one of the major providers of horizontal gene transfer (Salje, 2010) and thus, understanding the mechanism of plasmid maintenance is of crucial importance (Wang, 2017).

DNA replication is not sufficient to ensure stable plasmid maintenance in a growing host population (Actis, *et al*., 1999). Consequently, plasmids present in high copy-number are usually lost at a high rate in the absence of selective pressure. These plasmids undergo random diffusion to ensure that each dividing cell receives copies of the plasmid (Million-Weaver and Camps, 2014). Plasmids that replicate at the low-copy-number (> 10 copies per cell) are usually large in size and require a specific mechanism to ensure stable transmission to daughter cells during cell division. In fact, these plasmids harbour an active partitioning (*par*) system to ensure stable distribution of plasmids upon cell division (Summers and Sherratt, 1985). Some plasmids, such as F and ColE1 plasmids, ensure their maintenance in the cell line using a mechanism called post-segregational killing, in which plasmid-free cells are selectively killed (Sengupta and Austin, 2011).

1.2 Maintenance of plasmids

1.2.1 Random diffusion of high copy-number plasmids

High copy-number plasmids lack genes that encode active partitioning systems. These plasmids depend on random distribution to ensure that each daughter cell receives a suitable number of plasmids. In the random diffusion model, it is assumed simply that the plasmids diffuse freely throughout the cytoplasm and segregate randomly during cell division. The probability of a plasmid-free daughter cell in each round of cell division increases dramatically and it depends on the plasmid copy number (Summers, 1998). The loss rate is given by $L = (1/2)^{2n}$, where n is the plasmid copy number (Nordström and Gerdes, 2003). Based on this equation, if the bacterial cell harbours > 15 copies of a plasmid, the probability of the plasmid-free cell is vanishing small. Therefore, cells are capable of maintaining their plasmids for many generations without the need to develop an active segregation system (Summers and Sherratt, 1985; Summers, 1998).

Furthermore, it was shown that the high copy number plasmid ColE1, which distributes randomly, was more stable than any other ColE1 derivative plasmids. This instability was due to high order multimerisation of certain derivative plasmids. Therefore, after investigation, it was noted that the ColE1 plasmid encodes a determinant called *cer*, that is essential for recombination events that facilitate the resolution of multimers into monomers. Accordingly, it was concluded that random distribution is simplified by multimer resolution system, where site-specific DNA recombination is employed to resolve plasmid dimers and oligomers into monomers (Summers and Sherratt, 1984; Summers, 1998).

Many studies had provided evidence against a random distribution model. For instance, fluorescent microscopy experiments that used elongated bacterial cells to distinguish plasmids from chromosomal DNA had shown that plasmids form clusters at cell poles, which are excluded from the bacterial nucleoid. Significantly, the number of foci that were observed was less than the copy number, which strongly supports the idea of cluster formation by plasmids (Eliasson *et al.*, 1992). This idea was later confirmed by another study, which showed that clusters were either at cell poles or in the mid-quarter of the cell. Furthermore, it was shown that these clusters divided into two, in which one of them moved to the middle of the cell (Yao, *et al.*, 2007).

A recent study on ColE1 derivative plasmids using super-resolution fluorescence microscopy had shown that some plasmids form a cluster, whilst others diffuse randomly in the bacterial cell. However, not all plasmids were excluded from the nucleoid region,

as previously shown(Yao, *et al.*, 2007; Wang, *et al.*, 2016). Instead, some of the plasmids were found to diffuse within the chromosomal volume (Wang, *et al.*, 2016).

1.2.2 Post-segregational-killing system (Toxin/antitoxin)

To prevent plasmid loss in populations, many plasmids utilise a mechanism, which eliminates the cells that lost their plasmid after cell division through a system known as post-segregational killing (Gerdes, *et al.*, 1986; Jensen and Gerdes, 1995; Tsang, 2017). This system was first discovered in low copy number plasmids, F and R1 (Ogura and Hiraga, 1983; Gerdes, *et al.*, 1986). The system consists of two components: a stable toxin and a labile antitoxin. The toxin is a protein that prevents cell growth by disrupting vital molecular processes of the bacterial cell, including; transcription, translation, DNA replication, and cell wall synthesis, while the antitoxin inhibits the toxin's activity (Kedzierska and Hayes, 2016).

Toxin-antitoxin (TA) systems (Figure 1.1) are encoded on bacterial chromosomes as well as on plasmids (Díaz-orejas, *et al.*, 2017). The two adjacent genes are part of a single operon, with the antitoxin gene upstream of the toxin gene. The antitoxin gene is expressed more abundantly, and the protein or RNA is more unstable than the toxin. In the TA system, the toxin is always a protein, whereas the antitoxin is either a protein or a small regulatory RNA (Hayes, 2003; Kedzierska and Hayes, 2016). Most of the toxins are small with a compact structure, whereas structures of antitoxins showed some regions that lack proper folding. Thus, they are more prone to degradation. The degradation of antitoxins is caused either by endonucleases, when the antitoxin is a RNA or by proteases when it is a protein (Kedzierska and Hayes, 2016).

TA systems can be classified into six types (Figure 1.1) based on the antitoxin properties and the mechanisms by which the antitoxins inhibit the activity of the toxin. In type I systems, the toxin is a small hydrophobic protein, and the antitoxin is an anti-sense RNA. The latter binds the toxin mRNA and this leads to inhibition of the antitoxin mRNA translation (Kedzierska and Hayes, 2016; Yang and Walsh, 2017). The first discovered and best described type I module is the *hok-sok* system of *Escherichia coli* R1 plasmid. This system comprises three partially overlapping genes. The *hok* gene encodes an extremely toxic transmembrane protein that damages the cell membrane. The *mok* open reading frame overlaps with *hok* and is responsible for the expression and regulation of *hok* RNA translation. The *sok* gene is an antisense RNA that prevents *mok* mRNA translation, thus the translation of *hok* is inhibited indirectly (Hayes, 2014; Tsang, 2017).

In the type II system, both toxin and antitoxin are small proteins that together form a stable complex. The antitoxin inhibits the toxin activity by blocking the region responsible for the toxicity. An example of a type II system is the CcdA/CcdB system that was first identified as a maintenance system for the F plasmid in *E. coli* (Ogura and Hiraga, 1983; Jaffe, *et al*., 1985). Genes homologous to *ccdA*/*ccdB* are widely distributed on bacterial chromosomes and plasmids. The CcdB toxin targets the topoisomerase II DNA gyrase through interaction with the catalytic GyrA subunit of gyrase and entrapping the gyrase cleavage complex. CcdA antitoxin inhibits the reverse interaction of CcdB with the gyrase (Hayes, 2003; Sengupta and Austin, 2011; Hayes and Kedzierska, 2014; Hall, *et al.*, 2017).

In the type III system, the antitoxin, which is a RNA molecule, binds to the toxin, which is a protein, and blocks the activity of the toxin. Similar to the type II system, both elements in type IV and V systems are proteins. In type IV, the toxin is prevented from accessing the target by its cognate antitoxin. An example of this type is the CbeA antitoxin that acts as an inhibitor for its cognate toxin activity and stimulates polymerization of FtsZ and MreB during cell division (Kedzierska and Hayes, 2016). In contrast, in type V, the antitoxin, is an endoribonuclease, like GhoS, that inhibits toxin mRNA synthesis (Kedzierska and Hayes, 2016). Recently, a new type has been identified, in which the antitoxin works as a proteolytic adapter for the ClpXP protease, whereby the toxin undergoes degradation from the antitoxin action (Goeders and Melderen, 2014; Hayes and Kedzierska, 2014; Kedzierska and Hayes, 2016; Yang and Walsh, 2017). Currently, more than 100 TA models have been characterised and so far, the first two types of TA system, type II and I, seem to be the most abundant among prokaryotes compared to the other systems (Yang and Walsh, 2017).

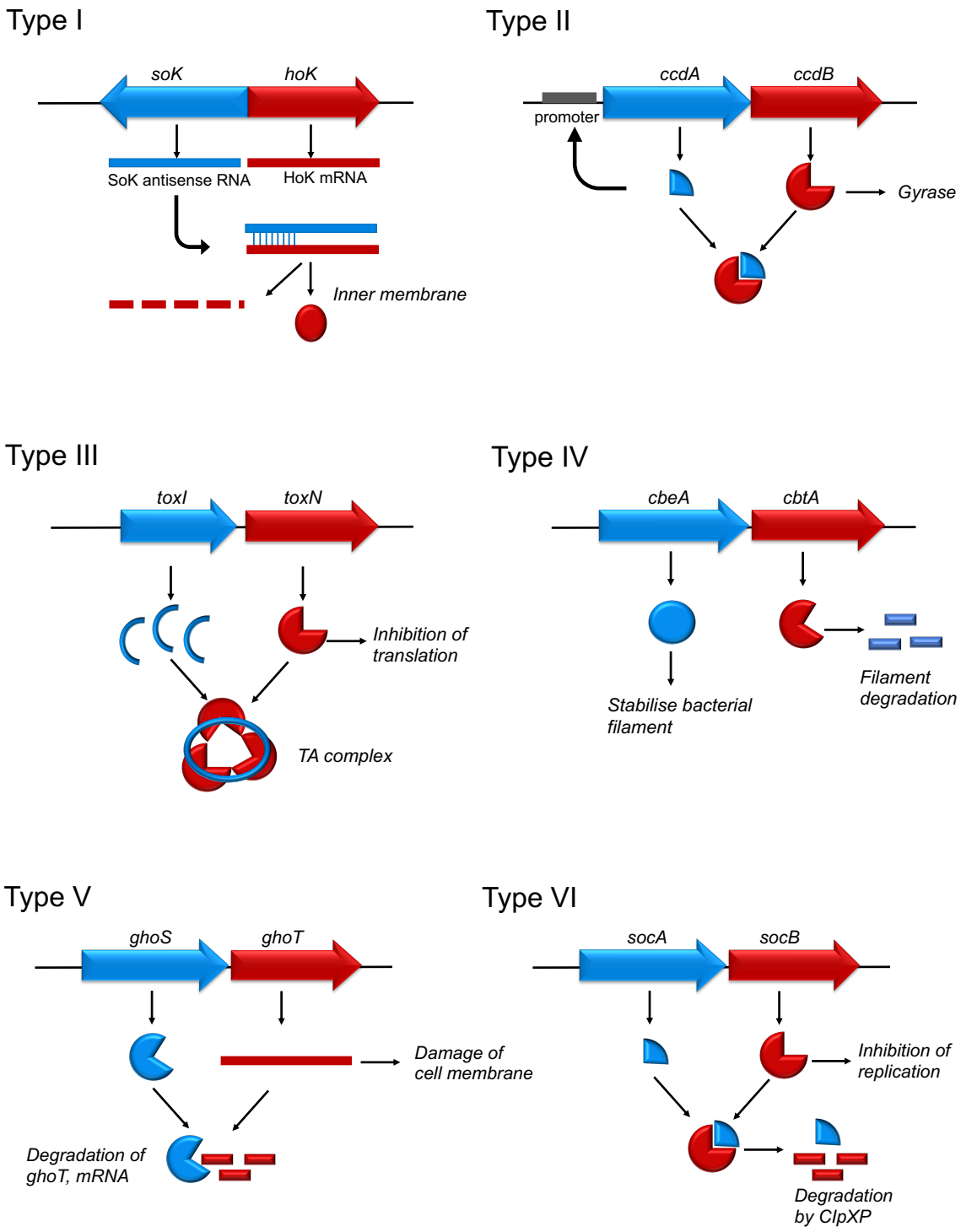


Figure 1.1 Genetic organisation and cellular mechanisms of different TA systems

The diagram shows six classes of TA systems (type I, II, III, IV, V and VI). The genes encoding toxins are indicated as red arrows and the genes encoding antitoxins are indicated as blue arrows (adapted from Page and Peti, 2016).

1.2.3 Active plasmid partition system

Bacterial low-copy-number plasmids requires an active partitioning mechanism that ensures stable maintenance of the genetic element after cell division. Partition systems are encoded on most bacterial chromosomes as well as low-copy-number plasmids (Hiraga, 2000; Baxter and Funnell, 2014; Misra, *et al.*, 2018). In active partitioning, the process is dynamic, and plasmids are moved and positioned at the one quarter and third quarter of the bacterial cell prior to cell division so that each daughter cell receives a copy of the newly replicated plasmid (Baxter and Funnell, 2014; Oliva, 2016).

The genetic organisation of bacterial plasmid partition (*par*) systems is quite simple and the components are conserved. The *par* systems consists of three elements; two proteins and one or more partition sites (Hiraga, 2000; Baxter and Funnell, 2014; Misra, *et al.*, 2018). The DNA partition sites represent *cis*-acting centromeric sites, consisting of tandem repeats that are present either in direct or inverted orientation. The number and the sequence of repeats vary between different *par* systems (Figure 1.2). The number and location of *par* sites is equally varied. They may be located either upstream, close to or overlapping with the promoter of the operon. Moreover, they can be positioned directly downstream of the *par* genes (Figure 1.3), and the system may include only one or multiple sites on the plasmid. In contrast, chromosomes often harbor several *par* sites mainly located close to the origin of replication (Hayes and Barillà, 2006; Baxter and Funnell, 2014; Million-Weaver and Camps, 2014; Funnell, 2016). One of the proteins of the system is a motor protein, either an ATPase or a GTPase, which utilises energy from the hydrolysis of nucleotides triphosphate to drive the movement of plasmid DNA within the bacterial cell. The other protein is a site-specific DNA-binding protein that associates to the centromeric site. In general, mechanisms of segregation involve sequence-specific interactions of centromere binding proteins and *par* sites to form the nucleoprotein complexes (Hayes and Barillà, 2006). The motor protein interacts with these complexes forming a larger structure called segrosome.

Par systems can be divided into numerous classes, based on the properties of the motor protein (Figure 1.2). The type I system encodes a Walker-type ATPase protein that contains the canonical Walker-A P-Loop motif for adenosine triphosphate (ATP) binding. Examples of type I systems are those encoded on the F and P1 plasmids in *E. coli*. The type II system encodes an actin-like ATPase protein that resembles eukaryotic actin. The type III system relies on a tubulin-like GTPase. Recently, a fourth type of segregation system has been identified, which encodes only a single protein serving as a motor and an adaptor protein (Baxter and Funnell, 2014; Million-Weaver and Camps, 2014; Oliva,

2016; Misra, *et al.*, 2018). Based on the size of the ATPase protein, type I systems can be further divided into two sub-groups, type Ia and type Ib systems. In type Ia systems the ATPase protein is large as it contains an additional site-specific DNA binding domain that is responsible for autoregulation of the expression of the *par* operon. The type Ib ATPase lacks this domain and is smaller. Therefore, in type Ib systems the centromere binding protein acts as a transcriptional repressor (Sengupta and Austin, 2011; Baxter and Funnell, 2014; Oliva, 2016).

Based on the structures of centromere binding proteins, they can be classified into two groups: helix-turn-helix (HTH) in type Ia system and ribbon-helix-helix (RHH) in type Ib and type II systems (Baxter and Funnell, 2014; Funnell, 2016).

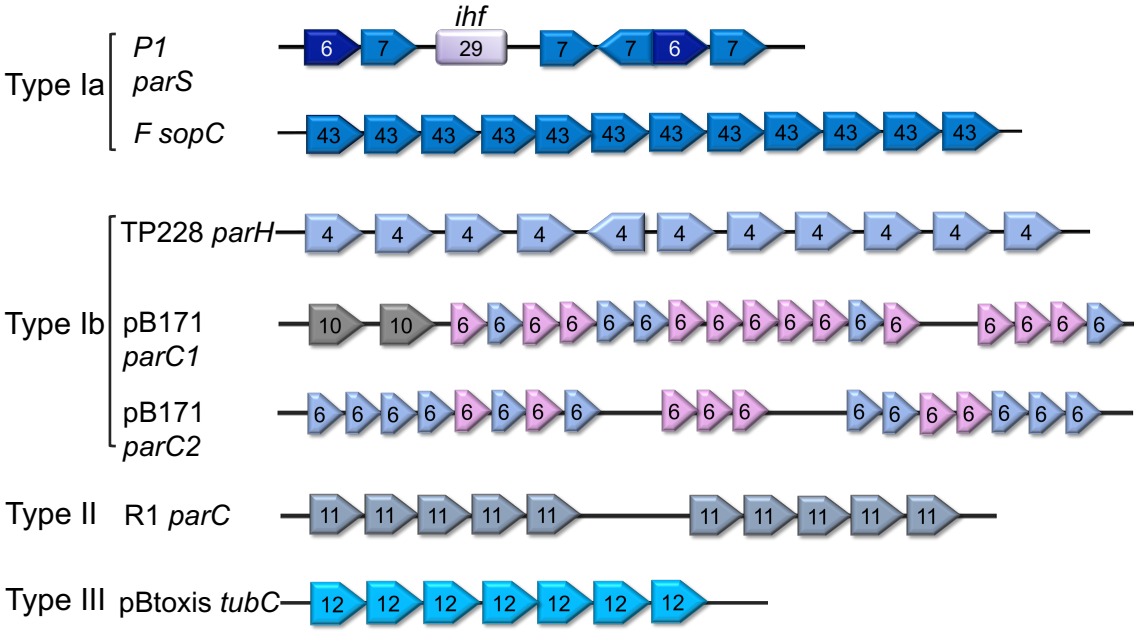


Figure 1.2 Different *par* centromere sites

The schematic represents the centromere sites for each specific type of *par* system (type Ia, type Ib, type II and type III). The arrows represent the number of tandem repeats within each individual *par* site. The arrows are schematic and do not represent the length of each repeat. The number of nucleotides in each repeat are indicated and the orientation (directed or inverted) of every single repeat is shown.

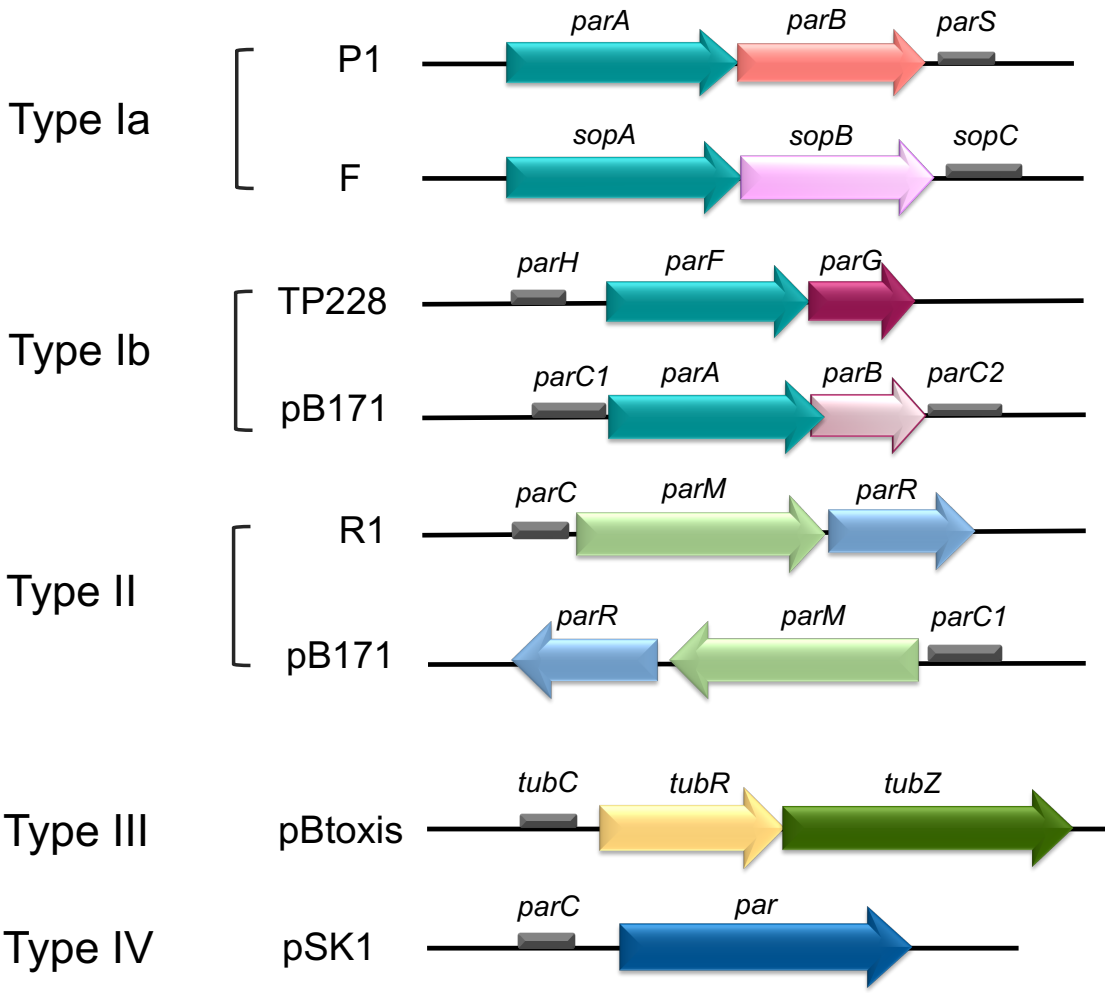


Figure 1.3 Genetic organisation of different *par* systems

The schematic represents the partition cassette of different classes of *par* systems (type Ia, type Ib, type II, type III and type IV). The genes encoding (NTPase) proteins are indicated with green arrows (teal for type Ia and Ib system, light green for type II and dark green for type III system). The genes encoding centromere binding proteins for type I systems are indicated with salmon colour for P1 plasmid, light purple for F plasmid, red for TP228 and pink for pB171. The gene encoding the centromere binding protein for type II system is indicated with a blue arrow, whereas for type III is indicated with a yellow arrow. The centromere sites for each *par* system are indicated with grey bars and are labelled. The gene encoding for type IV system is indicated with dark blue arrow. The *par* sites are located downstream of the cassette for type Ia system and upstream for type Ib, II, III and IV systems.

1.4 Types of segregation systems

1.4.1 Type I system

The type I systems (Figure 1.3) are ubiquitous in nature as they are present both on bacterial and archaeal chromosomes as well as on low copy number plasmids. This type is characterised by the Walker-type ATPase protein. The system consists of two proteins: a Walker-A P-loop ATPase (ParA) protein and a centromere binding protein (ParB), and one or more of centromere-like (*parS*) site. Based on the genetic organisation of the *par* operon and different sizes and sequences of both proteins, type I systems can be subdivided into two groups, types Ia and Ib systems (Figure 1.3). In both types, the location of genes encoding the ATPase proteins are upstream of the centromere binding protein. However, the location of the centromeric sites varies hence it is located downstream of *par* operon in type Ia, and upstream in type Ib (Schumacher, 2008).

The ATPase proteins in type Ia systems are larger than the type Ib homologues. The type Ib ParA Walker box proteins are small, consisting of 200-230 residues. In contrast, the ParA proteins of type Ia are large, consisting of 250-440 residues that besides the Walker box motifs, contain an additional HTH DNA-binding motif in the N-terminal domain. The DNA-binding domain in large ParAs allows sequence-specific interaction with the operator sites for transcriptional autoregulation of the *par* operon (Figure 1.4). In type Ib systems, the ATPase proteins are not involved in operon regulation. Transcriptional autoregulation of type Ib systems is achieved through centromere-binding proteins (Figure 1.4) (Hayes and Barillà, 2006a; Schumacher, 2008). Nevertheless, ATPases of both subgroups of Type I system share the same feature of binding non-specifically to the nucleoid to drive the plasmid movement (Baxter and Funnell, 2014).

Interestingly, the site-specific DNA-binding proteins of type Ia systems are larger than those of type Ib systems and they do not share sequence similarity. The most conserved region amongst the centromere-binding proteins of type Ia is the HTH motif. In contrast, in type Ib system, the site-specific DNA-binding protein is smaller and displays a RHH motif. The *parABS* and *sopABC* loci of P1 and F plasmids respectively are among the best-investigated cassettes of type Ia systems. The *par* system of TP228 and pSM19035 are among the best characterized type Ib systems (Sengupta and Austin, 2011; Baxter and Funnell, 2014; Oliva, 2016).

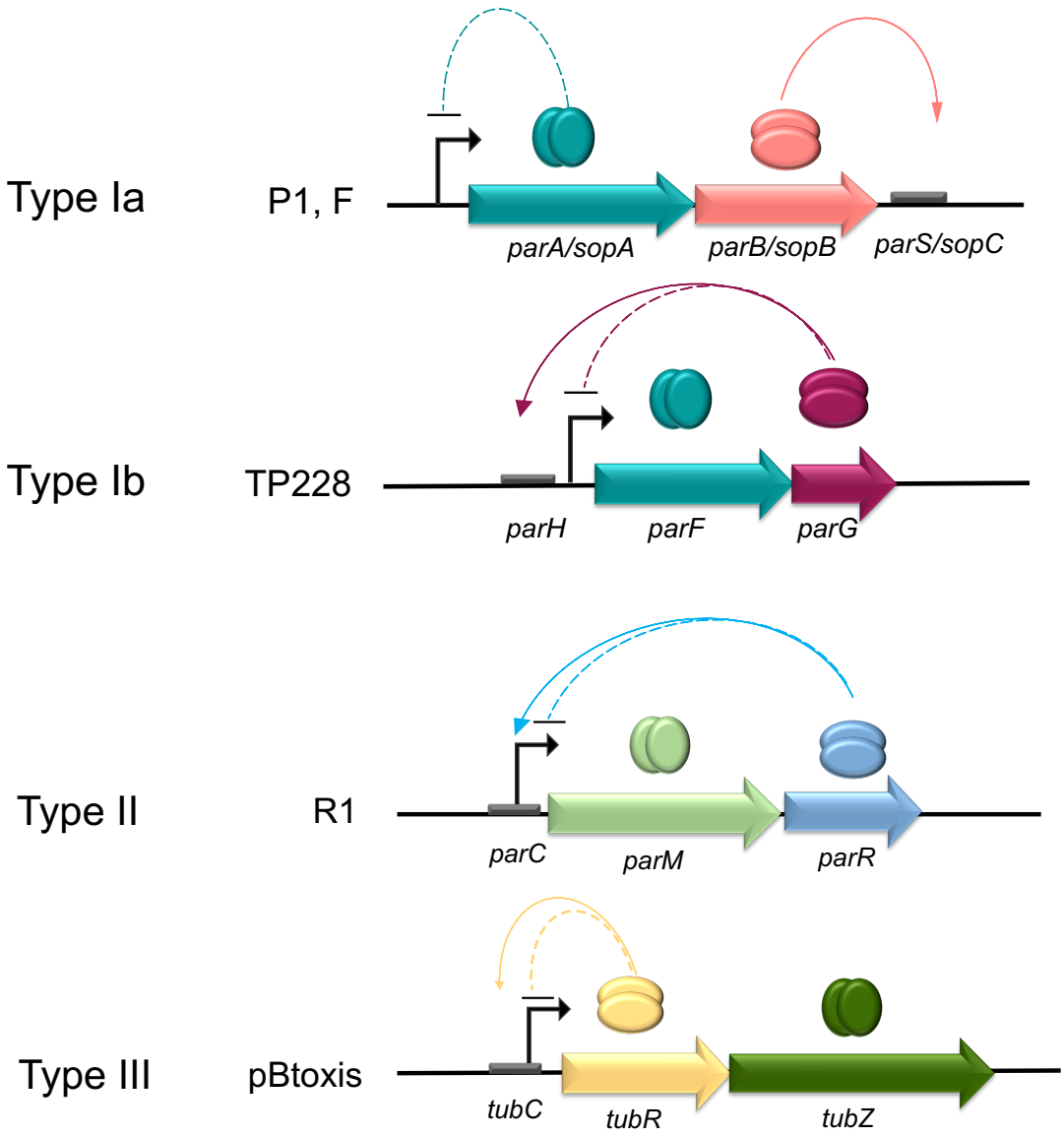


Figure 1.4 Genetic organisation and regulation of partition operon

The schematic shows the auto regulation of partition operon of type Ia, Ib, II and III partition systems. Dashed arcs represent regulation of the partition operon by the indicated proteins. In type Ia, the ATPase proteins (ParA and SopA) act as transcriptional regulator. In type Ib, II and III, the centromere binding protein acts as transcriptional repressor. Solid arcs represent the formation of the partition complexes by the centromere binding protein (adapted from Ebersbach and Gerdes, 2005).

1.4.1.1 Type Ia segregation system of P1 plasmid

The best-studied plasmid segregation system is that harboured on the P1 plasmid from *Escherichia coli* (Oliva, 2016). The partition system of the P1 plasmid is encoded by a 2.5 kb region (Austin and Abeles, 1983). The Par system consists of two *trans*-acting proteins, ParA, a Walker type ATPase protein and ParB, a unique DNA-Binding protein, and a centromere-like site or partition site *parS*, which is located downstream of the *parAB* cassette. Besides the partition elements, another protein was shown to be involved in plasmid partitioning, which is the Integration Host Factor (IHF) (Figure 1.5*A*) (Funnell, 1988b).

The centromeric site *parS* contains two types of repeats, four copies of a heptameric sequence named as A-box (G/T TGAAAT), and two copies of a hexameric sequence called the B-box (TCGCCA) (Figure 1.5*A*). Both boxes are recognised by ParB and are arranged into two clusters separated with 32 bp (Davis and Austin, 1988; Funnell, 1988a). The centromeric site, *parS*, can be divided into three regions: left, right and central regions. The left region consists of A1 and B1 boxes, whereas the right region comprises of A2, A3, and B2 boxes. The two A and B repeats flank a central region, a 29 bp sequence recognised by the IHF protein (Figure 1.5*A*). The arrangement and orientation of the *parS* motifs are essential for complex formation *in vitro* as well as *parS* activity *in vivo*. Structural studies showed that the IHF site is essential for DNA bending. This allows the ParB protein to bind as a dimer on the *parS* site bringing both arms of *parS* together, resulting in the formation of a high-affinity protein-DNA complex (Figure 1.5*B*) (Schumacher *et al.*, 2007b). When IHF is not present, ParB can still binds to these motifs site-specifically, however, the binding affinity is weaker, and the protein requires only the *parS* right region for activity. The so-called small *parS*, which consists of A2, A3, and B2 boxes, is sufficient for partition, but not optimal, as ParB can only form a weak complex with it. The binding of ParB to the left arm requires the presence of IHF, which allows ParB to extend from the box B1 to the B2 box (Martin *et al.*, 1991; Funnell and Gagnier, 1993; Hayes and Austin, 1994). Once the initial partition complex is formed through the binding of IHF to the central region of *parS,* a ParB dimer binds to the left and right arms of *parS* (Schumacher *et al*., 2007b). ParB has been shown to spread beyond its site-specific binding sites, extending at least 11 kb. ParB spreading leads the formation of a large nucleoprotein structure (Rodionov *et al.*, 1999).

P1 plasmid ParB is a 333 amino acid residues protein and was shown to share very little sequence homology with other ParB proteins. ParB crystal structure revealed that the N-terminal region consists of residues 1-141 and is highly flexible as limited exposure to

proteases leads to degradation of this part of the protein, as shown in Figure 1.6*A* (Schumacher and Funnell, 2005). The study showed that this flexible region is the domain bounded by the partner ParA. The most rigid and stable region corresponds to residues 142-333. This region contains the determinants needed for ParB to dimerise as well as to form a higher order complex structure on the *parS* site. Later, the crystal structure of ParB in complex with a short sequence of the *parS* site (small *parS*) was also solved (Figure 1.6*B*). Two DNA-binding domains have been determined, HTH domain (147-270), which is the essential for *parS* binding, and a second domain comprising residues 275-333 (dimerisation domain), which is also a DNA-binding domain. The two regions are separated by a short flexible linker, which consists only of 4 residues. This highly flexible linker between the two domains permits free movement and rotation of the domains, thus, allowing ParB to bind to a variety of A and B boxes within the *parS* site (Schumacher and Funnell, 2005). The ParB structure shows that the HTH motif binds to A boxes, whereas the dimerization domain binds to B boxes. Moreover, the structure shows how ParB bridges the left and the right arms (Schumacher and Funnell, 2005).

ParA is a weak Walker-type ATPase, that interacts with its partner ParB through the highly flexible N-terminal domain. Together, they form a complex structure that is enhanced by the presence of ATP and the ATPase activity of ParA (Davis *et al.*, 1996). ParA also autoregulates expression of the *par* operon by binding to the operator site upstream of the *parAB* cassette. Localization of the ParB-*parS* complexes in the cell is ParA dependent as the *in vivo* studies showed that the absence of ParA in the cell leads to misdistribution of the ParB-*parS* complexes (Erdmann *et al.*, 1999). The structure of ParA has three domains: an α-helix N-terminal domain, a winged-HTH central domain, and a C-terminal domain that contains a deviant Walker A motif (Figure 1.7*A* and 1.7*B*) (Dunham *et al.*, 2009). ParA binds to the chromosome, ParB and the operator site (Davis *et al.*, 1996; Bouet and Funnell, 1999; Vecchiarelli *et al.*, 2010). Studies have shown that ParA can switch between functions by binding to adenine nucleotides. More precisely, ParA binds non-specifically to the chromosomal DNA in the presence of ATP to direct plasmid segregation and binds in a site-specific manner to the operator (*parOP*) in the presence of adenosine diphosphate (ADP) to repress *parAB* transcription. Once ParA binds to ATP, it undergoes a slow conformational change in response to ParB (Vecchiarelli *et al.*, 2010). The extra 100 residues in the N-terminal region of ParA are consistent with its repressor function, that is absent in all ParAs of type Ib (Baxter and Funnell, 2014)

Based on results obtained from *in vivo* studies as well as the biochemical characterisation, a diffusion-ratchet model for ParA-mediated plasmid partition has been suggested. In this model, a concentration gradient of ParA on the nucleoid surface is suggested to provide the driving force for the plasmid cargo movement (Hwang *et al.*, 2013; Vecchiarelli *et al.*, 2014a; Vecchiarelli *et al.*, 2015). This mechanism of segregation will be discussed in more details under the “Mechanisms of plasmid segregation” Section 1.7.2.

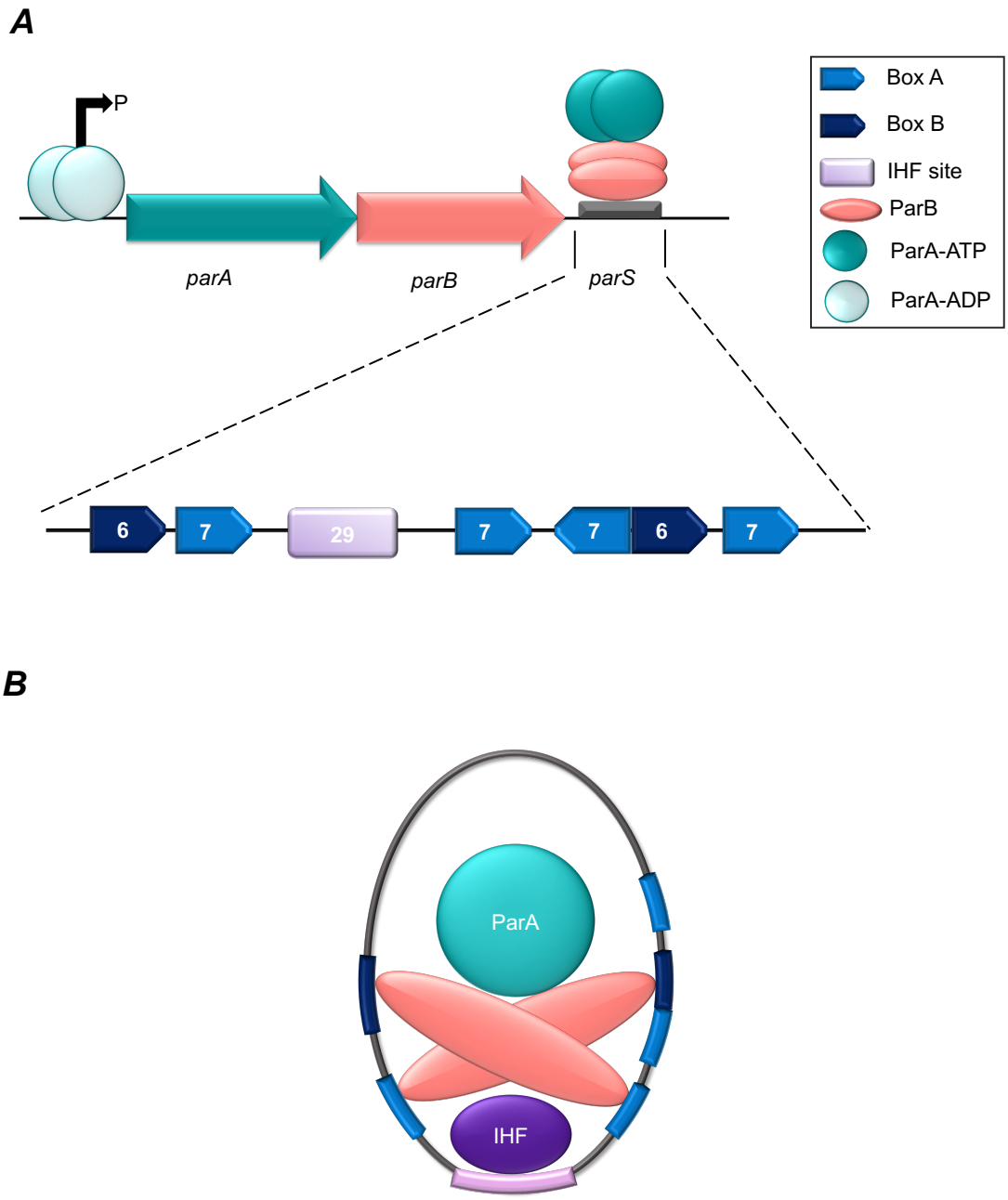


Figure 1.5 Genetic organisation of the *parABS* cassette and model of segrosome assembly of P1 plasmid partition complex

A) the diagram represents the *parABS* cassette, the teal arrow represents the gene encoding the Walker-type ATPase protein, ParA, and the gene encoding the CBP, ParB, is indicated with an orange arrow. The *parS* site is shown as a grey bar, located downstream of *parB*. ParB recognizes 7 bp box-A repeats and 6 bp box-B repeats. ParA-ADP binds the operator site for transcription regulation (light teal), whereas ParA-ATP binds ParB and the nucleoid for plasmid partition (teal). ***B)*** IHF (dark purple) binds the 29 bp *ihf* site (light purple) and enhances ParB binding to the *parS* site by bending the DNA and bringing the left and right arms of *parS* close together. ParB dimers can bind Box A and Box B at the same time to form the nucleoprotein structure.

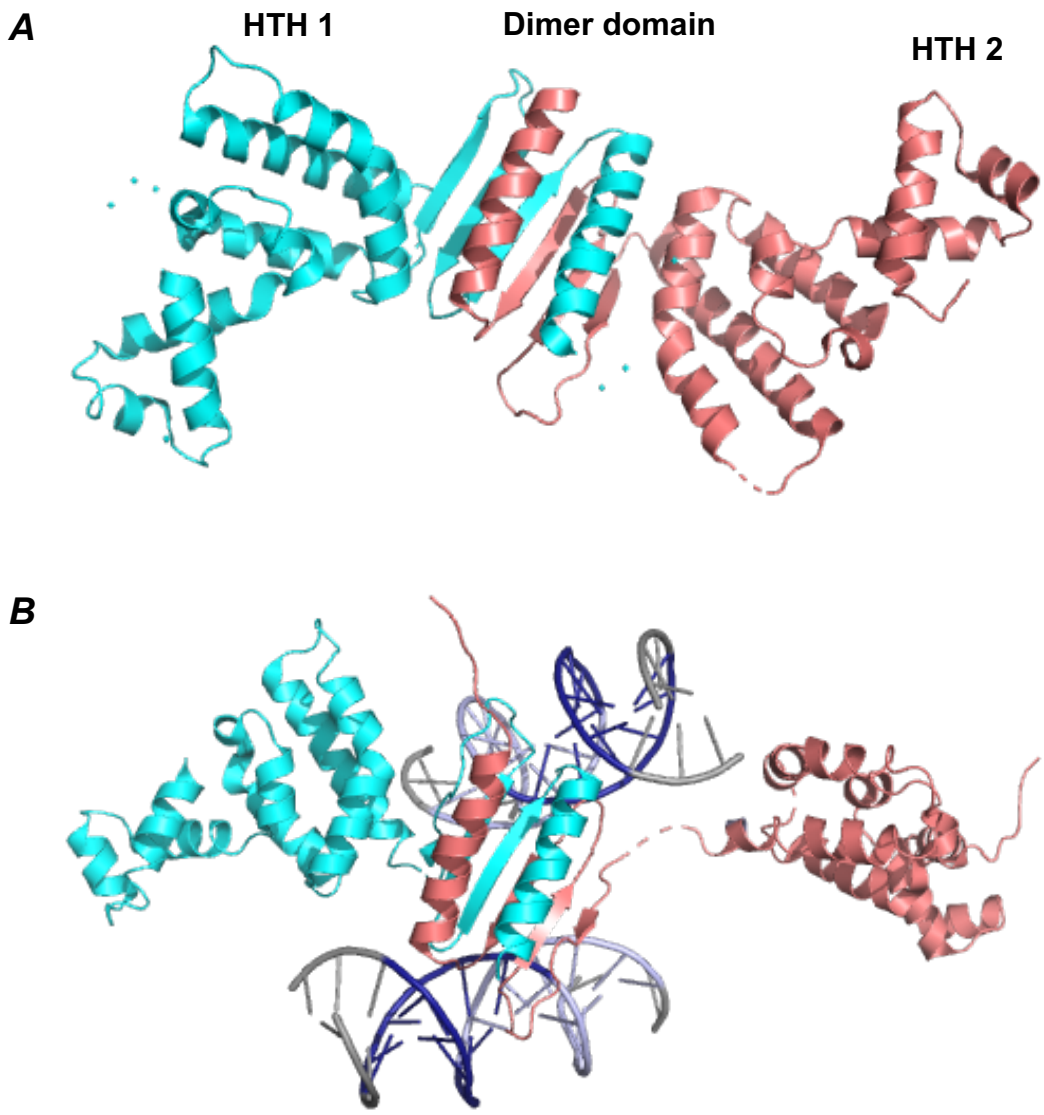


Figure 1.6 Crystal structures of P1 ParB

A) ParB dimer structure (code 1ZX4) with the dimer domain identically arranged and the two monomers are colored cyan and deep salmon. ***B)*** Ribbon diagram of ParB structure in complex with small *parS* (code 2NTZ) including box A3 (light blue) and B2 (deep blue) and shows the binding of dimer domain to boxes A and B. The two boxes motif are colored same as in Figure 1.5. The two figures are generated with PyMol (DeLano, 2002).

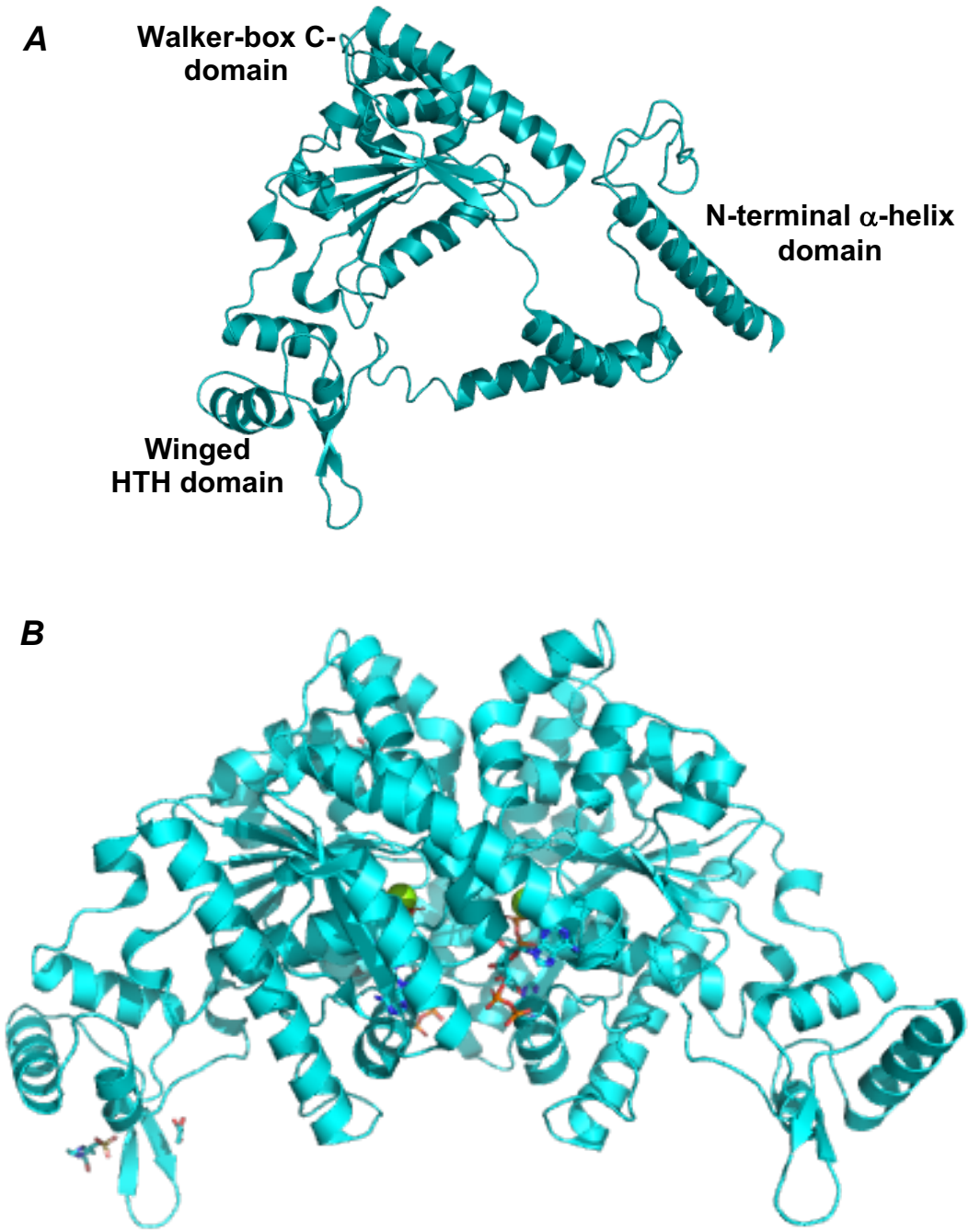


Figure 1.7 ParA crystal structure

A) A cartoon structure of apo ParA monomer (code 3EZ7) shows the three regions: an N-terminal α-helix domain, a winged HTH region and the Walker-box C-terminal domain. ***B)*** ParA dimer bound to ADP, which shows in red sticks (3EZ2). The two subunits are shown in deep cyan. The figures are generated using PyMol (DeLano, 2002).

1.4.1.2 Type Ia segregation system of F plasmid

One of the first segregation systems to be identified is that of the 100-kb conjugative plasmid F in *Escherichia coli*, belonging to type Ia subgroup. The partition system is located within the 3 kb region, outside the origin of replication (Ogura and Hiraga, 1983). Like other types I systems, the partition system of F plasmid is composed of two proteins, an ATPase protein, SopA, a centromere-binding protein, SopB, and a centromeric-like site named as *sopC* (Figure 1.8) (Ogura and Hiraga, 1983; Mori *et al.*, 1986).

Unlike the *parS* of P1 plasmid, *sopC* is a very simple centromeric region consisting of 12 tandem repeats of 43 bp (Figure 1.8). Each repeat of *sopC* consists of a 16 bp inverted repeat motif to which the SopB protein binds (Mori *et al.*, 1986). Biochemical assays have shown that all the 16 bp repeats are essential for optimal binding of SopB. Moreover, no nucleotides outside the 16 bp repeats is involved in specific binding, only five bases on both arms of the *sopC* site are essential (Pillet *et al.*, 2011).

SopB is a DNA-binding protein, that associates to the *sopC* site with high affinity through its HTH motif (Mori *et al.*, 1986, 1989; Bouet and Lane, 2009). Unlike, the partition complex of the P1 plasmid, SopB does not distort the DNA upon binding, as no host factor is involved in binding to *sopC* (Schumacher *et al*., 2010). The three-dimensional (3D) structure for SopB was determined, and the co-crystal structure of SopB-*sopC* was analysed as well (Figure 1.9*A*). Like ParB of P1 plasmid, SopB is composed of three regions: a N-terminal region or SopA-binding domain, an HTH or DNA-binding domain and the dimerisation domain. The SopB protein dimerises through the C-terminal domain (Figure 1.9*B*), which does not bind to the DNA, unlike the C-terminus of P1 ParB (Kim and Shim, 1999; Schumacher *et al.,* 2010). Moreover, structural studies have shown that SopB extends through the *sopC* site rather than wrapping around (Schumacher *et al*., 2010). The N-terminal domain is highly flexible and multifunctional. The first 45 amino acids of this domain are a sequence-specific SopA interaction region (Ravin *et al.*, 2003). In addition, the N-terminal region contains the arginine finger motif that stimulates the SopA ATPase activity (Kim and Shim, 1999; Ah-seng *et al.*, 2009). It is thought that this domain might also be involved in dimer-dimer binding. It is through the HTH domain that the SopB binds *sopC* and the SopB-*sopC* structure has shown that each dimer binds only one arm of *sopC*. Like other ParBs from type Ia systems, SopB not only forms complexes on the centromeric site but can also spread beyond the *sopC* region (Hanai *et al.*, 1996; Bouet and Lane, 2009). An *in vivo* study showed that SopB prevents negative supercoiling when the protein binds to the *sopC* site and shields the DNA (Bouet and Lane, 2009).

The function of SopB extends beyond the interaction with the *sopC* site and formation of the partition complex, it also interacts with its partner SopA. SopA is a deviant Walker-type ATPase, which works as an autoregulator and as a motor protein in the partitioning of the plasmid (Mori *et al.*, 1989). However, excessive production of SopA may disrupt plasmid maintenance (Kusukawa *et al.*, 1987). To maintain the concentration of SopA at suitable quantities, SopA binds to four repeats (5′-CTTTGC-3′) in the *sop* promoter region through the N-terminal domain to repress the *sopAB* transcription (Mori *et al.*, 1989; Ravin *et al.*, 2003). Although the repression is extremely weak, it is enhanced by the presence of SopB. SopB does not bind to the operator site directly, instead, it acts as a co-repressor, which strongly stimulates SopA binding activity in the presence of *sopC* in *cis* or *trans* (Mori *et al.*, 1989; Biek and Strings, 1995; Hirano *et al.*, 1998; Yates *et al.*, 1999; Ravin *et al.,* 2003).

Besides its autoregulatory function, SopA is one of the pivotal elements in mediating plasmid partitioning. SopA is a weak ATPase, which binds non-specifically to DNA through its ATP-dependent non-specific binding domain, and binds SopB, which provides the arginine finger-like motif (Arg-36) that is responsible for stimulating SopA ATPase activity (Ah-seng *et al.*, 2009). Lysine in position 340 has been shown to be directly involved in the binding of SopA to DNA as the ATP-dependent DNA binding activity decreased 100-fold when this lysine was changed to alanine. As a result, K340A has defected SopA activities that depends on it. Thus the plasmid stability has abolished meaning that non-specific interaction SopA with DNA is crucial for plasmid partitioning (Castaing *et al.*, 2008). Moreover, the ATPase activity of SopA is enhanced in the presence of the SopB-*sopC* complex as SopB alone failed to stimulate its ATPase activity (Watanabe *et al.*, 1992). The highest ATPase activity was observed when both non-specific DNA and the SopB-*sopC* complex were present (Castaing *et al.*, 2008; Ah-seng *et al.*, 2009).

SopA undergoes conformational changes upon binding to ATP or ADP through the Walker A motif (Libante *et al.*, 2001). SopA binds non-specifically to the DNA only in the ATP-bound form and polymerises only in the presence of ATP. *In vivo* experiments had shown that SopB stimulates SopA polymerisation and prevents DNA to breakdown the polymers by coating the DNA in a process called spreading. SopB binds to the *sopC* site and spreads beyond the region allowing the SopA-ATP polymerisation, hence initiating plasmid partitioning. SopA-ADP was found only when the protein binds to the sequence-specific in the operator site for repression (Bouet *et al.*, 2007).

Recently, Le Gall *et al.,* by using high resolution microscopy, have shown that SopA forms small batches within the nucleoid volume rather than filament formation that was observed previously (Le Gall *et al.*, 2016). The mechanisms of segregation of SopABS system is similar to that proposed for P1 plasmid and it relies on the a SopA gradient on the nucleoid, used as a medium to drive the two sister plasmids apart (Hwang *et al.*, 2013; Vecchiarelli *et al.*, 2013a; Le Gall *et al.*, 2016). This system will be discussed under the Section 1.7.2 “Mechanisms of plasmid segregations”

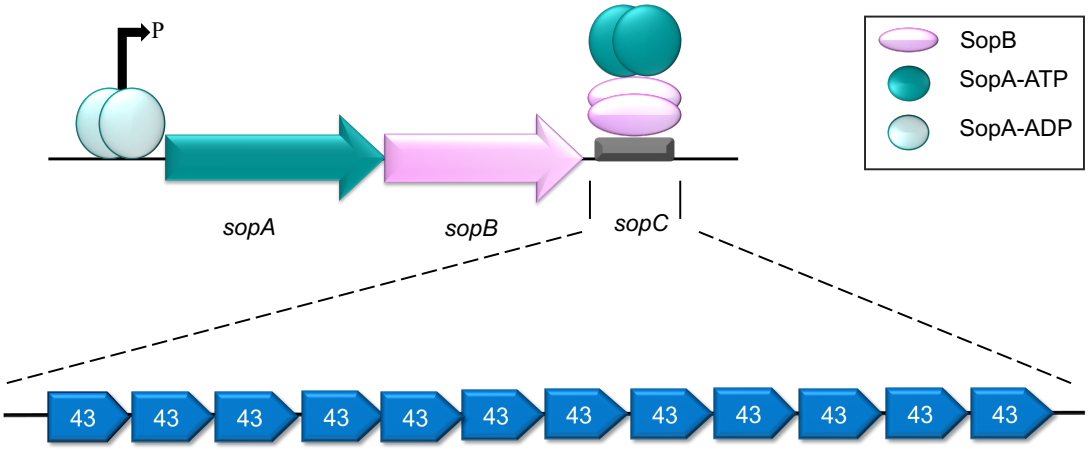


Figure 1.8 Genetic organisation of the *sopABS* cassette of F plasmid partition system

The diagram represents the *sopABS* cassette, the teal arrow represents the gene encoding the Walker-type ATPase, SopA, and the purple arrow indicates the gene encoding the CBP, SopB, and the *sopC* site is indicated with a grey bar, located downstream of *sopAB*. ParB recognizes twelve 43 bp repeats (blue arrows). SopA-ADP binds to the operator site for transcription regulation (light teal), whereas SopA-ATP binds SopB and the nucleoid for plasmid partitioning (teal).

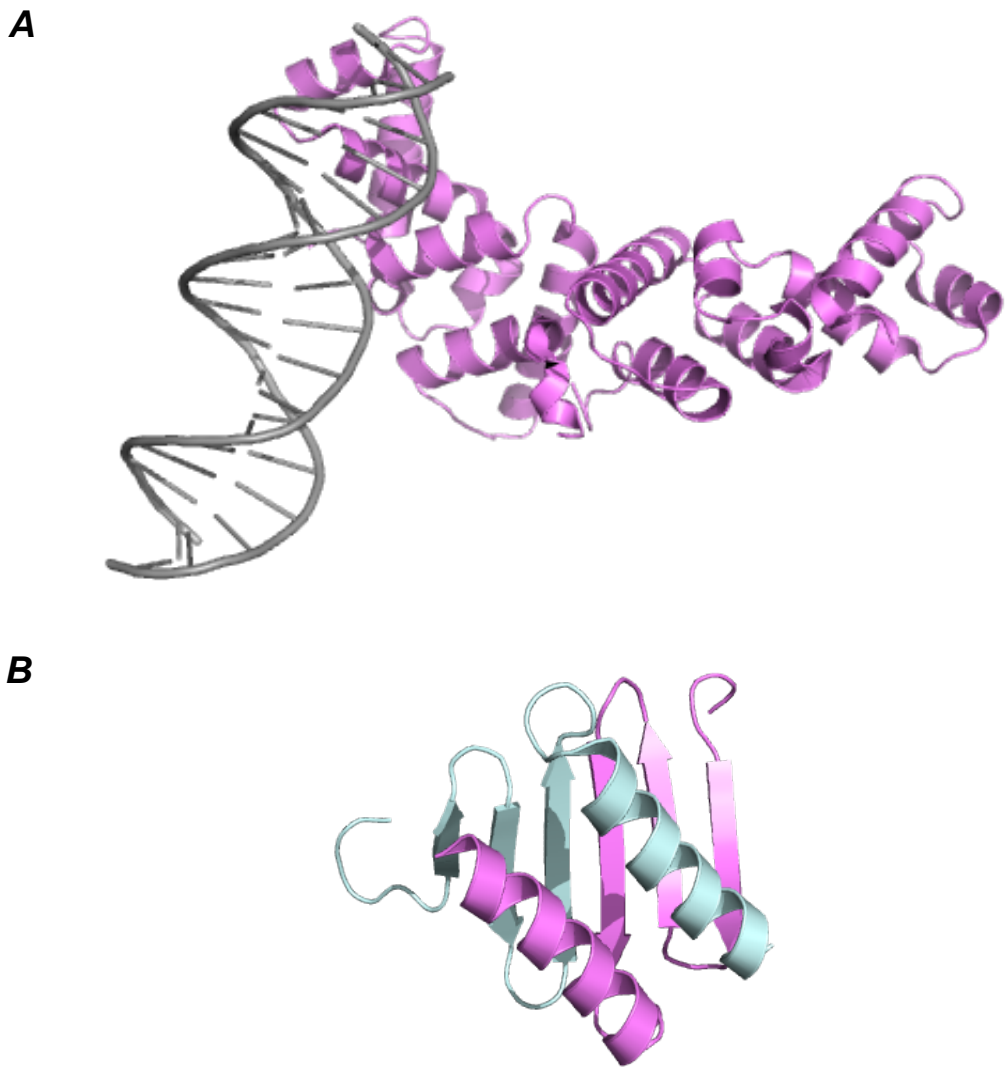


Figure 1.9 Crystal structure of F1 SopB

A) A cartoon shows the SopB dimer structure bound to DNA through the HTH domain (code 3MKY). ***B)*** A cartoon shows SopB central region (amino acids 271-319) dimer structure (code 3KZ5) with the dimer domain identically arranged and the two monomers are colored pale cyan and violet. The two figures are generated with PyMol (DeLano, 2002).

1.4.1.3 Type Ib segregation system of TP228 plasmid

The conjugative, 72 kb, low copy-number TP228 plasmid was first identified in *Salmonella newport*. This plasmid encodes resistance to multiple antibiotics, including kanamycin, neomycin, spectinomycin, streptomycin, sulphonamides and tetracycline. Besides these antibiotics, it was also shown to be resistance to mercuric ion (Hayes, 2000). TP228 plasmid replicates at low copy number and no obvious loss was detected after growing *E.coli* harbouring the plasmid for 25 generations in the absence of selective pressure (Hayes, 2000). The stable maintenance of TP228 plasmid is achieved by the presence of a type Ib segregation system. Like any other type I partition system, TP228 partition system is comprised of two proteins, a Walker-type ATPase protein, ParF (ParA homologue), a centromere binding protein, ParG (ParB analougue), and a centromere site named *parH* (Figure 1.10) (Hayes, 2000).

The region upstream of the *parFG* genes contains the centromere *parH* site, which comprises eleven direct repeats and one inverted repeat of 4 bp (5′-ACTC-3′). These tetrameric repeats are separated by a 4 bp AT-rich spacers (Wu *et al.*, 2011). Apart from these tetrameric repeats in the *parH* site, eight additional tetramer box motifs are located downstream of *parH* and are arranged in a directed and inverted orientation. This is an operator site, O_F. Three of these boxes are in direct orientation, the other five are inverted. The O_F site overlaps with the -10 region on promoter site (Figure 1.10) (Zampini *et al.*, 2009).

The repeats in *parH* and operator sites are recognized by ParG. ParG is a DNA-binding protein, which binds site-specifically to *parH* to mediate plasmid partitioning and to O_F to effect transcriptional repression (Barillà and Hayes, 2003; Carmelo *et al.*, 2005). ParG loads onto a single repeat of *parH* and O_F as a dimer, therefore, twelve ParG dimers are needed for partitioning and eight for transcriptional repression. Progressive deletion of repeats in *parH* leads to a gradual reduction in plasmid retention level (Wu *et al.*, 2011).

An *in vivo* experiment demonstrated that the O_F site is as effective as the centromere site in plasmid stability and it could be part of the partition site as deletion of one site, either *parH* or O_F, does not affect plasmid stability (Wu *et al.*, 2011). At the minimum two repeats are required for complex formation as one separate box was not sufficient to allow ParG to bind and form a complex (Zampini *et al.*, 2009).

The ParG protein is a homodimer (Barillà and Hayes, 2003; Golovanov *et al.*, 2003). The structure of ParG was solved using nuclear magnetic resonance (NMR) spectroscopy

and this revealed that ParG is a member of Arc/MetJ family of DNA-binding transcriptional repressors (Figure 1.11*A*). Each ParG monomer consists of two domains, the flexible N-terminal domain (1-32 amino acid residues), and the folded RHH C-terminal region (33-76). The two subunits are intertwined by β-strands (Figure 1.11*A*) (Golovanov *et al.*, 2003). The folded C-terminal domain is bifunctional, the ParG protein binds to *parH* through the folded domain to make the partition complex. Moreover, biochemical studies showed that the ParG protein dimerises through the folded region and deletion of the flexible N-terminal part had no effect on dimerisation (Barillà and Hayes, 2003; Carmelo *et al.*, 2005). It is presumed that the interaction of ParG with the operator site upstream of *parFG* occurs via the folded domain and that the β-strand is inserted into the major groove of DNA in a way similar to the one of Arc/MetJ transcriptional repressors (Golovanov *et al.*, 2003). ParG interacts with its partner ParF both through the flexible and the structural domains (Barilla *et al.*, 2007). Moreover, the unstructured flexible tail of ParG is crucial for transcriptional repression of the *parFG* promoter, as deletion of the flexible part of ParG is accompanied by a reduction in transcriptional repression (Carmelo *et al.*, 2005). The N-terminal domain of ParG also harbours an arginine finger-like motif with an arginine residue in position 19, which stimulates the ATP hydrolysis of ParF and it is crucial for plasmid partitioning (Barillà *et al.*, 2005; 2007).

ParF is a Walker-type ATPase protein. This protein can polymerise in the absence of the nucleotides, and this activity increases upon ATP binding, leading to the formation of filaments *in vitro* (Barillà *et al.*, 2005). The ParF crystal structure was solved and showed that the monomer is composed of only one domain with central seven twisted β-strands surrounded by four helices on each side (Figure 1.11*B*). The structure of ParF bound to nucleotides has also been solved and has shown that ParF is a monomer in ADP-bound form and a dimer in the ATP-bound state (Figure 1.11*B*). The structure of ParF bound to AMPPCP (Phosphomethylphosphonic acid adenylate ester, a non-hydrosable ATP analogue) was solved and showed that ParF assembles into a dimer of dimer units that forms a specific filament for polymer assembly. The nucleotide binding pocket is located in the N-terminus of the α-helix 1 of the Walker A motif (Schumacher *et al.*, 2012).

Electron microscopy experiments showed that *in vitro* ParF forms long filament bundles, characterised by a loose and a compact end. The formation of higher order structures upon ATP binding and dissolution of these structures following ATP hydrolysis are believed to be crucial processes for positioning plasmids to the opposite cell halves prior to cell division (Barillà *et al.*, 2005).

Recently, *in vivo* investigations have shown that ParF binds to and oscillates over the nucleoid in the presence of the ParG-*parH* complex. Based on these results, a Venus flytrap model was proposed for the segregation of plasmid TP228, according to which ParF forms a three-dimensional meshwork within the nucleoid volume. The ParG-*parH* complex is trapped inside the ParF matrix and oscillates in the wake of ParF before being deposited at opposite nucleoid poles (McLeod *et al.*, 2017). This mechanism will be discussed in detail under the Section 1.7.3 “Mechanisms of plasmid segregation”.

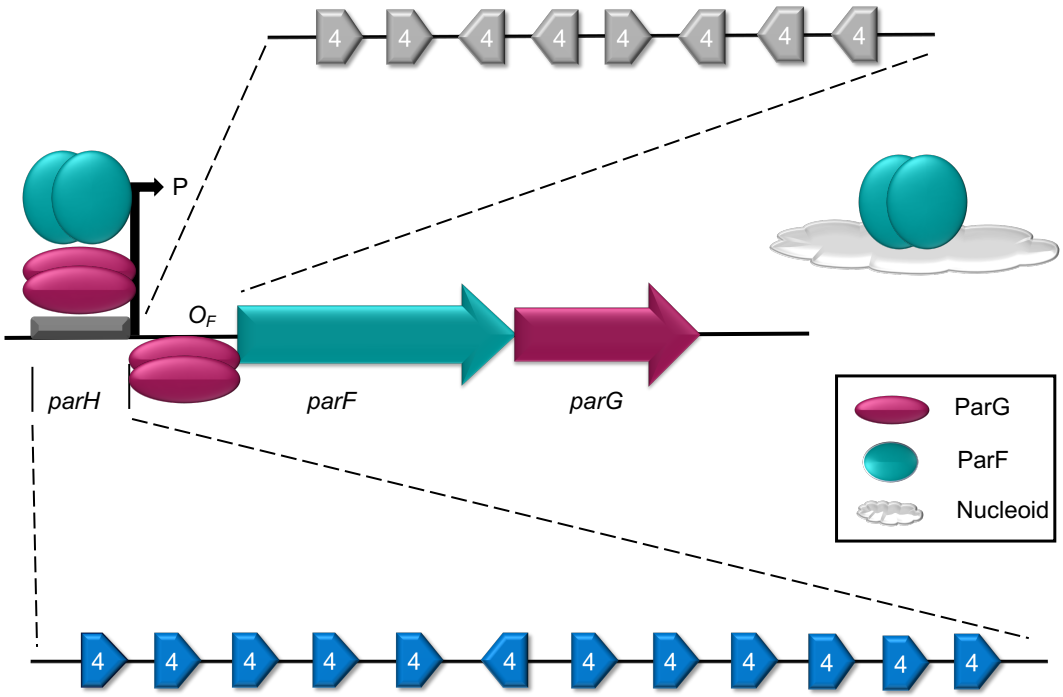


Figure 1.10 Genetic organisation of the *parFGH* cassette of TP228 plasmid partition system

The diagram represents the *parFGH* cassette, the teal arrow represents the gene encoding the Walker-type ATPase protein, ParF, and the red arrow indicate the gene encoding the DNA-binding protein ParG, and the *parH* site is indicated with a grey bar, located upstream of *parFG*. ParG recognises twelve 4 bp repeats (blue arrows) in the *parH* site and recognises eight 4 bp repeats in the operator site, indicated with the grey arrows (O_F). ParF binds to ParG and to the nucleoid.

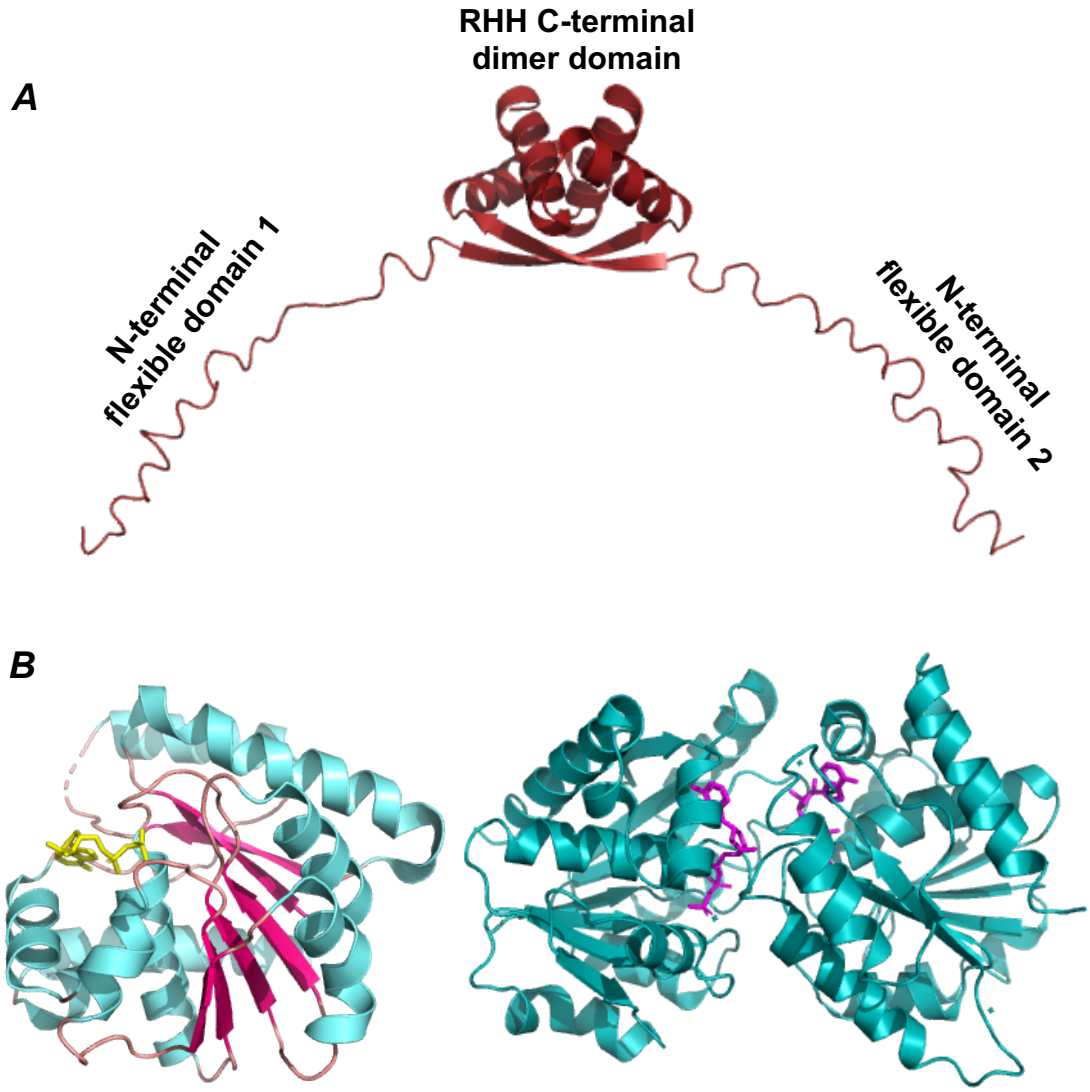


Figure 1.11 Cartoon structures of TP228 partition proteins ParG and ParF

A) NMR structure of ParG dimer (code 1P94) shows the N-terminal flexible domain and a RHH C-terminal domain. ***B)*** A cartoon of ParF monomer bound to ADP (yellow stick) on left (code 4E03). The central seven β-strands (hot pink) are surrounded with four helices (cyan). A ParF dimer (code 4E07) bound to AMPPCP (magenta sticks). The two subunits are shown in teal colour. All figures are generated using PyMol (Delano, 2002).

1.4.1.4 Type Ib segregation system of pSM19035 plasmid

The pSM19035 plasmid is present in 1-3 copies per cell in *Bacillus subtilis* and it is also harboured by a wide range of species in the phylum Firmicutes (Pratto *et al.*, 2008). The segregation system of pSM19035 belongs to the type Ib sub-family and consists of two *trans*-acting proteins, a ParA-like ATPase (δ), a ParB-unrelated protein (ω) and two sets of three centromere *parS* sites. The three *parS* sites consist of 9, 7 and 10 adjacent 7 bp heptads respectively, with the sequence 5′-WATCACW-3′; where W is either an A or a T and are arranged either in direct or inverted orientation. The *parS* sites overlap with the promoter region of δ, ω, and *copS* genes, respectively (Figure 1.12) (Dmowski and Kern-Zdanowicz, 2016).

Both *trans*-acting proteins have been thoroughly investigated. Both structures were solved and the interplay of partition elements *in vivo* and in *vitro* has been studied extensively. Omega is a DNA-binding protein and occurs as a homodimer (Misselwitz *et al.*, 2001). The monomer consists of 71 amino acid residues and includes two domains: the unstructured N-terminal domain and the ribbon-helix-helix C-terminal domain (Figure 1.13*A*) (Murayama *et al.*, 2001). The flexible N-terminal domain is essential for plasmid partitioning, but unnecessary for copy number control and TA (toxin-anti-toxin) expression. ω dimerises through the N-terminal domain and interacts with δ_2. The Omega RHH domain binds the 7 bp direct or inverted repeats with high affinity and cooperativity (Soberón *et al.*, 2011). ω was shown to be involved in many functions, including acting as a repressor of genes involved in copy number control, in plasmid maintenance and in the post segregational killing system (Pratto *et al.*, 2008).

Deletion of 8 or 18 amino acid residues in the N-terminal region has shown no or minor effects in the conformational of the core structure, plasmid stability and binding of Omega to the operator region, while substituting Thr-29 to alanine in the β-strand weakened the monomer-monomer interaction and the DNA binding affinity (Welfle *et al.*, 2005).

Both the structure of the ω-DNA complex and single molecule studies have shown that ω_2 binds continuously to DNA to form a left-handed helix that wraps around *parS* DNA without bending the *parS* site (Figure 1.13*A*) (Weihofen *et al.*, 2006; Pratto *et al.*, 2009). The Omega protein binds site-specifically as a dimer with high affinity (20 nM) to two or more of 7 bp DNA heptad of direct or inverted repeats, whereas the ω monomer binds non-specifically and with low affinity to single 7 bp repeat or subsite repeats (>1 μM) (de

la Hoz *et al.*, 2004). A very specific characteristic of ω, which contrasts with other RHH repressors, is that ω_2 can bind to non-palindromic direct repeats (Weihofen *et al.*, 2006). The motor protein δ is a Walker-type ATPase working as a homodimer and it shares sequence identity between 14 and 25% with both bacterial and archaeal Walker-box ATPases, namely *Tth* Soj of *Termus thermophilus* and *Pfu* MinD of *Pyrococcus furiosus* (Pratto *et al.*, 2008). The monomer consists of 284 amino acid residues. Each monomer is composed of eight β-strands surrounded by twelve α-helices (Figure 1.13*B*). The protein exhibits structural similarity to both Soj and MinD, although the sequence identity within these proteins is limited. The unique wide and open U-shape of δ.ATP.Mg^{2+} dimer complex allows the protein to free exchange of ATP and ADP without dissociation of the two monomers (Pratto *et al.*, 2008, 2009). Atomic Force Microscopy (AFM) studies have shown that δ binds DNA in a non-specific manner and dissociates from DNA upon binding to the partition complex (Pratto *et al.*, 2009).

In vitro analysis has shown that the interplay of δ/ω is dependent on the concentration of the Omega protein. At low 0.09-1.4: 1 molar ratio of ω:δ, ATP hydrolysis stimulation occurs, however the ATPase stimulation is declined when the ratio is between 2.8 and 4.2. Moreover, two Walker A mutants, Delta K36A and D60A have been shown to be involved in ATPase activity as its stimulation was abolished (Pratto *et al.*, 2008). Moreover, Volante and Alonso have shown that ATP, non-specific DNA, and partition complex formation promotes a structural transition in both proteins (Volante and Alonso, 2015).

Proteins interaction is a crucial stage for plasmid segregation. Different stages of interaction depend on the molar ratio of the two proteins. When the molar ratio is low (nanomolar range), ω binds to the centromeric site (*parS*) and enhances the ATPase activity of δ, which promotes pairing of two supercoiled plasmids. When the molar concentration is close to 1, ω stimulates ATP hydrolysis and promotes the dissociation of the paired plasmids. At a high molar ratio (millimolar range), ω enhances δ polymerisation onto DNA, which positions both plasmids at opposite poles prior to cell division (Pratto *et al.*, 2008, 2009).

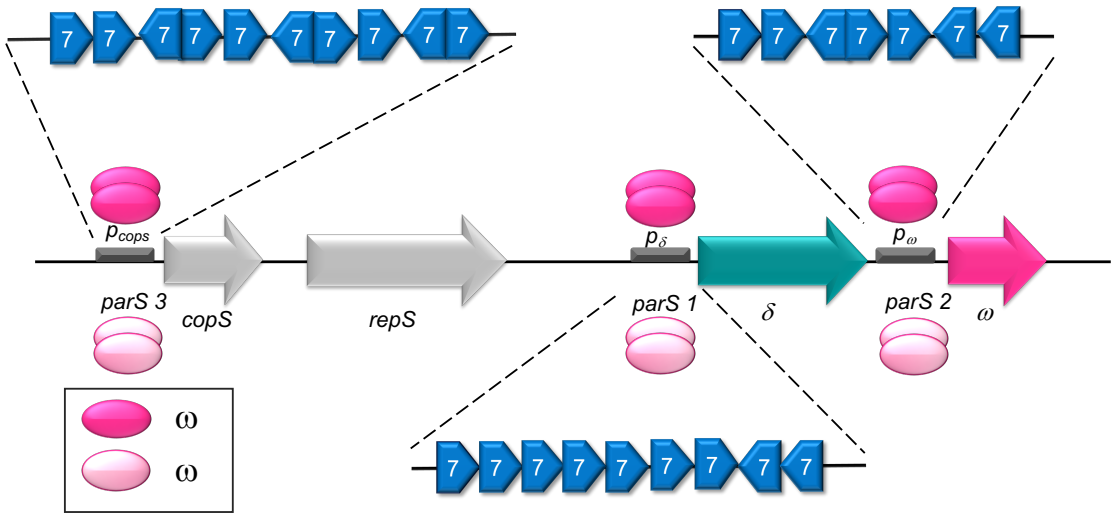


Figure 1.12 Genetic organisation of the pSM19035 plasmid partition system

The diagram represents the *par* cassette, the teal arrow represents the gene encoding the Walker-type ATPase δ, and the pink arrow indicates the gene encoding the DNA-binding protein, ω, and the regions upstream of δ, ω and *copS* are the three centromeric sites (grey bars) represent *parS1*, *parS2* and *parS3* respectively. The three *parS* sites are enlarged and the number and direction of contiguous 7 bp heptad repeats are indicated (blue arrows). ω_2 (oval pink) recognizes repeats in the *parS* sites. The promoters repressed by ω_2 (oval light pink) are indicated.

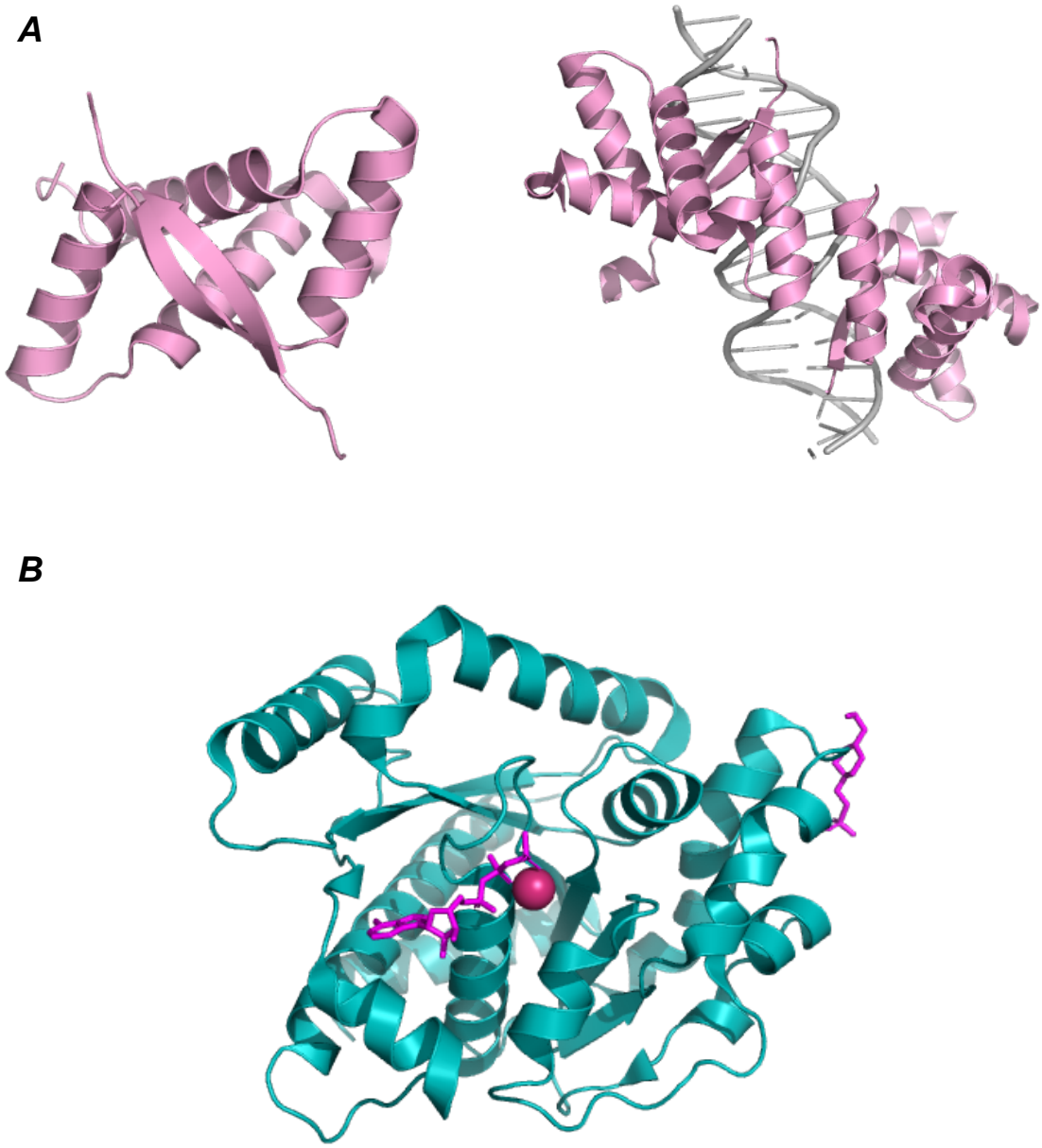


Figure 1.13 Crystal structures of pSM19035 partition proteins

A) A cartoon shows the Omega dimer C-terminal domain on left with pink color (code 1IRQ) and on the right a cartoon shows two Omega dimers bound to DNA (code 2BNW), which forms a left-handed helix in which the β-sheet inserted the major grove DNA. ***B)*** Delta bound to ATPγS (code 2OZE), which shows as magenta sticks and magnesium is shown as a red ball. The two proteins are colored same as in figure 1.5. The two figures are generated with PyMol (DeLano, 2002).

1.4.1.5 Type Ib segregation system of the *par2* system of pB171

The large 69 kb low-copy-number adherence factor-encoding plasmid pB171 is harboured by an enteropathogenic *Escherichia coli* strain (EPEC) (Tobe *et al.*, 1999). EPEC that carries adherence factor (EAF) plasmids are the main cause of diarrhoea in the developing countries. The EAF pB171 plasmid harbour a locus that encodes the bundle-forming pili (BFP). BFP occupies a 12 kb region and comprised of 14 genes that are required for BFP biogenesis and function (Bieber *et al.*, 1998; Tobe *et al.*, 1999). The stable maintenance of this large virulence plasmid is achieved by the presence of an unusual double partition system (Ebersbach and Gerdes, 2001). The pB171 plasmid is unique as it contains two adjacent opposite oriented partitioning loci from two distinct classes, type II and type I named *par1* and *par2* respectively. Both *par* systems share a common *cis*-acting site (Figure 1.14). Each of the *par* systems can mediate efficient plasmid segregation independently of the other. However, optimal plasmid partition efficiency requires the presence of both. *In vivo* experiments demonstrated that *par1* and *par2* gave 15- and 140- fold of stabilization respectively, but together they yielded 180-fold. This means that both systems support each other to ensure accurate plasmid partitioning (Ebersbach and Gerdes, 2001). The *par1* system encodes two proteins, ParM, an actin-like protein, and ParR, a DNA-binding protein (Figure 1.14). The *par1* system is homologous to that encoded on the R1 plasmid (Ebersbach and Gerdes, 2001).

The *par2* segregation system is the focus of this thesis work. Like any other partition system, *par2* of pB171 encodes two proteins, a Walker-type ATPase protein, ParA, a centromere binding protein, ParB, and includes two centromere sites, *parC1* and *parC2* (Figure 1.14) (Ebersbach and Gerdes, 2001). Like *parABS* of P1 plasmid, pB171 *par2* contains two IHF binding sites and this protein is the only host factor that might contribute to the partitioning system (Ringgaard, *et al.*, 2007).

1.4.1.5.1 Centromeric sites *parC1* and *parC2*

The *parC1* site is located upstream of the *parAB* genes, whereas the *parC2* site is located downstream (Ebersbach and Gerdes, 2001). The *parC1* and *parC2* sites are composed respectively of 17 and 18 of 6 bp repeats. All repeats are arranged in direct orientation only and denoted as B repeats. The repeats in the *parC1* site are arranged into two clusters with the first cluster consisting of 13 repeats, while the second cluster is composed of the remaining repeats. The two clusters are separated by 31 bp. The four repeats in the second cluster are located in the promoter region (P2) of the *par2*

system. The promoter of the *par1* system, P1, is located 6 bp upstream of the first cluster (Figure 1.14) (Ebersbach and Gerdes, 2001).

Similarly, the region downstream of *parB* is the *parC2* site and consists of 18 6 bp direct repeats organised into three clusters. The first repeats of the *parC2* site overlap with the stop codon of *parB*. All repeats, in *parC1* and *parC2*, are separated by one nucleotide. Moreover, the repeats in *parC1* and *parC2* are related and based on the sequences, they can be divided into two subclasses, named as class I and class II. The sequence of class I repeats is [AT][TA]CA[TC]A and that of class II TTAT[GT]A. The class I repeat is more frequent in *parC1*, whereas class II is more frequent in *parC2*, and both classes are randomly distributed (Figure 1.14) (Ebersbach and Gerdes, 2001). An *in vitro* experiment showed that ParB binds as a dimer to B repeats at the *parC1* and *parC2* sites (Fothergill *et al.,* 2005; Ringgaard *et al.*, 2007a).

Besides the B hexamer repeats in *parC1*, two additional identical 10 bp direct motifs, AATACTCAAT, are located 9 bp upstream of the B1 repeat, between the P1 promoter and the *parMR* cassette (Figure 1.14). The two repeats are separated by 31 bp and named R1 and R2 respectively. These two repeats are recognized by ParR, not ParB (Ringgaard, *et al.*, 2007a).

Similarly to the *parABS* system of the P1 plasmid of *E.coli*, the *parABC* system of pB171 plasmid contains two IHF sites, *ihf1* and *ihf2* (Ringgaard *et al.,* 2007a). IHF is the only host factor that has been identified and proven to be involved in plasmid partitioning so far (Funnell, 1988b). The two *ihf* sites are in opposite orientation with the *ihf1* site located within the *parA* gene in opposite direction, 504 bp downstream of the *parA* start codon. The *ihf2* site is located in the opposite direction of *ihf1*, overlapping with 2 bp of the second stop codon of *parB* (Ringgaard *et al.,* 2007a). An *in vitro* experiment demonstrated that the IHF protein binds to both sites. IHF has been shown to contribute to plasmid segregation, as deletion of the *IHF* gene from the host cell resulted in decreased plasmid stability compared to the wild type strain (Ringgaard *et al.,* 2007a).

1.4.1.5.2 Centromere-binding protein ParB

ParB is a site-specific DNA binding protein consisting of 91 residues (Ebersbach and Gerdes, 2001; Fothergill *et al.*, 2005). Despite its name, pB171 ParB does not display the HTH motif that is present in type Ia DNA-binding partition proteins. Secondary structure predictions of ParB suggest that the protein might harbour a RHH folded domain similar to that of TP228 ParG and pMV158 CopG, although the sequence

similarity among them is very limited (Ringgaard *et al.*, 2007b). Similarly to TP228 ParG, ParB is able to self-associate as a homodimer (Fothergill *et al.*, 2005). The structures of both ParG of TP228 and ω of pSM19035 from the type Ib segregation system have been solved and showed that both consist of two domains, the highly disordered N-terminal domain and the RHH C-terminal domain (Murayama *et al.*, 2001; Golovanov *et al.*, 2003). Biochemical assays have shown that ParB binds with high affinity and cooperatively to the B repeats in *parC1* and *parC2* sites. The binding of ParB to partition sites occurs through the C-terminal domain as deletion of the N-terminal part has no effect on binding. However, the pattern of complexes formed by truncated versions is not similar to the one observed with the full-length protein. ParB full length forms a high molecular weight nucleoprotein complex, but it exhibits an unusual pattern with ParBΔ19 and ParBΔ39, thus the N-terminal end is required for this phenomenon (Ringgaard *et al.*, 2007a; Ringgaard *et al.*, 2007b) ParB dimerises through the C-terminal domain. In fact, ParB, ParBΔ19 and ParBΔ39 proteins are all able to self-associate *in vitro* (Ringgaard *et al.,* 2007b). Furthermore, ParB binds to one single repeat as a dimer and *in vitro* experiments have shown that seventeen ParB dimers are required to bind the full *parC1* site, which consists of seventeen repeats (Ringgaard, *et al.*, 2007a).

One of the primary functions of the N-terminal domain is *parC* centromere pairing. The mechanisms underpinning plasmid segregation in the *par2* system is however not quite completely understood, although, plasmid pairing is considered an essential step in the segregation process (Ringgaard *et al.*, 2007b).

Apart from the crucial role of ParB in plasmid partitioning, ParB plays an additional role in autoregulation of *parAB*. Like many other ParBs of type Ib, such as ParG of TP228 and ω of pSM19035, ParB binds the region upstream of *parAB* operon (P2) in a site-specific manner to repress transcription. ParB works as a transcriptional repressor for both the *par2* and the *par1* system (Figure 1.15) (Ringgaard *et al.*, 2007a).

Importantly, it was shown that the N-terminus of TP228 ParG contains an arginine finger-like motif harbouring R19, which is essential for stimulation of ParF ATPase activity and plasmid segregation (Barilla *et al.*, 2007). Therefore, it was presumed that pB171 ParB N-terminal end of ParB may perform the same role in stimulating ParA ATP hydrolysis. The arginine in position 26 of ParB was changed to either alanine or lysine and the plasmid bearing this mutation was a three-fold decrease in stability (Ringgaard *et al.*, 2009). *In vivo* studies showed that ParA oscillates in a spiral-like structure over the nucleoid regions (Ebersbach and Gerdes, 2004). Moreover, time-lapse microscopy

experiments showed that the residue change in the ParB N-terminus affects ParA oscillation. These two effects indicate that the R26 in ParB N-terminal end may perform a similar role as that of R19 of TP228 ParG (Ringgaard *et al.*, 2009). However, it has never been demonstrated that ParB-R26A or ParB-R26K do not stimulate ParA ATPase activity.

1.4.1.5.3 Walker-type ATPase ParA

ParA is a weak Walker-type ATPase. As other types Ib ParAs, ParA of pB171 is relatively small consisting of 214 amino acids and it lacks the N-terminal DNA binding motif that is present in type Ia ParA proteins (Ebersbach and Gerdes, 2001). Microscopy studies showed that ParA colocalizes within the nucleoid and oscillates from one nucleoid pole to other multiple times per each cell cycle. The oscillation of ParA is ATP-dependent and was abolished when single amino acid changes were introduced in the Walker A box ATP-binding domain. Glycine at position 10 and lysine at position 14, which is conserved among most of Walker A boxes, were changed to valine and glutamine respectively. Moreover, ParA-G10V and ParA-K14Q mutations were revealed that the ParA protein plays two crucial roles, positioning the plasmid at mid-cell and separating the newly replicated plasmid to opposite poles prior to cell division (Ebersbach and Gerdes, 2001; 2004).

In vivo studies showed that ParA oscillates in a spiral-like structure over the nucleoid regions. This spiral-shaped structure is assembled only in the presence of the ParB-*parC* complex. Moreover, the oscillation of ParA is ParB-*parC* dependent as the absence of either *parC* or ParB eliminates the dynamics of the ParA movement. Instead, ParA assembled into stationary helical structures that extended from one nucleoid pole to the other, thus, plasmid stabilization was disrupted (Ebersbach and Gerdes, 2001; Ebersbach and Gerdes, 2004). Interestingly, both ParA-G10V and ParA-K14Q mutations in the ParA ATP binding site of ParA prevents the formation this helical structure both in the presence or absence of the ParB-*parC* complex. This indicates that ATP has a crucial role in the formation of the helical structure (Ebersbach and Gerdes, 2004). *In vitro* experiments showed that in the presence of ATP, ParA polymerises and forms bundles of filaments. Moreover, it was shown that ParA can also binds to ADP (Ebersbach *et al.*, 2006). The formation of filaments by ParA represents a feature common in type I motor proteins, like SopA of F plasmid, ParF of TP228, as well as the related MinD ATPase of *E. coli*, which also forms helical structures *in vivo* (Barillà *et al.*, 2005; Ebersbach *et al.*, 2006).

In vivo experiments showed that *par2* locus of pB171 distributes plasmid evenly along the cell axis. The same plasmid distribution were observed even in cells with multiple plasmid foci (Ebersbach *et al.,* 2005; 2006). However, cytological studies revealed that most of the *parABC* plasmids localise close to mid-cell. It was assumed that many host factors could contribute and provide the mechanical force to direct plasmid partitioning. However, investigation have shown that none of the *E. coli* division proteins interacted with ParA or ParB (Ebersbach *et al.*, 2006).

Based on all the observations from *in vivo* studies as well as the biochemical data that was obtained previously, a model was proposed, in which ParA provides the mechanical force through the polymerisation/depolymerisation process to drive the newly replicated plasmids apart (Ringgaard *et al.*, 2009). The mechanism of segregation of *par2* of pB171 plasmid will be discussed in detail under Section 1.7.1.

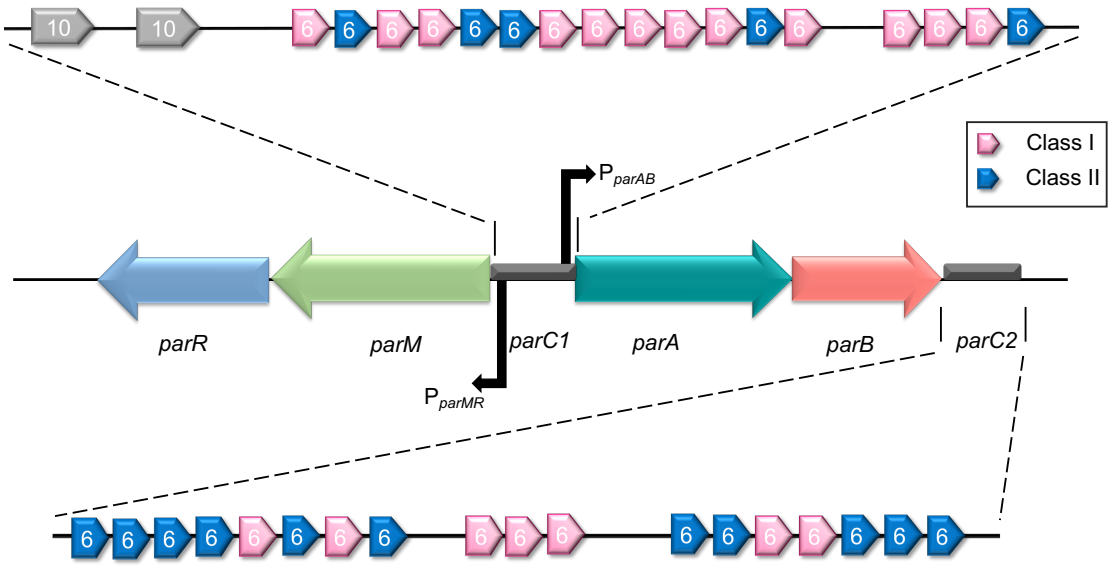


Figure 1.14 Genetic organisation of the double *par* loci of pB171 plasmid partition systems

The diagram represents the divergently oriented of double *par* loci, *par1* and *par2*. In the *par1* system, the light green arrow represents the gene encoding the actin-like ATPase ParM, and the blue arrow represents the gene encoding the DNA-binding protein ParR. In the *par2* system, the teal arrow represents the gene encoding the Walker-type ATPase, ParA, and the orange arrow indicates the gene encoding the DNA-binding protein ParB. The *parC1 and parC2* sites are indicated with a grey bar, with *parC1* located upstream and *parC2* located downstream of *parAB*. ParB recognises 6 bp repeats (pink and blue arrows for class I and class II repeats, respectively) in *parC1* and *parC2* sites, whereas, ParR recognises two 10 bp repeats (grey arrows) in the *parC1* site.

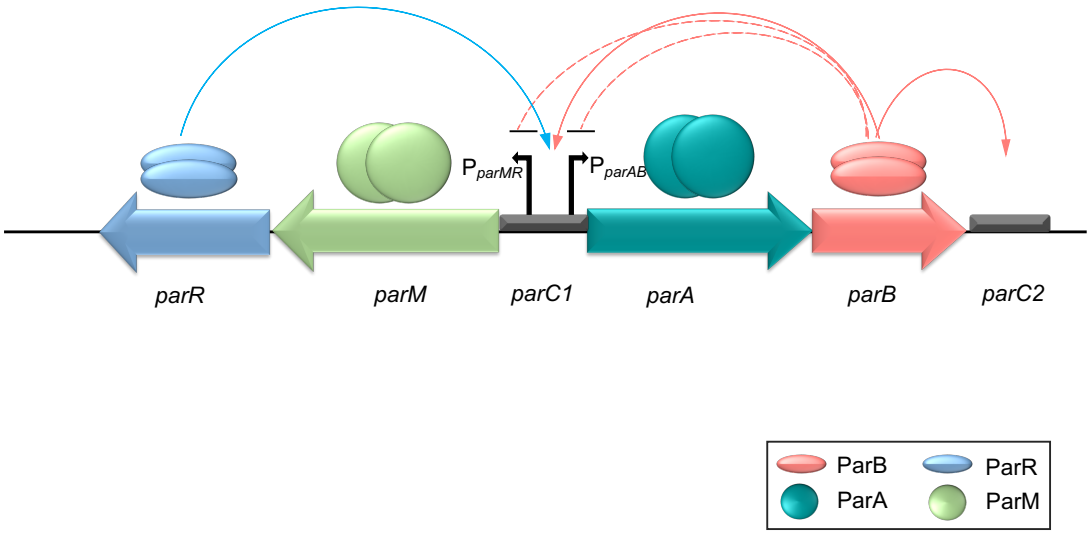


Figure 1.15 Genetic organisation and regulation of pB171 *par1* and *par2* systems

The diagram represents the autoregulation of partition operons of *par1* and *par2*. Dashed arcs represent regulation of the partition operons by the indicated proteins. In *par1* (type II) and *par2* (type Ib), the centromere-binding proteins (ParR and ParB, respectively) act as transcriptional repressors. Solid arcs represent the formation of the partition complexes by the centromere-binding protein.

1.4.2 Type II segregation system

Type II plasmid segregation systems are characterized by an actin-like ATPase that is a member of the actin/hsp70 superfamily of ATPases. The organisation and principles of plasmid segregation are conserved and involve an adaptor protein ParR binding in a site-specific manner to the centromere site and forming a plasmid partition complex. The two paired plasmids are transported and positioned in the cell by force given through the motor protein ParM. The most well-studied type II system is that of the antibiotic resistance plasmid R1 from *E. coli* (Salje, 2010; Sengupta and Austin, 2011).

1.4.2.1 Type II segregation system of R1 plasmid

The *par* locus of the drug-resistance R1 plasmid is located within a 1600 bp region. It consists of two genes, one encoding an actin-like protein, ParM, the second encoding the DNA-binding protein, ParR, and a centromeric site called *parC* (Gerdes and Molin, 1986; Dam and Gerdes, 1994) *.* The centromeric site is located upstream of *parMR*. The *parC* site consists of ten 11 bp direct iterons, which are organised into two clusters of five. The two clusters are separated by a 39 bp core region, which contains the *parMR* promoter (Figure 1.16*A*) (Dam and Gerdes, 1994). The *parMR* promoter is bound by ParR in its transcriptional repressor's role, while ParM is not involved in autoregulation (Figure 1.4). The highest level of repression and plasmid stabilisation is achieved when all 10 iterons are present (Jensen *et al.*, 1994; Breüner *et al.*, 1996).

Besides the role of ParR in transcriptional repression (Jensen *et al.*, 1994), ParR binds site-specifically to the *parC* site to mediate plasmid partitioning (Jensen *et al.*, 1998). Like ParBs of type Ib systems, ParR also belongs to the Arc/MetJ superfamily. ParR binds in a sequence-specific manner to *parC* repeats through the N-terminal RHH domain. ParR needs at least two 11 bp repeats to initiate binding. However, all the other 10 repeats are required for plasmid stabilisation (Møller-Jensen *et al.*, 2003). ParR is dimer in solution, and every repeat is bound by a dimer (Hoischen *et al.*, 2008). The N-terminus of ParR is required to form a stable ParR-*parC* scaffold. The binding of ParR to *parC* causes significant bending of DNA to form a U-shape structure with the DNA wrapped around (Figure 1.16*B*) (Hoischen *et al.*, 2008; Salje and Löwe, 2008). ParM is a GTPase protein that assembles into filaments, which are comprised of two parallel filaments along the cell axis (van den Ent *et al.*, 2002). The C-terminal domain of ParR is required for specific interactions with the ParM filament (Figure 1.16*B*). The affinity of interaction is low, but it increases with a high concentration of ParR (Salje and Löwe, 2008). Moreover, ParR-*parC* complexes bind the antiparallel ends of ParM filaments very

tightly where each end of a single polar ParM filament binds strongly to a ParR-*parC* complex (Garner *et al.*, 2007; Choi *et al.*, 2008).

ParM is an actin-like ATPase. A structural study has shown that ParM belongs to the actin/Hsp70 superfamily, although the sequence similarity is limited (van den Ent *et al.*, 2002). ParM forms filamentous structures along the cell axis, and these filaments are required for the directional movement of the plasmids to the opposite cell poles. The formation of filaments depends on ATP binding by ParM, and depolymerisation results from nucleotide hydrolysis. The ParM ATPase activity is enhanced by ParR and maximal ATPase activity was achieved when ParR-*parC* is present in the reaction (Jensen and Gerdes, 1997). The ParM protein forms filaments in the presence of both ParR and *parC*, which is considered as the nucleation site for ParM polymerisation (van den Ent *et al.*, 2002; Møller-Jensen *et al.*, 2003). Cryo-electron microscopy showed that ParM forms bipolar left-handed double helical filaments, which is different from other actin filaments that possess a right-handed twist (Gayathri *et al.*, 2012, 2013). The mechanisms of segregation of R1 plasmid depend on ParM polymerisation that provides the mechanical force to drive the plasmids apart (Møller-Jensen *et al.*, 2003).

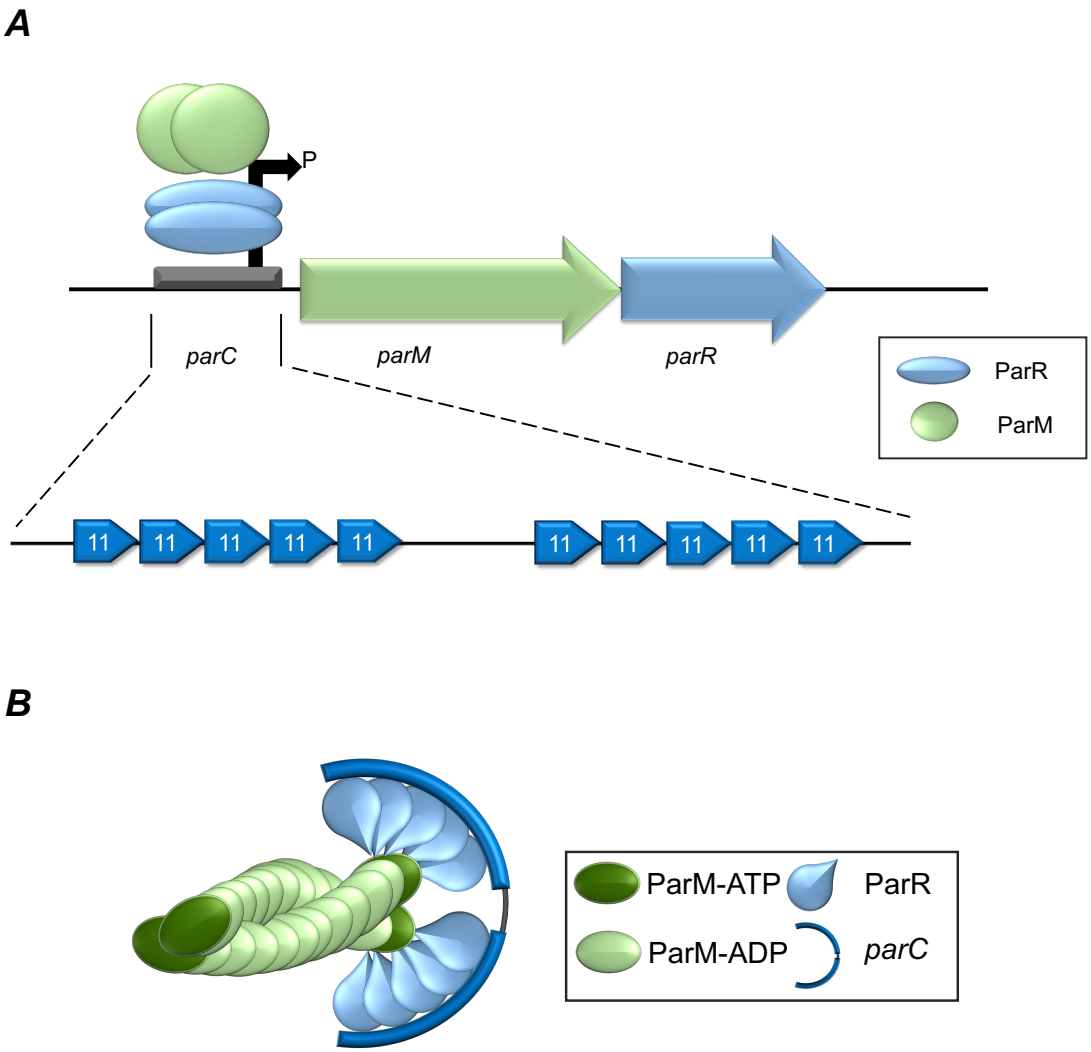


Figure 1.16 Genetic organisation of the *parMRC* cassette and model of segrosome assembly of R1 plasmid partition system

A) The diagram represents the *parMRC* cassette, the light green arrow represents gene encoding the actin-like ATPase ParM, the sky-blue arrow represents the gene encoding the CBP, ParR, and the *parC* site is indicated with a grey bar, located upstream of *parMR*. ParR recognises ten of 11 bp repeats (enlarged) indicated with the blue arrows. ParR (light green balls) binds the operator site for transcription regulation and *parC* for plasmid partitioning. ***B)*** the diagram represents the segrosome assembly model of *parMRC*. ParR (blue tear-drop shapes) binds *parC* and bends the DNA with the narrow end pointing towards the centre of the ring. ParM-ATP (dark green oval shape) binds to the flexible C-terminal domain of ParR (pointed ends) and polymerises (light green).

1.4.3 Type III segregation system

More recently, the type III partition systems have been identified on some of the large virulent plasmids of *Bacillus* species and in clostridial bacteriophages. Similar to type I and type II partition systems, the type III system consists of two proteins a Tubulin/FtsZ-like GTPase protein, namely TubZ, the DNA-binding protein TubR, and a centromeric site, which is located upstream of the *tubZR* genes (Larsen *et al.*, 2007; Oliva *et al.*, 2012). The best-characterized plasmid is the pBtoxis plasmid from *Bacillus thuringiensis*.

1.4.3.1 pBtoxis plasmid

The partition system of the large and low copy number virulence plasmid pBtoxis is composed of two proteins: TubZ is a GTPase distantly related to tubulin and FtsZ and TubR is the DNA-binding protein. The third element is a centromere site named *tubC*, which is located upstream of *tubZR* (Tang *et al.*, 2006; Larsen *et al.*, 2007). The centromere site consists of seven 12 bp direct repeats and is bound by the adaptor protein TubR (Figure 1.17). Together, the TubR-*tubC* complexes bind to TubZ, which later forms a bundle of filaments that drive the plasmids to opposite cell halves (Tang *et al.*, 2006; Fink and Löwe, 2015).

TubR is a dimer in solution. Each repeat of *tubC* is bound by a TubR dimer. The TubR structure consists of the intertwined dimer with a HTH motif. The crystal structure of TubR in complex with DNA also revealed that TubR interacts with DNA in a manner different from other HTH proteins. The N-terminus recognition helices of TubR insert into a single DNA groove, and the wings insert into the adjacent groove (Ni *et al.*, 2010). Besides the role of TubR in plasmid partitioning, TubR works as a transcription regulator by binding to the repeats upstream of *tubZR* (Tang *et al.*, 2006, 2007).

TubZ is a tubulin-like protein that was shown to be homologous to tubulin and FtsZ. One *in vivo* experiment showed that TubZ assembles into dynamic polymers and these polymers translocate by a treadmilling mechanism (Larsen *et al.*, 2007). TubZ forms polymers in the guanine triphosphate (GTP) bound state (Ni *et al.*, 2010). TubZ forms right-handed double filaments *in vivo* and the subunits are weakened upon GTP hydrolysis (Aylett *et al.*, 2010). Unlike other partition systems, the interaction of TubZ with TubR-*tubC* complexes does not require nucleotide binding and is very strong. Moreover, TubZ binds TubR through the C-terminal domain, which was shown to be a highly flexible domain (Ni *et al.*, 2010). Besides the role of the unstructured C-terminal tail of TubZ in binding to the partner TubR, it was shown that the TubZ C-terminal tail of *Clostridium botulinum* has multiple functions, including filament arrangement, polymer

remodelling into tubulin-like rings and the filament disassembly process (Fuentes-Pérez *et al.*, 2017). Interestingly, in a *Clostridium botulinum* phage, a third conserved protein was identified and denoted as TubY, whose gene is located downstream of *tubZ*. TubY has a HTH motif with a coiled-coil C-terminal end, and is assumed to have a regulatory function for the partition system. Importantly, it was shown that TubY binds to TubZ filaments alone through the C-terminal region, and it also binds to the TubZRS complexes (Oliva *et al.*, 2012). TubY in pBtoxis is located upstream of *tubZR,* but in the opposite orientation. Although TubY is highly conserved among Bacilli and Clostridia species, it is still unclear whether it has the same role in the pBtoxis plasmid (Oliva *et al.*, 2012).

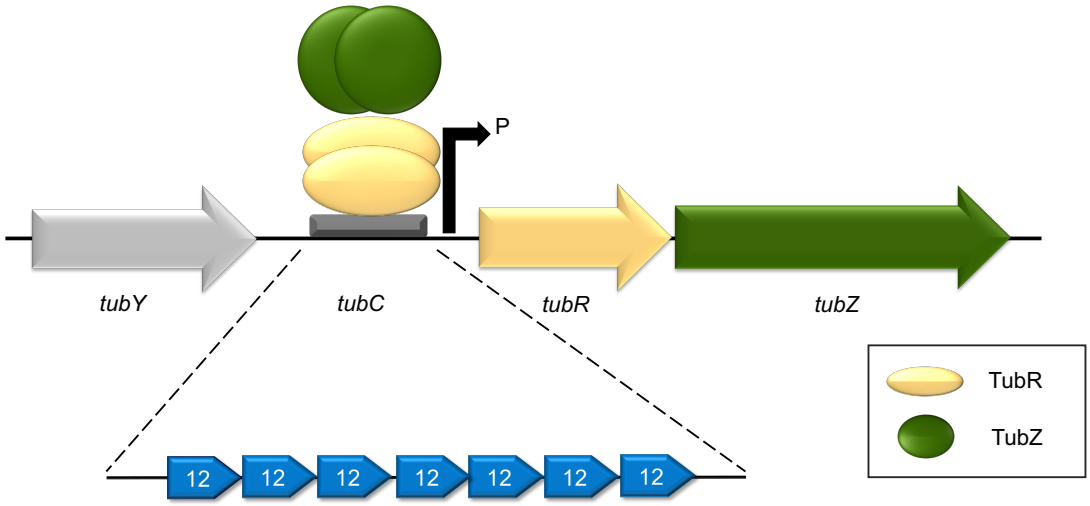


Figure 1.17 Genetic organisation of the *par* locus of the pBtoxis plasmid partition system

The diagram represents the *tubZRC* cassette, the dark green arrow represents the gene encoding the tubulin-like GTPase, TubZ, the yellow arrow represents the gene encoding the DNA-binding protein, TubR, and the *tubC* site is indicated with a grey bar, located upstream of *tubZR*. TubR recognises seven of 12 bp repeats (enlarged) indicated with the blue arrows. TubR (yellow balls) binds to the operator site for transcriptional regulation and binds to *tubC* for plasmid partitioning. The light grey arrow upstream of tubZR genes represents the gene encoding TubY, whose function is still unclear in pBtoxis

1.4.4 Type IV segregation system

The partitioning system of the 28.4 kb pSK1 plasmid from *Staphylococcus aureus* is distinct from all known partitioning systems and thus this plasmid is classified as a type IV partitioning system. The stable maintenance of the pSK1 plasmid is mediated by a single *par* gene, located upstream of the *rep* gene. The centromeric site is located upstream of the *par* gene, and it consists of seven 12 bp direct repeats and one of 7 bp inverted repeat. The coiled-coil domain in the Par protein could play a dual role, allowing the protein to act both as a DNA-binding protein and as a motor protein. However, unlike other partition systems, no ATPase domain was found in the Par protein (Simpson *et al.*, 2003; Dmowski and Jaguara-Burdzy, 2013). Nevertheless, many questions await answers, for example: how the newly replicated plasmids are driven apart, or whether other factors are involved in the partitioning system.

1.5 Types of centromere-binding proteins

The stable maintenance of low copy number plasmid is achieved by the action of active partitioning systems. Two proteins are involved in the segregation process, ParA, an NTPase protein, and ParB, a centromere-binding protein (Baxter and Funnell, 2014; Funnell, 2016; Oliva, 2016). The first step in the partition process involves recognition of the *par* site repeats by ParB. Unlike the NTPase protein, the sequence similarity between centromere-binding proteins CBPs is not high, even within a given family. Despite the fact that the specificity of DNA binding differs among CBPs, they share some common roles. Firstly, many studies showed that multiple CBPs bound to their centromeric site to form the partition complex. Consequently, this large nucleoprotein complex provides multiple binding sites for the NTPase protein. This protein-protein interaction between NTPase protein and CBP is likely to be crucial for the dynamic movement of the cargos resulting in strong anchoring of the plasmids and minimizing loss during their subsequent movement. Secondly, it was demonstrated that the CBP-plasmid partition complex stimulates the NTPase activity of the partner protein. Finally, the pairing of sister plasmids is facilitated by the CBP and is considered to be the first step of the segregation process. Based on the structure of CBPs, they can be classified into two main groups, HTH- and RHH-harbouring proteins (Baxter and Funnell, 2014).

1.5.1 Helix-turn-helix centromere-binding proteins

HTH DNA binding motifs are present in all type Ia CBPs. Despite the lack of sequence identity among the ParBs, they share the same domain arrangement (Funnell, 2016).

Structural information is available for only three type Ia system plasmid CBPs: ParB of P1 plasmid, KorB of RP4 plasmid and SopB of F plasmid. Structures of these proteins bound to their cognate DNA site have also been solved. Both structural and biochemical studies revealed that these proteins consist of three domains: the highly flexible N-terminal domain, which is called the ATP-binding domain, a central HTH domain, which is the only domain that is conserved among ParBs, and the C-terminal dimerisation domain. Moreover, the crystal structures of ParB, KorB and SopB, have shown that the HTH domain and the dimerisation domain are separated by a short flexible linker (Delbruck *et al.*, 2002; Schumacher and Funnell, 2005; Schumacher *et al.*, 2007b; Schumacher *et al.,* 2010; Oliva, 2016). The structure of P1 ParB revealed that this highly flexible linker permits free movement and rotation of the domains. Thus, it allows ParB to bind a variety of A and B boxes within the *parS* site (Schumacher and Funnell, 2005).

In SopB, the highly flexible N-terminal domain carries out several roles. The first 45 amino acid residues of this domain are a site-specific SopA interaction region (Ravin *et al.*, 2003). In addition, SopB stimulates SopA ATPase activity via its arginine finger-like motif (Kim and Shim, 1999; Ah-seng *et al.*, 2009). The central ParB domain, which has a canonical HTH motif, represents the only conserved domain among type Ia ParBs. Importantly, the HTH regions of ParB, SopB, and KorB bind to the inverted repeats in the centromeric site, which is a common feature of HTH CBPs. However, the HTH domain of each protein recognises the centromeric site repeats in a very different manner. The interaction of P1 ParB with the cognate site is different from that of KorB and SopB, which occurs mainly through the recognition helix of the HTH domain (Schumacher *et al.*, 2007b). In SopB, the interaction occurs through the recognition helix as well as via a residue, Arginine 219, outside the HTH motif, which was shown to interact with guanine (Schumacher *et al.*, 2010; Oliva, 2016). In addition, SopB uses the HTH recognition region to make non-specific interactions with the DNA (Schumacher *et al*., 2010). In contrast, KorB utilises residues in the region outside the HTH motif only to make a specific interaction with the inverted repeat sequence DNA (Khare *et al.*, 2004; Schumacher *et al.*, 2010; Oliva, 2016). It is believed that the ParB HTH central domain contains all the information needed for ParB to dimerise as well as to form high order complex structures on the *parS* site (Ravin *et al.*, 2003; Oliva, 2016).

The C-terminal dimerisation domain in P1 ParB consists of three antiparallel β-strands and a C-terminal alpha helix. The ParB structure shows that the dimerisation domain binds in a sequence-specific manner to the B box in which the loops between β1 and β2, as well as β2 and β3, insert into the major groove of B boxes (Schumacher and Funnell,

2005). Interestingly, the structure of the SopB C-terminal dimerisation domain is similar to the one of P1 ParB, however, it lacks the loops between the adjacent β-strands that could explain the specificity of this region to bind DNA (Schumacher *et al.*, 2010). In contrast, the dimerisation domain of KorB is similar to the Src homology 3 (SH3) domain of c-Crk. It consists of five antiparallel β-strands arranged with intervening loops (Delbruck *et al.*, 2002). The structure of P1 ParB shows that ParB binds to both arms of *parS* and brings them together with the help of the IHF protein, which recognises the sequence located in the middle of the *parS* site and thus bends the DNA (Schumacher and Funnell, 2005). Unlike P1 ParB, neither SopB nor KorB bends the DNA. Alternatively, the structure of SopB-DNA revealed that non-specific DNA binding is required to allow it to spread beyond the centromere site region to coat the DNA and prevent the dissociation of SopA. The ability of ParBs to spread few kilobases away from the centromeric site is a characteristic feature of all ParBs of type Ia partition systems (Schumacher *et al.*, 2010). The structure of full length ParB has not yet been solved, but the structure of some ParB domains with and without DNA provided insights into the domain organisation of ParB alone or when forming a complex with DNA.

1.5.2 Ribbon-helix-helix (RHH) centromere-binding protein

A RHH centromere-binding protein is commonly present in type Ib and type II systems. The RHH fold was originally characterised in the Arc/MetJ family of transcriptional repressors. All the RHH CBPs that have been studied so far are also transcriptional repressors, and act by binding to the operator site and repressing transcription (Baxter and Funnell, 2014). The RHH CBP of type Ib and type II systems differ mainly in domain organisation. The RHH DNA-binding domain is present in the C-terminal domain of type Ib CBPs, and the interaction with the cognate ATPase protein mainly occurs via the N-terminal tail. The RHH type II CBPs interacts with the NTPase protein via the C-terminal domain and, the interaction with DNA occurs through the N-terminus. Accordingly, these can be classified as type I CBPs of type Ib systems and type II CBPs of type II systems (Baxter and Funnell, 2014).

Structural information has been provided for three type Ib CBPs, including ParG of TP228, Omega of pSM19035, and ParB of pCXC100 (Figure 1.18*A*, *B* and *C*) (Murayama *et al.*, 2001; Golovanov *et al.*, 2003; Huang *et al.*, 2011). The three proteins show structural similarities with the highly flexible N-terminal domain and a folded RHH in the C-terminus, although the sequence similarity among them is very low (Murayama *et al.*, 2001; Golovanov *et al.*, 2003; Huang *et al.*, 2011). The RHH CBPs of type Ib are dimers and the folded domain consist of two monomer subunits, which are tightly

intertwined to form a symmetric homodimer. Each subunit consists of one β-strand followed by two alpha helices, α1 and α2. The β-strand from both subunits pair with each other to form an antiparallel β-sheet. The number of amino acids forming the β-strand varies between ParG, Omega and pCXC100 ParB, consisting of 9, 5 and 3 amino acids residues respectively. The two helices are connected by a short flexible loop. The C-terminal folded domain contains the minimal DNA binding domain as well as the dimerisation domain. Dimerisation is mediated by the two helices (Murayama *et al.*, 2001; Golovanov *et al.*, 2003; Huang *et al.*, 2011).

The N-terminal domains of both Omega and ParB are very likely to be flexible, as they are sensitive to proteolytic digestion (Murayama *et al.*, 2001; Huang *et al.*, 2011). The ParG structure was solved using solution NMR, and the N-terminal end is unstructured (Figure 1.18*B*) (Golovanov *et al.*, 2003). The C-terminal end of the folded domain of ParB from pCXC100 that consists of 11 amino acid residues (129-139) is likely to be a flexible part (Huang *et al.*, 2011). Importantly, the N-terminal domain of the CBP was found to be crucial for the interaction with the ATPase protein as well as enhancing its ATPase activity via the arginine finger-like motif (Barillà *et al.*, 2007).

The crystal structure of Omega protein from pSM19035 bound to its DNA site is the only crystal structure that is available for RHH protein of type Ib (Figure 1.18*D*) (Weihofen *et al.*, 2006). The crystal structure of Omega bound to two tandem repeats, both in a direct and inverted orientation, was solved. It revealed nucleotide base-specific contacts in which the β-strand is inserted into the major groove of the DNA. Solving the protein-DNA complex led to building a model in which seven Omega dimers bind to heptad repeats by wrapping around the DNA as a left-handed helix (Weihofen *et al.*, 2006). Unlike P1 ParB, Omega does not bend or distort the DNA (Weihofen *et al.*, 2006).

The only structural data available for RHH CBPs of type II systems are from pB171 ParR, and pSK41 ParR in a complex with their centromeric DNA (Møller-Jensen *et al.*, 2007; Schumacher *et al.*, 2007a). The structure of pB171 ParR contains two monomers arranged in the asymmetric unit, with each of the subunit consisting of a short β-strand followed by five α-helices. The two monomers come together to form a tight homodimer; however, one monomer is distinguished from the other by a small α-helix 5 at the C-terminus, which is absent from the other monomer (Figure 1.18*E*). The ParR protein assembles to form a helical structure consisting of 12 ParR molecules per full 360º turns. The dimers of ParR are arranged with the N-terminal fold facing outward and the C-

terminal domain pointing towards the centre of the helix. The dimers of dimers are held together mainly with residues from helices 3, 4 and 5 (Møller-Jensen *et al.*, 2007).

pSK41 ParR structure was solved with a 20-mer DNA (Figure 1.18*F*) (Schumacher *et al.*, 2007b). The most significant feature of the structure was that the 20-mer DNA fragments, which corresponds to two 10 bp repeats, are packed together to form a pseudo-centromere-like site. Like ParR from pB171, the N-terminal folded domain consists of a RHH dimer. Each monomer consists of a β-strand followed by two helices. The two β-strands fit tightly together in an antiparallel fashion and the two alpha helices from both subunits form a tight dimer (Schumacher *et al.*, 2007). Interestingly, the protein binds to 20 bp repeats as a dimer of dimers in which each dimer binds to a 10 bp repeat. Therefore, the dimer of dimers interacts to form a super-helical-like structure with distinct positive and negative surfaces. The positive surfaces wrap around the DNA, whereas the negative surfaces point inwards towards the centre of the helical structure (Schumacher *et al.*, 2007). This structure is consistent with the R1 ParR superstructure that was observed with electron microscopy (Møller-Jensen *et al.*, 2007). The distortion and the widening of the major groove of DNA allows the antiparallel β-strand to insert into the major groove and to make base contacts using Lys-7, Leu-9 and Lys-11 (Schumacher *et al.*, 2007). Therefore, the open ParR super-helical structure allows ParM to insert to the convex structure of ParR-*parC* and interact with the free C-terminal domain of ParR to form actin-like filament structures (Schumacher *et al.*, 2007; Baxter and Funnell, 2014).

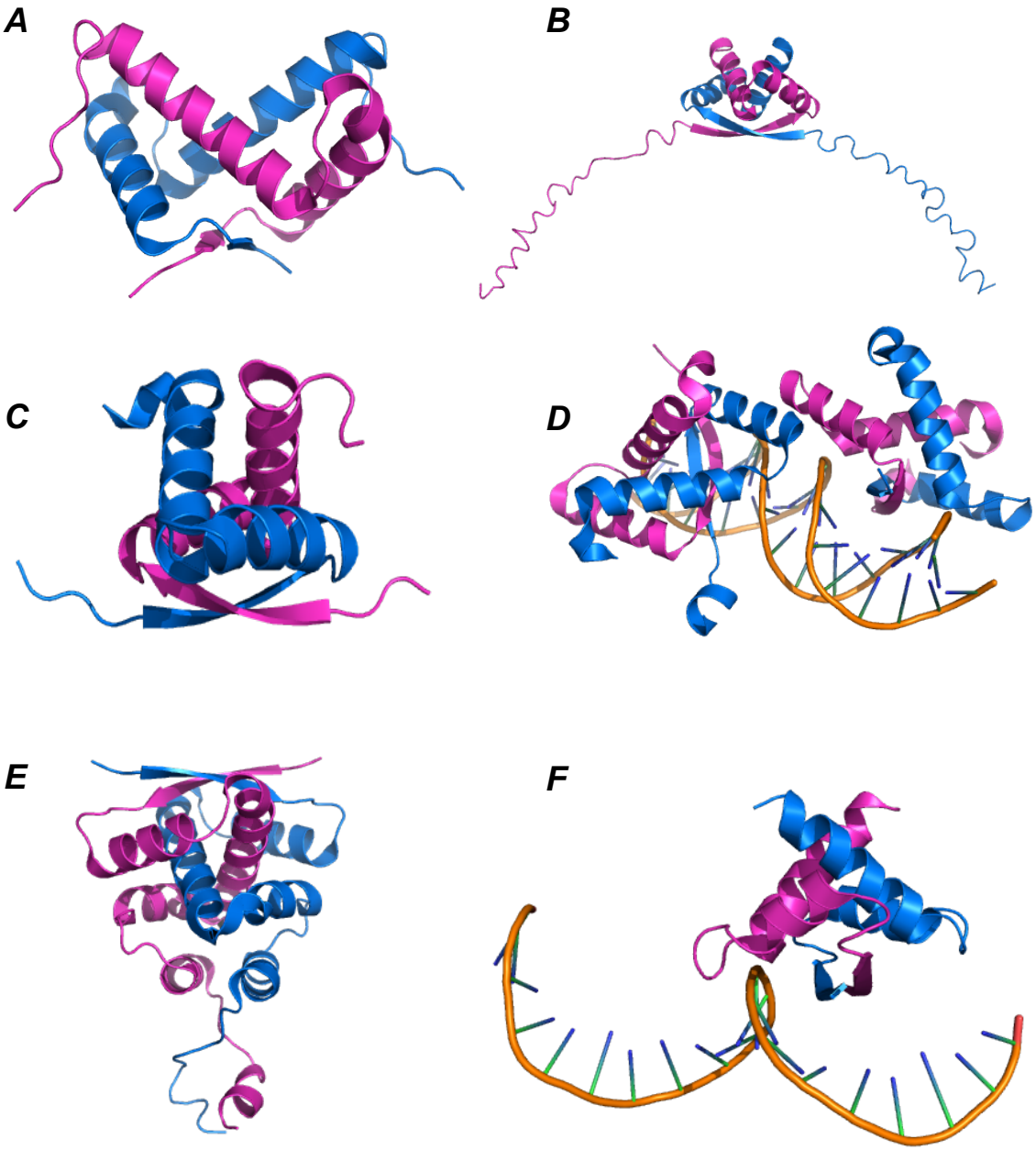


Figure 1.18 Crystal structure of RHH centromere-binding proteins of type Ib and II partition systems

(***A***)-(***C***) cartoons of DNA-binding proteins of type Ib partition system. ***A)*** pCXC100 ParB dimer shows the C-terminal RHH domain (code 3NO7). ***B)*** TP228 ParG dimer shows the flexible N-terminal domain and a RHH C-terminal domain (code 1P94). ***C)*** pSM19035 Omega dimer shows the C-terminal RHH domain (code 1RIQ). ***D)*** A cartoon shows Omega bound to DNA by inserting β-sheet into the major groove of DNA (code 2BNW). ***E)*** A cartoon shows pB171 ParR RHH N-terminal domain of type II partition systems (code 2JD3). ***F)*** A cartoon shows pSK41 ParR RHH N-terminal domain bound to DNA of type II partition systems (code 2Q2K). The CBP in all figure are dimers and the two subunits are distinguished with magenta and blue colours. The figures are generated with PyMol (DeLano, 2002)

1.6 Ribbon-helix-helix type Ib CBP structure homologs

1.6.1 RHH transcriptional repressor superfamily

In prokaryotes, most of the transcriptional regulator proteins adopt a HTH motif in which the recognition alpha helix inserts into the DNA major groove to bind operator DNA site-specifically. A RHH DNA binding motif was first identified when the structure of the methionine repressor protein MetJ from *E. coli* was solved (Rafferty *et al.*, 1989; Schreiter and Drennan, 2007). After structure solution of the Arc and MetJ proteins in their apo state as well as in a complex with their cognate operator DNA, a new transcription factor RHH superfamily was recognised. A RHH superfamily is named as the Arc/MetJ superfamily after the two protein structures were solved and shown that they share the same secondary structure elements that comprise a small domain of these proteins (Schreiter and Drennan, 2007).

All transcriptional repressor harbouring RHH motifs that have been identified are dimers. Their two subunits are tightly intertwined to form a stable homodimer. Nevertheless, the conjugal transfer TraY protein that is encoded by the F plasmid, is a monomer in solution. TraY contains two non-identical RHH motifs in a single polypeptide chain, one is present in the C-terminus and the other present at the N-terminus (Figure 1.19*A*). Moreover, the TraY protein makes base contacts via residues in β-strand (Schildbach *et al.*, 1998; Lum and Schildbach, 1999). Another significant feature of RHH DNA-binding proteins is that the two monomers are paired via β-strands, which are located at the N-terminus of the RHH motif. These two β-strands form an anti-parallel β-sheet, which sits into the major groove of DNA and makes sequence-specific base contacts. Three amino acid side chains from each β-strand are involved in making sequence-specific contacts with the DNA. In addition to the specific interaction, a backbone amide nitrogen at the N-terminus of the second helix makes non-specific contact with the DNA phosphate backbone on both sides of the DNA major groove. Accordingly, these contacts allow optimal interaction of β-sheets (Schreiter and Drennan, 2007). The RHH motif can either be the only motif harboured by a protein or can be found together with other domains in larger proteins (Figure 1.19*A*) (Schreiter and Drennan, 2007).

The *met* repressor protein MetJ regulates the expression of several *E. coli* proteins, which are involved in the biosynthesis of methionine. MetJ binds site-specifically DNA that spans from two to five 8 bp sequences called metboxes. Metboxes are located in the promoter regions of genes in the *met* regulon. The binding of MetJ to tandem repeats is enhanced by the presence of its co-repressor, the methionine derivative, *S*-

adenosylmethionie (SAM) that binds at the C-terminus of the RHH motif (Schreiter and Drennan, 2007). The structure of the MetJ dimer in its apo (Figure 1.20*A*) and holo forms as well as in a complex with metbox DNA were determined. The structure in all three states showed that the MetJ protein is a member of RHH family of DNA-binding proteins. It binds DNA by positioning of the antiparallel β-strands into the major groove DNA, where Lys-23 and Thr-25 make specific nucleotide base contacts (Rafferty *et al.*, 1989; Somers and Phillips, 1992; Augustus *et al.*, 2006). Solving MetJ structure in a complex with two metboxes allowed visualisation of the conformational changes occurring in the higher order complex structure by modelling the MetJ bound to five of metboxes. This complex model revealed the formation of a compact super-helical structure resulted from the interaction of the adjacent MetJ dimers (Augustus *et al.*, 2006).

The Arc repressor that regulates the lytic growth of bacteriophage P22 is an example of a RHH DNA-binding protein, which contains only this motif. Arc binds in a site-specific manner to a 21 bp operator site as a dimer (Figure 1.20*C*) (Raumann *et al.*, 1994). Similar to other RHH DNA-binding proteins, the two monomers intertwine to form a homodimer with β-strands paired to form an antiparallel β-sheet. The specific interaction of Arc with DNA occurs through the amino acid residue in both β-strands similarly to MetJ (Raumann *et al.*, 1994). However, the protein undergoes conformational changes in β-sheets upon DNA binding, forming less symmetrical subunits, differently from MetJ, despite the fact that the Arc protein is a homodimer in its apo state (Raumann *et al.*, 1994).

CopG also consists of a single RHH domain containing 45 amino acid residues (Figure 1.20*B*). CopG is encoded by the streptococcal plasmid pMV158 that controls plasmid copy-number. The CopG protein belongs to the Arc/MetJ transcriptional repressor superfamily and it binds to the DNA operator site in a manner similar to other RHH DNA-binding proteins. No significant differences were observed in two CopG structures, neither in its apo state nor in a complex with 19 bp DNA and were shown that the two CopG structures form RHH folds. This protein was shown to be homodimeric, however the structure revealed that when two dimers bound to DNA, bending DNA around 60° was observed, with both the major and minor grooves facing the protein (Gomis-rüth *et al.*, 1998).

The global regulator Omega protein from the low copy number plasmid pSM19035, binds site-specifically to DNA heptad repeat sites. These sites are located upstream of the promoter regions of the *copS*, γ genes and of the ε -γ -ζ operon. The Omega protein

works as a direct transcriptional repressor for the *copS*, δ and ω genes. Furthermore, Omega indirectly controls the expression of the ε and ζ genes (de la Hoz *et al.*, 2000). The crystal structure of Omega protein was solved, and it showed that the protein is a member of the MetJ/Arc superfamily, although the sequence similarity among them is very low. The crystal structure of Omega consists of two symmetric subunits that form a homodimer. Each monomer consists of β-strand and two alpha helices. The two helices A and B are connected with a small flexible loop that includes two glycine residues, Gly-48 and Gly-49 (Murayama *et al.*, 2001). It was shown that deletion of either 8 or 18 amino acid residues in the N-terminal region produced no or minor effects in terms of the conformational change of the core structure, the plasmid stability and Omega binding to the operator region. Changing threonine 29 in β-strand to alanine weakened the monomer-monomer interaction as well as the binding affinity to DNA (Welfle *et al.*, 2005).

The RHH transcription factors share similar RHH motifs, in spite of low sequence identity among them (Figure 1.19*B*). They bind to their DNA operator site in a similar way using amino acid side chain in the β-strands and make non-specific contacts via residues in α-helix 2. However, each individual RHH transcription repressor recognises different operator sequences and makes base contacts using different amino acids. Each RHH transcription repressor uses unique protein-protein interactions that stabilize their interaction with the DNA (Schreiter and Drennan, 2007).

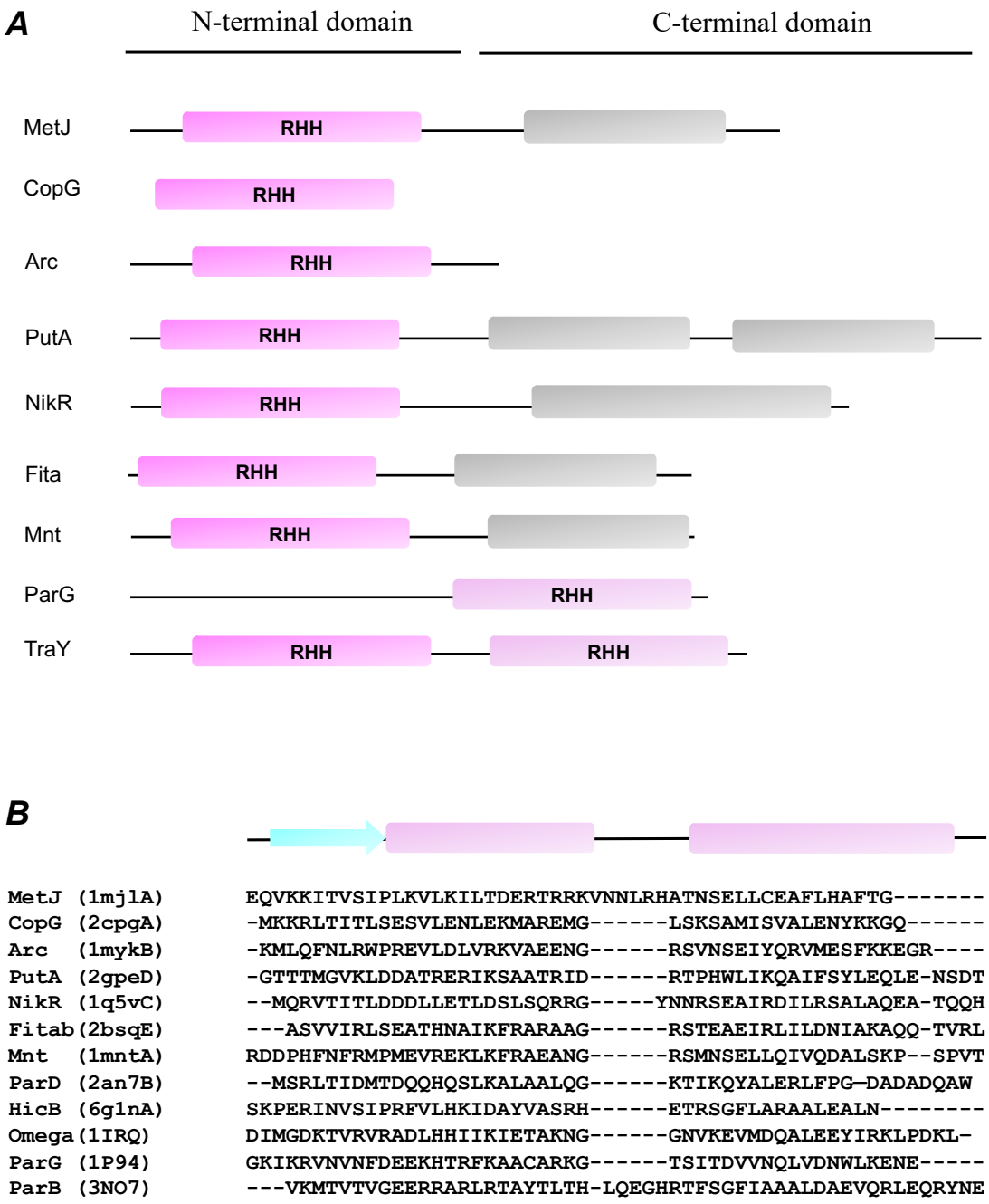


Figure 1.19 Domain organisation and sequence alignment of RHH transcriptional repressors Arc/MetJ family

A) Domain organisation of the RHH transcriptional repressor DNA-binding protein. The RHH motif is shown with magenta bars and the other domains are shown in grey bars. The N-terminal and C-terminal domains are displayed on top. ***B)*** Sequence alignment of some of RHH transcriptional repressor DNA-binding domain.

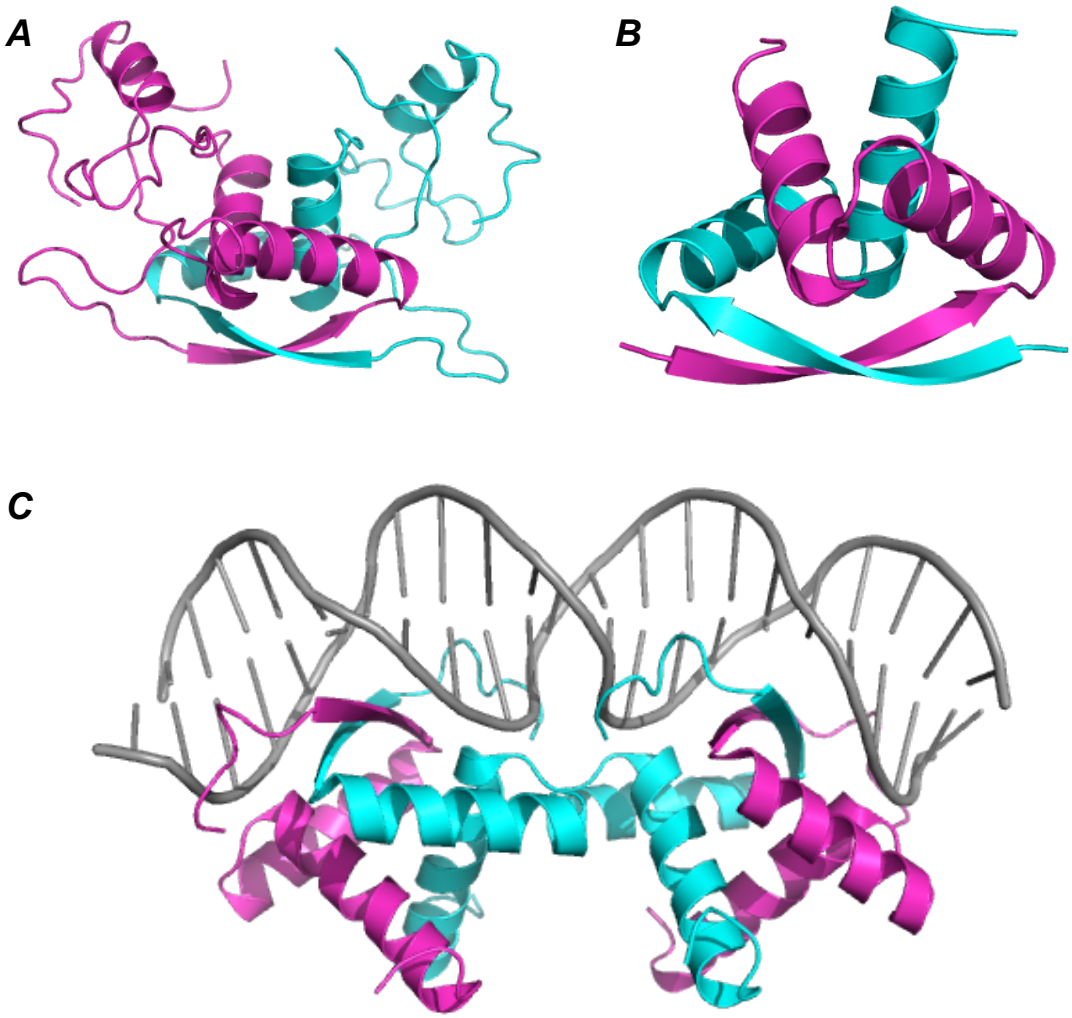


Figure 1.20 Crystal structure of RHH transcriptional repressors Arc/MetJ family

A) pMV158 CopG dimer shows the RHH domain (code 2CPG). ***B)*** Apo MetJ dimer shows the RHH N-terminal domain (code 1CMB). ***C)*** P22 Arc dimers bound to DNA (code 1PAR), in which the β-sheet insert the major groove DNA and make base contact. The transcriptional repressors are dimers and the two subunits are distinguished with magenta and cyan colours. The figures are generated with PyMol (DeLano, 2002)

1.7 Mechanisms of plasmid segregation

1.7.1 Filament-pulling mechanisms of type I system

Many studies have shown that ParA filament formation in type I systems mediates an essential step, which provides the force for driving the newly replicated plasmids apart (Barillà *et al.*, 2005; Ebersbach *et al.*, 2006; Barillà *et al.,* 2007; Bouet *et al.*, 2007; Hatano *et al.,* 2007; Machón *et al.*, 2007; Pratto *et al.*, 2008; Ringgaard *et al.*, 2009). Recently, a cytological study showed that the dynamic movements of ParA over the nucleoid is fundamental for dragging the plasmid to opposite nucleoid poles (Havey *et al.,* 2012; Hwang *et al.*, 2013; Vecchiarelli *et al.*, 2013a, 2013b, 2014a, 2014b; Hu *et al.*, 2017a; McLeod *et al.*, 2017). Both *in vivo* and *in vitro* studies have shown that many ParAs form cytoskeletal structures (Ebersbach and Gerdes, 2001, 2004; Lim *et al.*, 2005; Pratto *et al.*, 2008; Dunham *et al.*, 2009). In addition, it was shown that pB171 ParA oscillates in spiral-like structures between the two nucleoid poles (Ebersbach and Gerdes, 2004; Ebersbach *et al.*, 2006), whereas TP228 ParF forms a meshwork of polymers and oligomers through the nucleoid volume (McLeod *et al.*, 2017). Importantly, *in vitro* studies showed that ParA binds DNA in a non-specific manner, consistent with the *in vivo* observation where ParA oscillates over the nucleoid (Ebersbach and Gerdes, 2001; Hester and Lutkenhaus, 2007). ParA non-specific interaction with DNA stimulates its polymerisation and bundling. However, the presence of non-specific DNA inhibits SopA polymerisation in the F plasmid segregation system (Bouet *et al.*, 2007).

Microscopy observations of the *par2* system of pB171 plasmid allowed Ringgaard *et al* to shed light on type I segregation systems. Combining cytological observations and biochemical studies on other type I systems allowed them to propose a filament pulling model (Figure 1.21) (Ringgaard *et al.*, 2009). The model that is proposed by the authors supports the idea of ParA assembly and disassembly may provide the force that pulls the sister plasmid away from mid-cell towards the cell pole (Ringgaard *et al.*, 2009). ParA is dimeric in the ATP-bound form, but monomeric in the ADP-bound form. Once the plasmid is replicated, ParB binds specifically to the tandem repeats in the centromere site to form the partition complex. ParA-ATP dimers bind cooperatively to the nucleoid forming long ParA filaments. The growing ParA filament then contacts the plasmid through interaction with ParB-*parC* complexes (Ringgaard *et al.*, 2009). Walker-type ParAs have weak ATPase activity that is enhanced by the N-terminus of the partner ParB protein (Davis *et al.*, 1996; Ebersbach *et al.*, 2006). Interestingly, the N-terminus of TP228 ParG contains an arginine finger-like motif that stimulates ParF ATPase activity and this arginine has been shown to be highly conserved among ParB proteins (Barillà *et al.*, 2007). Once the ParA-ATP filament binds the ParB-*parC* complex, the ParA-ATP

tip of the filament hydrolyses ATP and is converted to a ParA-ADP molecule that detaches from the nucleoid. Thus, a new ParA-ATP filament end becomes accessible and free for interaction with the ParB-*parC* anchoring site of the plasmid. This cycle of depolymerisation continues until all the ParA-ATP molecules are converted to ParA-ADP, leading to filament disassembly and release of the plasmid at the nucleoid pole. A new ParA filaments builds up at the opposite nucleoid pole and continues to grow till it finds the released plasmid and moves it in the opposite direction (Figure 1.21) (Ringgaard *et al.*, 2009). However, a recent *in vivo* study combined with mathematical modelling demonstrated that ParA oscillations, structure formation, and plasmid positions, relies on the nucleoid architecture and morphology. Moreover, it was shown that ParA oscillates less as it was assumed previously (Ietswaart *et al.*, 2014).

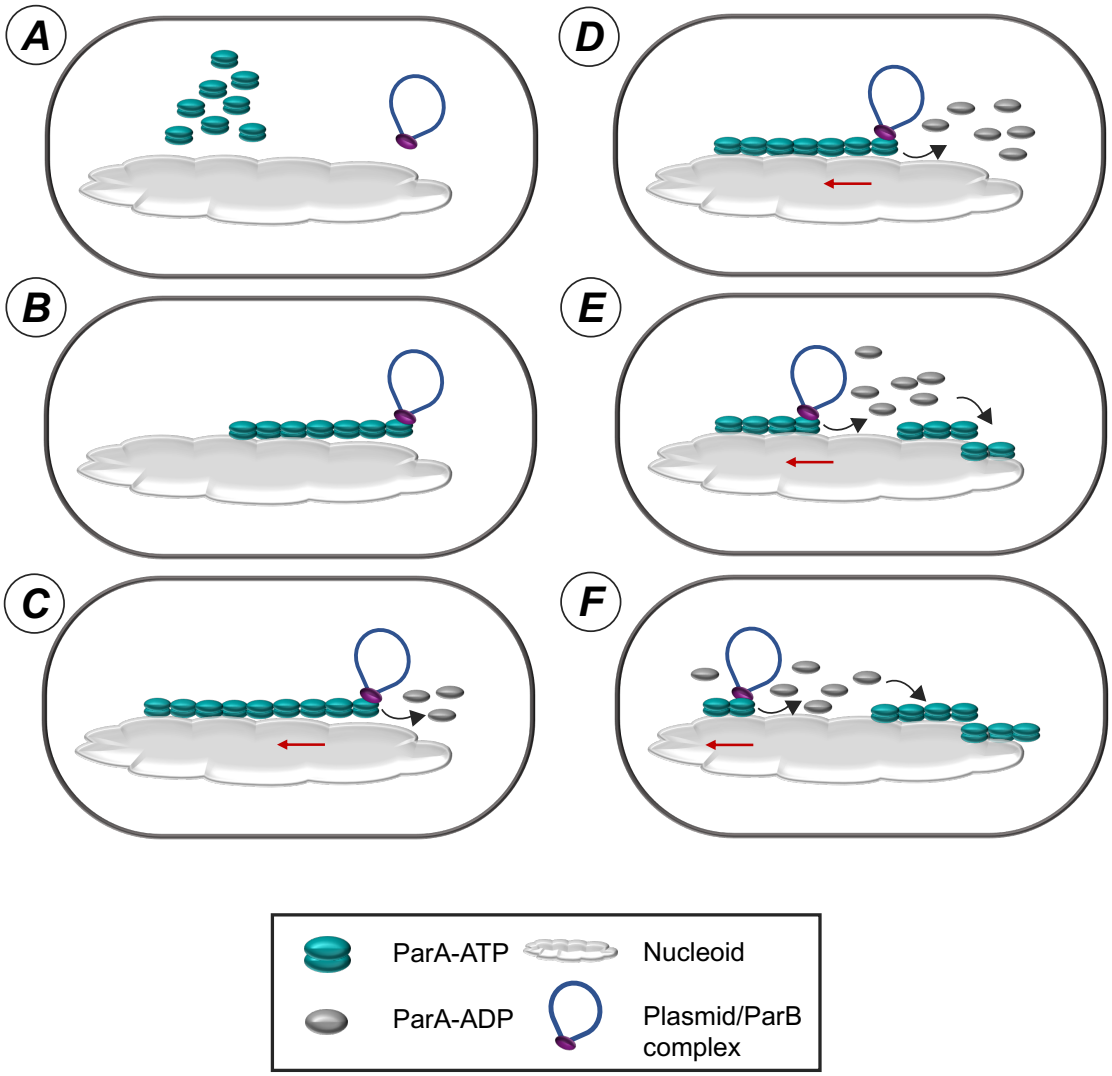


Figure 1.21 Model of the pulling mechanisms of plasmid segregation mediated by type I systems

A) ParB loads on the centromere site to form a nucleoprotein complex on the plasmid. The ParA-ATP dimers bind the nucleoid non-specifically and start polymerising. ***B)*** Subsequently, this non-specific interaction leads to filament formation. The ParA-ATP filaments continues to grow and contacts the plasmid via ParB bound to the centromeric site. ***C)*** Once ParB is bound to ParA, it stimulates the ATPase activity of ParA. Thus, this activity converts ParA-ATP to ParA-ADP, releasing it from the DNA and leaves the ParB/plasmid complex free to contact a new ParA-ATP at the tip of the filament. ***D)*** Polymerisation and depolymerisation events lead to the plasmid being pulled in one direction. ***E)*** The plasmid continues to move in the same direction towards one of the nucleoid poles by the action of ParA-ATP hydrolysis, leaving a ParA-free nucleoid zone behind. ParA released from the ParB/*parC* complex builds up to form a new filament. ***F)*** Once the plasmid reaches the extremity of the disassembled filament, it is dropped. A new ParA-ATP filament grows at the opposite side nucleoid pole and the cycle re-starts. The diagram is adapted from (Ringgaard *et al.*, 2009).

1.7.2 Diffusion ratchet mechanism of type I system

The diffusion ratchet segregation mechanism of type I systems was first proposed by Vecchiarelli *et al* (Vecchiarelli *et al.*, 2010) (Figure 1.22). This model challenges the filament-based model, which was proposed earlier, proposing that a ParA concentration gradient on the nucleoid surface could provide the driving force for the cargo movement (Ringgaard *et al.*, 2009; Vecchiarelli *et al.*, 2010). A combination of cytological, biochemical studies and mathematical modelling on P1 and F1 plasmids, as well as *in vivo* study on *Caulobacter crescentus* chromosome segregation provided strong evidence against the filament-based model and in favour of a Brownian ratchet mechanism (Havey *et al.,* 2012; Hwang *et al.*, 2013; Vecchiarelli *et al.*, 2013a, 2014a, 2014b; Lim *et al.*, 2014; Hu *et al.*, 2017b, 2017a). Moreover, a structural study showed that the interaction of ParA with the nucleoid leads to trap the plasmid and ParA, thus, preventing them from floating away into the cytosol (Zhang and Schumacher, 2017).

Super-resolution microscopy studies revealed that the localisation and oscillation of the DNA cargo depends on ParA-ATPase activity as well as the dynamic DNA-binding characteristics of ParA (Le Gall *et al.*, 2016). Inconsistent with the diffusion ratchet model, an *in vivo* study on *Caulobacter crescentus* chromosome segregation showed that the elastic properties of the chromosome provide the mechanical force needed to translocate the DNA cargo. Consequently, a DNA-relay mechanism was proposed (Lim *et al.*, 2014).

Nevertheless, both the filament-based and the diffusion ratchet models support the view according to which ParA dimerises upon binding to ATP, which in turn promotes its non-specific interactions with the nucleoid DNA (Leonard *et al.*, 2005; Castaing *et al.*, 2008). Studies conducted *in vitro* and *in vivo*, showed that ParA-ATP dimers can assemble into filaments (Barillà *et al.*, 2005; Ebersbach *et al.*, 2006; Barillà *et al.*, 2007; Hatano *et al.*, 2007; Machón *et al.*, 2007; Ringgaard *et al.*, 2009).

In the diffusion ratchet model (Figure 1.22), the nucleoid is used as a matrix, in which the plasmid is moved by the action of ParB and ParA. Moreover, ParA was found in two states; active, when it is bound to DNA, and inactive, when ParA is free, and not bound to DNA. ParA dimers in the ATP-bound form bind dynamically to the nucleoid DNA and many ParB dimers bind the plasmid through the centromere site, *parS*, to form the partition complex (Vecchiarelli *et al.*, 2010). ParA has a weak ATPase activity that is enhanced by the ParB-*parS* complex. Once ParB stimulates ParA ATPase activity, the ParA-ATP dimer is converted to monomer and moves away from the nucleoid. The

continuous conversion of ParA-ATP to ParA-ADP acts to move the partition complex towards low concentrations of ParA. The dissociated inactive ParA cannot re-bind DNA immediately. Instead, it diffuses throughout the cell cytosol, where ParA undergoes a slow transition in which ParA-ADP is converted to ParA-ATP. This time delay allows an active ParA-ATP to bind again to the nucleoid DNA (Vecchiarelli *et al.*, 2010, 2013b; Hwang *et al.*, 2013). The partition complex changes its direction once it reaches one of the nucleoid poles. The uneven distribution of ParA on the nucleoid matrix revives the plasmid movement back towards the other nucleoid end and positions the plasmid at mid-cell (Vecchiarelli *et al.*, 2010, 2013a, 2013b; Hwang *et al.*, 2013). An end to end oscillation pattern of ParA and ParB has been observed for many partition systems (Ringgaard *et al.*, 2009; Lim *et al.*, 2014; Surovtsev *et al.*, 2016). Moreover, the interaction of ParA with the nucleoid matrix could explain the helical movement of ParA that was observed for the segregation complex of pB171 plasmid (Ringgaard *et al.*, 2009). Once the plasmid is duplicated, ParA builds up in between the two-sister plasmids. Thus, the partition complexes start to remove ParA by hydrolysis and each partition complexes continues to move along the nucleoid in the opposite direction towards the nucleoid poles (Vecchiarelli *et al.*, 2010).

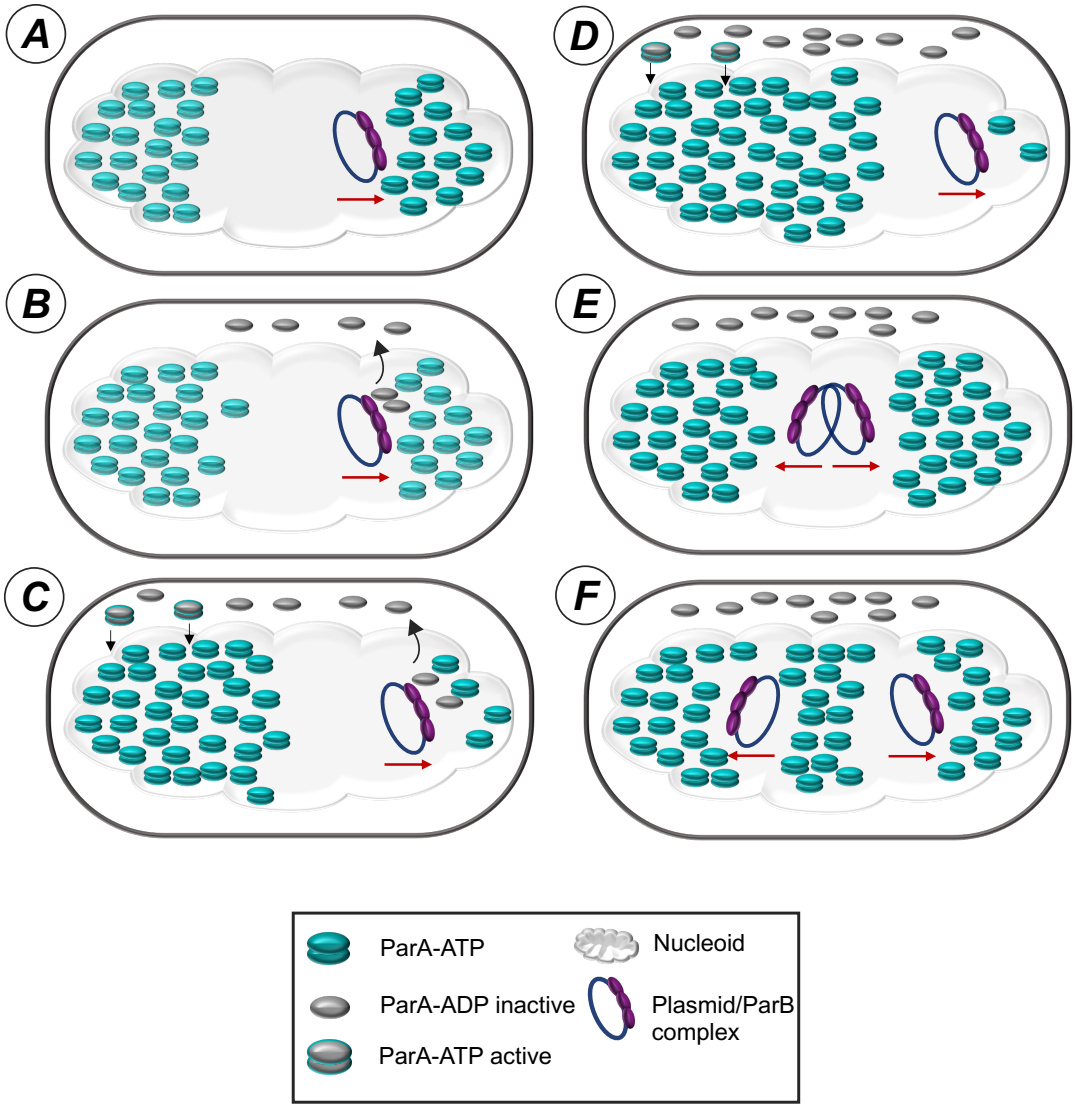


Figure 1.22 Model of the diffusion ratchet segregation mechanisms in type I systems

A) ParB loads on the plasmid centromere site to form a nucleoprotein complex. The ParA-ATP dimer binds non-specifically to the nucleoid. ***B)*** Once ParB has bound to ParA, it stimulates the ATPase activity of ParA. This activity converts ParA-ATP to its ADP form that is released from the nucleoid. The continuous cycles of ParA hydrolysis lead to plasmid movement towards the low ParA-ATP concentrations. ***C)*** Once ParA-ADP is released from the nucleoid, it exchanges ADP with ATP and undergoes a slow conformational change. This time delay allows ParA-ATP to diffuse in the cytosol and becomes activate to rebind to the nucleoid. ***D)*** When the plasmid reaches the nucleoid pole, it changes its direction to the ParA-ATP gradient leaving the depleted area behind. ***E)*** When the plasmid is duplicated, each individual plasmid moves into opposite direction due to ATP hydrolysis. ***F)*** The plasmids continue their movement until they reach the ¼ and ¾ position. The diagram is adapted from (Vecchiarelli, *et al*., 2010).

1.7.3 Venus flytrap of TP228 plasmid

In 2017 a study that used a combination of 3D super-resolution microscopy and confocal microscopy allowed McLeod and co-workers to propose a Venus flytrap model for the type Ib segregation system of TP228 plasmid (Figure 1.23). In this model, ParG-*parH* complexes are captured and engulfed by a 3D meshwork formed by ParF (McLeod *et al.*, 2017). The segregation system of the TP228 plasmid consists of three main elements: ParF, a Walker-type ATPase protein, ParG, a DNA-binding protein, and a centromeric site, *parH* (Hayes, 2000). Following plasmid replication, ParG binds site-specifically to the tandem repeats in the *parH* site to form a partition complex that mediates plasmid pairing (McLeod *et al.*, 2017). Pairing newly replicated plasmid is a fundamental step in plasmid segregation (Edgar *et al.*, 2001).

Like many ParAs, ParF forms higher order structures (bundles) *in vitro* (Barillà *et al.*, 2005). The crystal structure of the ParF-ATP bound state showed that ParF is monomeric in its ADP-bound state. In the ATP-bound state ParF forms dimers, which assemble into dimers of dimers that subsequently form polymers (Schumacher *et al.*, 2012). ParF has a weak ATPase activity that is enhanced by ParG through its flexible tail containing an arginine finger-like motif (Barillà *et al.,* 2007). An *in vitro* study showed that the tail of ParG promotes the formation of ParF polymers (Barillà *et al.*, 2005). Assembly and disassembly of ParF higher order oligomers is required for the oscillation of the protein over and through the volume of the nucleoid (McLeod *et al.*, 2017). A previous model that was proposed for type I systems was based on ParA filaments formation over the nucleoid, which grow into long filaments. These ParA filaments dissociated via ATP hydrolysis and thereby act to pull the plasmid out of the centre toward the nucleoid pole before cell division (Ringgaard *et al.*, 2009). However, 3D super-resolution microscopy showed that ParF does not form a nucleoid-long filaments as was assumed previously, instead, ParF forms a meshwork that permeates the entire nucleoid volume (McLeod *et al.*, 2017).

The ParF 3D meshwork observed in live cells was a mix of long polymers and short oligomers that branch and are distributed within the nucleoid surface and volume. The ParF meshwork possesses two ends, with a leading compact edge in front and the loose, lagging end that contains fewer oligomers (Figure 1.23). Once the plasmid is duplicated and ParG binds to *parH*, the two sister plasmids are engulfed in the ParF matrix and transported from one end of the nucleoid to the another. ParF oligomers start to grow in between the two plasmids leading one of the plasmids towards one end. One of the plasmids is then dropped at one of the nucleoid poles due to the ATP hydrolysis of ParF,

leading the lagging edge to dissociate. The other plasmid remains attached to the compact edge of the meshwork and is moved to the opposite cell pole where it is eventually released from the meshwork due to ATP hydrolysis of ParF (McLeod *et al.*, 2017).

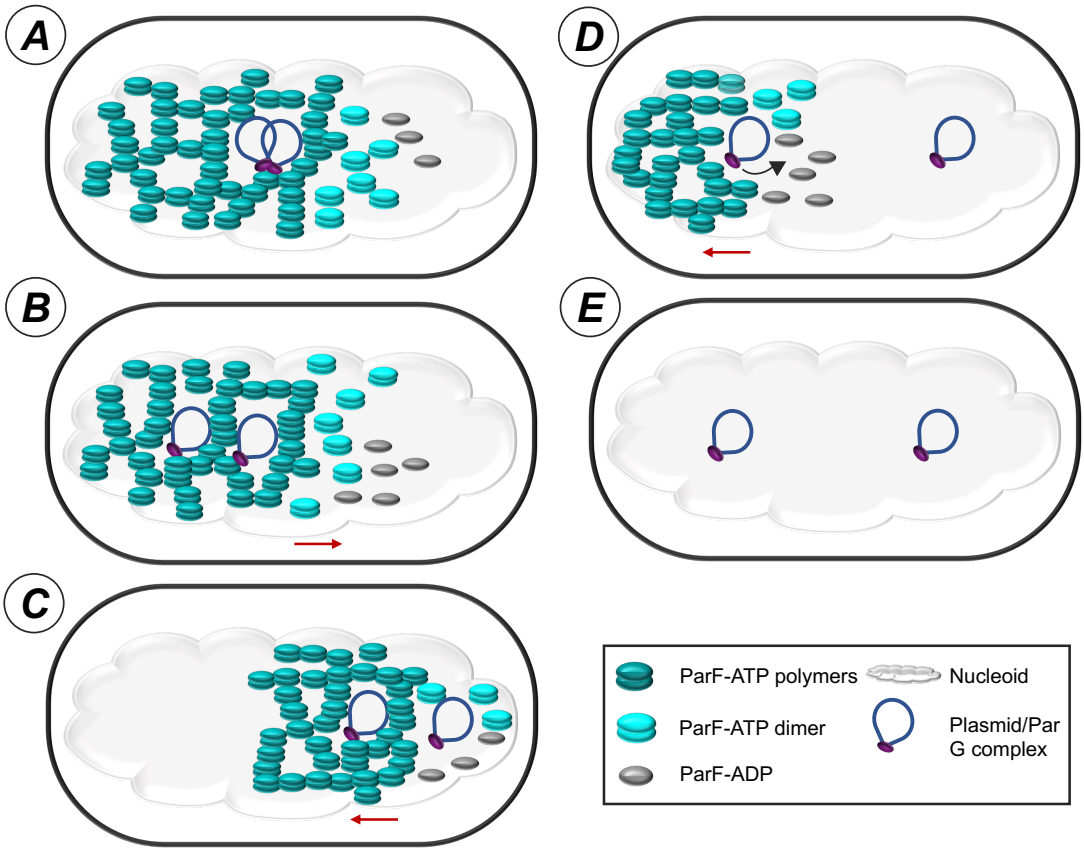


Figure 1.23 Venus flytrap model illustrating the segregation mechanisms of the TP228 plasmid

A) ParG loads on the centromere site to form a nucleoprotein complex structure on the plasmid. ParA-ATP dimers bind non-specifically to the nucleoid and form meshwork throughout the chromosome. The ParF meshwork possesses two ends, with a leading compact edge on the left and the loose, lagging end on the right. ***B)*** The ParF meshwork engulfs the newly replicated plasmids and transports them toward one nucleoid poles. During the transport stage, ParF oligomers start to grow in between the sister plasmids. ***C)*** When one of the sister plasmids reaches the nucleoid pole, it detaches from the ParF meshwork and is released due to the disassembly of the matrix caused by ParF ATPase activity. ***D)*** The other sister plasmid remains attached to the ParF meshwork and is moved to the opposite nucleoid pole. ***E)*** Finally, the sister plasmid is released from the meshwork, and repeated oscillations of ParF meshwork leads to repositioning of the two sister plasmids. The diagram is adapted from (McLeod, *et al*., 2017).

1.7.4 Pushing (insertional polymerisation) mechanisms of type II system

The segregation dynamics mediated by type II *parMRC* rely on ParM polymerization that is responsible for positioning the newly replicated plasmids to the opposite cell poles (Figure 1.24). However, the structure and arrangement of the filaments formed by R1 ParM and pB171 ParM are different from the one of pSK41 ParM (Møller-Jensen *et al.*, 2003; Popp *et al.*, 2010; Rivera *et al.*, 2011). *In vivo* studies showed that the duplicated R1 plasmid can be transported equally into daughter cells after cell division (Møller-Jensen *et al.*, 2007). ParR binds site-specifically to the tandem repeats in the *parC* site and mediates plasmid pairing (Jensen *et al.*, 1998). Each ParR dimer binds to a single repeat, and the binding of ParR to *parC* bends the DNA to form a U shape-like structure with the DNA wrapped around (Jensen *et al.*, 1998; Hoischen *et al.*, 2008). The actin-like protein, ParM, forms a bipolar left-handed double strand filament (Gayathri *et al.*, 2012, 2013). ParM forms filaments in the ATP-bound state. Binding experiments have shown that ParM binds to ATP and GTP, but with a preference for ATP, although the protein is more stable in its GTP-bound form (Galkin *et al.*, 2009; Rivera *et al.*, 2011). ParR stimulates the ATPase activity of ParM and maximal ATPase activity is achieved when ParR-*parC* is present (Jensen and Gerdes, 1997). The ParM protein forms filaments in the presence of both ParR and *parC*, which is the nucleation site for ParM polymerisation (Møller-Jensen *et al.*, 2002; van den Ent *et al.*, 2002). ParM-ATP monomers are continuously added to the ends of the filaments and dissociate from the ends upon hydrolysis. Polymerisation and depolymerisation occur at an equal rate from either end. Additionally, ParM monomers are added to the ends of ParR-bound filaments forming a bipolar spindle. The bipolar spindles of ParM filaments elongate from both ends of ParR, thereby pushing the R1 plasmid out of the centre towards opposite cell poles (Garner *et al.*, 2007; Choi *et al.*, 2008; Gayathri *et al.*, 2012; Bharat *et al.*, 2015).

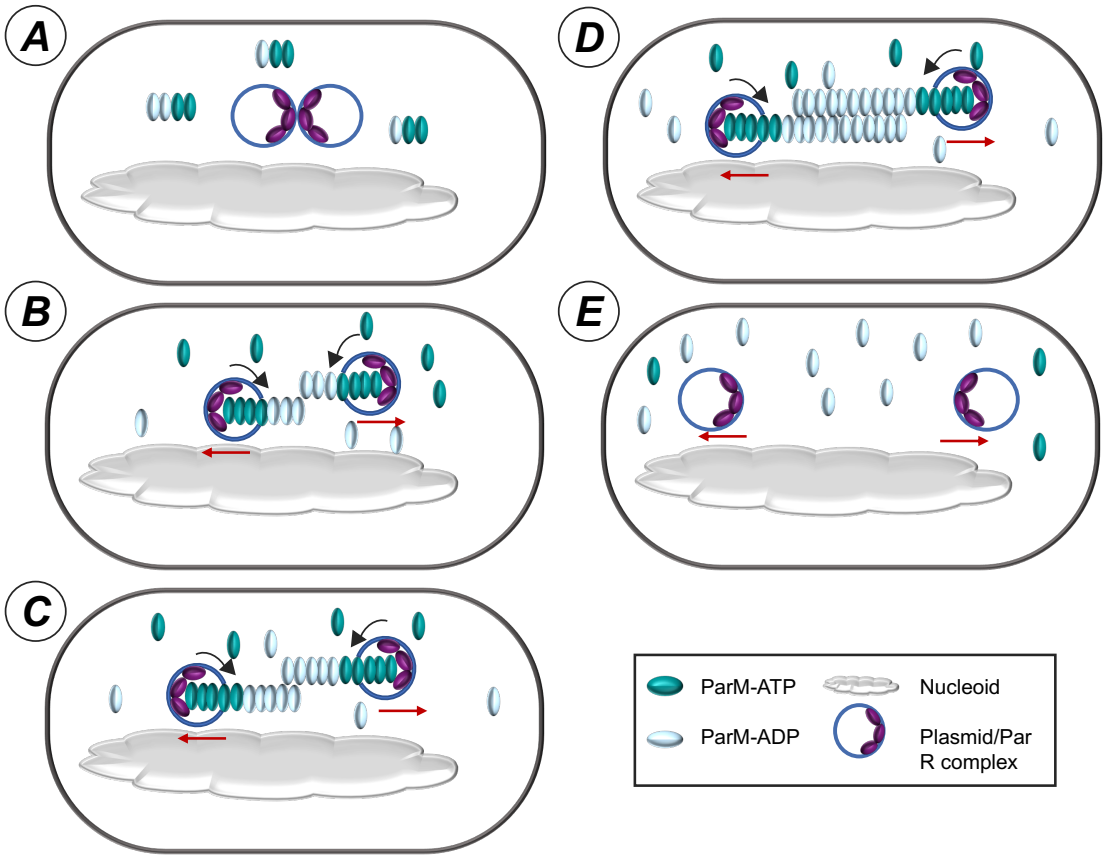


Figure 1.24 Model of the pushing (insertional polymerisation) mechanism of type II system

A) After plasmid replication, ParR dimers bind to *parC* site and bend the DNA. ***B)*** The ParR- *parC* complexes are the nucleation sites for ParM. ParM monomers are continuously added to the nucleation sites and the filaments start to grow. ParM within the filaments convert ATP to ADP and the bipolar spindle filaments continue to elongate as long as ParM is bound to ParR-parC complex. ParM quickly dissociated from the filaments. ***C)*** The bipolar spindle continues to elongate on both ends at an equal rate. ***D)*** The continuous growing of ParM filaments leads to push the two sister plasmids apart toward the cell poles. ***E)*** The ParM filaments eventually disassemble.

1.7.5 Pulling-Treadmilling mechanisms of type III system

Segregation mechanisms underpinning the pBtoxis plasmid are still not fully understood. However, Larsen *et al* used fluorescence recovery after photobleaching (FRAP) experiment to show a treadmilling mechanism for the first time in a prokaryotic tubulin (Larsen *et al.*, 2007). Microscopy studies using TubZ-GFP showed that TubZ forms long filaments *in vivo*, which extend from one cell end to the other. Additionally, the investigations showed that the translocation is very dynamic, as one end of the filament elongates, whereas the other end shrinks (Larsen *et al.*, 2007). Later, the dynamics of elongating (plus) ends and shrinking (minus) ends were shown by Fink and Löwe using single-molecule imaging and they proposed a pulling treadmilling mechanism (Fink and Löwe, 2015).

Unlike the actin-like ParM proteins, TubZ does not form bipolar filaments. Instead, it polymerises continuously at the plus end and depolymerises from the minus end. Thus, by unknown mechanism, it positions the plasmid towards the cell pole (Fink and Löwe, 2015).

In analogy to other partition systems, the first step of plasmid partition is the binding of TubR, DNA-binding protein, to the tandem repeats in the centromeric site, *tubC* to form a partition complex (Ni *et al.*, 2010; Fink and Löwe, 2015). The tubulin protein TubZ forms polymers upon binding GTP (Larsen *et al.*, 2007; Ni *et al.*, 2010). TubZ binds to TubR through the long, flexible, C-terminal tail. The centromeric TubR-*tubC* complex was found to bind to the shrinking minus end rather than the elongating plus end (Fink and Löwe, 2015). Therefore, it was assumed that the minus end exposes multiple free TubZ C-terminal ends rather than the plus ends, to allow TubR binding (Montabana and Agard, 2014; Fink and Löwe, 2015). When TubZ depolymerises, the minus end exerts force to pull the TubR-*tubC* complex while treadmilling (Fink and Löwe, 2015). The cycle of adding subunits to the growing end and losing subunits from the shrinking end leads to transportation of TubR-*tubC* to the cell pole. Once the TubZ polymers reaches the cell pole, it bends around the curved cell pole and continues to grow towards the opposite direction (Ni *et al.*, 2010). It is still unclear how the plasmid is dropped at the cell pole and whether a third protein TubY, could be involved in the mechanisms of partitioning (Fink and Löwe, 2015). However, it was assumed that the force of the interaction with the membrane leads to the release of the TubR-plasmid complex (Larsen *et al.*, 2007). Moreover, TubY was shown to interact with TubZ polymers and the TubZR-*tubC* complex (Oliva *et al.*, 2012).

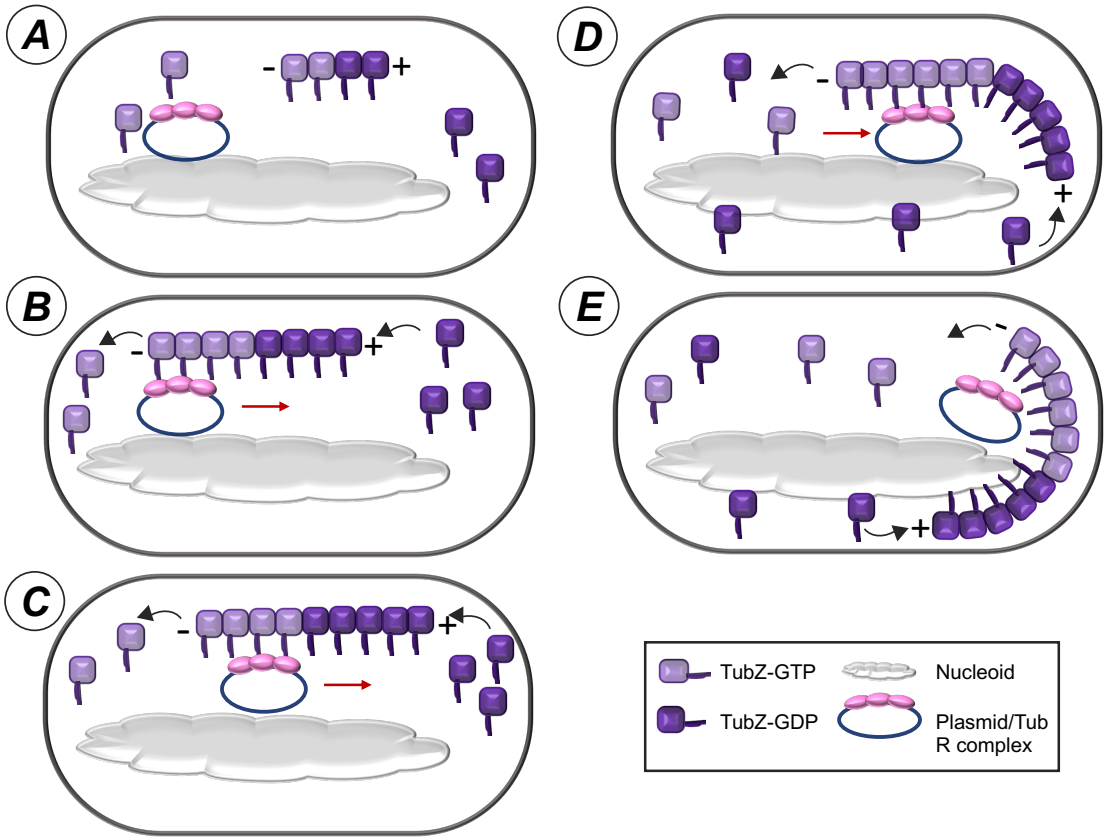


Figure 1.25 Model of the pulling-treadmilling mechanisms mediated by the type III system

A) TubR binds the plasmid via the partition site, *tubC*. TubZ monomers associate to form a dynamic filament, in which TubZ-GTP is continuously added to the plus (+) end, and TubZ-GDP dissociates from the minus (-) end. ***B)*** The TubZ filament grows and associates with TubR through the C-terminal tail. ***C)*** The plasmid is handed from one molecule to another toward the direction of plus end due to the depolymerisation of the minus end. As a result of TubZ depolymerisation, TubR-*tubC* complex is pulled while treadmilling. ***D)*** Once the filament reaches at the cell pole, it bends. ***E)*** The interaction of the filament with the curved cell pole leads to detachment of the plasmid from the filament. The TubZ filament then moves towards the opposite cell pole.

1.8 Aims of the project

The stable maintenance of genetic information, which is crucial for all organisms, is achieved by genome partition. DNA segregation in bacterial cells is a highly accurate process that ensures the precise transmission of chromosomes and plasmids at cell division. Large and low copy number plasmids play a fundamental role in pathogenicity as well as in the spread of multidrug resistance. Therefore, understanding the molecular mechanisms mediating the stable maintenance of plasmids has significant impacts.

The enteropathogenic *E. coli* strain B171-8 cause acute and persistent diarrhoea in children under the age of five years. The strain has medical importance as it harbours a ~69 kb and low-copy number pB171 plasmid that encodes a cluster of 12 genes encoding BFP. The set of virulence genes are responsible for localised adherence exhibited by EPAF. Therefore, understanding the stable maintenance of the virulence plasmid is of the medical important. The segregation of pB171 plasmid from EPEC B171 requires the centromere-binding protein, ParB, the ATPase protein, ParA, and two centromere sites, *parC1* and *parC2*. However, no structural information on either of the proteins was available at the time this study began, and very little was known about the functional role and interactions of the repeats in the partition sites. Therefore, a combination of biochemical, biophysical and structural biology experiments allowed us to solve the structure of the DNA-binding protein, ParB. The C-terminal folds into a RHH DNA-binding domain, whereas the N-terminal domain is highly flexible. In addition, biochemical and biophysical investigations allowed us to understand the interaction of ParB with the centromeric sites. Throughout this project, it was possible to identify three clusters of amino acid residues in the N-terminal domain that one of them was shown to interact with a single repeat, while the amino acids of the second cluster are believed to undergo conformational transition upon interaction with a single repeat. In addition, an *in vivo* plasmid stability assay was performed to investigate the stable maintenance of mini plasmid that harbour the *parABC* partition cassette and shown that one of the centromere sites is indispensable for plasmid segregation than the other. Finally, the interaction of the Walker-type ParA with ParB studies have identified certain amino acids that could be crucial for the interplay between both segregation proteins.

The overall aim of this project has been to investigate the interaction between the components of the pB171 partition system. Specifically:

1. *In vivo* analysis of the functional importance for plasmid segregation of all the repeats in the centromere sites *parC1* and *parC2*, by truncating some of them and assessing plasmid stability
2. Characterising the properties of the binding of ParB to the different repeats of the centromere sites
3. Determining the structure of ParB and investigating the role of the N-terminal domain in DNA binding
4. Investigating the interaction and interplay of the two segregation proteins ParA and ParB in the presence and absence of DNA

Chapter 2

Materials and Methods

2.1 Bacterial strains and plasmids

2.1.1 Bacterial strains

The *E. coli* strains used in this study were obtained from laboratory glycerol stocks stored at -80°C. The bacterial cells were streaked from glycerol stocks on Luria-Bertani (LB) agar and grown at 37°C overnight. One fresh colony was taken from the overnight culture and inoculated into 5 ml LB. The culture was grown overnight at 37°C for competent cells preparation as described in Section 2.3.1. The bacterial strains used in this work with genotype descriptions are outlined in Table 2.1.

Table 2.1 List of *E. coli* strains used in this study

***E. coli* strain**	**Genotype**	**Application**	**Reference**
DH5α	F^- φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17(r_K^-, m_K^+) *pho*A *sup*E44 λ^- *thi*-1 *gyr*A96 *rel*A1	Cloning, plasmid mini-prep and storage	-
BL21(DE3)	F^-, *omp*T, *hsd*S_B (r_B^-, m_B^-) *gal*, *dcm* (DE3)	Protein overproduction	-
BR825	It contains a mutation in *polA* that prevents DNA polymerase from starting replication from the medium copy number	Partition assays	(Ludtke *et al.*, 1989)

2.1.2 Plasmids

Plasmids were extracted from *E. coli* cultures as described in Section 2.3.3. Plasmids used in this work were either constructed during this study or were available in the laboratory collection. The list of plasmids used in this study is outlined in Table 2.2.

Table 2.2 List of plasmids used in this study

Plasmid	**Description**	**Selection marker**	**Source**
pFH450	A pBR322 derivative plasmid harbouring P1 and ColE1 origins of replication	Chloramphenicol	(Hayes, 1998)
pET-22b (+)	Overexpression vector harbouring the T7 promoter and codons fro a (His)×6-tag	Ampicillin	Novagen

	that becomes fused to the C-terminus of the protein		
pCC02	A pBR322 derivative plasmid harbouring P1 and ColE1 origins of replication and the full *parABC* partition cassette	Chloramphenicol	(Fothergill *et al.*, 2005)
pET-ParB	The plasmid contains *parB* cloned into pET-22b (+) vector	Ampicillin	(Fothergill *et al.*, 2005)
pET-ParA	The plasmid contains *parA* cloned into pET-22b (+) vector	Ampicillin	(Fothergill *et al.*, 2005)
pET-ParBΔ30N	*parBΔ30N* allele cloned into pET-22b (+) vector using *Nde*I and *Xho*I restriction sites	Ampicillin	This study
pET-ParB-K3A	*parBK3A* allele cloned into pET-22b (+) vector using the *Nde*I and *Xho*I restriction sites	Ampicillin	This study
pCC02-parBK3A	The *parABC* partition cassette with *parBK3A* allele cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study
pC1ΔR1R2	The *parABC* partition cassette with deletion of R1 and R2 repeats from the *parC1* site, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study
pC1ΔB1	The *parABC* partition cassette with the deletion of R1, R2 and B1 repeats from the *parC1* site, cloned into	Chloramphenicol	This study

	pFH450 vector using the *Eco*RV and *Xho*I restriction sites		
pC1ΔB2	The *parABC* partition cassette with the deletion of R1, R2 and B1-B2 repeats from the *parC1* site, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study
pC1ΔB3	The *parABC* partition cassette with the deletion of R1, R2 and B1-B3 repeats from the *parC1* site, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study
pC1ΔB4	The *parABC* partition cassette with the deletion of R1, R2 and B1-B4 repeats from the *parC1* site, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study
pC1ΔB5	The *parABC* partition cassette with the deletion of R1, R2 and B1-B5 repeats from the *parC1* site, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study
pC1ΔB6	The *parABC* partition cassette with the deletion of R1, R2 and B1-B6 repeats from the *parC1* site, cloned into pFH450 vector using	Chloramphenicol	This study

	the *Eco*RV and *Xho*I restriction sites		
pC1ΔB7	The *parABC* partition cassette with the deletion of R1, R2 and B1-B7 repeats from the *parC1* site, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study
pC1ΔB8	The *parABC* partition cassette with the deletion of R1, R2 and B1-B8 repeats from the *parC1* site, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study
pC1ΔB9	The *parABC* partition cassette with the deletion of R1, R2 and B1-B9 repeats from the *parC1* site, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study
pC1ΔB10	The *parABC* partition cassette with the deletion of R1, R2 and B1-B10 repeats from the *parC1* site, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study
pC1ΔB11	The *parABC* partition cassette with the deletion of R1, R2 and B1-B11 repeats from the *parC1* site, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study

pC1ΔB12	The *parABC* partition cassette with the deletion of R1, R2 and B1-B12 repeats from the *parC1* site, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study
pC1ΔB13	The *parABC* partition cassette with the deletion of R1, R2 and B1-B13 repeats from the *parC1* site, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study
pC2ΔB35	The *parABC* partition cassette with the deletion of B35 repeat from the *parC2* site, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study
pC2ΔB34-35	The *parABC* partition cassette with the deletion of B34-35 repeats from the *parC2* site, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study
pC2ΔB33-35	The *parABC* partition cassette with the deletion of B33-35 repeats from the *parC2* site, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study
pC2ΔB32-35	The *parABC* partition cassette with the deletion of B32-35 repeats from the	Chloramphenicol	This study

	parC2 site, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites		
pC2ΔB31-35	The *parABC* partition cassette with the deletion of B31-35 repeats from the *parC2* site, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study
pC2ΔB30-35	The *parABC* partition cassette with the deletion of B30-35 repeats from the *parC2* site, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study
pC2ΔB29-35	The *parABC* partition cassette with the deletion of B29-35 repeats from the *parC2* site, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study
pC2ΔB28-35	The *parABC* partition cassette with the deletion of B28-35 repeats from the *parC2* site, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study
pC2ΔB27-35	The *parABC* partition cassette with the deletion of B27-35 repeats from the *parC2* site, cloned into pFH450 vector using the	Chloramphenicol	This study

	*Eco*RV and *Xho*I restriction sites		
pC2ΔB26-35	The *parABC* partition cassette with the deletion of B26-35 repeats from the *parC2* site, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study
pC2ΔB25-35	The *parABC* partition cassette with the deletion of B25-35 repeats from the *parC2* site, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study
pC2ΔB24-35	The *parABC* partition cassette with the deletion of B24-35 repeats from the *parC2* site, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study
pC2ΔB23-35	The *parABC* partition cassette with the deletion of B23-35 repeats from the *parC2* site, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study
pC2ΔB22-35	The *parABC* partition cassette with the deletion of B22-35 repeats from the *parC2* site, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study

pC2ΔB21-35	The *parABC* partition cassette with the deletion of B21-35 repeats from the *parC2* site, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study
pC2ΔB20-35	The *parABC* partition cassette with the deletion of B20-35 repeats from the *parC2* site, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study
pC2ΔB19-35	The *parABC* partition cassette with the deletion of B19-35 repeats from the *parC2* site, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study
pCC02Δ*parC2*	The *parABC* partition cassette with the truncation of the *parC2* site, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study
pCC02-B14-B18	The *parABC* partition cassette with the deletion of R1, R2 and B1-B13 from *parC1* and B19-B35 from *parC2*, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study
pCC02ΔB18	The *parABC* partition cassette with the deletion of the B18 repeat from *parC2*,	Chloramphenicol	This study

	cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites		
pCC02ΔB18-B21	The *parABC* partition cassette with the deletion of the B18-B21 repeats from *parC2*, cloned into pFH450 vector using the *Eco*RV and *Xho*I restriction sites	Chloramphenicol	This study

2.2 Microbiological media and antibiotics

2.2.1 Media

2.2.1.1 Luria-Bertani (LB)

All *E. coli* cells used in this work were grown either in Miller LB broth (Fisher Scientific) or on LB agar (Formedium) and were supplemented with antibiotics when required. The media were prepared following the manufacturer's instructions by dissolving an adequate amount of media powder into a suitable volume of distilled water. The media were sterilized by autoclaving for 20 minutes at 121°C. Antibiotics and any nutritional supplements were added when required after the media were cooled down to ~40°C. The compositions of both media are listed in Table 2.3.

Table 2.3 Compositions of Luria-Bertani media

Composition	Concentration g/L
Tryptone	10
Yeast extract	5
Sodium Chloride	10
Agar (Solid medium only)	12

2.2.1.2 Minimal medium

The minimal medium was used for expression of the isotopically enriched recombinant ParB protein in *E. coli* cells for Nuclear Magnetic Resonance (NMR) studies. The 1X base solution was prepared by mixing an appropriate amount of each component into an adequate volume of deionized Milli-Q water (Table 2.4), followed by sterilisation in an autoclave. The isotopic labelled (U-^{15}N) NH_4Cl was used for ParB overexpression only.

All trace elements were added after the 1X base solution was cooled. The trace elements used for minimal medium preparation are listed in Table 2.5. All trace elements were filter sterilized using 0.22 µm syringe filters (Milex GP, Millipore) before being used.

Table 2.4 Compositions of 1X Base Solution

Components	**1X Base Solution (mM)**
Na_2HPO_4	43
KH_2PO_4	25
NaCl	10
(U-^{15}N) NH_4Cl	15

Table 2.5 List of trace elements used for minimal medium preparation

Trace element stock	**Volume added to 1 L 1X Base solution**
20% W/V (U-^{13}C) Glucose	10 ml
1 M $MgSO_4$	1 ml
1 M $CaCl_2$	1 ml
100 mM $MnCl_2$	1 ml
50 mM $ZnSO_4$	1 ml
100 mM $FeCl_3$	0.5 ml
100 mg/ml Ampicillin	1 ml
100X Vitamins	10 ml

2.2.2 Antibiotics

Antibiotic stock solutions were prepared by dissolving the required amount of powder in either 10 ml of 100% ethanol for chloramphenicol or in 10 ml sterile Milli-Q water for ampicillin. All antibiotics were filter sterilised using 0.22 µm syringe filters (Milex GP, Millipore) and stored at -20°C. Both stock solution and the working solution concentrations for each antibiotic are shown in Table 2.6.

Table 2.6 List of antibiotics and concentrations

Antibiotic	Stock solution mg/ml	Working solution μg/ml
Ampicillin (Amp)	100	100
Chloramphenicol (Cm)	30	30 for cloning experiments 10 for partition assays

2.3 Recombinant DNA techniques

2.3.1 Competent cells preparation

E. coli cells were streaked directly from glycerol stocks on LB medium plates and were grown overnight at 37°C. A single colony from freshly cultured cells was inoculated into 10 ml LB broth and incubated overnight at 37ºC with 250 rpm shaking. A 300 µl aliquot was taken from the overnight culture and inoculated into 60 ml LB broth in a 250 ml conical flask. The cells were grown at 37ºC with 220 rpm shaking until the Optical Density at 550 nm wavelength (OD_{550}) attained value between 0.4-0.6. When the cells reached the required growth level, the culture was incubated on ice for 10 minutes. Following incubation, the cells were harvested by centrifugation at 7,000 x g for 5 minutes at 4ºC. The supernatant was then discarded, and the cell pellets were resuspended in 20 ml of RF1 buffer (Table 2.7), equivalent to 1/3 of the initial volume (60 ml). The resuspended pellets were incubated for 1 hour on ice. The cells were pelleted following the one-hour incubation at 7,000 x g for 5 minutes at 4ºC. The supernatant was discarded, and the pellets were resuspended in 5 ml of RF2 buffer (Table 2.7), equivalent to 1/12.5 of the initial culture volume (60 ml). The cells were incubated on ice for 15 minutes. The chemically competent cells were aliquoted into 1.5 ml eppendorf tubes and flash frozen using liquid nitrogen. The cells were stored at -80°C.

Table 2.7 The compositions of RF1 and RF2 buffers

Chemical	RF1 pH 5.8	RF2 pH 6.8
Glycerol	15%	15%
RbCl	100 mM	10 mM
$MnCl_2$	50 mM	-
CH_3CO_2K	30 mM	-
CaCl2	10 mM	75 mM
MOPS	-	10 mM

Both buffers were filter sterilised using 0.22 µm Millipore and stored at 4ºC.

2.3.2 Bacterial transformation

Chemically competent *E. coli* cells were thawed on ice for 5 minutes prior to transformation. A 1 μl volume of plasmid DNA was added to a 1.5 ml Eppendorf tube along with an empty tube to use as a control. 100 μl of competent cells was added to each tube, then the mixture was incubated on ice for 40 minutes. After which, the cells were heat-shocked at 42ºC for 90 seconds. Before adding the 400 μl of sterilised liquid LB, the cells were left on ice to cool for 2 minutes. The cells were then incubated for at 37ºC for 1 hour (this amount of time is required for the bacterial cells to generate the antibiotic resistance proteins encoded on plasmid). Then an aliquot of 100 μl was spread on LB plates with a suitable antibiotic and then the plates were incubated overnight at 37ºC.

2.3.3 Small scale plasmid DNA extraction

DH5α *E. coli* cells were transformed with the plasmid to be isolated. One colony from the freshly transformed cells was picked and inoculated into 5 ml sterile LB supplemented with the appropriate antibiotic. The *E. coli* cells were grown at 37°C overnight with shaking (250 rpm). The grown *E. coli* cells were harvested at 11,000 × g for 2 minutes. The plasmid DNA was extracted using the Macherey-Nagel Nucleospin Plasmid miniprep kit, following the manufacturer's instructions. To lyse the cells, the pellets were resuspended with 250 μl of buffer A1 and then 250 μl of buffer A2 were added. Once the A2 buffer was added, the mixture was mixed by inverting 6-7 times. The cell lysate was then incubated at room temperature for 5 minutes. Following incubation, 300 μl of buffer A3 were added to the cell lysate and mixed by inverting several times (6-8 times) followed by centrifugation for 10 minutes at 11,000 × g to separate the cell debris from nucleic acids. The supernatant containing the plasmid was removed carefully and loaded onto a specific column, which contains a resin that binds DNA. Later the column was centrifuged at 11,000 × g speed for 1 minute. The flow-through solution was discarded, and the column was washed twice, first with 500 μl of washing buffer and then with 600 μl of buffer A4 (containing ethanol). The column was centrifuged between each wash step at 11,000 × g for 1 minute. An additional centrifugation step was applied at 11,000 × g for 2 minutes to dry the membrane before the plasmid sample was eluted. The plasmid was eluted by loading 30-50 μl of elution buffer (AE) to the column and was centrifuged at 11,000 × g for 1 minute. The isolated plasmid was stored at -20°C.

2.3.4 Primer design

Forward and backward primers that have been used in this study are listed in Table 2.8. The primers were used to amplify the required DNA fragment using Polymerase Chain Reaction (PCR) and were designed following standard rules. Each single primer length was between 20-40 nucleotides. Plasmid pCC02 was used as template to amplify regions of the *parABC* segregation cassette. The melting temperatures was between 45-60°C, and primers GC content was comprised between 35-70%. For cloning purposes, a restriction endonuclease site was included at the 5′ end of primers. In addition, 6 bp flanking arm was included on both 5′ end of the primers to allow the restriction enzyme to efficiently digest the DNA. All primers were manufactured by SIGMA-ALDRICH. The potential formation of secondary structures, dimers and hairpin formation were also assessed by using the Sigma primer software and the sequence was changed when necessary. For protein-DNA interaction studies, the primers were designed similarly and subjected to High-Performance Liquid Chromatography (HPLC) purification.

Table 2.8 List of primers used in this study

Primer names	**Sequences (5′ → 3′)**
ParA forward	ATGATTACTGTAGTTGGTGGGACC
ParA backward	AAGTGTGATTTCATAAGTAAT
Pb171 backward	GTAGACCTCCGTTGAGCTGA
Pcc02 forward	CAATGCGCTCATCGTCATCCT
Pcc02 backward	AGCAGCCCAGTAGTAGGTTGA
ParC1-1 Forward	CACACAGATATCTTGAGCACTATCATAGTT
ParC2 Backward	ATATATCTCGAGTCAGCTCAACGGAGGTCT
ParC1-2 forward	CACACAGATATCGTTATGAAATCACAAAAC
ParC1-3 forward	CACACAGATATCATCACAAAACACAAGTAT
ParC1-4 forward	CACACAGATATCAACACAAGTATTATTTATTAC
ParC1-5 forward	CACACAGATATCAGTATTATTTATTACATCACA
ParC1-6 forward	CACACAGATATCTTTATTACATCACATGAC
ParC1-7 forward	CACACAGATATCATCACATGACACACTTCA
ParC1-8 forward	CACACAGATATCGACACACTTCATATTTCA
ParC1-9 forward	CACACAGATATCCATCATATTTCATACTTCATA
ParC1-10 forward	CACACAGATATCTATTTCATACTTCATAAATATGAA
ParC1-11 forward	CACACAGATATCTACTTCATAAATATGAATACA
ParC1-12 forward	CACACAGATATCAATATGAATACATATTGACAC

ParC1-13 forward	CACACAGATATCATACATATTGACACCTCA
ParC1-14 forward	CACACAGATATCTTGACACCTCAGGTTACC
ParC2-1 backward	ATATATCTCGAGTTCATAAATCATAATTATGAA
ParB3-K3A forward	CCCATGGTGGCGAAACCCAGC
ParB2-K3A backward	GCTGGGTTTCGCCACCATGGG
ParC1-Btn forward	GCAGTATACGTTCATCTATAGCCC
ParC1-Btn backward	ACCAACTACAGTAATCATTGCAAC
Pcc02-EcoRV	CACACAGATATCTTGCTAACTTCATCATACGTG
Pcc02-delta parC2	ATATATCTCGAGACGTTAGTGTTTTTAAGTATT
ParC2-35 backward	ATATATCTCGAGTTCATAAATCATAATTATGAA
ParC1-B1-B6-Btn forward	ATCATAGTTATGAAATCACAAAACACAAGTATTATTTATTAC
ParC1-B1-B6-backward	GTAATAAATAATACTTGTGTTTTGTGATTTCATAACTATGAT
Parc2-34 backward	ATATATCTCGAGATCATAATTATGAATTGTGAT
ParC2-33 backward	ATATATCTCGAGTTATGAATTGTGATTTATTCT
ParC2-32 backward	ATATATCTCGAGTTGTGATTTATTCTGTAAAAA
ParC2-31 backward	ATATATCTCGAGTTATTCTGTAAAAAAAGAGAC
ParC2-30 backward	ATATATCTCGAGGTAAAAAAAGAGACCACTGCA
ParC2-29 backward	ATATATCTCGAGAAGAGACCACTGCAATATGTG
ParB-K4A forward	CCCATGGTGAAGGCACCCAGCCAG
ParB-K4A backward	TTGCTGGCTGGGTGCCTTCACCAT
ParB-PET forward	CACACACATATGGTGAAGAAACCCAGCCAGCAA
ParB-PET backward	CACACACTCGAGCGTTAGTGTTTTTAAGTATTG
ParC1-Cy5 forward	GCAGTATACGTTCATCTATAGCCC

ParB-K3A-PET forward	CACACACATATGGTGGCGAAACCCAGC
ParC2-Btn forward	CGTAATAATCAATAAGTTATT
ParC2-28 backward	ATATATCTCGAGATCTCTTGTATGCAAGGGTGC
ParC2-27 backward	ATATATCTCGAGGTATGCAAGGGTGCTTAAACA
ParC2-26 backward	ATATATCTCGAGAGGGTGCTTAAACAGTATGAA
ParC2-Cy5 forward	GTAATAATCAATAAGTTATTA
ParC2-22 backward	ATATATCTCGAGTTCATAAGTAATAACTTATTG
ParC2-20 backward	ATATATCTCGAGCTTATTGATTATTACGTTAGT
ParC2-25 backward	ATATATCTCGAGGTATGAATTCACAAGTGTG
ParC2-24 backward	ATATATCTCGAGTTCACAAGTGTGATTTCATAA
ParC2-23 backward	ATATATCTCGAGGTGTGATTTCATAAGTAATAA
ParC2-21 backward	ATATATCTCGAGGTAATAACTTATTGATTATTA
ParC2-19 backward	ATATATCTCGAGATTATTACGTTAGTGTTTTTA
ParC2-18 forward	ATATATCTCGAGCAATAAGTTATTACTTAT
ParC2-delta B18-B21 forward	ATATATCTCGAGATCACACTTGTGAATTCATAC
ParC2-Btn backward	AGTTTTCGATGCCTGAGTCCG
ParC1-BamHI backward	ATATATGGATCCTGCAACCTCATAAATGTGATG
ParC2ΔB18-SacI forward	ATATATGAGCTCTCAATAAGTTATTACTTATGA
ParC2ΔB18-B21-SacI forward	ATATATGAGCTCAATCACACTTGTGAATTCATA

ParBΔ30-PET forward	CACACACATATGCCATACGGCGCGCCGGAAAAA
ParB-SacI backward	ATATATGAGCTCTGACGTTAGTGTTTTTAAGTATTG
ParB forward	ATGGTGAAGAAACCCAGCCAG
B18 forward	GCGCGCTAATAAGGGCCC
B18 backward	GGGCCCTTATTAGCGCGC
ParB-R12A forward	GCATTAAACGCAGCTGCAGTT
ParB-R12A backward	AACTGCAGCTGCGTTTAATGC
ParB-N11A forward	CAAGCATTAGCCAGAGCTGCAGTT
ParB-N11A backward	AACTGCAGCTCTGGCTAATGCTTG
ParB-R26A forward	CTGGCTCAGGCCCTTGCGGAT
ParB-R26A backward	ATCCGCAAGGGCCTGAGCCAG
ParABC-SacI backward	ATATATGAGCTCTCAGCTCAACGGAGGTCTACA

2.3.5 Polymerase chain reaction (PCR)

Amplification of specific DNA fragments was carried out using Polymerase Chain Reaction (PCR). The master mix was prepared on ice as per Table 2.9 with a final volume of 60 µl. A 5 mM stock of the deoxynucleotide triphosphates (dATP, dTTP, dCTP and dGTP) was prepared separately from a 100 mM stock (Roche) using sterile Milli-Q water. Primers were designed to amplify the required region as described in Section 2.3.4 and synthesised by SIGMA-ALDRICH. The lyophilized primers were resuspended by adding 1 ml of Milli-Q water to give the concentration that the manufacture provided and then a working stock of 5 (pmol/µl) was prepared. The DNA template (pCC02) was diluted 1:10 (to achieve a concentration equal to ~1-10 ng/µl) before being added to the reaction. The PCR programme used for each DNA fragment is shown in Table 2.10. The annealing temperature was changed based on the GC content of the forward and backward primers. A fixed extension temperature of 72ºC was used and the time was

altered based on the size of the amplified DNA (on average 1 minute per 1 kb). The number of PCR cycles was varied between 25 and 35 cycles.

Table 2.9 PCR master mix

Component	**Volume (µl)**
10X PCR buffer	6
DNA template	1 (~1-10 ng/µl)
dNTPs (dATP, dTTP, dCTP and dGTP) 5 mM	2.4 each
Forward primer 5 pmol/µl	3
Reverse primer 5 pmol/µl	3
Sterile Milli-Q water	36.4
DNA *Taq* polymerase (Promega)	1 (0.1 U/µl)
Total reaction volume	**60**

Table 2.10 Thermocycler program used for amplifying the DNA fragments

Step	**Thermocycler**	**Temperature (°C)**	**Time (minute)**
1	Initial denaturation	94	3
2	Denaturation	94	1
3	Annealing	45-60 (Depending on the melting temperature (Tm) of the primers)	1
4	Extension	72	0.5-2 (depending on the size of the DNA fragment)
5	Repeat step (2-4) 25-35 times		
6	Final extension	72	6
7	Hold	10	Unlimited (until used)

2.3.6 Restriction endonuclease digestion

Restriction enzyme digestions were used to verify the plasmid constructs, to linearize vectors as well as to generate the required DNA fragments for cloning procedure. The

reaction mixture was prepared as shown in Table 2.11. The DNA was digested in a buffer supplied by the enzyme manufacturers. All the restriction enzyme digestion reactions used in this study were incubated at 37°C for 2 hours and inactivated at 65°C or 80°C for 20 minutes, if it was not to be used on the same day. Double digests were performed in either 20 μl for confirming the positive clones and to linearize vectors for cloning, or in 60 μl for digestion of amplified DNA. In case of double digestion, a buffer suitable for both enzymes was used. The reaction mixture set up for different volumes is shown below (Table 2.11). The reaction mixture was then purified when required using either ethanol precipitation (Section 2.3.7) or a DNA clean up kit. Restriction enzymes were purchased from different sources such as Promega, New England Biolabs and Thermo Fisher Scientific.

Table 2.11 Reaction mixture of restriction enzymes digestions

Component	**Single digestion**	**Double digestion (cloning)**	**Double digestion (PCR)**
Restriction enzyme 1	1 μl (2.5 U)	1 μl (2.5 U)	2.5 μl (2.5 U)
Restriction enzyme 2	-	1 μl (2.5 U)	2.5 μl (2.5 U)
10X buffer	2 μl	2 μl	6 μl
DNA	15 μl (~ 1-3 μg)	15 μl (~ 1-3 μg)	30 μl (~ 1-5 μg)
Milli-Q water	2 μl	1 μl	19 μl
Total volume	**20 μl**	**20 μl**	**60 μl**

2.3.7 DNA precipitation (ethanol precipitation)

DNA precipitation was used either to purify DNA from other contaminants or to concentrate the DNA sample after PCR, restriction enzyme digestion and annealing oligonucleotides. The volume of DNA sample was adjusted to 200 μl with Milli-Q water. A 1/10 (20 μl) volume of 3 M sodium acetate, pH 5 and 2 volumes (440 μl) of cooled absolute ethanol were added to the DNA sample. The mixture was incubated at low temperature, either overnight at -20ºC or 1-2 hours at -80ºC. The DNA precipitated during the incubation; this precipitate was collected by centrifugation at 12,000 × g at 4ºC for 30 minutes. The supernatant was removed carefully, and the pellets were suspended gently with 500 μl of 70% cooled ethanol. To pellet the DNA, the suspension was centrifuged at 12000 × g for 20 minutes. The supernatant was removed carefully, and the pellet air-dried to remove the remaining traces of ethanol either by incubating for 20-30 minutes

at room temperature or 5-10 minutes at 37°C. The pellet was then resuspended in 20-50 µl of Milli-Q water or TE buffer. The DNA sample was stored at -20°C until use.

2.3.8 Alkaline phosphatase treatment (DNA dephosphorylation)

Alkaline phosphatase was used to prevent re-ligation of the digested plasmid vectors. The linearized plasmid was subjected to alkaline phosphatase treatment by removing the phosphate group from the 5′-end prior to ligation. A 10 µl aliquot of 10X alkaline phosphatase buffer was added to the 20 µl reaction of digested plasmid and 2 µl of alkaline phosphatase enzyme (5U/µl) were added to the mixture. The volume of the reaction was adjusted with sterile Milli-Q water to 100 µl. The reaction mixture was incubated at 37°C for 30 minutes. Another 2 µl of phosphatase enzyme (5U/µl) were added to the reaction mixture, and a further incubation was carried out at 37°C for 30 minutes. Finally, following incubation with phosphatase enzyme, 10 µl of 0.5 M ethylene glycol-bis (β-aminoethyl ether)-N,N,N',N'-tetraacetic acid (EGTA) were added to the mixture and incubated for 10 minutes at 75°C. To remove enzymes and buffer, the linearized plasmid was cleaned up using a Macherey-Nagel PCR clean-up kit, following the manufacturer instructions. 200 µl of NTI buffer, which is 2X the reaction volume, was added to the 100 µl reaction and mixed thoroughly. The mixture was loaded onto the column that is provided with the kit and centrifuged for 1 minute at 11,000 × g. The silica membrane was washed twice with 700 µl of NT3 buffer and centrifuged for 1 minute at 11,000 × g. An additional centrifugation step was carried out for 2 minutes at 11,000 × g to remove any remaining buffer residue from the column and to dry the column before the DNA sample was eluted. The DNA sample was eluted by adding 30 µl of NE buffer to the dry column and was centrifuged for 1 minute at 11,000 × g. To avoid any contamination, the column with the DNA sample was air-dried for 5 minutes at room temperature before the elution step was carried out. Finally, the DNA sample was stored at -20°C and was used when required.

2.3.9 DNA ligation

For cloning experiments, the vector backbone and the desired insert were ligated using T4 DNA ligase (New England Biolabs). For a standard reaction, one molar ratio of dephosphorylated vector to 10-15 molar ratios of the insert was used. DNA concentration was quantified using either Nanodrop 1000 (Thermo Fisher Scientific) or Qubit 3.0 Fluorometer (Invitrogen/Thermo Fisher Scientific). The ligation reaction was performed in 30 µl total volume, in which 3 µl of 10X T4 DNA ligase buffer and 1 µl (2.5 U) of T4

DNA ligase (NEB) were added to an appropriate volume of DNA insert and vector. The reaction volume was adjusted to the final volume with sterile Milli-Q water. In parallel to the cloning reaction, another reaction was prepared which contained only the digested (linearized) vector to test whether the plasmid has been efficiently digested and dephosphorylated (i.e. if the plasmid vector that have been digested and dephosphorylated is able to re-ligate again with itself). This reaction mixture was prepared as described above but without the DNA insert. Both reactions were incubated at room temperature for 4 hours followed by transformation into DH5α *E. coli* competent cells as described in Section 2.3.2. For transformation, two negative controls were used along with the two ligated reactions; the circular plasmid (uncut) was used to test the efficiency of the transformation and competent cells, and *E. coli* competent cells alone were used for testing the sterilisation conditions.

2.3.10 Screening of recombinant plasmid clones by colony PCR

To screen for positive plasmid clones, colony PCR was performed followed by restriction digestion (Section 2.3.6). A PCR reaction was prepared for each plasmid clone following the same protocol as detailed previously (Section 2.3.5). Bacterial colonies were used as a template for PCR reactions. After the components of the PCR reaction were added, a single *E. coli* colony was taken off from the transformation plate with a sterile pipette tip and streaked on LB plate first and then resuspended in the PCR mix. For each individual clone, the same forward and backward primers used to amplify the insert were added to the PCR reaction. The final mixture volume was 20 µl. The colony PCR product was checked on a 1% (w/v) agarose gel (Section 2.3.11) to verify that the DNA insert had been cloned. For further verification, the recombinant plasmid was extracted as described in Section 2.3.3 and digested with the same enzymes that had been used for the cloning, as described in Section 2.3.6. The sequences of the recombinant plasmids were confirmed by Sanger sequencing.

2.3.11 Agarose gel electrophoresis

Plasmid minipreparations, DNA digestion products and all PCR products were verified by agarose gel electrophoresis. Different concentrations of agarose gel ranging within 0.8-2% (w/v) were used based on the size of the DNA fragment. The gel was prepared by dissolving the agarose powder in 50 ml of 1X TAE [40 mM Tris-HCl, 20 mM acetic acid, 1 mM Ethylenediaminetetraacetic acid (EDTA)]. The melted agarose was allowed to cool down to ~50ºC. Before pouring the agarose into the cast, an aliquot of 0.05% (v/v) SYBR Safe DNA gel dye (Invitrogen/Thermo Fisher Scientific) was added to the

agarose solution to stain the DNA. Once the gel was set, 6X loading buffer (New England Biolabs or Thermo Fisher Scientific) was mixed with the DNA sample at 1X final concentration before the mixture was loaded into the well. Together with the DNA sample, an appropriate molecular weight DNA marker was run to estimate the size of DNA fragments. The molecular weight markers were either GeneRuler 1 kb DNA ladder (Thermo Fisher Scientific and NEB) or PCR marker (New England Biolabs). The 6X loading dye contains coloured dyes that are used to monitor the progress of the electrophoresis. 1X TAE buffer was used to run the gel. For standard agarose gels, 5 to 8 V/cm was used. Once the dye reached ¾ of the gel, the electrophoresis was stopped, and the DNA visualised using either a BIORAD Gel doc EZ imager or an ultraviolet (UV) transilluminator. The agarose gel images were taken using Image lab 4.0.1 software. All the images were converted to grey scale in which the saturated bands (the more DNA concentration) were transferred from red colour to dark colour outlined with a white colour.

2.3.12 Extraction of DNA from agarose gel

PCR products, annealed oligonucleotides and digested plasmids were subjected to a purification step prior any further experiments. The DNA fragment was separated using a 1-2 % of (w/v) TAE agarose gel (Section 2.3.11). The DNA fragment was visualized using a UV transilluminator after SYBR safe staining. The DNA band of interest was cut using a sterile scalpel and transferred into a pre-weighed 1.5 ml Eppendorf tube. The DNA was isolated using the Macherey-Nagel PCR clean-up and gel extraction kit following the manufacturer instructions. The gel slice was weighed and two volumes of NT1 buffer were added to the slice. The mixture was incubated at 50°C for 10 minutes and was vortexed every 2 minutes until the gel was dissolved completely. The solution was then transferred to the NucleoSpin column and centrifuged at high speed 11,000 × g for 1 minute. Next the flow-through was discarded from the collection tube and the DNA was washed with 700 µl of NT2 buffer that contains ethanol and centrifuged at 11,000 × g for 1 minute. The previous step was repeated twice and the flow-through was discarded. The column was centrifuged for an additional 2 minutes to eliminate any residual ethanol. The NucleoSpin column was later placed into a new clean 1.5 ml Eppendorf tube and the DNA was eluted by adding 50 µl of elution (TE) buffer. The tube was incubated at room temperature for 1 minute before it was eluted with 30 µl elution buffer using 11,000 × g speed for 1 minute. The DNA sample in the flow-through was verified by running 2 µl of the sample in 1% (w/v) agarose gel and was visualized using a BIORAD Gel doc EZ imager. The sample was then stored at -20°C until use.

2.3.13 Polyacrylamide gel electrophoresis (PAGE)

2.3.13.1 Polyacrylamide gel preparation

A polyacrylamide native gel was used to extract the small DNA fragments spanning 16-40 bp such as annealed oligonucleotides. It was also used to distinguish the single-stranded DNA from the double-strand DNA (annealed DNA fragments). Two concentration of acrylamide gel were used for the gels, 6% for large DNA fragments between 30-40 bp and 12% for small size of DNA fragments ranging from 10 to 20 bp. The native gel was prepared using the components listed in the Table 2.12. The last two components, ammonium persulfate (APS) and N,N,N'N'-tetramethyl-ethylenediamine (TEMED) were added last, immediately before pouring the gel to avoid polymerisation. All the components were added to a 50 ml falcon tube and mixed gently to avoid any air bubbles, before the solution was poured into the gel casting apparatus (Bio-Rad), between two plates, filled up to the top of the cassette. A comb was placed carefully from the top to avoid any air bubbles forming inside the gel. The gel was allowed to polymerise for 50-60 minutes. Once the gel solidified, the comb was removed carefully, and the gel was transferred immediately to the tank (Bio-Rad). The tank was filled up with 1 X TBE buffer prepared from 10X TBE buffer (Table 2.13). 5 μl of DNA sample and 1 μl 6X loading dye were mixed before loading into the well. 2 μl of PCR molecular weight marker were loaded next to the sample to estimate the size of the investigated DNA fragments. The gel was subjected to electrophoresis for 60 minutes at 100 V. The gel was later stained with SYBR Safe dye (5 μl of dye in 50 ml distilled water) for approximately 30 minutes. The DNA bands were then visualized using a BIORAD Gel doc EZ imager. Gel images were taken using Image lab 4.0.1 software.

Table 2.12 Components of 12% polyacrylamide gel

Component	Volume added
30% acrylamide	4.8 ml
Milli-Q water	6 ml
10X TBE pH 8.3	1.2 ml
10% ammonium persulfate	200 μl
TEMED	10 μl

Table 2.13 Compositions of 10X TBE buffer

Components	Quantity
Tris-base	108 g

Boric acid	55 g
0.5 M EDTA, pH 8.0	40 ml
Milli-Q water	To make 1 L volume

2.3.13.2 Extraction of DNA from acrylamide gel

To excise small DNA fragments obtained from annealing oligonucleotides and to separate double-stranded DNA from the single-strand DNA, acrylamide gel electrophoresis was performed. Based on the fragment size, either 6% or 12% acrylamide gels were used. 5 µl of the DNA fragment was mixed with 1 µl of 6X loading dye and the sample was loaded along with the PCR markers (NEB). The native gel was run for approximately one hour at 100 volts. The gel was later stained with SYBR Safe (5 µl of dye in 50 ml distilled water) for 30 minutes. The DNA bands were then visualized on a UV transilluminator and the DNA band of interest was cut using a clean sharp scalpel. The excised gel slice was put in 500 µl of 1X TBE buffer. 5 cm of SnakeSkin dialysis tubing (Thermo Scientific) was cut and immersed into 1X TBE buffer for several minutes. The mixture (1X TBE buffer with gel slice) was introduced into the dialysis tube. The dialysis tube ends were closed, and the tube was inserted into a gel electrophoresis tank and subjected to electrophoresis for 30 minutes at 100 volts. Once the run was finished, 500 µl of TBE buffer containing the DNA sample were collected. The sample was cleaned up using ethanol precipitation as mentioned in Section 2.3.7.

2.3.14 Annealing oligonucleotides

For NMR and X-ray crystallography studies, two single-stranded oligonucleotides ranging between 17 and 33 nucleotides were used. The two single-stranded oligonucleotides with complementary sequences were annealed. First, both oligos were resuspended into 1X annealing buffer (10 mM Tris-HCl, pH 8.0, 50 mM NaCl and 1 mM EDTA) at 100 µM final concentration. Equal volume of equimolar oligonucleotides was mixed in an Eppendorf tube. The mixture was incubated for 7 minutes at 95°C followed by cooling at room temperature for approximately one hour. The double-stranded oligonucleotides was run along with the single-stranded one on a 12% acrylamide gel. The annealed oligonucleotides were cleaned-up using ethanol precipitation (Section 2.3.7) and then resuspended into the required buffer.

2.3.15 DNA sequencing

Sanger sequencing was performed by GATC Biotech (currently Eurofins Genomics), Ebersberg, Germany. The reaction was prepared following the company instructions, by mixing 5 μl plasmid (80-100 ng/μl) and 5 μl of primer (5 pmol/μl). The list of primers used is given in Table 2.8. DNA sequences were analysed using Chromas Software (Technelysium).

2.3.16 Site-directed mutagenesis using the overlapping PCR method

Site-directed mutagenesis is used to generate mutants either by substitution, addition or deletion of a specific base in DNA sequences. PCR based method site-directed mutagenesis was used to create specific mutations in the *parB* sequence. The K3, K4, N11, R12 and R26 codons in the recombinant *parB* sequence were replaced by alanine codons.

Overlap extension PCR requires three PCR steps and four primers (Figure 2.1). The mutation is introduced by mutagenic primers that contain one or more mismatched bases. The method involves using four primers, two flanking external primers and two internal primers. The external primers, A and D, were designed to contain suitable restriction sites at the 5′ ends, *Eco*RV and *Xho*I, respectively. The two restriction sites are flanked by 6 bp AT- or CG-rich arms for restriction digestion purposes. These two primers were complementary to the ends of the target DNA sequence. The two internal complementary primers, B and C, harboured the site of the desired mutation contained the mutation. The mutation is located in the middle of the primers. The two primers, B and C, are complementary to each other. The first fragment obtained with PCR1 was amplified using A and B primers, while the second fragment obtained with PCR2 was generated using C and D primers. The reaction mixtures for PCR1 and PCR2 were carried out as described in Section 2.3.5. The PCR programme was designed as described in Section 2.3.5. The pCC02 plasmid was used as DNA template.

The size of PCR1 and PCR2 products was verified by running the samples on 1% agarose gel. The gel extraction for PCR1 and PCR2 products was carried out as described in Section 2.3.12. After gel extraction, PCR1 and PCR2 products were mixed at equal molar ratio and a pre-cycle reaction was performed to anneal the overlapping sequences. The pre-cycle reaction mixture was prepared by mixing equal molar ratio of PCR1 and PCR2, 2.4 μl of each of 5 mM dNTPs, 6 μl of 10X *Pfu* polymerase buffer and sterile Milli-Q water to make volume of 50 μl. The pre-cycle programme consists of a

denaturation step at 94°C for 3 minutes followed by 20 cycles of 40 seconds at 94°C and 40 seconds at 72°C. Before the initial denaturation step was completed, the programme was paused and 1 μl of *Pfu* polymerase (0.1 U/μl) was added. Once the pre-cycle was completed 5 μl of each flanking primers (5 pmol/μl) were added and the PCR3 programme was started as described in Section 2.3.5. The PCR3 product was digested using *Eco*RV and *Xho*I enzymes. The pFH450 vector was also cut using the same enzymes. Both the PCR3 product and the cut pFH450 vector were run on a 1% agarose gel and the bands were extracted using a Macherey-Nagel PCR and gel extraction kit. A ligation of the PCR insert and vector was carried out and the nucleotide sequences and the mutation in the recombinant plasmid was verified by Sanger sequencing by GATC biotech (Section 2.3.15).

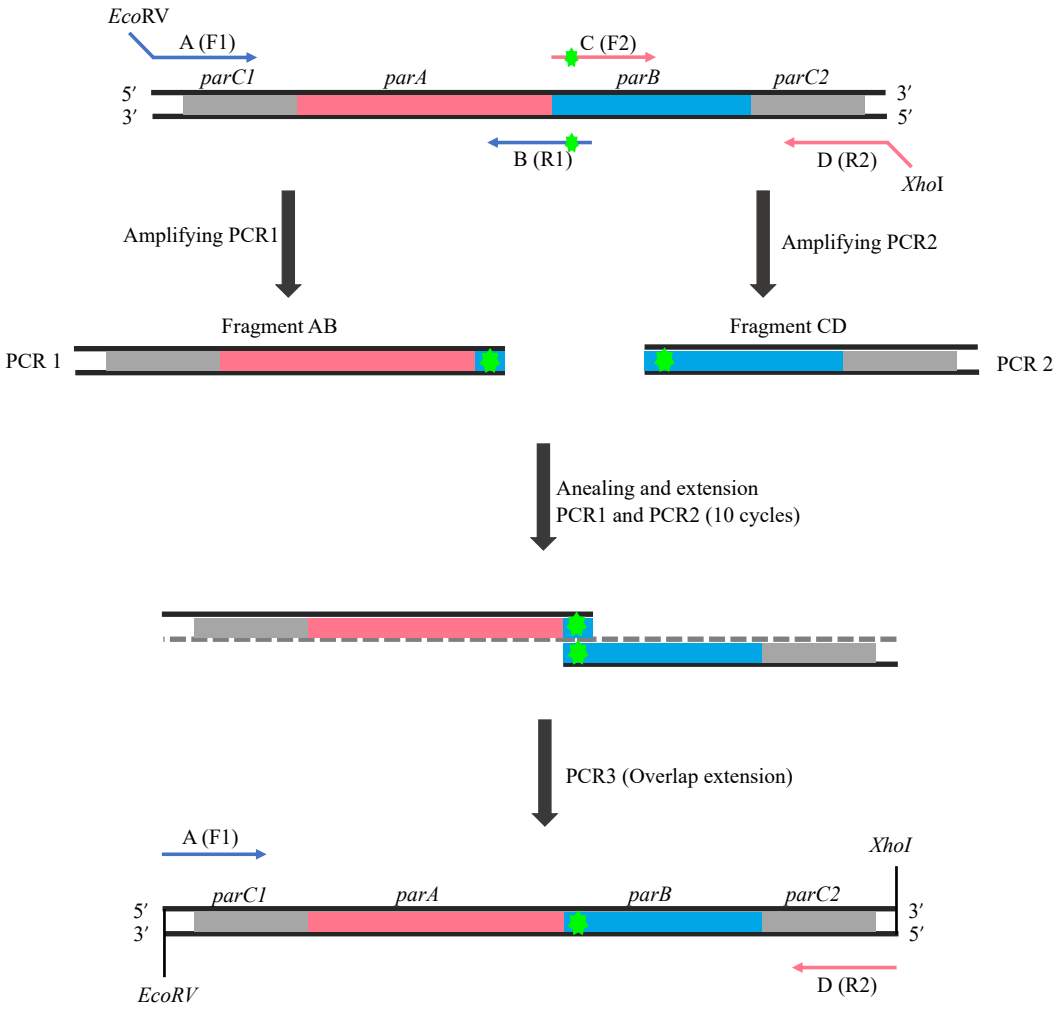


Figure 2.1 Diagram illustrating site-directed mutagenesis coupled with overlap extension PCR

The diagram shows the steps to generate *parABC* cassette encoding ParB mutants. In the first step, two PCR fragments are produced by using four primers, two externals primers containing the site for *Eco*RV (blue arrow) and *Xho*I (red arrow) restriction enzymes and two internal primers (forward red arrow and reverse blue arrow) that harbour the mutation. PCR1 is generated using primers A (F1) and B (R1) (blue arrows), while PCR2 is generated using primers C (F2) and D (R2) (red arrows). In the next step, the two fragments (AB and DC) are incubated together and allowed to pair using a pre-cycles reaction consisting of annealing followed by extension steps. The last step is the overlap extension PCR to generate the full *parABC* cassette encoding the required mutant.

2.4 Protein production methods and techniques

2.4.1 Protein overexpression in LB medium

All the $(His)_6$-tagged recombinant proteins, ParB, ParBΔ30N and ParA were overexpressed using the same protocol. All the recombinant plasmids were transformed into BL21 (DE3) *E. coli* cells. 5-7 colonies were picked with a sterile loop from the transformation plate and were inoculated into 10 ml LB medium supplemented with ampicillin (100 μg/ml). The cells were grown for two hours at 37ºC. After two hours, the culture was transferred into either 50 ml for small scale overexpression tests or 200- 500-ml LB medium for medium scale protein production. The culture was grown for further until the OD_{600} reached a value between 0.6-0.7. A 1 ml aliquot of the culture was collected and centrifuged at 11,000 x g for 1 minute. The supernatant was discarded, and the pellet was resuspended in 100 µl of 1X binding buffer (see composition in Table 2.14). The suspension was used as uninduced control. The remaining culture was induced with 1 mM Isopropyl β-D-1-thiogalactopyranoside (IPTG) and was grown for over three hours at 37ºC for the ParB proteins and at 30ºC for the ParA protein. For overexpression tests, 100 μl was collected at each hour post induction with IPTG in a 1.5 ml Eppendorf tube and centrifuged at 11,000 x g for 1 minute. The supernatant was discarded, and the pellets was resuspended in 20 µl of binding buffer. The level of protein overexpression was assessed by Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis (SDS-PAGE). The pellets for protein purification were stored at -20°C.

2.4.2 Solubility assay

To assess the solubility of the recombinant proteins, the pET-parB and pET-parA plasmids were transformed into BL21 (DE3) *E. coli* cells and the transformants were incubated overnight at 37ºC. 3 colonies from the transformation were picked with a sterile loop and inoculated into 10 ml LB supplemented with ampicillin (100 µg/ml). The bacterial cells were grown for 2-3 hours at 37ºC till the OD_{600} reached a value between 0.65-0.75. A 100 µl aliquot was collected and centrifuged at high speed (13,000 x g) for 1 minute. The supernatant was discarded, and the pellets was used as uninduced sample. The remaining culture was induced with 1 mM IPTG. After one hour of induction, 1 ml was collected from the culture and centrifuged at 13,000 x g for 1 minute. The pellet was resuspended with 1 ml of 1X binding buffer (Table 2.14). The crude extract was sonicated 6 times with 5 seconds on and 1 minute off. 20 µl from the crude extract was collected and used as a total sample. The remaining was pelleted at 11,000 x g for 1 minute. The supernatant was collected and used as soluble fraction and the pellet was resuspended in 1 ml of 1X binding buffer (Table 2.14) and used as insoluble fraction. The previous

steps were repeated for samples collected after two and three hours of induction. The solubility of the protein was assessed by SDS-PAGE (Section 2.4.8).

2.4.3 Small scale overexpression of ParB in minimal medium

To test the expression level of the $(His)_6$-tagged ParB protein in the isotopic labelled media, the pET-ParB plasmid was transformed into BL21(DE3) cells. Few colonies (3-4) were taken from the fresh transformation and inoculated into 10 ml LB broth supplemented with ampicillin (100 μg/ml). The cells were incubated for 6-8 hours on a 200-rpm shaker at 37°C. The OD_{600} of the culture was recorded after the incubation time and it was ~1.5. An aliquot of the culture was pelleted at high speed (14,000 x g) for 5 minutes. The pellet was then resuspended into 10 ml of minimal medium (Section 2.2.1.2) with the 100 μg/ml ampicillin to obtain a culture with an OD_{600} =0.05. The culture was grown overnight on a 200-rpm shaker at 37°C. The overnight culture should grow between OD_{600}1 and 1.5. An aliquot of the cells was centrifuged at 14,000 x g for 5 minutes. The cells were then resuspended into 50 ml of fresh minimal medium (250 ml flask) to achieve an OD_{600} = 0.05. The cells were grown on a 200-rpm shaker at 37 °C till the OD_{600} reached a value between 0.8-1. A 100 µl aliquot of the culture was collected and pelleted at 14,000 x g for 1 minute. This was used as the uninduced control. The culture was then induced with 1 mM IPTG and incubated for 4 hours on a 120-rpm shaker at 37°C. A 100 µl aliquot was taken from the culture and centrifuged at 14,000 x g for 1 minute. This is the induced sample. The overexpression level of ParB was tested by resuspending both samples into 20 µl of 1X binding buffer (Table 2.14), and running them on a SDS-polyacrylamide gel.

2.4.4 Large scale overexpression of ParB in minimal media

To produce $(His)_6$-tagged ParB and ParBΔ30N in isotopic label media, the pET-ParB and pET-ParBΔ30 plasmids were transformed into BL21(DE3) cells. 5-7 colonies were taken from the transformation and were inoculated into 10 ml LB broth supplemented with 100 μg/ml ampicillin. The cells were incubated for two hours on a 200 rpm shaker at 37°C. After incubation, the culture was transferred into 500 ml LB broth supplemented with 100 μg/ml ampicillin (2 L flask) and was grown with shaking at 120 rpm at 37°C for 3-4 hours until the OD_{600} reached a value around 0.7. The cells were harvested at 8000 x g for 5 minutes. The supernatant was discarded, and the pellets was washed twice with M9 salt (Table 2.4) and was centrifuged at 8,000 x g for 5 minutes. The washed pellet was resuspended in 1 L isotopic label M9 minimal medium (Table 2.5) at 37°C with 1 g/L $^{15}NH4Cl$ (Cambridge Isotopic Laboratories) and 2 g/L ^{13}C glucose (Cambridge Isotopic

Laboratories) as a source of nitrogen and carbon, respectively. The cells were grown for an hour, which is the time required for the bacterial cells to adapt to the new medium and during that period the OD_{600} reached 0.8. Subsequently, the culture was induced with 1 mM IPTG and was incubated either for 17 hours at 20°C or three hours at 37°C, on a 120-rpm shaker. A 1 ml sample was collected in a 1.5 ml Eppendorf tube before and after induction of the culture with IPTG. The samples were centrifuged for 1 minute at 13,000 x g, the supernatants were removed, and the pellet was resuspended with 100 μl of 1X binding buffer (Table 2.14). 15 μl of the samples was mixed with 15 μl of 2X SDS-loading dye and were loaded onto a 15% SDS-polyacrylamide gel. The gel was run along with Unstained PageRuler protein ladder (Thermo Scientific) for 80 minutes at 150-190 V. The 500 ml cell culture was harvested at 14,000 x g for 40 minutes. The pellets were stored at -20ºC.

2.4.5 Protein purification

The ParB and ParA proteins were purified using Ni^{+2} affinity chromatography (GE Healthcare Peristaltic Pump). The pellets from the overexpressed culture were resuspended with 15 ml of 1X binding buffer (Table 2.14) for 200 ml culture and 35 ml of 1X binding buffer for 500 ml culture. 10 µl/ml of 10 mg/ml lysozyme was added, one and two tablets of protease inhibitor cocktail (cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail, Roche) for 200 ml and 500 ml starting culture, respectively were added. The tablet was mashed and dissolved inside the resuspension. The lysate was incubated at 30ºC for 15 minutes. After incubation another aliquot of 10 µl/ml of 10 mg/ml lysozyme was added and the incubation continued for 15 minutes at 30 ºC. To completely lyse the cells resuspended in the 15 ml volume, the lysate was sonicated on ice for 12 times alternating 15 seconds at 60% sonicator power and with 1-minute pause intervals using a micro-tip. For the 35 ml lysate, the cells were sonicated 6 times alternating 7 seconds at 70% sonicator power and with1-minute pause intervals using a macro-tip. The cell lysate was centrifuged at 10,000 x g for 40 minutes. The supernatant was removed carefully and filtered through a 0.45 μm Milipore filter before starting the purification. A 6 ml aliquot of His-bind resin (Merck) was packed into a column and washed 5 columns volume (CVs) of filtered, sterile Milli-Q water. The resin beads were charged with 50 mM $Ni^{+2}SO_4$ and equilibrated with 5 CVs of 1X binding buffer. The cell lysate was flowed over the column and re-circulated through the column for ~90 minutes to allow the protein to bind to the column. Unbound proteins were washed first with 5 CVs of 1X binding buffer then with 10 CVs of 1X wash buffer (Table 2.14). The proteins were eluted with 1X elution buffer (Table 2.14) and 12 x 1 ml fractions were collected. The protein concentration was

estimated by using the Bradford assay (Section 2.4.7). The proteins were buffer exchanged immediately into storage buffer (Table 2.14) as described in Section 2.4.6.1.

Table 2.14 Protein purification buffers

Buffer	ParB proteins	ParA protein
1X Binding buffer	20 mM Tris-HCl, pH 7.5, 500 mM NaCl, 15 mM Imidazole and 10% glycerol	50 mM Tris-HCl, pH 7.5, 500 mM NaCl, 15 mM Imidazole and 10% glycerol
1X Wash buffer	20 mM Tris-HCl, pH 7.5, 1 M NaCl, 80 mM Imidazole and 10% glycerol	50 mM Tris-HCl, pH 8.0, 1 M NaCl, 75 mM Imidazole and 10% glycerol
1X Elution buffer	20 mM Tris-HCl, pH 7.5, 500 mM NaCl, 400 mM Imidazole and 10% glycerol	50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 300 mM Imidazole and 10% glycerol
Storage buffer	50 mM HEPES pH 8.0, 50 mM KCl, 10% glycerol and 2 mM DTT	30 mM Tris-HCl, pH 8.0, 100 mM KCl, 10% glycerol and 2 mM DTT

2.4.6 Buffer exchange

2.4.6.1 Buffer exchange using 5 ml Hi-Trap Desalting columns

ParB and ParA proteins were buffer-exchanged into storage buffer using 5 ml Hi-Trap desalting columns (GE Healthcare). The column is prepacked with Sephadex G-25 superfine resin. The 5 ml column was first washed with 5 CVs of sterile Milli-Q water before it was equilibrated with 5 ml CVs of storage buffer. 1.5 ml of purified protein was loaded through the column and the protein recovered by loading twice of 1 ml of storage buffer. The previous step was repeated for all fractions. The column was then stripped using 4 CVs of strip buffer (20 mM Sodium Phosphate, 0.5 M NaCl and 50 mM EDTA, pH 7.4) followed by washing with 4 CVs of sterile Milli-Q water. The column was stored with 20% ethanol and was kept at 4 ºC for further use. After buffer exchange, the concentration of protein fractions was assessed by Bradford assay (Section 2.4.7). The protein sample was aliquoted into 200 μl volumes, flash-frozen in liquid nitrogen and stored at -80°C.

2.4.6.2 Buffer exchange using 500 µl Zeba™Spin Desalting column

To buffer exchange a small volume of protein (70-200 µl) into the required buffer before it was used for any further experiments, a 7,000 MWCO 500 µl volume Zeba™Spin Desalting column (Thermo Scientific) was used. The column was placed into a 2 ml Eppendorf collection tube after the bottom closure was removed and the cap lid was loosened. The storage buffer was removed by centrifuging for 1 minute at 1,500 x g. The column was equilibrated by adding 300 µl of equilibration buffer and was centrifuged for 1 minute at 1,500 x g. The flow-through was discarded and the column was placed back to the collection tube. This step was repeated 3-4 times. The column was placed into a clean 1.5 ml collection tube and the protein sample (ideally 70-150 µl) was applied slowly to the middle of the prepacked column. The protein sample was then eluted at 1,500 x g for 2 minutes.

2.4.7 Estimation of protein concentration by Bradford Assay

To estimate the concentration of total protein in a sample, the Bradford protein assays was used. A standard curve was obtained using 2 mg/ml of Pierce™ Bovine Gamma Globulin Standard to calculate the unknown protein concentration. 6 standard reactions of 1 ml of each containing 0, 0.4, 1, 2, 4 and 10 µg BSA were prepared in triplicate using a 0.2 mg/ml of BSA stock, as shown in the Table 2.15. In parallel, 10 µl of unknown protein sample was mixed with 790 µl of Milli-Q water. Finally, 200 µl of Bradford reagent dye (Bio-Rad) were added to each of the reactions to make up a final volume of 1 ml. All the tubes were mixed thoroughly and were incubated for 5-60 minutes at room temperature. The OD_{595} of the samples was measured using a UV/vis spectrophotometer (Jenway 6315). The average absorbance values of the standard were calculated and plotted against BSA concentration and the points obtained were fitted linearly. The concentration of protein was determined from the standard curve as shown in Figure 2.2.

Table 2.15 Reaction components for the Bradford assay

BSA 0.2 mg/ml (µl)	Milli-Q water (µl)	Bradford reagent dye (µl)	BSA concentration (µg)
0	800	200	0
2	798	200	0.4
5	795	200	1
10	790	200	2
20	780	200	4
50	750	200	10

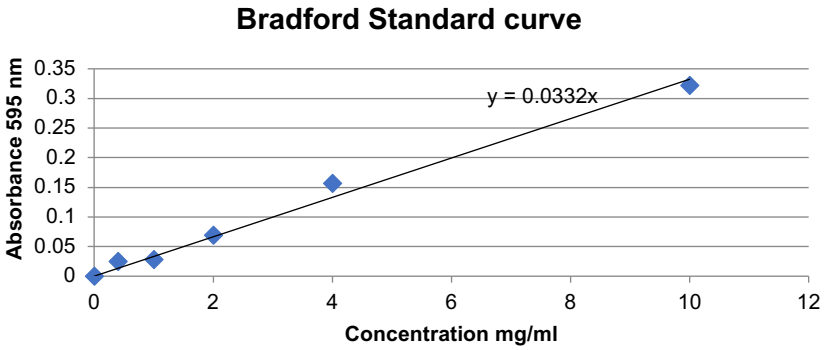


Figure 2.2 Bradford standard curve.

The figure shows the average absorbance of 6 of the protein standard BSA at different concentrations.

2.4.8 Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis (SDS-PAGE)

2.4.8.1 SDS-polyacrylamide gel preparation and gel electrophoresis

The resolving gel was prepared using the components listed in Table 2.16. The last two components, ammonium persulfate (APS) and N,N,N'N'-tetramethyl-ethylenediamine (TEMED) were added immediately before pouring the gel to avoid polymerisation. All the components were added to a 50 ml falcon tube and mixed gently to avoid any air bubble formation, before being poured into gel casting apparatus (Bio-Rad), between the two plates, up to 1 cm bellow the comb. 1 ml of isopropanol was added onto the top of the gel to prevent the gel from being exposed to air. The gel was allowed to polymerise for 30-40 minutes. Once the polymerisation was completed, the isopropanol was drained from the top of the gel using a filter paper and washed with Milli-Q water before adding the stacking gel. The stacking gel was prepared by mixing the components listed in Table 2.16 and again APS and TEMED were added just before pouring the gel. Once the stacking gel was mixed gently and applied up to the top of the gel, a comb was placed carefully to avoid the formation of air bubbles inside the gel. The gel was left for extra 30 minutes to allow polymerisation.

Once the gel solidified, the comb was removed carefully, and the gel was transferred immediately into the protein tank (Mini-PROTEAN, Bio-Rad). The tank was filled up with 1X SDS running buffer (Table 2.17). A 1:1 ratio of protein sample and SDS-loading dye (15 µl each) were mixed and a reducing agent for example, dithiothreitol (DTT) or β-mercaptoethanol (β-ME) was added. Additionally, the sample was denatured for 10

minutes at 95°C. Subsequently, the sample was either vortexed or centrifuged for 30 seconds to mix thoroughly before being loaded onto the well. 5 µl of PageRuler unstained protein molecular weight marker (Thermo Fisher Scientific) was loaded alongside the sample to estimate the size of the proteins. The gel was electrophoresed for ~80 minutes at a constant voltage of 150 V. The gel was run till the blue dye front reached the bottom of the gel.

Table 2.16 Solutions used for SDS-polyacrylamide gel preparation

Component	**12% resolving gel solution (10 ml)**	**15% resolving gel solution (10 ml)**	**5% Stacking gel solution (5 ml)**
Milli-Q H_2O	3.2 ml	2.3 ml	3.5 ml
30% 29 Acrylamide :1 Bisacrylamide	4 ml	5 ml	0.825 ml
1.5 M Tris-HCl pH 8.8	2.6 ml	2.5 ml	-
1 M Tris-HCl pH 6.8	-	-	0.625 ml
10% SDS	0.1 ml	0.1 ml	0.05 ml
10% APS	0.1 ml	0.1 ml	0.05 ml
TEMED	0.004 ml	0.004 ml	0.005 ml

Table 2.17 Buffers used for SDS-PAGE gel electrophoresis

Buffer	**Component**
10X Running buffer*	144 g Glycine, 30.3 g Tris base and 10 g SDS
2X SDS-PAGE loading buffer**	0.62 M Tris-HCl, pH 6.8, 10% (w/v) SDS, 20% (v/v) Glycerol, 0.1% (w/v) Bromophenol blue and 10% (v/v) DDT

*The buffer was diluted to 1:10 with dH2O before use.

**The buffer was stored at -20°C

2.4.8.2 SDS-polyacrylamide gel staining and protein visualization

After the blue dye reached the bottom of the gel, the gel was removed from the tank and the two plates were opened carefully. The gel was transferred into a plastic tray and was washed with water first before it was immersed with a Coomassie blue stain (Table 2.18). The gel was left to stain for 30-40 minutes with gentle shaking followed by de-staining

the gel with de-staining solution (Table 2.18) for 1 hour. The de-stain was changed until the background of the gel became clear and the bands became visible. To visualize the protein, the BIORAD Gel doc EZ imager was used, and the protein size was determined using the protein molecular weight marker. Images were taken using image lab 4.0.1 software.

Table 2.18 SDS-polyacrylamide gel visualisation solutions

Component	Coomassie Blue stain	Coomassie destain
Coomassie blue	0.1% (w/v)	-
Methanol	5% (v/v)	5% (v/v)
Glacial acetic acid	10% (v/v)	10% (v/v)

2.4.9 Size-exclusion chromatography (SEC)

Gel filtration chromatography was used as a final purification step for the labelled ParB and ParBΔ30N and to exchange the buffer of the sample with a different buffer. Furthermore, the gel filtration chromatography was also employed to determine the oligomeric state of ParB in solution. The protein sample (concentration range 2-8 mg/ml) was passed through a 0.22 µm filter. The protein sample was then concentrated into a 1 ml volume using a viva spin column 6 with 5,000 MWCO and was centrifuged at 8,000 x g speed. A Hiload superdex 75 16/600 (120 ml volume) column with a molecular weight range of 3-70 kDa (GE Healthcare) was used. The column was equilibrated with two column volumes of the desired buffer at a flow rate of 0.3 ml/minute overnight using an AKTÄ purifier. The 1 ml protein sample was injected into a 2 ml loop and was eluted with at least one column volume of the elution buffer at a flow rate of 0.5 ml/minute. The peak of the protein sample was identified and was collected. The molecular weight of the sample was then tested using SDS-PAGE and was stored at -20°C for further use.

2.4.10 Size-Exclusion Chromatography-Multi Angle Laser Light Scattering (SEC-MALLS)

The oligomeric state of ParB in solution was characterized using SEC-MALLS. 1 ml of purified ParB fraction (2.5 mg/ml) was injected into the pre-equilibrated Hiload superdex 75 16/600 (GE Healthcare) column using the AKTÄ purifier. The sample was eluted with 20 mM Tris-HCl, pH 7.5 and 150 mM NaCl at a flow rate of 0.5 ml/minute and 1 ml fractions were collected. All the fractions were collected and run on a SDS-

polyacrylamide gel before further use. Three different concentrations of the ParB protein were used (2, 0.6 and 0.2 mg/ml) (120 µl volume each) along with Bovine Serum Albumin (BSA) (2.5 mg/ml) as a control, which has a molecular weight of 66 kDa. The samples were run at 0.5 ml/minute and detected using UV absorbance at 280 nm and light-scattering signal from Wyatt Dawn HELEOS-II. The data were analysed for the calculation of molecular mass using Astra V software.

2.4.11 Western blot

2.4.11.1 Sample preparation and protein transfer

Western blot was used to characterize the extra band visible in purified ParA fractions. Before the Western blot was performed, a 12% SDS-polyacylamide gel was run with 10 µM of purified ParA sample as described in Section 2.4.8. 10 µl of ParA sample was mixed with 10 µl of 2X SDS-loading dye and heated at 95°C for 10 minutes. Next to the ParA sample, 5 µl of a pre-stained Page Ruler (Thermo Fisher Scientific) protein marker were loaded to identify the band separation and to estimate the size of the ParA protein. Once the electrophoresis run was completed, the protein in the gel was transferred onto a Hybond ECL nitrocellulose membrane (GE Healthcare) using the Bio-Rad system. One piece of the nitrocellulose membrane and four pieces of 3 mm Whatman filter paper were cut to the same size as the SDS-gel. The membrane, Whatman filter papers and two sponge pieces were soaked for 10 minutes in 1X transfer buffer (Table 2.19). The transfer sandwich (Bio-Rad system) was assembled first by placing the nitrocellulose membrane on top of two Whatman filter papers and the polyacrylamide gel was placed on top of the membrane carefully, avoiding air bubbles formation, covered by two pieces of Whatman filter paper. The stack was placed between two sponges before the cassette was placed into the tank. The protein transfer was performed at 30 mA overnight in 1X transfer buffer (Table 2.19) at 4°C. Once the transfer was finished the next day, the cassette was disassembled, and the position of the lane was marked.

2.4.11.2 Detection of protein

Once the membrane was removed from the cassette, it was rinsed with 1X Phosphate Buffer Saline (PBS) (Table 2.19) before being blocked in 50 ml of blocking buffer (Table 2.19). The membrane was incubated for 1 hour at room temperature with gentle shaking. The membrane was then rinsed and washed twice with PBST buffer (Table 2.19) and was incubated for a 10 minutes washing step at room temperature with gentle shaking. The membrane was transferred into a clean dry box and incubated with the primary anti-

his (anti mouse) antibody used at 1:2000 dilution in 10 ml of blocking buffer and then incubated for one hour at room temperature with shaking. The membrane was rinsed and washed three times with PBST for 10 minutes with shaking at room temperature before being incubated with the secondary antibody (anti-Rabbit) (Life Technologies) diluted 1:10000 in 10 ml of blocking buffer. The membrane was incubated either overnight at 4°C or at room temperature for two hours with moderate shaking. The membrane was rinsed twice and washed for three times with PBST buffer for 10 minutes each time of washing step at room temperature with gentle shaking. The membrane was dried using clean Whatman filter paper before carrying out the detection using the Pierce ECL Western blotting detection solutions (Thermo Fisher Scientific). 2 ml of luminol enhancer solution was mixed with 2 ml of peroxide solution in a clean 15 ml tube and then the solution was added on top of the membrane and incubated for 5 minutes at room temperature. The membrane was dried by placing on top of a Whatman filter paper and was then inserted into an autoradiography cassette. The membrane was covered with an acetate sheet and an X-ray film was exposed to the membrane for 1 minute at first. Then different length exposures were tried according to the strength of the signal.

Table 2.19 Solutions used in western blot

Buffer	Composition
10X PBS	685 mM NaCl, 13 mM KCl, 50 mM Na_2HPO_4 and 8.8 mM KH_2PO_4, pH 7.4
10X Transfer	380 mM Glycine, 230 mM Tris-Base and 12 mM SDS
Blocking	1X PBS, 0.1 % Tween 20 (Sigma) and 5% dried milk
PBST	1X PBS buffer and 0.1 % Tween 20 (Sigma)
1X Transfer	10% 10X transfer buffer, 70% distilled water and 20% methanol

2.5 Nuclear Magnetic Resonance Spectroscopy

Nuclear Magnetic Resonance (NMR) spectroscopy was used to determine the backbone amide of the ParB protein as well as to determine the interaction and the structural change of ParB upon binding to its cognate repeat. Solution NMR is a commonly used technique in structural biology to identify the structure of small proteins (with a molecular weight lower than 20 kDa). It is also commonly used to obtain information of protein dynamics in solution, interactions with other molecules and the chemical environment of molecules.

The basic principle behind NMR experiments is the quantum mechanical properties of the nuclei that display a magnetic property called spin. These specific nuclei behave as a dipole and can be excited once placed in a magnetic field. However, not all nuclei have this property. Atomic nuclei with an even number of protons and neutrons have a zero spin, whereas, atomic nuclei with odd numbers of neutrons and protons have ½ spin and can display a magnetic moment (Wider, 2000). This technique is useful for the biology field because some atomic nuclei in biological molecules such as proteins, possess ½ spin magnetic characteristics. The predominant naturally occurring isotopes with odd numbers, such as ^{13}C, ^{15}N and ^{1}H makes NMR a useful tool to study macromolecules. This magnetic moment can take one of two orientations or spin state once placed in the magnetic field. The spin is either lined up with low energy state (parallel) or down with high energy state (antiparallel) to the direction of the applied external magnetic field. The two states correspond to slightly different energies that the spins are allowed to jump from one state to another and is proportional to the strength of the applied magnetic field (Wider, 2000). For example, a spinning proton in an antiparallel state can be moved up to the excited state by applying a pulse of electromagnetic radiation, also named as a radiofrequency or RF pulse. When the spin returns to its equilibrium state (relaxation state), energy is emitted at the same frequency. Once the system returns to its equilibrium state, the signal will decay. This state is called the Free Induction Decay (FID) and the NMR signal is recorded. As a result, the resonance spectrum of a nuclei will be obtained by a Fourier transformation (Wider, 2000).

The resonance frequency of the energy differences between the two states depends on the local environment of the magnetic nucleus. These properties can be used to determine different nuclei in the molecule and can easily be distinguished by producing a specific peak in the NMR spectrum. The difference between frequency of the resonance expressed with references of a nuclei in its environment is called Chemical shift and is always expressed as part per million (ppm) and it depends on the spectrometer frequency (Marion, 2013).

2.5.1 An overview of NMR experiments

2.5.1.1 Two-dimension (2D) heteronuclear NMR spectroscopy

2D heteronuclear NMR spectroscopy is frequently used experiments for protein NMR in solution. These experiments correlate the proton resonance with nitrogen (or carbon) resonance that are directly bonded (Kay *et al.*, 1992). In a 2D [^{1}H ^{15}N] heteronuclear single quantum correlation (HSQC) spectrum, each cross-peak represents the proton attached to a nitrogen with the proton in one axis and nitrogen on the other axis. Thus,

each residue in the protein (except proline) will yield a cross-peak (Williamson, 2013). As such, the 2D (^{1}H, ^{15}N) HSQC is referred to as a fingerprint of the protein backbone and is a pre-requisite for assigning protein backbone resonance and structure determination.

A 2D [^{1}H ^{15}N] heteronuclear multiple quantum correlation (HMQC) spectrum provides similar information as to a 2D [^{1}H ^{15}N] HSQC spectrum. Selective Optimized Flip-Angle Short-Transient Heteronuclear Multiple Quantum coherence (SOFAST-HMQC) is a fast-pulsing experiment that records 2D correlation spectra in just a few seconds of data acquisition time. SOFAST-HMQC increases signal to noise ratio (Schanda and Brutscher, 2005). SOFAST-HMQC was used for ParB-DNA interaction and also to assess the folding state of ParB and ParBΔ30N.

HADamard-encoded AMino-ACid-type editing (HADAMAC) is a 2D [^{1}H, ^{15}N] NMR experiment that employs a Hadamard editing step to encode residue-type information. It is a powerful tool that categorises the twenty amino acids into seven individual groups, which are: Gly (group I), Val and Ile (group II), Ala (group III), Thr (group IV), Asn and Asp (group V), Phe, Tyr, Trp, His, Cys, Ser (*Cys-Arom* group) and Arg, Glu, Lys, Pro, Gln, Met and Leu (*Rest*) (Lescop *et al.*, 2008). This experiment was used as part of the process of sequential protein backbone resonance assignment of PaBΔ30N. A set of six 2D [^{1}H, ^{15}N] spectra were recorded in which each cross-peak represents the amino acid type of the previous residues in the ParB sequence.

[^{1}H, ^{15}N] heteronuclear Nuclear Overhauser Effect (hetNOE) is a 2D NMR experiment that reports protein dynamics on nanoseconds to picoseconds (ns-ps) timescale (Ferrage *et al.*, 2010). Those that undergo faster movement than the overall tumbling rate of the protein gives low NOE signal intensity. The decreased values are usually located in the N-terminal or C-terminal ends of the protein or regions of higher flexibility, such as loops.

2.5.1.2 Three-dimensional (3D) heteronuclear magnetic spectroscopy

3D experiments are widely used for sequential assignment of proteins. In these experiments, three different nuclei are correlated: ^{1}H, ^{15}N and ^{13}C where the protein is double labelled [^{15}N and ^{13}C] (Ikura *et al.*, 1990). These experiments are named by the nuclei that are correlated in the experiment, such as H, N and C. CO denotes the carbonyl carbon, whereas CA and CB represent the Cα and Cβ nuclei, respectively. This

representation also applies to HA and HB, which refer to Hα and Hβ nuclei, respectively. The order of nuclei in the name of each experiment typically provides information on the magnetisation transfer pathway. The nuclei within parentheses are also involved in magnetisation transfer although these chemical shifts are not recorded. A commonly used triple resonance experiments for backbone assignment of proteins is CBCA(CO)NH spectrum. This experiment correlates the ^{1}H and ^{15}N resonances of the amide group of a residue with the Cα and Cβ chemical shifts of the residue preceding. This experiment is typically paired with a CBCANH experiment, which correlates the chemical shifts of amide ^{1}H, ^{15}N nuclei with the Cα and Cβ chemical shifts of the same residue as well as with the residue preceding. A 3D HNCO experiment correlates amide ^{1}H, ^{15}N nuclei of residue with the carbonyl carbon of the previous residue. This experiment is typically paired with the HN(CA)CO. Lastly, HNCA is a 3D NMR experiment that correlates the backbone amide with the Cα in the same residue as well as in the preceding residue. This experiment is typically paired with HN(CO)CA. HN(CO)CA correlates the chemical shift of amide with Cα within the same residue. Another 3D pair NMR experiment used is HANH and HA(CO)NH that connects a resonance amide with the Hα in the same residue and the previous residue respectively. All the sets of 3D NMR experiments that are mentioned earlier were used for resonance backbone assignment of ParB and ParBΔ30N. 3D NMR experiments were processed by Dr Alex Heyam using either NMRpipe or using TopSpin version 3 (Bruker Biospin). Spectra of triple resonance assignment experiments were analysed, and the backbone assignments of proteins were obtained suing CcpNMR Analysis version 2.4 (Vranken *et al.*, 2005).

2.5.2 NMR data collection

2.5.2.1 Sample preparation

Purified samples of [^{15}N and ^{13}C] labelled ParB and ParBΔ30N were passed through a second purification step and buffer exchanged into 20 mM Bis-Tris-HCl, pH 6.0, 50 mM NaCl and 1 mM DTT, using SEC, which was performed on the day of experiments, as described in Section 2.4.9. The purity of the protein samples was verified by SDS-PAGE (Section 2.4.8) before using it for any further experiments. 600 µl of 200-700 µM ParB and ParBΔ30N were used for NMR experiments. 700 µM protein concentration was the maximum concentration that could be obtained without precipitation or degradation of ParB. The samples were prepared by mixing 570 µl of protein and 30 µl of D_2O (5%). The sample was centrifuged at high speed 14,000 x g for 5 minutes before it was loaded carefully into the 5 mm NMR tube (Wilmad) to avoid any air bubbles forming during

loading the sample. The composition of the buffer therefore changed upon addition of 5% D_2O and became 19 mM Bis-Tris-HCl, pH 6.0, 47.5 mM NaCl, 0.95 mM DTT and 5% D_2O. This buffer was the standard buffer used for all NMR experiments. For buffer optimisation, the protein sample volume was reduced to 300 μl of 120 μM and loaded into a small NMR tube (Shigemi tube). The sample was then transferred into a Bruker Avance Neo 700 MHz NMR spectrometer equipped with N_2-cooled triple resonance cryoprobe and NMR experiments were recorded. All the experiments were performed at 25°C with a recycle delay of 1 second.

2.5.2.2 (^{1}H-^{15}N) HSQC and SOFAST-HMQC

All the HSQC and SOFAST-HMQC experiments were carried out in the standard NMR buffer (20 mM Bis-Tris-HCl, pH 6.0, 50 mM NaCl and 1 mM DTT in 5% D_2O) using the pulse sequence hsqctf3gpsi and sfhmqcf3gpph respectively. The concentration of ParB samples ranged from 0.1 to 1 mM (monomer). The spectra were recorded using parameters shown in Table 2.20. Both HSQC and SOFAST-HMQC were set up and processed by Dr Alex Heyam using TopSpin version 3.5 pl 7. The 2D (^{1}H-^{15}N) HSQC was used as a first step for the backbone assignment of the ParB protein.

Table 2.20 Parameters used for HSQC and SOFAST-HMQC NMR experiments

Experiment	Pulse sequence	Nuclei	Offset (ppm)	Spectral width (ppm)	# of real and imaginary points
HSQC	hsqctf3gpsi	^{1}H ^{15}N	4.7 118	16 30	2048 256
SOFAST-HMQC	sfhmqcf3gpph	^{1}H ^{15}N	4.7 118	16 30	2048 128

2.5.2.3 Backbone resonance assignment of ParB N-terminal domain

A 0.65 mM aliquot of [^{15}N, ^{13}C] labelled ParB was prepared in standard NMR buffer and used to collect data to enable backbone assignment of ParB N-terminal domain. A CBCANH (cbcanhgpwg3d) and CBCA(CO)NH (cbcaconhgpw3d) set of 3D triple resonance experiments were used to assign and link all the signal to H-N of 2D ^{1}H, ^{15}N HSQC. The spectra were recorded using the parameters given in Table 2.21.

Table 2.21 Experimental parameters used for backbone assignment of ParB N-terminal domain.

Experiment	Pulse sequence	Nuclei	Offset (ppm)	Spectral width (ppm)	# of real and imaginary points
CBCANH	cbcanhgpwg3d	^{1}H ^{15}N ^{13}C	4.7 118 43	16.2 30 80	2048 1 32
CBCA(CO)NH	cbcaconhgpw3d	^{1}H ^{15}N ^{13}C	4.7 118 43	16 30 80	2048 1 32

2.5.2.4 Backbone resonance assignment of ParBΔ30N

A 1.2 mM sample of [^{15}N, ^{13}C] labelled ParBΔ30N was prepared in standard NMR buffer and used for the backbone assignment of the ParB C-terminal domain. 3D HNCA, 3D HN(CO)CA, 3D HNCO, 3D HN(CA)CO and 3D HNCACB, 3D HANH and 3D HA(CO)NH sets of 3D triple resonance experiments were used to assign and link all the signal to H-N of 2D ^{1}H, ^{15}N HSQC. The non-uniform sampling parameter was used for HN(CA)CO, HNCA and HANH. In addition, 2D [^{1}H,^{15}N] HADAMAC experiment was used together with 3D triple resonance experiments. The spectra were recorded using the parameters given in Table 2.22.

Table 2.22 Experimental parameters used for backbone assignment of ParB C-terminal domain.

Experiment	Pulse sequence	Nuclei	Offset (ppm)	Spectral width (ppm)	No. of points (real and imaginary)
HNCA	hncagp3d	^{1}H ^{15}N ^{13}C	4.7 118 53.2	13.7 30 30	2048 56 128
HN(CO)CA	hncocagp3d	^{1}H ^{15}N ^{13}C	4.7 118 53.2	16 30 30	2048 56 128
HNCO	hncogp3d	^{1}H ^{15}N	4.7 118	13.7 30	2048 56

		^{13}C	173.5	14	128
HN(CA)CO	hncacogp3d	^{1}H	4.7	13.7	2048
		^{15}N	118	30	56
		^{13}C	173.5	14	128
HNCACB	hncacbgp3d	^{1}H	4.7	13.7	2048
		^{15}N	117	30	56
		^{13}C	43	80	138
HANH	hanhgpwg3d	^{1}H	4.7	13.7	2048
		^{15}N	118	30	64
		^{1}H	4.7	4	128
HA(CO)NH	haconhgpwg3d	^{1}H	4.7	13.7	2048
		^{15}N	118	30	64
		^{1}H	4.7	4	128
HADAMAC	seqHADAMAC_sel	^{1}H	4.7	16.2	1220
		^{15}N	118	30	128

2.5.2.5 Heteronuclear Nuclear Overhauser Effect (HetNOE)

[^{1}H, ^{15}N] hetNOE experiment was carried out in the standard NMR buffer (20 mM Bis-Tris-HCl, pH 6.0, 50 mM NaCl and 1 mM DTT in 5% D_2O) using the pulse sequence hsqcnoef3gpsi. The ParB sample concentration was > 0.65 monomer. The nitrogen and proton offsets were 118 and 4.7 ppm respectively; spectral widths of 30 ppm were used for nitrogen and approximately 16 ppm for proton. The collected real and imaginary points were 256 for nitrogen dimension and 2048 for proton. The hetNOE experiment was set up and processed by Dr Alex Heyam using TopSpin version 3.5 pl 7.

2.6 Structure determination of ParB C-terminal domain using X-ray crystallography

2.6.1 Principle of protein X-ray crystallography

X-ray crystallography is another approach to determine the high-resolution three-dimensional structure of the protein. Three components are required for 3D structure analysis of the protein of interest: a protein crystal, a source of X-ray and a detector. The use of X-rays provides the highest resolution because the X-ray beam has a wavelength ranging between 0.5-1.54 Å, which is similar to the distance between atoms in the molecule. The technique requires a highly ordered crystal that is obtained by slowly adding any salt to a concentration in which the protein reduces its solubility in solution.

This process allows the molecules to pack together to form a three-dimension crystal. A single crystal contains a huge number of identical protein molecules that align together to make data collection more sensitive. The crystal is positioned in an orientation that the X-ray beam can detect. Some of the X-ray goes through the crystal while the rest scatters in a different direction. The scattering of X-rays is named X-ray diffraction. The crystal is repeatedly exposed to the X-ray beam by rotating the crystal one degree per exposure. The crystal is rotated many times so the X-rays can strike the crystal from many directions. A single exposure provides an image that contains spots, which represents a diffracted X-ray beam. Thousands of spots are collected from hundreds or thousands of images to solve the 3D structure of the protein of interest. The intensity of the spots is used to determine the electron density of molecules within the protein crystals. This information is used to build up a model of the molecule.

2.6.2 Protein sample preparation

For structure determination of the C-terminal domain of ParB, the truncated version pET-ParBΔ30N was overexpressed in a 500 ml LB culture, as described in Section 2.4.1 and purified using Ni^{+2} affinity chromatography (Section 2.4.5). The purified protein sample was concentrated to 6 mg/ml using 5 kDa MWCO vivaspin 6 (Sartorius) concentrator. The concentrated sample was dialysed overnight against 20 mM Bis-Tris pH 6.0 and 50 mM NaCl at 4°C using 7 kDa MWCO SnakeSkin dialysis tubing (Thermo Scientific). Then buffer exchange was carried out. A 1 ml aliquot of purified ParBΔ30N (6 mg/ml) was loaded onto 7 kDa MWCO dialysis tube after one end and then sealed in the dialysis sachet. The sample was dialysed against 500 ml of freshly prepared 20 mM Bis-Tris pH 6.0, and 50 mM NaCl buffer in a1 L beaker. A magnetic stirrer was added into the beaker. The sample was buffer exchanged at 4°C for two hours at very low stirring speed. After 2 hours exchange, the buffer in the beaker was replaced with other 500 ml of freshly prepared buffer and incubated overnight at 4°C at the same stirrer speed. The following day the dialysis sachet was taken out from the buffer and one end was opened. The sample was collected carefully using a pipette and transferred into a 1.5 ml Eppendorf tube. The concentration of the ParBΔ30N protein was measured using Bradford assay (Section 2.4.7) and then the sample was concentrated down to 200 μl of 6 mg/ml with a 5 kDa MWCO vivaspin 6 (Sartorius) concentrator. Before the crystal tray was prepared, 1 mM Tris-(2-carboxyethyl) phosphine (TCEP) was added to the concentrated ParBΔ30N sample at a final concentration.

2.6.3 Crystal plates preparation and crystal growth

The ParBΔ30N protein crystallisation trials were performed in 96-wells using the sitting-drop vapor-diffusion methods (Swissci plates). PACT and X-tall-screen plates were used and contained two protein drop positions per reservoir well. The reservoir wells contained a 100 μl volume. ParBΔ30N protein samples were prepared at a concentration of 3 and 6 mg/ml in 20 mM Bis-Tris pH 6.0, 50 mM NaCl and 1 mM TCEP buffer for each 2 drop wells. A 1:1 ratio of protein and crystal screening reagents from the reservoir well were mixed to the sitting drop well, making up a total reaction volume of 200 nl. First a 100 nl from 6 mg/ml and 3 mg/ml were dispensed using Mosquito LCP liquid handling robot (TTP Labtech Ltd) into the first and the second protein well positions, respectively, followed by the crystal reagents. Two identical plates were prepared (2 for each plate). Plates were sealed immediately with flat upper surface and checked under the microscope to verify drop coalescence. Plates were incubated at 18° and 4°C until the crystals started to grow. Plates were checked every two days to monitor crystals growth.

2.6.4 Crystal mounting

ParBΔ30N crystals were obtained from condition A6 of the Hampton Research Crystal Screen Kit after growing for two weeks at 4°C. The A6 condition contains 0.2 M Magnesium Chloride hexahydrate, 0.1 M Tris-HCl pH 8.5, 30% w/v Polyethylene glycol 4,000. The crystals were harvested under the microscopy using a small circular loop of nylon fibre and flash-frozen in liquid nitrogen. The crystals were kept in liquid nitrogen until being sent for X-ray diffraction.

2.6.5 X-ray diffraction

ParBΔ30N crystals were sent for X-ray diffraction. The high-resolution diffraction data were collected at beamline I03 of the Diamond Light Source synchrotron using Pilatus 6M. A total of 2200 images with 0.5° rotation per image were collected with 1.93 Å resolution

2.6.6 Data processing and refinement

Data processing and structure determination was carried out by Dr Clément Dégut. X-ray diffraction data were processed using autoPROC and scaled with SCALA. The structure of the ParB C-terminal domain (ParBΔ30N) was determined by molecular replacement in PHASER using the Omega (ω_2) from pSM19035 (*Streptococcus*

pyogenes, PDB code 2bnw) as a model. In the model, the non-conserved residues were changed into alanine. ARP/WARP was used first for automated model building procedure. Restrained refinements of the structure were performed with refmac in the Phoenix suite. Coot program was used for manual reconstructions of map visualisation.

2.7 ParB/DNA interaction techniques

2.7.1 Electrophoretic Mobility Shift Assays (EMSA)

2.7.1.1 Preparation of biotinylated DNA fragment

EMSA was used to assess the interaction between ParB and the two centromere sites, *parC1* and *parC2*. The gel retardation assays were performed using biotinylated DNA fragments. The biotinylated DNA fragments were obtained by amplifying the *parC1* and *parC2* regions of the *parABC* cassette of pB171, which were 260 and 240 bp long, respectively. The *parC1* and *parC2* fragments were generated using two primers, forward and reverse. Primers were designed following standard methods, and the forward primers were tagged with biotin at the 5′ end, while the reverse primers were not modified. The amplified DNA fragments were examined using 1.5% (w/v) agarose gel followed by extraction (Section 2.3.11 and 2.3.12) to remove any secondary structure that was formed because of biotin. DNA concentrations were measured using a Nanodrop 1000 (Thermo Fisher).

2.7.1.2 Sample preparation and gel electrophoresis

Once the DNA fragment was purified, the reaction mixture was prepared as shown in Table 2.23. A working stock solution of 2 µM of protein was prepared. Reaction mixtures with a total volume of 20 µl were prepared with increasing concentrations of ParB and with fixed 0.6 nM for *parC1* and 1.6 nM for *parC2* final concentration of biotin-labelled DNA fragment. The reaction contains 2.5 µl of 10X binding buffer (100 mM Tris-HCl, pH 7.5, 500 mM KCl and 10 mM DTT), 2 µl of 50% Glycerol, 1 µl of 100 mM $MgCl_2$ and 1 µl of poly-dI-dC (Poly(deoxyinosinic-deoxycytidylic) acid) (Thermo Fisher). ParB and 2.5 µl of 5 nM DNA were added to the mixture as last. Along with the reactions, a control sample was also prepared that contains all the component except the ParB protein. All the reactions were incubated at room temperature for 20 minutes. Meanwhile, an agarose gel, 1% (w/v) was prepared in 0.5X TBE buffer (Section 2.3.11) and pre-run for 30 minutes at 60 Volts at 4ºC. Once the incubation time was over, the reactions were loaded carefully and 2 µl of 6X loading dye (NEB) was added to the first sample (control). The gel was run for ~2.5 hours at 60 Volts at 4ºC until the red dye migrated ¾ of the gel.

Table 2.23 Reaction components for EMSA

Component	No protein (µl)	50 nM (µl)	100 nM (µl)	200 nM (µl)	300 nM (µl)	400 nM (µl)	500 nM (µl)	700 nM (µl)
Sterile Milli-Q water	12	11.5	11	10	9	8	7	6
10X Binding buffer	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5
Glycerol 50%	1	1	1	1	1	1	1	1
100 mM $MgCl_2$	1	1	1	1	1	1	1	1
Poly-dI-dC	1	1	1	1	1	1	1	1
DNA	2.5	2.5	2.5	2.5	2.5	2.5	2.5	2.5
Protein	_	0.5	1	2	3	4	5	7
Total	**20**	**20**	**20**	**20**	**20**	**20**	**20**	**20**

2.7.1.3 DNA transfer onto positively charged membrane

Immediately after the electrophoresis was completed, the DNA was transferred onto positively charge nylon membrane (Roche) overnight. One piece of the nitrocellulose membrane and four pieces of Whatman 3 mm filter paper were cut to the same size as the agarose gel. The membrane and Whatman filter papers were soaked for 10 minutes in 0.5X TBE buffer before the sandwich was assembled. Eight pieces of paper towel were folded and used as a cassette by placing four at the base and the other four on the top of the sandwich. The transfer sandwich was assembled manually first by placing two Whatman filter papers on top of four folded paper towels followed by nitrocellulose membrane. The agarose gel was placed on top of the membrane carefully, avoiding air bubble formation, covered by two pieces of Whatman filter paper and the remaining four folded blue tissues were placed on top of the sandwich. A heavy book was placed on top of the sandwich to create pressure and to effectively transfer the DNA. The DNA was transferred overnight at room temperature (Figure 2.3).

Once the DNA was transferred onto the positively charge nylon membrane, it was cross-linked onto the membrane. The nylon membrane was wrapped first with cling film and

exposed to UV light using a UV transilluminator by placing the DNA side down for 5 minutes to covalently bind DNA to the positively charged nylon membrane.

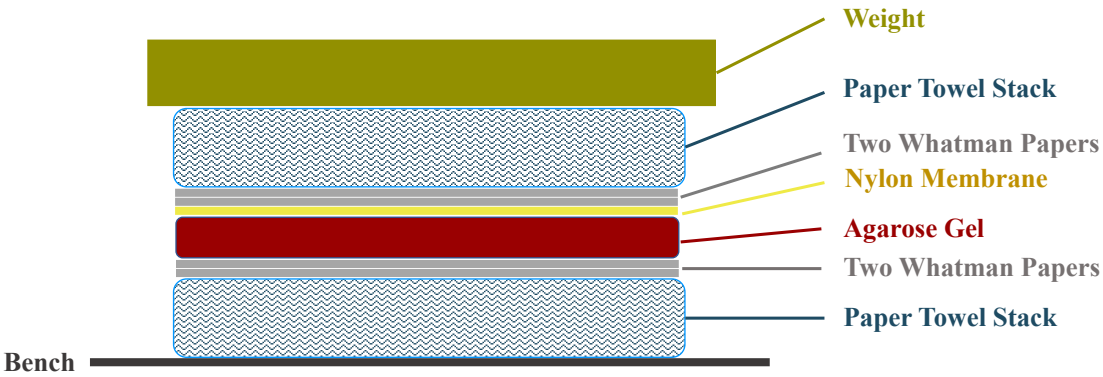


Figure 2.3 DNA blot assembly on nylon membrane

The diagram shows the manual transfer of DNA onto the positively charged nylon membrane.

2.7.1.4 Detection of DNA

The detection of biotinylated DNA was carried out using the LightShift Chemiluminescent EMSA kit (Thermo Fisher Scientific) and the buffers were prepared when required (Table 2.24). Blocking buffer and 4X wash buffer were incubated for 10 minutes at 40ºC using a water bath to ensure the SDS was completely dissolved before starting the detection. The membrane was transferred into a clean tray and blocked using 20 ml of blocking buffer with the DNA side was faced up and was incubated on the shaker for 15 minutes. The blocking buffer was discarded, and the membrane was incubated for a second time with 10 ml blocking buffer. 1 µl of streptavidin-horseradish peroxidase conjugate (Sigma-Aldrich) was added to the blocking buffer and incubated for 15 minutes on the shaker. Once the incubation was over, the blocking buffer was discarded, and the membrane was washed four times with 1X wash buffer for 5 minutes each time on a shaker. Next, the membrane was transferred into a new clean tray and equilibrated with 30 ml equilibration buffer. The membrane was later incubated at room temperature for 5 minutes. The membrane was removed carefully with forceps and was drained using filter paper. 2 ml of each of luminol/enhancer and stable peroxide detection solutions were mixed and the membrane was placed with the DNA side faced down in the tray. The membrane was incubated with the detections for 5 minutes at room temperature. The

membrane was then taken out from the box and the excess detection solution was drained from the membrane using the filter paper. Once the membrane was dry, it was transferred into an autoradiographic cassette with DNA sided up and was covered with an acetate sheet. The membrane was then exposed to the X-ray film for at least 1 minute and then additional exposures were taken based on the signal's intensity.

Table 2.24 Buffers used in EMSA detection

Blocking buffer	**4X wash buffer**	**Equilibration buffer**
100 mM Tris-HCl, pH 7.5	100 mM Tris-HCl, pH 7.5	100 mM Tris-HCl, pH 7.5
1% BSA	5% SDS	Distilled water
5% SDS	Distilled water	
Distilled water		

2.7.2 Microscale Thermophoresis (MST)

2.7.2.1 Overview

Microscale thermophoresis (MST) is a powerful biophysical technique that is used to quantify the interaction of biomolecules in solution. This technique is based on the thermophoresis phenomenon, which is the direct movement of molecules in a temperature gradient. This movement depends on many properties of a molecule, such as molecular weight, charge and conformation changes in the solvation shell (Jerabek-Willemsen *et al.*, 2014). Therefore, by using this phenomenon, it is possible to identify the changes in a biomolecule upon ligand binding.

MST is a fluorescence-based technique that requires one of the molecules under investigation be fluorescently labelled. A fluorescence detector allows to monitor the pathway of the molecule through the temperature gradient. Moreover, this technique allows the user to analyse the binding events of molecules in a reaction of a few microliters. In the MST experiment, the labelled molecule is usually kept at a constant low concentration (usually in the nanomolar range). Whereas, the unlabelled binding partner is titrated at increasing the concentrations up to the micromolar range. 12-16 concentrations are used to identify accurately the interaction of molecules as well as to build the binding curve. The concentrations are selected based on the expected dissociation constant (K_D) value: the starting concentration should be 10-fold lower than the expected k_D value (it is possible to avail of the NanoTemper Concentration Finder tool).

The Monolith NT.115 (NanoTemper Technology Gmbh) instrument is designed with an infrared (IR)-Laser combined with excitation/emission using an IR dichroic mirror. The fluorescently labelled molecules within a small glass capillary are excited using the IR-Laser and the thermophoresis is detected. The molecules are homogenously distributed within the sample where the initial fluorescence is detected. Once the laser is activated in the first second, a T-jump (Temperature jump) is observed, which represents the immediate changes in fluorophores within the sample due to the fast temperature changes. Subsequently, the movement of the labelled molecules out of the heated area is detected, which typically takes 30 seconds. After the IR-Laser is turned off the fluorescent molecules diffuse back into the depleted area (Jerabek-Willemsen *et al.*, 2011).

The MST experiment was used to investigate the interaction between ParB with its centromeric sites, *parC1* and *parC2*. MST was also used to calculate the binding cooperativity between the ParB and centromere site using the hill coefficient.

2.7.2.2 Preparation of DNA fragment for MST

For MST experiments, one of the molecules under investigation needs to be labelled to track the movement of molecules through the temperature gradient. For investigating ParB binding to *parC1* and *parC2* using MST, the two DNA fragments were generated and labelled with a fluorescent dye. The *parC1* and *parC2* regions were amplified using PCR. The forward primers were tagged with Cyanine5 (Cy5) dye at the 5′ ends. The forwards and reverses primers used to amplify the two fragments are listed in Table 2.8. The PCR reactions were set up as described in Section (2.3.5). The amplification products of both *parC1* (260 bp) and *parC2* (240 bp) were checked on a 1% agarose gel. The size of the DNA fragments was identified using a 1 kb ladder. Both fragments were extracted from 1.2% agarose gel (Section 2.3.12). The concentration of DNA fragments was determined using Qubit 3.0 (Thermo Fisher Scientific). The samples were kept in the dark at -20°C until used. However, it was preferable to use the purified fragments on the day to avoid dye bleaching.

2.7.2.3 Sample preparation and DATA analysis for MST

For MST experiments a fixed concentration of *parC1*-Cy5 and *parC2*-Cy5 fragments was used. Fluorescence intensity of both DNA fragments were determined using the MST instrument before an optimal concentration was selected. Ideal fluorescent values for MST binding experiments are between 200 and 1200 units. The labelled *parC1* and

parC2 were diluted using the 1X MST buffer (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 10 mM MgCl2 and 0.05% Tween 20). The optimal concentrations of *parC1* and *parC2* tagged with Cy5 were 25 and 29 nM stock, respectively. The final *parC1* and *parC2* concentration were 12.5 and 15 nM, respectively, upon addition of an equal volume of ParB to the reaction. The labelled DNA fragments were centrifuged for 10 minutes at 12,000 x g at 4°C to remove any precipitate before setting up the reactions. The solution was removed carefully using a pipette without touching the bottom of the tube and moved into a new clean 1.5 ml Eppendorf tube.

The ParB protein was purified as previously (Section 2.4.5). One aliquot of the purified ParB was taken from -80°C and thawed on ice at room temperature before preparing the reaction mixture. The concentration of ParB was determined using Bradford assay (Section 2.4.7) and was usually above 2 mg/ml. ParB was buffer exchanged into the 1X MST buffer using a desalting column as it is described in Section 2.4.6.2. ParB was then concentrated to 200 μM using 7 kDa MWCO concentrator at 10,000 x g for 30 minutes. ParB was centrifuged for 10 minutes at 12,000 x g at 4°C prior to setting up the reactions. 16 PCR tubes were used to prepare the reactions. 10 µl of the 1X MST buffer was added to tubes 2 to 16. Then 10 µl of the 200 µM unlabelled ParB were added to the first empty PCR tube and to the second PCR tube. The solution in the second tube was mixed several times and 10 μl of the solution was transferred into the third tube. The previous dilution step was carried out for the remaining tubes followed by discarding the last 10 µl. The concentration of ParB protein started at 100 µM and, by halving each time, reached the value of 3.05 nM. 10 µl of labelled DNA was added into each tube and mixed thoroughly by pipetting several times. The reaction mixtures were then incubated at room temperature for 10 minutes before being transferred into standard glass capillaries. 4-5 µl of each reaction was introduced into a standard capillary starting from reaction number 1 to 16 and the capillaries were put in the slots on the sample tray. The sample tray was then inserted into the Monolith NT.115 MST instrument (NanoTemper Technologies Gmbh). The LED power (LED colour red for excitation of Cy5) was adjusted according to the DNA concentration. The fluorescence signal was tested using a capillary scan and based on the signal intensity, the percentage of LED power was changed accordingly. The fluorescence signal was between 600-800 at 80 % LED power. Three MST measurements were carried out: 10, 20 and 40% (IR-laser) power using MO.Control software. The experiment was repeated in triplicate. The MST data were analysed for calculation of K_D using the MO. Affinity Analysis software (NanoTemper Technologies Gmbh). For data analysis the change in thermophoresis is described as the change in the normalized fluorescence (ΔF_{norm}). The ParB increasing concentrations resulted in a

gradual change in thermophoresis and plotted as ΔF_{norm} to yield a binding curve, where ΔF_{norm} is:

$$\Delta F_{norm} = F_{hot}/F_{cold}$$

GraphPad Prism 7.04 software was used to identify the binding curve using the one-site specific binding equation:

$$Y = \frac{B_{max}X}{K_D + X}$$

Where Y is the fraction bound, X is the concentration of the labelled DNA, B_{max} is the maximum binding and K_D is the equilibrium dissociation constant.

2.7.3 Chemical shift mapping

ParB was ^{15}N labelled as described in Section (2.4.4) and purified (Section 2.4.5). The backbone assignment of ParB was carried out as shown previously (2.5.2.3 and 2.5.2.4). The 18 bp DNA was designed and annealed, as described previously in Section 2.3.14. Chemical shift mapping of the interaction between ParB with short sequence 18 bp DNA that harbours a single repeat from its cognate site was performed using SOFAST-HMQC NMR experiment. The binding experiments were obtained using the ^{15}N labelled ParB at fixed concentration 0.14 mM and titrated with different ratios of DNA (1:0, 1:0.2, 1:0.4, 1:0.5, 1:0.6, 1:0.8, 1:1 and 1:2) in a 5 mm NMR tube (Wilmad). The NMR chemical shift mapping was performed in NMR standard buffer that was used for ParB backbone assignment. For each titration point, a SOFAST-HMQC was recorded using the pulse sequence sfhmqcf3gpph. The nitrogen and proton offsets were 118 and 4.7 ppm respectively; spectral width of 30 ppm was used for nitrogen and approximately 16 ppm for proton. The collected real and imaginary points were 128 for nitrogen dimension and 2048 for proton. The data were processed and analysed using TopSpin version 3.5 pl 7. The chemical shift changes were derived from the ^{1}H and ^{15}N shift differences using the formula:

$$\Delta\delta = \sqrt{(\delta H^2\ (\delta N/6.5))^2}$$

2.8 Chemical Cross-Linking

Chemical cross-linking was employed to investigate the interaction between ParA and ParB *in vitro*. It was also used to identify the interaction of both proteins when different ligands were present or absent, such as ATP and specific DNA sites. Two cross-linking reagents were used in this study: dimethyl pimelimidate (DMP) and bissulfoscuccinimidyle substrate (BS3). DMP contains of an amine-reactive imidoester

at both ends of a 7-carbon spacer arm with spacer arm length of 9.2 Å. BS3 contains an amine-reactive *N*-Hydroxysuccinimide (NHS) ester at both ends of an 8-carbon spacer arm with spacer arm length of 11.4 Å. Both cross-linking reagents form an amidine bonds with amine groups in the side chain of lysine or at the N-terminus of polypeptide chains at alkaline pH 8.0-9.0.

2.8.1 DMP cross-linking

DMP cross-linking reagent was used to probe the interaction of ParA with ParB in the presence and absence of one centromere site and ATP. ParA and ParB were buffer exchanged into cross-linking buffer (50 mM HEPES, pH 8.5, 50 mM KCl, 5 mM MgCl2) using the Zeba^{TM}Spin desalting column (Thermo Scientific) as described in Section 2.4.6.2 and the protein concentration was measured using the Bradford assay (Section 2.4.7). A 1:1 molar ratio of ParA and ParB (10 µg in 20µl) was used. DMP powder was diluted in sterile Milli-Q water just before being used and added to the reactions at a final concentration of 10 mM in a total reaction total volume of 20 µl. Five cross-linking reactions were prepared. In the first two reactions, ParA and ParB were crossed-linked on their own at 10 µg in 20µl and were used as controls for the experiment. The third reaction was prepared by adding equal molar ratio of ParA and ParB. ParA and ParB were also crossed-linked in the presence of the centromere site *parC1* (fourth reaction). The *parC1* site was amplified using PCR as mentioned in Section 2.3.5. The fragment was amplified using primers listed in Table (2.8) and was ~260 bp long. The fragment was gel extracted and purified as described in section 2.3.12. The reaction was prepared by adding equal molar ratio of ParA and ParB (10 µM final concentration) and the cognate DNA *parC1* was added at a final concentration of 0.6 µM. The final reaction was identical to the fourth reaction, but ATP (1 mM final concentration) was added. Cross-linking buffer and freshly prepared DMP (10 mM final concentration) were added to the all the reactions. The reaction mixtures were incubated at 22°C for 1 hour and then quenched by adding 50 mM Tris-HCl, pH 6.8. The cross-linking reaction were assessed by 12% SDS-PAGE. The 2X SDS loading dye at 1:1 ratio was added to all reactions. The samples were heated at 95°C for 5 minutes before being loaded on the SDS-polyacrylamide gel. Gels were analysed using a BIORAD Gel doc EZ imager and the Image Lab 4.0.1 software.

2.8.2 BS3 cross-linking combined with Liquid Chromatography-tandem Mass Spectrometry (LC-MS/MS) and data analysis

BS3 cross-linking experiments coupled with liquid chromatography-tandem spectrometry (LC-MS/MS) were used to determine the amino acid residues are involved in the interaction of ParA with ParB. This interaction was investigated using different ligands, such as the cognate DNA site and ATP. Several reactions were prepared. ParA and ParB were crossed-linked individually and used as controls for the experiments. ParA and ParB were crossed-linked together, in the presence of *parC1* and in the presence of *parC1* and ATP. First ParA and ParB were buffer exchanged into the cross-linking buffer (50 mM HEPES, pH 8.5, 50 mM KCl, 5 mM $MgCl_2$) as described in Section 2.4.6.2. After buffer exchange, the ParA and ParB samples were centrifuged at 13,000 x g for 10 minutes at 4°C to remove any aggregates. The samples were removed carefully without touching the bottom of the tube and transferred into a clean 1.5 ml Eppendorf tube. ParA and ParB concentration were measured by Bradford assay (Section 2.4.7). The cross-linking reactions were prepared in 50 μl volume and equal molar ratio of ParA and ParB at 10 μM final concentration was added to the reactions. When the cognate DNA was present, *parC1* was added at 0.6 μM final concentration. When present, ATP was added at a 1 mM final concentration. BS3 was prepared after all the reaction mixtures were prepared. BS3 was dissolved in 70 μl of sterile Milli-Q water to obtain a 50 mM stock solution. It was added to the reactions at a final concentration of 1 mM as the last reagent. The reactions were incubated at 22°C for 45 minutes. Once the incubation time was finished, the reaction was quenched for 15 minutes by adding 50 mM ammonium bicarbonate pH 8.5. The samples were run on 12% SDS-polyacrylamide gel, as described in Section 2.4.8. The gel was stained with Coomassie staining solution overnight followed by de-staining for few hours or until the background became clear. The ParA and ParB cross-linked bands were identified.

The next step LC-MS/MS and data analysis were performed by Dr Chris Taylor. Gel tryptic digestion was performed after reduction with dithioerythritol and S-carbamidomethylation with iodoacetamide. Gel slices were washed two times with aqueous 50% (v:v) acetonitrile containing 25 mM ammonium bicarbonate, then with acetonitrile and dried in a vacuum concentrator for 20 minutes. Sequencing-grade modified porcine trypsin (Promega) was dissolved in 50 mM acetic acid, then diluted 5-fold with 25 mM ammonium bicarbonate to give a final trypsin concentration of 0.02 mg/mL. Gel slices were rehydrated by adding 10 mL of trypsin solution, and after 10 minutes enough 25 mM ammonium bicarbonate solution was added to cover the gel

pieces. Digests were incubated overnight at 37°C. Peptides were extracted by washing three times with aqueous 50% (v:v) acetonitrile and once with acetonitrile before drying in a vacuum concentrator and reconstituting in aqueous 0.1% (v:v) trifluoroacetic acid.

Samples were loaded onto an UltiMate 3000 RSLCnano HPLC system (Thermo) equipped with a PepMap 100 Å C18, 3 µm trap column (75 µm x 20 mm Thermo) and a PepMap, 2 µm, 100 Å, C18 EasyNano nanocapillary column (75 µm x 500 mm, Thermo). The trap wash solvent was aqueous 0.05% (v:v) trifluoroacetic acid and the trapping flow rate was 4 µL/minute. The trap was washed for 5 minutes before switching flow to the capillary column. Separation used gradient elution of two solvents: solvent A, aqueous 1% (v:v) formic acid; solvent B, aqueous 80% (v:v) acetonitrile containing 1% (v:v) formic acid. The flow rate for the capillary column was 300 nL/minute and the column temperature was 40°C. The linear multi-step gradient profile was: 3-10% B over 7 minutes, 10-35% B over 30 minutes, 35-99% B over 5 minutes and then proceeded to wash with 99% solvent B for 4 minutes. The column was returned to initial conditions and re-equilibrated for 15 minutes before subsequent injections.

Mass spectra were acquired in an Orbitrap Fusion Tribrid mass spectrometer supplied by Thermo Fisher scientific. The spray voltage was set to 1.9 kV and the ion transfer tube temperature was held at 275°C. Precursor scans were acquired in profile in the Orbitrap at a resolution of 120K and with a scan range of between 375 to 1500 m/z. Injection times were limited to 100 milliseconds and an AGC (automatic gain control) target of 106 was set.

Peptides with charge state greater than 3 were preferentially selected for fragmentation. Dynamic exclusion allowed two repeats of the same precursor with 10 ppm mass tolerance followed by exclusion for 50 seconds. A minimum intensity of 4500 counts was required for selection with priority given to the highest charge state present.

To improve sequence coverage for cross-linked peptide candidates, consecutive centroid CID and EThcD MS2 scans were acquired in the Orbitrap at a resolution of 30K and a maximum injection time of 54 milliseconds. Isolation was carried out in the quadrupole and an AGC target of 5e4 was set for both fragmentation modes. Normalised collision energy of 35% was applied for CID acquisitions.

Following acquisition, RAW files were imported into the Progenesis QI software for inter-run alignment and peak picking. Concatenated MS2 peak lists were exported in Mascot generic format (mgf) for database searching using Byonic. Search settings specified a maximum of four missed cleavages, trypsin specificity, precursor mass tolerance of 3 ppm and a fragment mass tolerance of 0.01 Da. BS3 cross-linker and oxidation of

methionine residues were set as variable modifications and carbamidomethylation of cysteine was set as fixed. Identification results were imported into the Progenesis software for a cross-run comparison based on area under the curve of the precursor ions identified by the search engine.

Cross-linked peptide spectral identification quality was assessed by Byonic and expressed as posterior error probabilities. The PEP 2D value includes confidence inferred and weighted from other peptides matched for the same protein. The PEP 1D p-value is a better metric for the confidence of the individual spectrum/peptide identification as this value is derived only from the quality of the individual peptide match. The cross-linked peptides have also been further manually validated for spectral match quality, leaving four cross-linked peptides, two inter ParA-ParB and two intra ParB.

2.9 Partition Assays

In vivo partition assays were used to assess the stability of plasmid that harbours the partition cassette *parABC*, when the minimal partition sites were present. Partition assays were also performed to identify the effect of the *parB* mutations on plasmid stability. BR825 *E. coli* cells were transformed with pCC02 derivative plasmids that harboured the mutated *parB* genes or that contained different repeats of the partition sites. The stability of the mutated plasmids was compared with two controls, positive and negative. BR825 *E. coli* cells were also transformed with the wild type plasmid pCC02 (positive control) and the empty plasmid pFH450 (negative control). BR825 *E. coli* cells were transformed following the protocol described in Section 2.3.2 and were grown overnight at 37°C. Two LB agar plates supplemented with chloramphenicol (Cm) (10 μg/ml) for each construct were prepared and each plate was divided into five sectors. Ten single colonies were taken from the transformation plates using sterile toothpicks and streaked on the ten sectors of LB+Cm plates (Figure 2.4). The plates were incubated overnight (preferably 16-18 hours) at 37°C. The following day, a single colony was taken from each sector and streaked on LB agar with no selective pressure (no antibiotic) and the cells were grown overnight at 37°C. The following day, the step was repeated to allow cells to grow for 25 generations without selection (Figure 2.4). The day after, ten colonies were taken from each sector of LB agar plates and stabbed, first on LB-Cm agar plate then on LB agar with no antibiotics. The plates were incubated overnight at 37°C. 100 colonies for both plates per construct were analysed and the percentage of plasmids retained was determined (Figure 2.4). The experiment was repeated in triplicate.

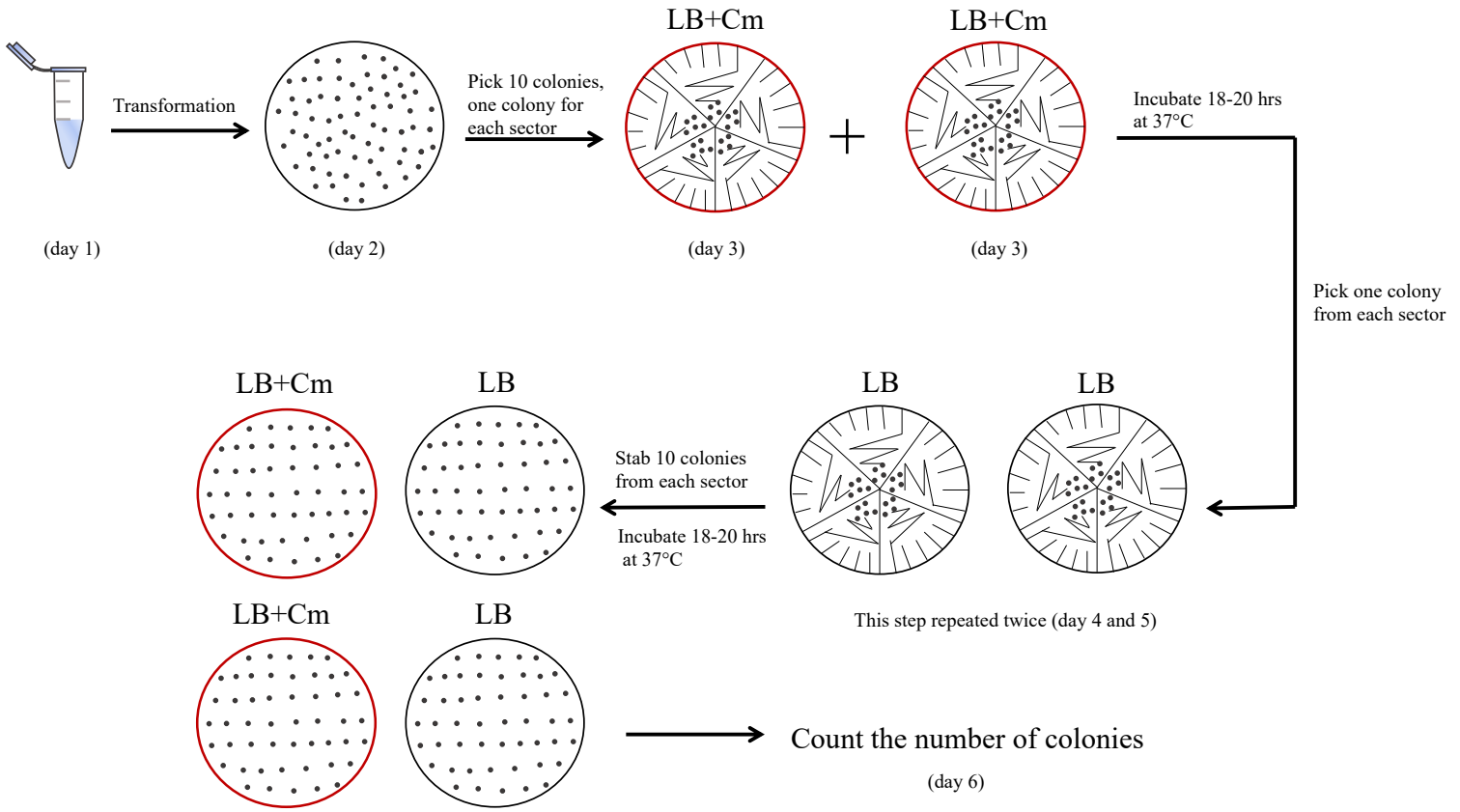


Figure 2.4 Plasmid partition assays diagram

The plasmid containing the wild-type partition cassette, the plasmid with no partition cassette and the plasmid that harbours a mutant allele are transformed into *E. coli* strain BR825 strain on day one. On day two, 10 colonies are streaked on LB agar plate supplemented with chloramphenicol. On day three, 10 colonies are picked from LB+Cm plates and streaked on LB agar plate. The previous step is repeated on day four allowing cells to grow for ~ 25 generations. On day five, 100 colonies are stabbed with a toothpick on both LB medium and LB+Cm plates. On the last day, colonies that have grown on LB+Cm plates are counted.

Chapter 3

Defining the minimal partition site of the *parAB* cassette

3.1 Introduction

The transmission of genetic information requires an accurate biological process that ensures each daughter cell receives the correct amount of genetic materials (Baxter and Funnell, 2014; Bouet and Funnell, 2019). Bacteria that harbour large and low-copy number plasmids tend to lose these plasmids as random distribution is not sufficient to maintain them. Therefore, those plasmids encode proteins responsible for ensuring their stable maintenance. These proteins constitute the plasmid segregation system (Bouet and Funnell, 2019). Three distinctive types of plasmid segregation systems have been characterised, and their mechanisms simply rely on three main elements: a NTPase protein, a centromere-binding protein and a centromere site or partition site. The molecular interactions among partition elements are crucial to driving the newly replicated plasmids apart before cell division (Baxter and Funnell, 2014; Million-Weaver and Camps, 2014; Oliva, 2016). The best studied plasmids of type I system are the P1 and F1 plasmids (Oliva, 2016). Little is known about the segregation of virulence plasmids that belong to the same type as pB171 from enteropathogenic *E. coli*. The pB171 plasmid is unique as it carries two sets of partitioning systems from two different types, type Ib and type II, whose operons are divergently arranged (Ebersbach and Gerdes, 2001; Ringgaard *et al.*, 2007a). The most prevalent between the two *par* systems is type Ib, as an *in vivo* plasmid stability experiment showed that *par1* (type II) and *par2* (type Ib) ensured 15- and 140- fold of stabilisation, respectively, when tested individually, but together they yielded 180-fold stability (Ebersbach and Gerdes, 2001).

The segregation cassette of *par2* from type Ib systems comprises the Walker-type ATPase, ParA and the centromere-binding protein ParB which binds to the *parC1* and *parC2* centromere sites to form the plasmid partition complex (Figure 3.1) (Ebersbach and Gerdes, 2001). ParA was shown to oscillate in an apparent spiral-like structure in the presence of ParB-*parC* complex through the nucleoid areas (Ebersbach and Gerdes, 2001; 2004). ParA binds to ATP and ADP. ParA was shown to assemble into long filaments within the cell in the presence of ATP (Ebersbach *et al.*, 2006). A model was proposed according to which the ParA filament bound to ParB-*parC* would drive the separation of the two newly replicated plasmids prior to cell division (Ringgaard *et al.*, 2009). However, more recent studies have provided evidence that goes against the single ParA filament model (Vecchiarelli *et al*., 2013; Lim *et al.*, 2014; McLeod *et al.*, 2017)

The *par2* cassette has two partition sites. The *parC1* site is located upstream of the *parAB* genes, whereas the *parC2* site is located downstream (Figure 3.1) (Ebersbach

and Gerdes, 2001). The *parC1* and *parC2* sites contain two types of repeats and are denoted as B repeats. Both types of repeats are hexameric but possess distinct sequences. Based on their sequence differences, they can be classified into class I and class II (Figure 3.1). The sequence of the class I repeat is [AT][TA]CA[TC]A and that of class II TTAT[GT]A, where the alternative nucleotides are between the brackets. The organisation and distribution of both classes of repeats are random (Ebersbach and Gerdes, 2001). However, the class I motif is more present in *parC1*, while class II is more frequent in *parC2*. Prior to this thesis, no work had investigated the importance of repeat arrangement and organisation in plasmid pB171 stability.

The *parC1* site consists of 17 direct repeats that are arranged into two clusters. The first cluster consists of 13 repeats (B1-B13) and the repeats are separated by a single nucleotide, while the second cluster is composed of the remaining repeats (B14-B17) also separated by only a single nucleotide. The two clusters are separated by 31 bp. The B14-B17 repeats are located in the promoter region (P2) of the *par2* system (Figure 3.1) (Ebersbach and Gerdes, 2001).

The centromeric site, *parC2*, is located downstream of the *parAB* genes and consists of eighteen hexameric direct repeats organised into three clusters. Clusters I and III are composed of 8 (B18-B25) and 7 (B29-B35) repeats respectively, while cluster II consists of only three repeats (B26-B28). Moreover, clusters I and II are separated by 8 bp, while clusters II and III are separated by 15 bp. The first repeat of the *parC2* site overlaps with the stop codon of *parB* (Figure 3.1) (Ebersbach and Gerdes, 2001).

Besides the B hexamer repeats in *parC1*, two additional identical 10 bp direct motifs, AATACTCAAT, are located 9 bp upstream of the B1 repeat. The two repeats are separated by 31 bp and named R1 and R2, respectively (Figure 3.1). These two conserved repeats are distinct from B repeats and are not recognised by ParB (Ringgaard, *et al.*, 2007a). In addition to centromere sites, the *parABC* cassette contains two additional repeats, *ihf1* and *ihf2*, which are believed to be sites recognised by the Integration Host Factor (IHF) protein (Ringgaard *et al.,* 2007a). For many years, researchers believed that many host factors could interact with components of the segregation system. However, to date, IHF has been the only host factor that has been recognised and proven to be involved in P1 plasmid partitioning (Funnell, 1988b) and likely in the segregation of pB171 plasmid (Ringgaard *et al.*, 2007a). The two *ihf* sites are in opposite orientation with the *ihf1* site located within the *parA* gene and is inverted, 504 bp downstream of the *parA* start codon. The *ihf2* site is located in the direct

orientation, overlapping with 2 bp of the second stop codon of *parB* (Figure 3.1) (Ringgaard *et al.,* 2007a). The two IHF sites that have been identified are shown to contribute to plasmid stability, as deleting the gene encoding the IHF protein from the host cell resulted in a reduction in plasmid stability (Ringgaard *et al.,* 2007a). However, the direct contribution of the IHF protein is yet to be elucidated and many questions have remained unanswered.

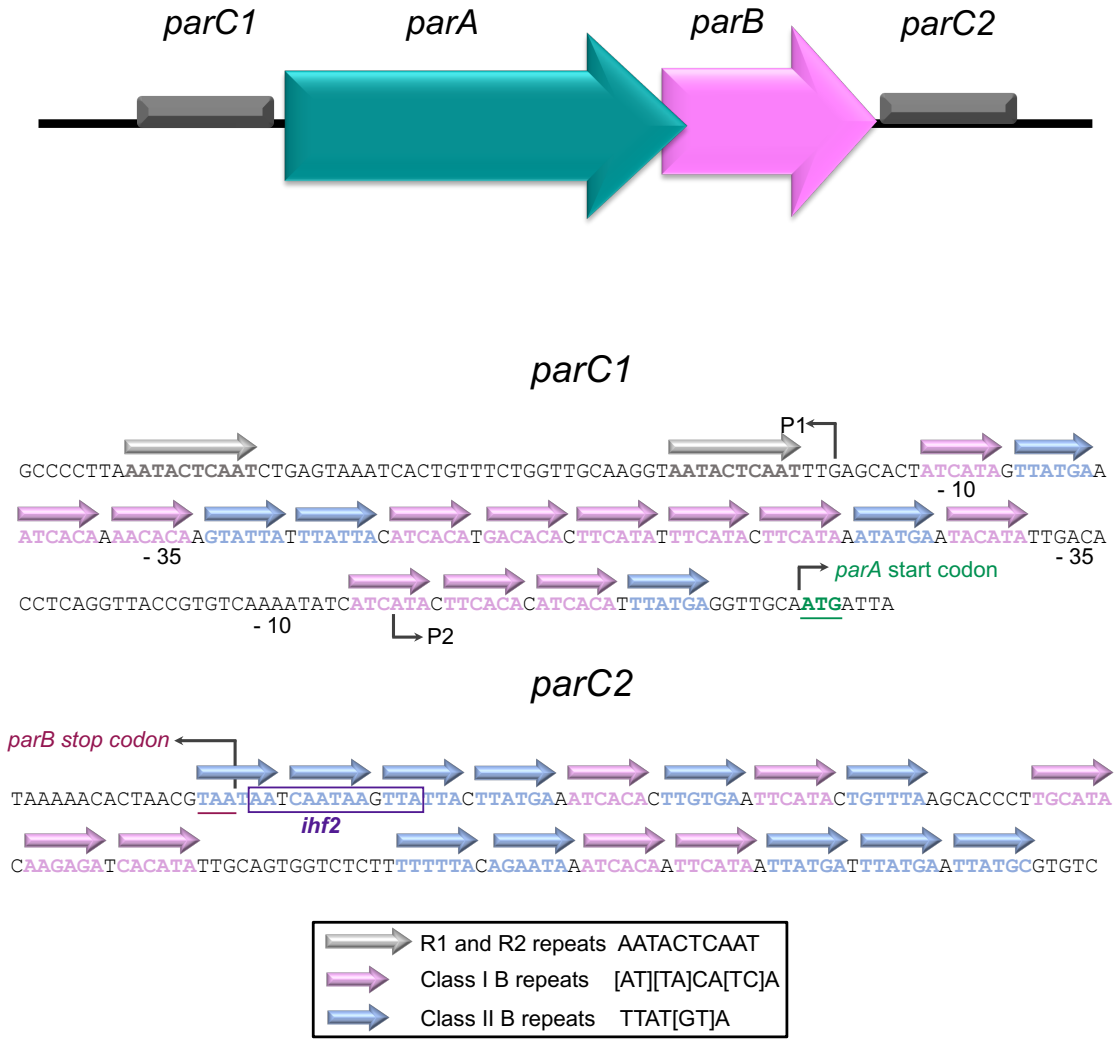


Figure 3.1 Genetic organisation of the pB171 *parABC* cassette

Genes encoding the ATPase ParA and the DNA-binding protein ParB are indicated with teal and magenta colours, respectively. The *parC1* site is located upstream and the *parC2* site is located downstream of the *parAB* genes. The two centromeres sites are shown as grey bars. ParB recognises the 6 bp direct orientation repeats. The repeats are categorised as class I (pink arrows) and class II (blue arrows). The sequences of each class of repeats are shown. The R1 and R2 repeats are indicated as grey arrows and their sequences are highlighted. The *parA* start codon in *parC1* and the *parB* stop codon in *parC2* are indicated with broken arrows and their sequences are highlighted with green and red lines respectively. The P1 and P2 promotors are indicated with broken arrows. The *ihf2* site is highlighted with a purple rectangle.

ParB is a DNA-binding protein that binds site-specifically to the B repeats in the *parC1* and *parC2* sites. *In vitro* experiments showed that ParB binds as a dimer to B repeats at the *parC1* and *parC2* sites (Fothergill *et al.,* 2005; Ringgaard *et al.*, 2007a). Prior to this thesis, ParB was predicted to have a Ribbon-Helix-Helix (RHH) C-terminal fold similar to that observed for TP228 ParG (Golovanov *et al.*, 2003; Fothergill *et al.*, 2005). During our study, the structure of pB171 ParB was solved, and details of the structure information will be discussed in the following chapter. Biochemical studies have shown that ParB dimerises and binds to the *par* sites through the C-terminal RHH fold, whereas the N-terminal region is crucial for plasmid pairing (Ringgaard *et al.*, 2007a; Ringgaard *et al.*, 2007b). In addition, ParB binds the region upstream of *parAB* operon (P2) in a site-specific manner to repress transcription (Ringgaard *et al.*, 2007a).

ParA is a Walker-type ATPase protein. *In vivo* studies showed that ParA oscillates over the nucleoid forming a helical like structure in the presence of the ParB-*parC* complex (Ebersbach and Gerdes, 2001; 2004). *In vitro* experiments showed that in the presence of ATP, ParA polymerises and forms bundles of filaments. Moreover, it was demonstrated that ParA can also bind to ADP (Ebersbach *et al.*, 2006; Machón *et al.*, 2007).

Few plasmids harbour two separate partition sites that belong to the same system and pB171 is one of them. Therefore, understanding the role played by the repeats in *parC1* and *parC2 in vivo* represents an essential early step that is crucial for understanding plasmid segregation. In this chapter, thorough investigations are carried out to identify the minimal number of repeats that make the plasmid stable *in vivo. In vivo* stability assays were employed to identify the minimal partition site. Both *parC1* and *parC2* sites have been shown to be recognised by ParB *in vitro* (Ringgaard *et al.*, 2007b), however, no *in vivo* studies have reported the contribution of the repeats to plasmid stability. The *parABC* cassette was cloned into the partition vector pFH450 plasmid before this study began (Fothergill *et al.,* 2005) generating plasmid pCC02. The pFH450 plasmid replicates at low copy number in the *E. coli* BR825 strain (Figure 3.2*A*) (Hayes, 1998). The *parABC* cassette is ~1.6 kb long and consists of *parA*, *parB* and the two partition sites. The pCC02 plasmid was shown to be quite stable *in vivo* (Figure 3.2*B*) (Fothergill *et al.,* 2005). Here, *in vivo* experiments have shown that *parC2* is crucial for plasmid segregation, but together with *parC1*, the plasmid stability is more pronounced. Moreover, these *in vivo* experiments have not revealed any preferences of the two classes of B repeats, suggesting that both repeats contribute equally to plasmid stabilisation.

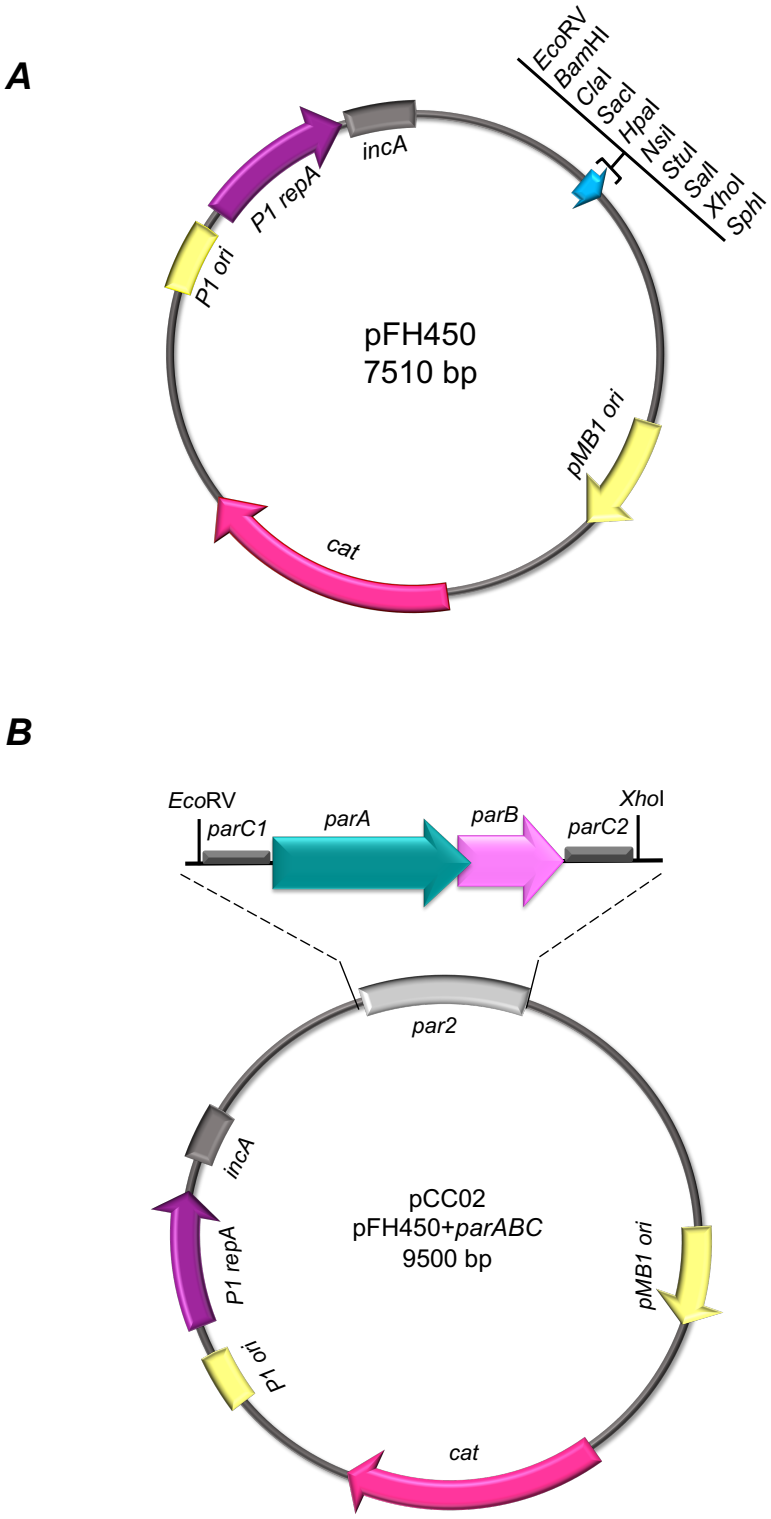


Figure 3.2 Maps of pFH450 and pCC02 plasmids

A) A pALA136 derivative plasmid containing multiple cloning sites, chloramphenicol resistance gene (*cat*), P1 (*repA*) and ColE1 (*pMB1*) replicons. ***B)*** A pALA136 derivative plasmid in which the *parABC* pB171 partition cassette is cloned into the pFH450 vector using *Eco*RV and *Xho*I restriction sites. The genetic organisation of *parABC* cassette is shown on top of the plasmid map. The genes encoding ParA and ParB are shown in teal and purple, respectively. The *parC1* is located upstream and the *parC2* site is located downstream of *parAB*.

3.2 Results

3.2.1 Construction of different truncations of the *parC1* site

The *parC1* site is located upstream of the *parAB* genes and is composed of seventeen repeats. The repeats are hexameric and arranged into two clusters with the first cluster consisting of 13 repeats, while the second cluster contains the remaining four repeats. Cluster II is located within the *parAB* promoter region (P2), as shown in Figure 3.1. Because of the critical location of the last four repeats (B14-B17) in the *parABC* cassette, these repeats have not been investigated. To assess the importance of each repeat in the *parC1* site for the plasmid partitioning system *in vivo*, several constructs were made (Figure 3.3*A*). To create each individual construct, truncated versions of the partition cassette were amplified by PCR (Section 2.3.5) using pCC02 plasmid DNA (Fothergill *et al.*, 2005) as a template, in order to systematically delete partition site repeats. The pCC02 plasmid consists of ~1.6 kb *parABC* partition cassette cloned into the pFH450 vector. The pFH450 vector is a pALA136 derivative plasmid that carries the chloramphenicol resistance gene (*cat*) and harbours two replicons, the moderate-copy number ColE1 (*pMB1*) and low-copy number P1(*repA*). The pFH450 plasmid replicates at low copy number under the control of P1 replicon when it is introduced into a *polA E. coli* host (Figure 3.2*A*) (Hayes, 1998). The forward primers for all constructs were designed at the site of deletion to remove the B repeat of interest, whereas, the reverse primer was designed downstream of the *parC2* site. The forward primers contain an *Eco*RV restriction site (GATATC), whereas the reverse primer contains a *Xho*I restriction site (CTCGAG). The two restriction sites in the primers are tagged with 6 bp flanking end (either CA×3 or AT×3) for restriction digestion purposes. The amplified fragments contain the *Eco*RV and *Xho*I sites, upstream and downstream respectively, of the fragment of interest, as shown in Figure 3.3*B*. The list of primers used for this study is reported in Table 2.8 in Materials and Methods. The amplified fragment was then verified using agarose gel electrophoresis. The 1% agarose gels in Figure 3.3*B* show the amplification products of all the *parC1* truncations. The size of each fragment which corresponds to the *parABC* truncated versions is indicated.

The PCR fragment and the pFH450 vector were double digested with *Eco*RV and *Xho*I restriction enzymes, as described under Section 2.3.6 in Materials and Methods and ligated (Section 2.3.9) to generate the recombinant plasmids.

Fourteen constructs were made following a similar strategy by inserting the relevant PCR fragments, which contain the truncated *parC1* in the *parABC* partition cassette, into the pFH450 vector. The successful clones were then grown in LB medium, supplemented

with chloramphenicol for plasmid selection. The extracted plasmids were then subjected to restriction analysis with *Eco*RV and *Xho*I to excise the truncated versions of *parABC* cassette. The sizes of the digested fragments were verified by electrophoresis on a 1% agarose gel and by comparison with a 1 kb ladder (Figure 3.4). Once the size of all digested fragments was confirmed, the nucleotide sequences of all constructs were verified by Sanger sequencing to ensure no misincorporation had occurred during the cloning procedure.

All constructs were made successfully for cluster I of the *parC1* site. However, it was impossible to delete the four remaining repeats in cluster II as this cluster is located between the promoter region and the *parA* starting codon, as shown in Figure 3.1.

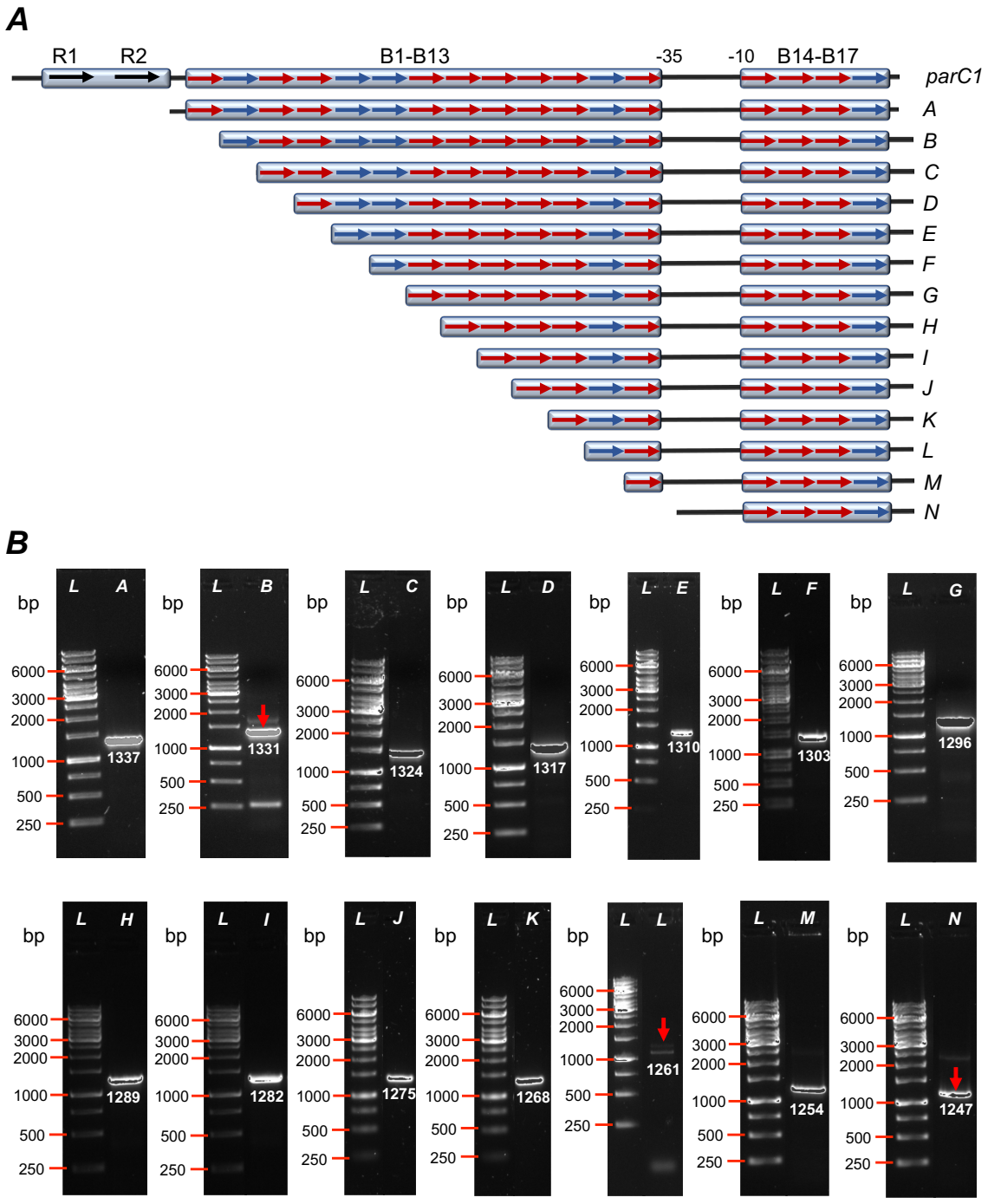


Figure 3.3 Diagram of the truncated repeats in the *parC1* site and agarose gels showing the corresponding PCR products for each individual construct

A) Diagram representing the serial truncations of the *parC1* repeats. The full *parC1* is shown on the top. The deletion of repeats starts from the left, R1 and R2, and ends with B13. The *parC1* deletions are labelled as A to N. The red and blue arrows representing the class I and class II repeats, while the two black arrows representing the R1 and R2 repeats. ***B)*** The 1% agarose gels show the PCR products generated for each individual construct. In all agarose gels, the first well contains the 1 kb ladder (Thermo Fisher Scientific) and the size of fragments are indicated on the left, while the second lane shows the PCR product. The PCR fragments named A to N, which refers to the name of each *parC1* truncated site indicated in the diagram. The higher intensities (Saturated), the more DNA concentration. The size of each fragment is indicated.

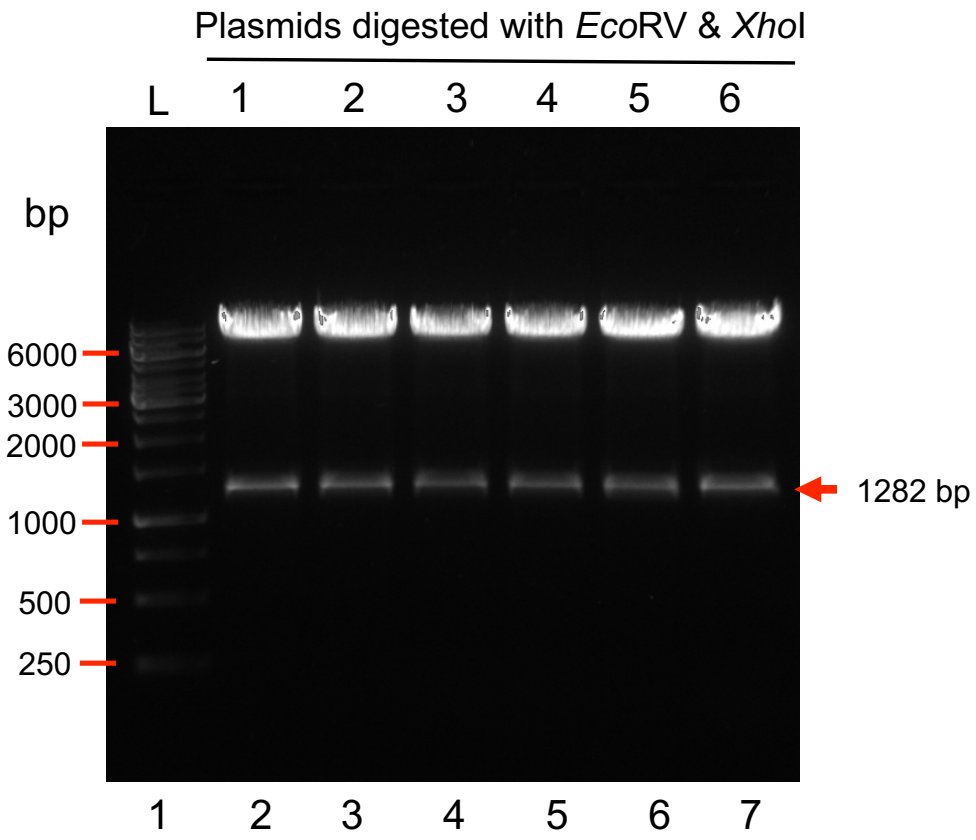


Figure 3.4 An example of the restriction digestion analysis of the positive clones

A 1% agarose gel showing the double digestion of pC1ΔB9, which corresponds to the I truncation in Figure 3.3. The lane numbers are shown at the bottom. The first lane contains the 1 kb ladder (Thermo Fisher Scientific). The size of fragments is indicated on the left. Lanes 2-7 show the restriction digestions analysis of positive clones using *Eco*RV and *Xho*I restriction enzymes. The insert is indicated by the red arrow and the size is 1282 bp.

3.2.2 *In vivo* analysis of the minimal *parC1* site of the *parABC* cassette

The pB171 virulence plasmid that is harboured by an enteropathogenic *E. coli* replicates at low-copy number. The pB171 plasmid carries two sets of partition cassettes, one from type II (*par1*) and the other from type Ib (*par2*) systems (Figure 3.1). The *par2* (*parABC*) cassette was cloned into pFH450 vector and the plasmid was named as pCC02 (Fothergill *et al.,* 2005). The plasmid map is shown in Figure 3.2*B*. The pCC02 plasmid replicates at low copy number using the P1 replicon when it is introduced into a *polA E. coli* host strain, such as BR825. The pCC02 plasmid is relatively stable and exhibits no noticeable loss after growing in non-selective pressure for ~25 generations (Fothergill *et al.,* 2005). The pFH450 plasmid vector alone does not harbour any stabilising system (Hayes, 1998) was shown to be maintained at a level of < 2% compared to pCC02 plasmid, whose stability was 50%, when it was grown for ~25 generations under non-selective pressure (Fothergill *et al.,* 2005).

The *parAB* cassette comprises two partition sites, *parC1* and *parC2*. Initially, the B hexameric repeats in the *parC1* site were probed *in vivo*. However, only repeats in cluster I were assessed as the repeats (B14-B17) in cluster II are located within the *parAB* promoter. Fourteen truncated constructs were generated for the *parC1* site, in which the repeats were deleted progressively, as shown in the previous section (Figure 3.3*A*). These fourteen truncations include the two 10 bp two identical repeats, R1 and R2, and the thirteen hexameric B repeats that form cluster I. To establish which repeats in *parC1* are crucial for plasmid retention, *in vivo* stability assays were performed. The plasmid stability assays were performed for all truncations along with two controls pCC02 and pFH450. The pCC02 plasmid was used as a positive control, as it harbours the full *parABC* partition cassette, while the pFH450 plasmid was used as a negative control, which does not harbour any stabilisation system. Assays were performed in triplicate.

The first deletion construct tested was pC1 in which the first two 10 bp identical repeats were deleted. The plasmid showed a high level of stabilisation *in vivo* and showed a retention level (79%) similar to that of the wild type pCC02 (Figure 3.5). This result suggests that these two repeats, R1 and R2, are not involved in plasmid stabilisation and they are thus dispensable. This result is consistent with an *in vitro* experiment, in which ParB was unable to recognise these two repeats (Ringgaard, *et al.*, 2007), implying that these repeats are not crucial for plasmid stability. When the first two repeats B1 and B2, were deleted from the *parC1* site, the plasmid pC1ΔB1 and pC1ΔB1-B2 were maintained at a level similar to that of the plasmid containing the full *parABC* cassette, with 77% and 78% retention level respectively (Figure 3.5). When three repeats (B1-B3) are deleted

together, the pC1ΔB1-B3 plasmid was shown to be slightly less stable than the wild type. The percentage of plasmid retention level for this mutant was 72%, which is ~9% less stable than the pCC02 plasmid (Figure 3.5). However, when four B repeats were deleted (B1-B4), the pC1ΔB1-B4 plasmid was shown to be more stable (~5%) than the plasmid harbouring B4-17 of *parC1* (ΔB1-B3) (Figure 3.5). A reason could be that minor changes in the growth environment such as temperature, incubation times and the composition of media might cause a slight effect on plasmid retention. This variation was not only seen in the mutants but also in the level of stability of the pFH450 plasmid that lacks any stabilising system and the pCC02 plasmid that harbours full partition cassette.

As the analysis of further deletions progressed, each subsequent plasmid showed lower stability compared to the previous one, i.e. the plasmid containing one additional repeat in the *parC1* partition site. Specifically, the plasmid showed a slight reduction in retention level when seven hexameric B repeats were deleted. The reduction in plasmid stability became more pronounced, when the following six hexameric B motifs (B1-B14) were truncated. Precisely, the last six mutants (pC1ΔB1-B8, pC1ΔB1-B9, pC1ΔB1-B10, pC1ΔB1-B11, pC1ΔB1-B12 and pC1ΔB1-B13) showed similar retention level approximately ~20% less than that of the plasmid that harbours the full *parC1* (Figure 3.5). However, we could not assess the contribution of the remaining B repeats in the *parC1* site since they are located in the *parAB* promoter region. These results suggest that the partition cassette is either relying on the *parC2* partition site to maintain the plasmid or the remaining repeats in *parC1* are sufficient for plasmid stabilisation. To investigate these hypotheses, new constructs were generated, and the results will be described in the following section.

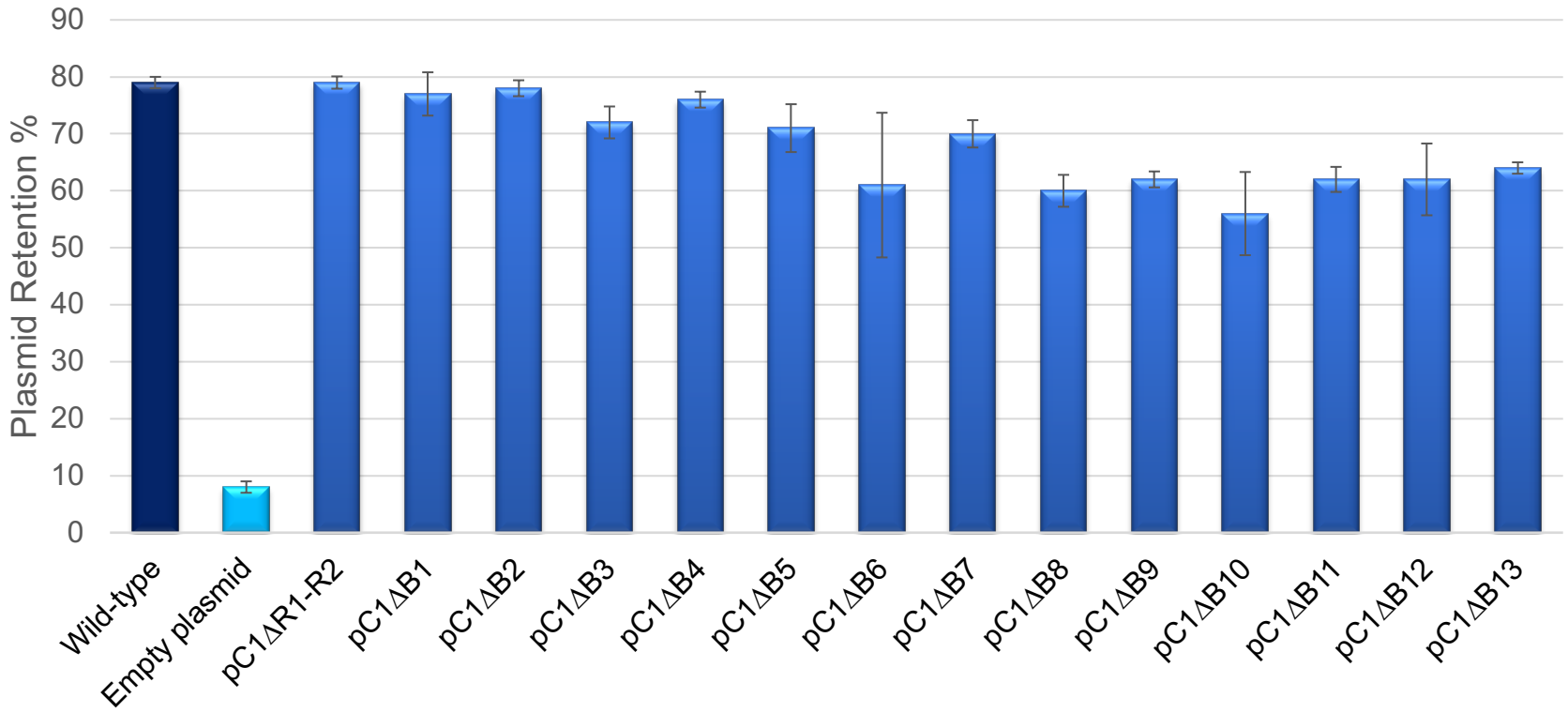


Figure 3.5 Percentages of plasmid retention of the truncated *parC1* constructs

Wild type (dark blue bar) represents the pCC02 plasmid that harbours the full *parABC* cassette and represents the positive control for the experiment. The empty plasmid (cyan bar) represents the pFH450 plasmid that lacks the stabilising systems and represents the negative control. Retention of the low copy number *parC1* mutant plasmids (blue bars) is measured via plasmid partition assays. The number of repeats in each pC1 plasmid, from left to right, represent the same order of *parC1* mutants shown in Figure 3.3*A*. The percentages represent the average of at least three independent experiment performed along with pFH450 and pCC02. Error bars represent the standard error of the mean.

3.2.3 Construction of different truncations of the *parC2* partition site

The *parC2* site is located downstream of the *parAB* genes and comprises eighteen repeats. The repeats are hexameric and arranged into three clusters with the first cluster consisting of eight repeats (B18-B25), while the middle cluster contains three repeats from class II and the third cluster consists of seven repeats (B29-B35). The repeats in *parC2* and *parC1* are related and all are in direct orientations. The first repeat B18 overlaps with the *parB* stop codon. The diagram of B repeats in *parC2* are shown in Figure 3.1. Similarly to the *parC1* site, the importance of each repeat in the *parC2* site for plasmid stability *in vivo* was assessed. To perform this analysis, a set of constructs were generated (Figure 3.6). The same sequential deletion strategy as for creation of *parC1* mutant plasmids was employed. The cassette that contains the truncated version missing repeats was amplified by PCR (Section 2.3.5) using pCC02 plasmid (Fothergill *et al.*, 2005) as a template in a similar way to the constructs made for truncation repeats in *parC1*. The amplified fragment that contains the truncated *parC2* in *parABC* partition cassette was cloned into the pFH450 vector. Reverse primers for all constructs were designed at the site of deletion to remove the B repeat of interest, whereas, the forward primer was designed upstream of *parC1* site. The forward primer contains an *Eco*RV restriction site (GATATC), whereas the reverse primers contain a *Xho*I restriction site (CTCGAG). The amplified fragments contain the *Eco*RV and *Xho*I sites, upstream and downstream, respectively. All the generated fragments are shown in Figure 3.7. The list of primers used for this study is shown in Table 2.8 in the Materials and Methods chapter. The amplified fragments were then verified using agarose gel electrophoresis. The 1% agarose gels in Figure 3.7 show the amplification products of all truncation repeats in the *parC2* and the size of each fragment is also indicated.

The gene was inserted using *Eco*RV and *Xho*I sites followed by ligation as described in Section 2.3.9. The resulting colonies were screened using colony PCR as described in Section 2.3.10. Positive clones were then grown in LB medium supplemented with chloramphenicol overnight for plasmid isolation. The extracted plasmids were then subjected to restriction analysis with *Eco*RV and *Xho*I (Figure 3.8). The size of the inserted fragments was verified by 1% agarose gel (Figure 3.8). Once the size of all digested fragments was confirmed, the nucleotide sequences of all constructs were verified by Sanger sequencing to ensure no mistake had occurred during the DNA amplification procedure.

In the previous section, it was mentioned that deleting the last four B motifs from *parC1* site is not possible due to their critical location, which will affect the *parAB* transcription.

Furthermore, we hypothesized that the repeats in cluster II could be sufficient for plasmid partitioning. To prove this, a new construct was generated. A PCR fragment was amplified, which contained four repeats in *parC1* (B14-B17) and only one repeat in *parC2* (B18) (Figure 3.9*A*). Two primers were designed, a forward primer generated downstream of B13 containing the *Eco*RV site. The reverse primer was designed upstream of B18 repeat containing the *Xho*I site. Figure 3.9*A* shows the *parAB*-B14-B18 partition cassette. Once the PCR fragment was verified on a 1% agarose gel (Figure 3.9*B*), it was digested along with pFH450 with *Eco*RV and *Xho*I. The PCR product and digested vector were ligated. Positive clones were confirmed using restriction digestion (Figure 3.9*C*) and the sequences were verified by Sanger sequencing.

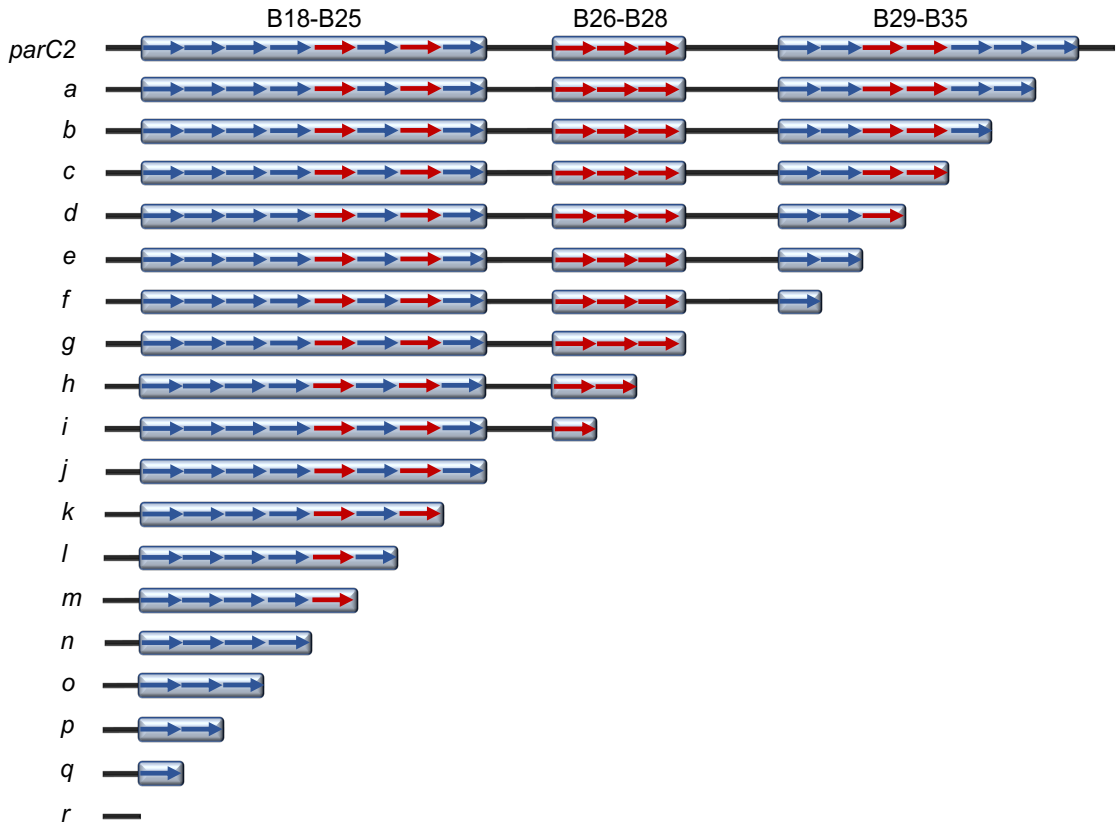


Figure 3.6 Diagram showing the truncation of repeats in the *parC2* site

The full *parC2* is shown on the top. The deletion of repeats starts from right B35 on the right end of the partition site and ends with B18. The *parC2* truncations are labelled as a to r. The red and blue arrows representing the class I and class II repeats, respectively.

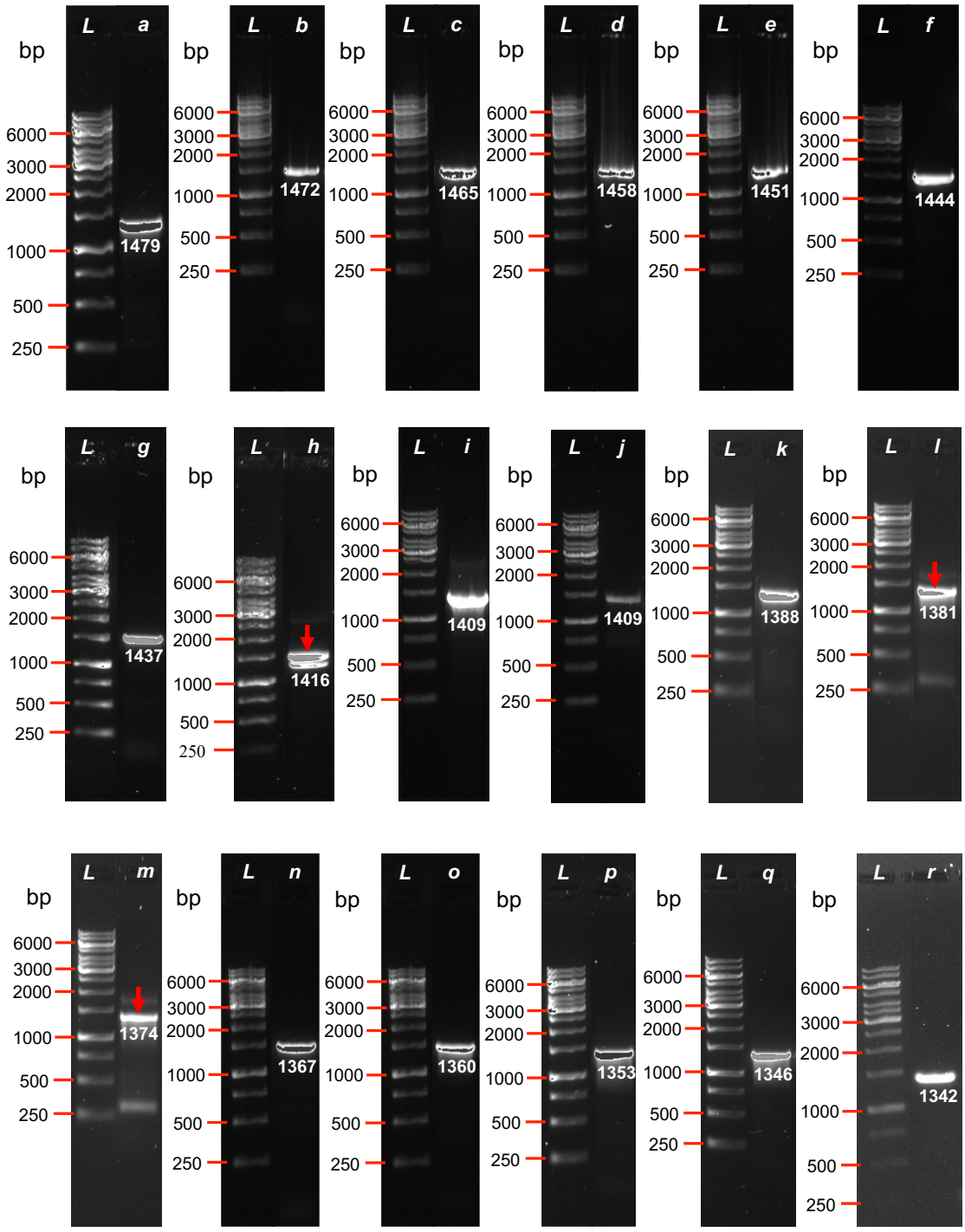


Figure 3.7 Agarose gels showing the PCR products of the different *parC2* constructs

The 1% agarose gels show the PCR products generated for each construct. In all agarose gels, the first well contains the 1 kb ladder (Thermo Fisher Scientific) and the size of fragments are indicated, while the second lane shows the PCR product. The PCR fragments are named a to r, which represents the name of each *parC2* mutant showed in Figure 3.6. The higher intensities (Saturated), the more DNA concentration. The size of each fragment is indicated.

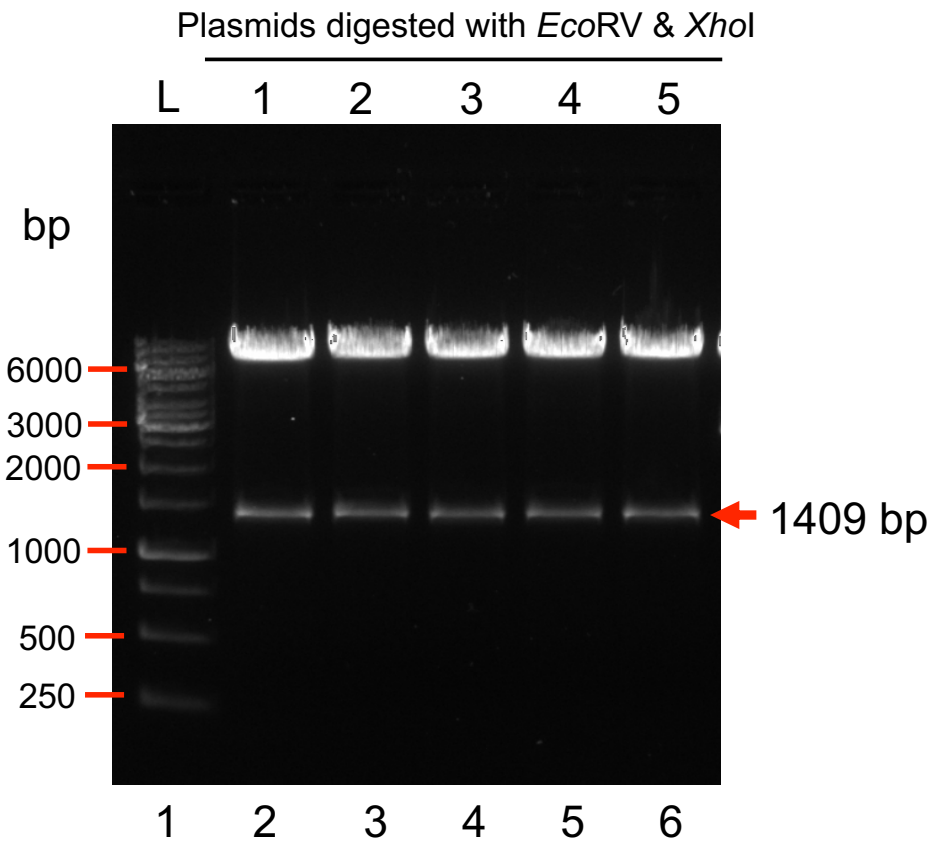


Figure 3.8 An example of positive clones restriction digestion analysis

A 1% agarose gel showing the double digestion of the pC2ΔB27-35, which corresponds to the i truncation in Figure 3.6. The first lane shows the 1 kb ladder (Thermo Fisher Scientific). Lanes 2-6 represent the restriction digestions analysis of positive clones using *Eco*RV and *Xho*I restriction enzymes. The size of the insert is 1409 bp and is indicated by the red arrow.

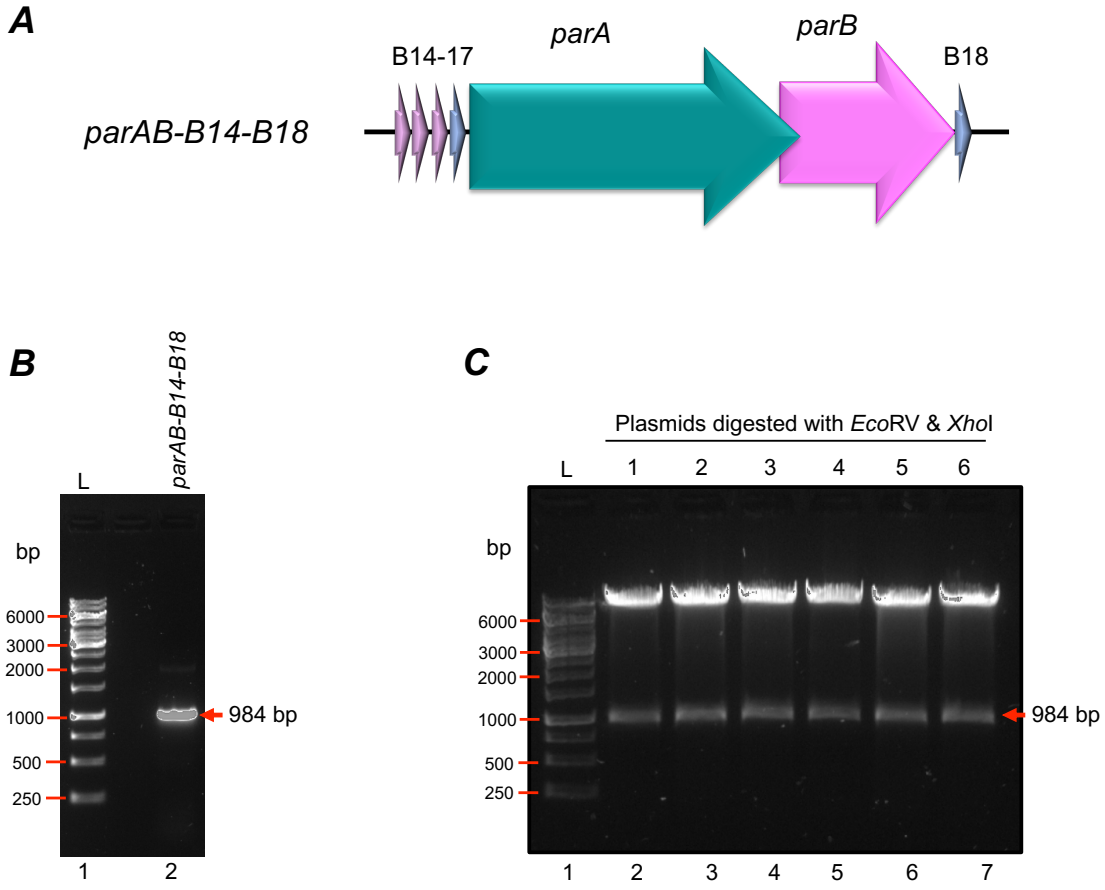


Figure 3.9 Cloning steps for generation of the pCC02-B14-B18 plasmid

A) Genetic organisation of the *parAB-B14-B18* cassette. Genes encoding the ATPase ParA and the DNA-binding protein ParB are shown in teal and magenta arrows, respectively. B14 to B18 hexameric repeats from *parC1* are located upstream, while B18 hexameric repeat from *parC2* is located downstream. ***B)*** 1% agarose gel showing the parAB-B14-B14 PCR product. The lane numbers are shown at the bottom. Lane 1 represents the 1 kb ladder (Thermo Fisher Scientific). Lane 2 represents the 984 parAB-B14-B18 PCR products and the fragment is indicated by the red arrow. ***C)*** 1% agarose gel showing the double digestion of the pCC02-B14-B18. The lane numbers are shown at the bottom. The first lane contains the 1 kb ladder (Thermo Fisher Scientific) and the size of fragments are indicated. Lanes 2-7 show the restriction digestions analysis of the positive clones using *Eco*RV and *Xho*I restriction enzymes. The size of the insert is 984 bp and is indicated by the red arrow.

3.2.4 *In vivo* analysis of the minimal *parC2* partition site of the *parABC* cassette

The hexameric B repeats in the *parC1* site were investigated *in vivo*. However, only repeats in the cluster I were assessed. It was shown that the plasmid bearing only four repeats in *parC1* was maintained with a 60% retention level, which is ~20% lower than that observed for the full *parC1*. The plasmid that carries only four repeats in *parC1* showed a reasonable stability *in vivo*. Based on these results, it is difficult to conclude whether *parC1* is necessary or not for plasmid stabilisation. To investigate further, the repeats in *parC2* were analysed by generating eighteen constructs, in which the repeats from B18-B35 were deleted in sequential order starting from B35 (Figure 3.6). It should be noted that all truncations generated are still contained the full *parC1* site. Each truncated cassette is assessed individually along with pCC02, that contains the full *parABC* cassette and pFH450 that lacks any stability system. The plasmid stability assays were performed in triplicate.

The first truncation construct examined was pC2Δ35, in which the last repeat (B35) was deleted from the *parC2* partition site. The plasmid showed a high level of stabilisation *in vivo* and a retention level of 75% that is ~5% lower than that of the wild type pCC02 (Figure 3.10). A similar retention level was observed when B34 was deleted from *parC2* (Figure 3.10). The plasmid stabilisation decreased to 63%, when the last three hexameric boxes (B33-B35) were deleted from *parC2* (Figure 3.10). Surprisingly, the plasmid was retained with higher stabilisation when the next three repeats were deleted, compared to the first three hexameric (Figure 3.10). As previously discussed, this variation in plasmid stabilisation could be associated with minor changes in growth factors that may have slightly affected plasmid stabilisation. Progressive deletion of hexameric B motifs from the centromere site *parC2* was accompanied by gradual reductions in plasmid retention level. Specifically, the plasmid was maintained at 57% retention level, when seventeen B motifs (B19-B35) were deleted compared to the full *parC2* (Figure 3.10). Notably, the mutant plasmid bearing only a repeat from *parC2* and full *parC1* was maintained at 57%, which is ~30% less than the stability mediated by the full *parABC* cassette (Figure 3.10). Overall, this result may indicate that both centromere sites *parC1* and *parC2* could work together more efficiently to effect plasmid stability. However, when the *parC2* site was completely deleted, the pCC02ΔparC2 plasmid could not be maintained and the retention level was negligible (0%) although it contains the full *parC1* partition site (Figure 3.10). This result suggests that the plasmid segregation system relies mainly on the *parC2* site. As previously described, there are two classes

of hexameric repeats, although due to the random distribution between the two centromeres, class II repeats are more frequent in *parC2* and their DNA sequence is more conserved. A plausible hypothesis is that class II repeats could be essential in centromere sites. However, this possibility required further investigations.

Moreover, the plasmid stability deficiency of pCC02∆parC2 could also answer another question, which is whether the four remaining repeats in *parC1* are sufficient for the plasmid to be maintained *in vivo*. This also indicates that *parC1* could not be involved in centromere activity when *parC2* is absent. It was assumed that the two partition sites could contribute equally to plasmid partitioning. Based on this hypothesis, a new construct was generated to assess the *parC1* repeats that are located in the promoter region. The construct comprises four repeats (B14-B17) from *parC1* and one repeat (B18) from *parC2* (Figure 3.9*A*). The plasmid displayed a level of stability of 52% (Figure 3.11). This result also suggests that the first repeat in *parC2* and the four repeats of *parC1* are sufficient to mediate the plasmid stability moderately. However, the role played by the *parC2* site in centromere activity is still conclusive.

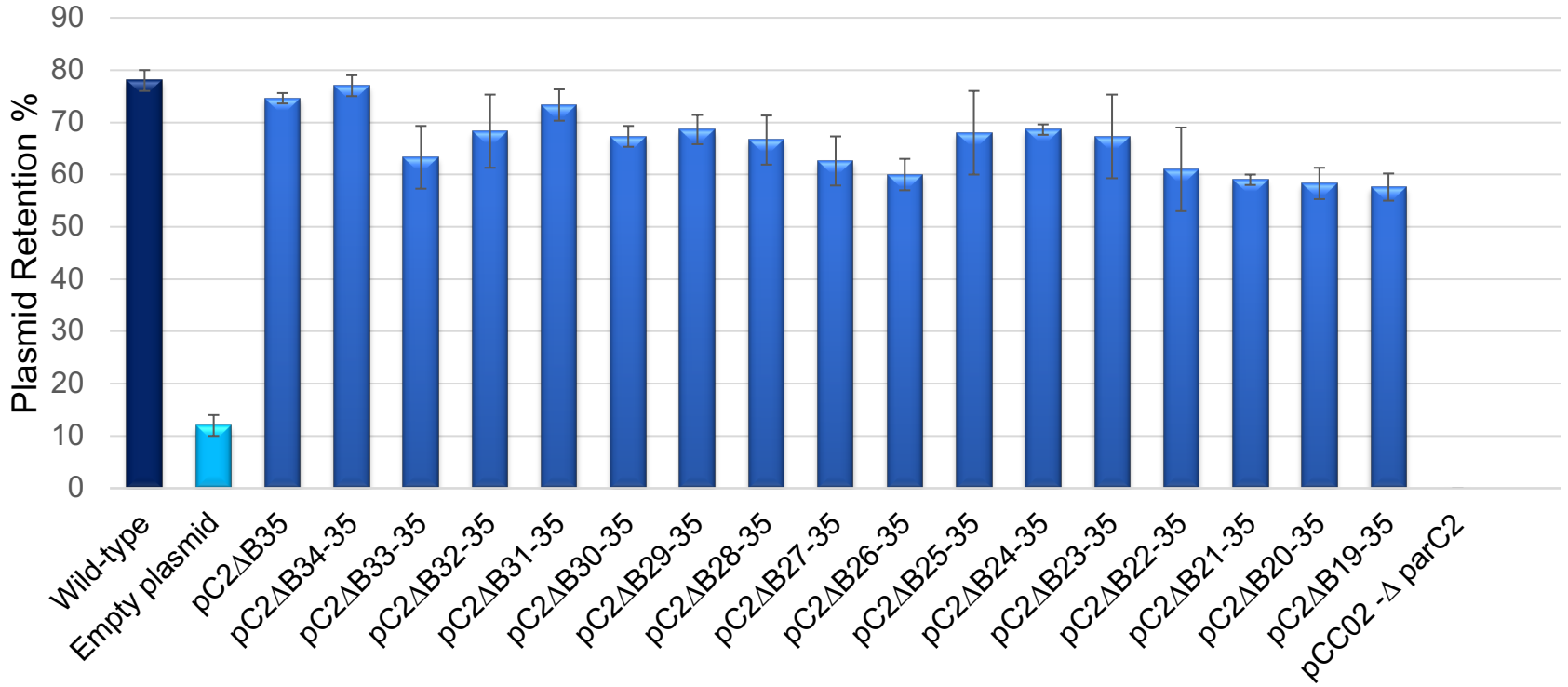


Figure 3.10 Percentages of plasmid retention of the truncated *parC2* constructs

Wild type (dark blue bar) represents the pCC02 plasmid that harbours the full *parABC* cassette and represents the positive control for the experiment. The empty plasmid (cyan bar) represents the pFH450 plasmid that lacks the stabilising systems and represents the negative control of the experiments. Retention of the low copy number pC2 mutant plasmids (blue bars) that contain the full *parC1* site, is measured via plasmid partition assays. The number of repeats in each pC2 plasmid, from left to right, represent the same order of *parC2* mutants shown in Figure 3.6. The percentages represent the average of at least three independent experiments that were performed along with pFH450 and the pCC02. Error bars represent the standard error of the mean.

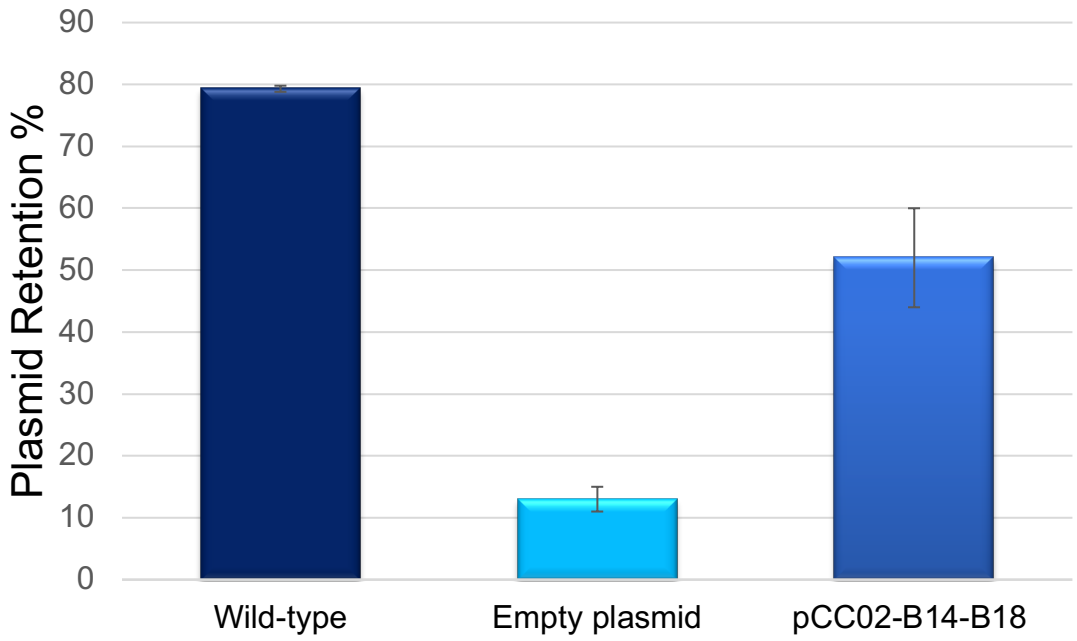


Figure 3.11 Stability assays of the pCC02-B14-B18 plasmid

Wild type (dark blue bar) represents the pCC02 plasmid that harbours the full *parABC* cassette and represents the positive control for the experiment. The empty plasmid (cyan bar) represents the pFH450 plasmid that lacks the stabilising systems and acts as the negative control of the experiments. Retention of the low copy number mutant pCC02-B14-B18 plasmids that harbours four repeats from *parC1* (B14-B17) and a single repeat from *parC2* (B18) (blue bars) is measured by plasmid partition assays. The percentage represents the average of at least three independent experiments that performed along with pFH450 and the pCC02. Error bars represent the standard error of the mean.

3.2.5 The importance of class II repeats position in the *parC2* partition site

3.2.5.1 The pCC02ΔB18 and pCC02ΔB18-B21 plasmids construction for *in vivo* analysis

We hypothesised that class II repeats in *parC2* could be necessary and fundamental for plasmid partitioning. The first four B motifs (B18-B21), which are class II repeats, are highly conserved. Based on this observation, two constructs were designed and generated. The B18 repeat is missing from the first construct, while in the second construct, the first four B repeats (B18-B21) from *parC2* are deleted (Figure 3.13*A*). To generate these two mutants' plasmids, two pieces of fragments were amplified using PCR. The two fragments were generated using pCC02 plasmid as a DNA template. The first PCR fragment was amplified for both constructs by using two primers. The forward primer was designed upstream of *parC1* and includes an *Eco*RV site while the reverse primer was designed downstream of the *parB* gene and contains a *Sac*I site. Figure 3.12*B* shows the parABC1 partition fragment. Once the PCR product was verified on a 1% agarose gel (Figure 3.12*B*), it was digested along pFH450 vector with *Eco*RV and *Sac*I. The *parABC1* fragment and digested vector pFH450 were ligated, and the genetic organisation of the *parABC1* partition cassette is shown in Figure 3.12*A*. Positives clones were verified using restriction digestion (Figure 3.12*C*) followed by Sanger sequencing to ensure no misincorporation had occurred during the cloning procedure. Later this plasmid will be used as a vector for generating two constructs.

The second two pieces of PCR fragments were generated by amplifying the region downstream of the *parAB* genes (*parC2* only). To generate these two fragments, two forward primers were designed. The first one from B19 repeats to eliminate the B18 motif from the *parC2* site, while the second forward primer starts from B22 to remove the first four hexameric boxes from *parC2* (Figure 3.13*A*). The two forward primers include the *SacI* restriction site. The same reverse primer that has been used to create constructs for truncation repeats in *parC1*, as mention in section (3.2.1), was used. The two PCR products are ~260 (*parC2*ΔB18) (Figure 3.13*B*) and 250 bp (*parC2*ΔB18-B21) respectively. The PCR fragments are subjected to restriction endonuclease analysis along with pCC02Δ*parC2* using *Sac*I and *Xho*I. The two truncated *parC2* fragments were cloned into two pCC02Δ*parC2* separately. The successful clones were verified by restriction analysis using *Eco*RV and *Xho*I (Figure 3.13*C*). The nucleotide sequences were then confirmed by Sanger sequencing.

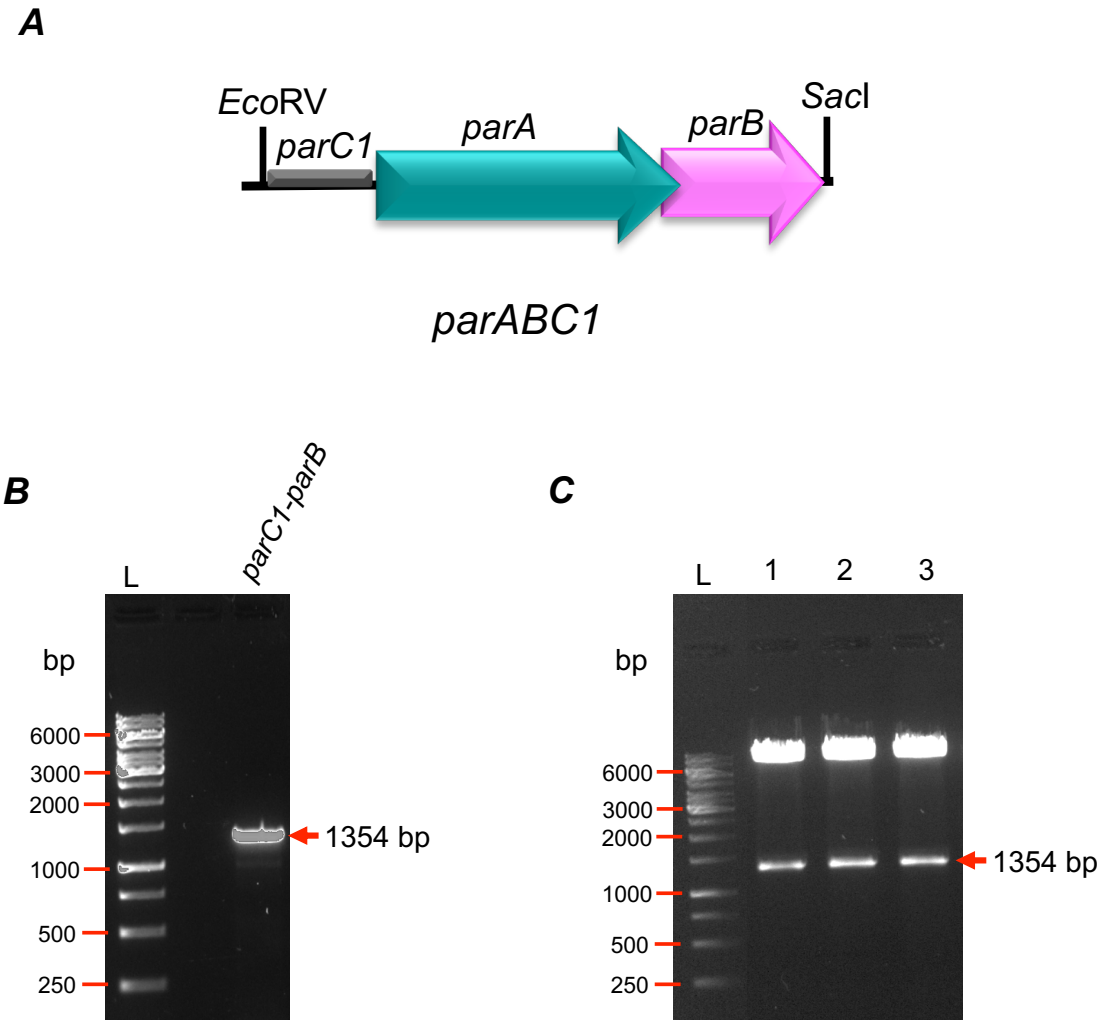


Figure 3.12 Construction of the pCC02ΔparC2 plasmid

A) Genetic organisation of the *parABC1* cassette. Genes encoding the ATPase ParA and the DNA-binding protein ParB are shown in teal and magenta arrows respectively. The *parC1* site locates upstream of the *parAB* genes. ***B)*** 1% agarose gel shows the *parABC1* product. Lane 1 contains the 1 kb ladder (Thermo Fisher Scientific) and the size of fragments are indicated on the left. Lane 2 shows the 1354 bp *parABC1* PCR products and the fragment is indicated by the red arrow. ***C)*** 1% agarose gel showing the double digestion of the pCC02Δ*parC2*. The first lane represents the 1 kb ladder (Thermo Fisher Scientific) and the size of fragments are indicated. Lanes 2-4 represent the restriction digestions analysis of the positive clones cut with *Eco*RV and *Sac*I *restriction* enzymes. The size of the insert is 1354 bp and is indicated by the red arrow.

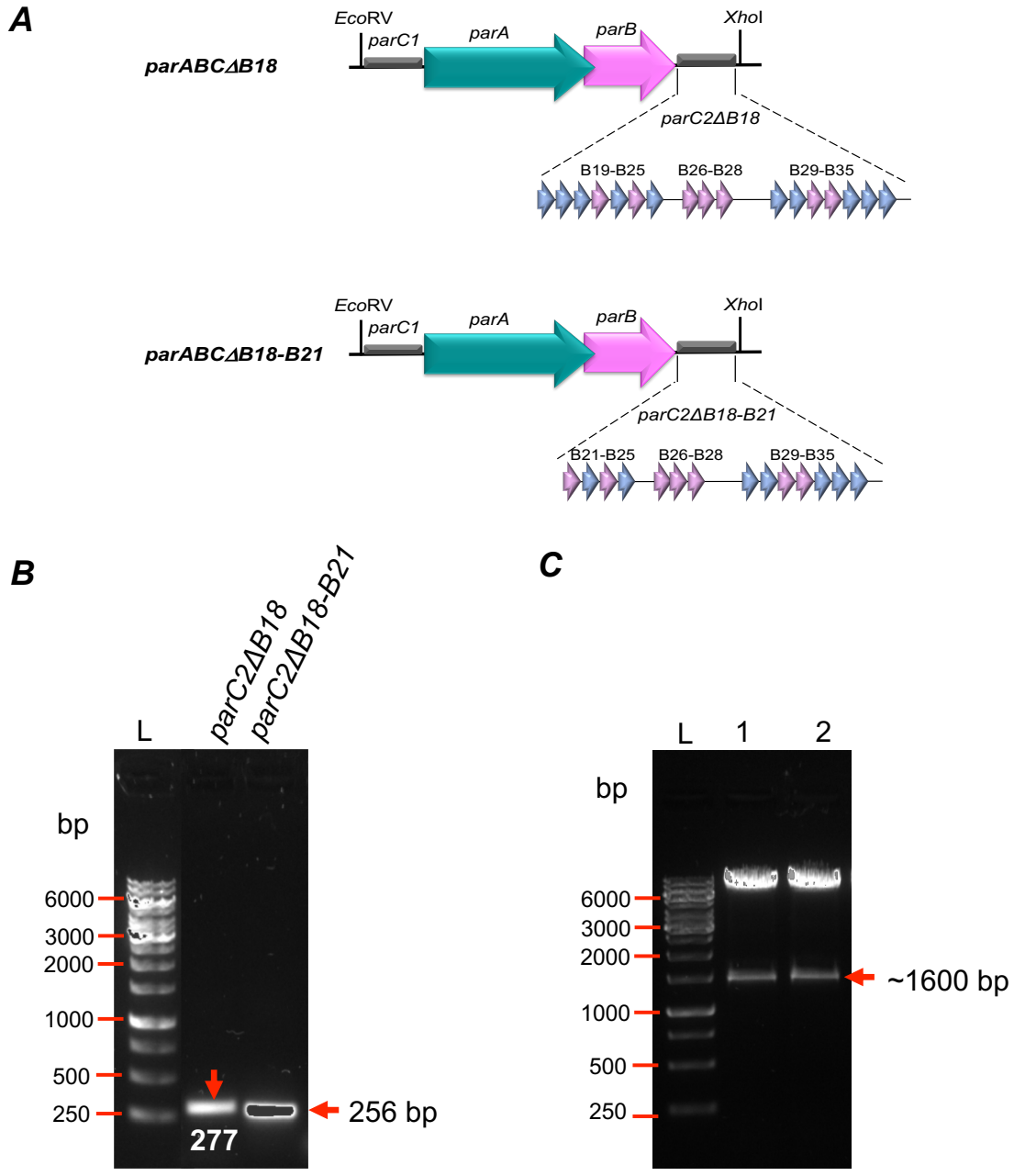


Figure 3.13 Construction of the pCC02ΔB18 and pCC02ΔB18-B21 plasmids

A) Diagrams showing the genetic organisations of *parABCΔB18* and *parABCΔB18-B21*. The arrangement of repeats in the two *parC2* deletion, *parC2ΔB18* and *parC2ΔB18-B21*, are indicated with pink arrows (class I) and blue arrows (class II). The two *parC2* deletions are shown below the full *parC2* site. ***B)*** 1% agarose gel showing the *parC2ΔB18* and *parC2ΔB18-B21* PCR products. Lane 1 contains the 1 kb ladder (Thermo Fisher Scientific) and the size of fragments are indicated. Lanes 2 and 3 show the 277 bp *parC2ΔB18* and the 256 bp *parC2ΔB18-B21* PCR products, respectively. The fragments are indicated by red arrows. ***C)*** 1% agarose gel showing the double digestion of the pCC02ΔB18 and pCC02ΔB18-B21 plasmids. The first lane contains the 1 kb ladder. Lanes 2 and 3 show the restriction digestions analysis of the positive clones cut with *Eco*RV and *Xho*I restriction enzymes. The size of both inserts is ~1600 bp and are indicated by the red arrow.

3.2.5.2 *In vivo* analysis of pCC02ΔB18 and pCC02ΔB18-B21

We speculated that the first four repeats could be essential for plasmid stability. We have also shown that deleting the *parC2* site abolishes plasmid segregation, possibly due to the B18 motif being crucial for plasmid stability. To test these two hypotheses, further *in vivo* investigations were carried out to assess the importance of the B18 and B18-B21 repeats. Stability assays were performed in triplicate for each plasmid and in parallel with two controls, pFH450 and pCC02, as negative and positive controls, respectively. Neither the deletion of the B18 repeat nor that of the B18-B21 repeats showed a significant reduction in stability (Figure 3.14). The plasmid carrying all repeats but B18 displayed a similar plasmid stability as that of the plasmid bearing the full partition sites (Figure 3.14). The retention level of pCC02ΔB18 was identical to that observed for the plasmid with the full partition cassette and it was 76%. The second plasmid pCC02ΔB18-B21 was shown to be less effective centromere. Specifically, the pCC02ΔB18-B21 plasmid that is missing four repeats in *parC2* was maintained at 66.6%, which is 15% lower than the level of stability provided by the full centromere sites (Figure 3.14). Although the plasmid showed lower stability, this does not indicate these repeats are crucial. The results indicate that the B18 and B18-B21 of *parC2* are not essential repeats. However, it was shown that deletion of the *parC2* partition site severely affects plasmid stability. Thus, these observations suggest that any repeat in the *parC2* site is important for plasmid stability.

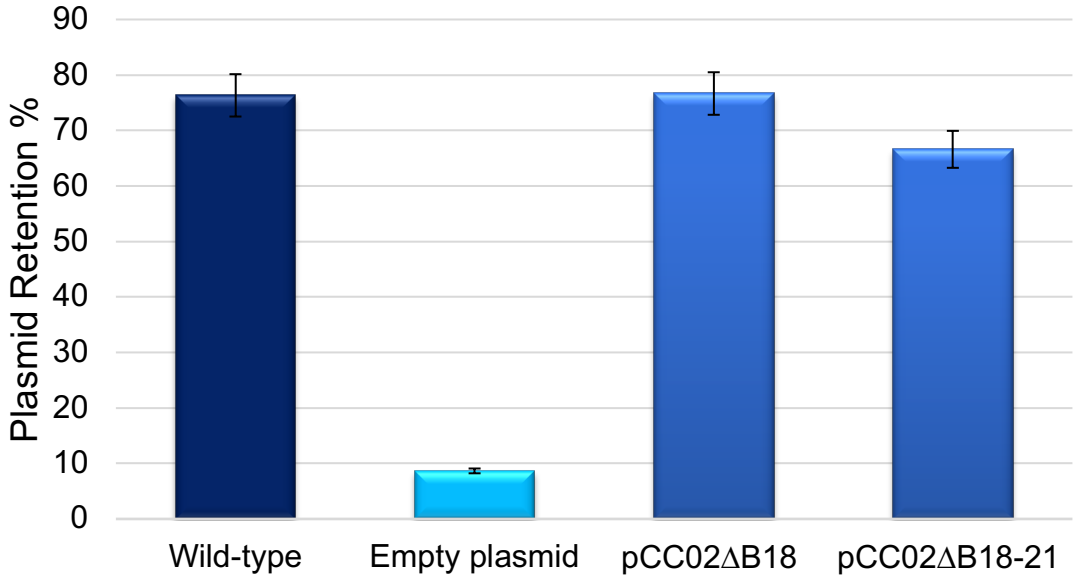


Figure 3.14 Percentages of plasmid retention of the class II mutant constructs

Wild type (dark blue bar) represents the pCC02 plasmid that harbours the full *parABC* cassette and represents the positive control for the experiment. The empty plasmid (cyan bar) represents the pFH450 plasmid that lacks the stabilising systems and represents the negative control of the experiments. Retention of the low copy number pCC02ΔB18 and pCC02ΔB18-21 plasmids (blue bars) are measured via plasmid partition assays. The percentages are the average of at least three independent experiments. Error bars represent the standard error of the mean.

3.3 Conclusions and discussion

Plasmid centromere sites are highly divergent in different partitioning systems. Typically, the centromere site consists of tandem repeats, though their sequences, lengths, numbers and orientations are extremely variable among plasmid partition systems (Bouet and Funnell, 2019). The centromere can be located upstream or downstream of the *par* genes (Hayes and Barillà, 2006a). Most of the plasmid partition systems possess one centromere site, although there are exceptions. The plasmid segregation system in the pSM19035 plasmid is composed of three centromeres (Dmowski and Kern-zdanowicz, 2016), while the *par2* of pB171 comprises two (Ringgaard, *et al.*, 2007a). In addition to repeats in the centromere site, plasmids like P1 possess an additional site, which is the IHF site (Funnell, 1988b). In P1 plasmid, the centromere site *parS* comprises three regions, left, middle and right regions. The IHF recognising site is located in the middle region between the two arms, left and right (Schumacher *et al.*, 2007b). It was shown that the IHF site is involved in the partition system by bending the DNA and permitting the ParB protein to bind multiple sites and form partition complexes with distinctive structure. However, in the absence of IHF site, ParB can still bind to the right arm with high affinity but cannot bind to the left arm (Schumacher *et al.*, 2007).

The *par2* system of the pB171 plasmid comprises two centromeres, *parC1* and *parC2*. The *parC1* is located upstream of the *parAB* genes, whereas *parC2* is located downstream (Ebersbach and Gerdes, 2001; Ringgaard *et al.*, 2007a). The two sites are composed of seventeen and eighteen hexameric motifs respectively. All repeats are arranged in direct orientation. The repeats in *parC1* site are arranged into two clusters of 14 and 4 boxes respectively, while the repeats in *parC2* are organised into three groups of eight, three and seven motifs, respectively (Ebersbach and Gerdes, 2001). Gel retardation assays have shown that the ParB protein similarly binds to *parC1* and *parC2* with high affinity (Ringgaard, *et al.*, 2007a). However, no evidence has been provided yet to show the involvement of centromeres in plasmid segregation in living cells.

In this chapter, we investigated the importance of the repeats for plasmid partitioning systems *in vivo*. Plasmid stability assays were employed to understand the importance of centromeres in plasmid stability and the contribution of each repeat was assessed individually. The *par2* partition cassette cloned into plasmid pFH450 conferred a 50-fold stabilisation compared to the empty vector pFH450 plasmid that lacks any maintenance system (Fothergill *et al.,* 2005). *In vivo* study reveals that the two repeats R1 and R2, are not crucial. Remarkably, a plasmid harbouring only four hexameric B repeats (B14-B17)

in *parC1* had ~20% lower stability compared to the plasmid harbouring the full *parC1* partition site, meaning that the remaining four repeats in *parC1* are sufficient for centromere activity. However, the optimal stabilisation is achieved when all the repeats are present. Moreover, ParB perhaps requires only limited number of repeats to assemble into partition complexes. In contrast, a plasmid harbouring only two clusters of iterons in *parC2* was maintained at the same level as the full *parC2*. However, a plasmid harbouring only one single repeat from *parC2* was modestly less stable than a plasmid harbouring the full *par* cassette. This indicates that there is a synergy between the two centromeres and that they may depend on one other for more effective plasmid stabilisation activity. Surprisingly, the plasmid is unable to be maintained in the absence of the whole *parC2* partition site and showed a retention level lower than that of the empty pFH450 that does not harbour any partition system. This suggests that the *parC2* site could work as a recognition nucleation site for ParB before it binds to *parC1*. Alternatively, an interaction between the two partition sites *parC1* and *parC2*, possibly mediated by ParB, might be necessary to form a partition complex proficient for plasmid segregation. In the absence of *parC2*, the partition complex is not properly assembled and thus the plasmid is lost. Overall, the results indicate that *parC2* is essential for plasmid stability and, as long as there is at least one repeat, the plasmid is maintained. Another explanation could be that the class II repeat, which is more abundant in *parC2*, is key to trigger the centromere activity and that the class I motif may play a supporting role. In the TP228 plasmid, the centromere site comprises of the *parH* and O_F locus. It was demonstrated that *parH* alone provides efficient centromere activity, although optimal plasmid stabilisation was achieved when both *parH* and O_F are present (Wu *et al.*, 2011). Additionally, we hypothesised the first four repeats in *parC2* might be essential as their sequences are more conserved than the others. *In vivo* observations indicated that none of the four repeats were crucial. This suggests that any repeat in *parC2* could work as a nucleation site for ParB assembly. This feature has been observed previously for the plasmid F1 partition system *sopABC*. In this plasmid, the *sopC* centromere site comprises 12 repeats (Pillet *et al.*, 2011) and it was shown that one repeat is as efficient in promoting plasmid partition as the full centromere with 12 repeats (Biek and Shi, 1994).

It should be noted that besides the two centromere sites, *parAB* system has two additional sites, *ihf1* and *ihf2*, which are binding sites for the IHF protein. A biochemical study showed that IHF binds to *parC2* that harbours the *ihf2* site but could not bind to *parC1* that lacks any recognition site (Ringgaard, *et al.*, 2007a). The *ihf2* site overlaps with the *parB* gene stop codon, the first B repeat and the first two bases from the second

B repeat in the *parC2* site. When *parC2* is deleted, the IHF site becomes missing as well. This suggests that the IHF host factor could be involved in plasmid partition as it has been shown for the P1 plasmid. This hypothesis requires further investigation.

Chapter 4

Structure determination of the centromere-binding protein ParB

4.1 Introduction

DNA partition represents an essential process that ensures stable maintenance of genetic information during the cell cycle (Baxter and Funnell, 2014). Low copy number plasmids are relatively small in size and are present in few copies in the cell, thus represent an ideal model to study DNA segregation at molecular level. Plasmid partition cassettes are characterised by only three elements: two genes encoding a NTPase protein and a DNA-binding protein, and a centromeric-like site also known as partition site (Salje, 2010; Schumacher, 2012). Based on the NTPase characteristics, the segregation systems are divided into three main types: Walker type ATPases, actin-like ATPases, and tubulin-like GTPases (Million-Weaver and Camps, 2014; Oliva, 2016; Misra *et al.*, 2018). Unlike the ATPase protein, the centromere-binding protein (CBP) generally lacks sequence homologies even within the same family. Despite differences in protein sequences, CBP shares the same function, they recognise the repeats in the centromere-like site and bind site-specifically forming a partition complex (Baxter and Funnell, 2014; Funnell, 2016). The binding of the CBP to different tandem repeats in the partition site is considered to be the first step in the process of plasmid segregation. Structural studies have shown that the CBP in type Ia systems contains a HTH DNA-binding motif, whereas, in type Ib and type II systems, the CBP contains a RHH fold (Funnell, 2016). To date, three CBPs structures belonging to the type Ib segregation system have been determined: ParG of TP228, Omega of pSM19035 and ParB of pCX101. The CBP structures of the three proteins share a similar RHH C-terminal fold (Murayama *et al.*, 2001; Golovanov *et al.*, 2003; Huang *et al.*, 2011). However, little is known about the N-terminal domain as no information is available for this domain with the exception of TP228 ParG, which possesses a flexible N-terminal tail (Golovanov *et al.*, 2003). The N-terminal domain of the CBP plays a crucial role in recruiting the ATPase partner protein and mediating the separation of the two-sister plasmids, which is considered the subsequent step of plasmid partitioning after the formation of the partition complex (Barillà *et al.,* 2007; Pratto *et al.,* 2008). However, no structure is available for the larger size CBPs of type Ib and no information is available as to whether the N-terminal domain adopts the same fold as in the smaller CBPs.

The virulence plasmid pB171, harboured by an enteropathogenic *E. coli* (Tobe *et al.*, 1999) utilises a type Ib system (*par2*) for its segregation (Ebersbach and Gerdes, 2001). The partition cassette consists of two genes encoding ParA, a Walker type ATPase, and ParB, a DNA-binding protein and two *cis*-acting regions *parC1* and *parC2* located upstream and downstream of the *parAB* operon, respectively (Ebersbach and Gerdes, 2001). ParA is a weak ATPase protein consisting of 214 residues that binds to the

nucleoid non-specifically. ParA performs multiple roles: positioning the plasmid at mid-cell and separating the newly replicated plasmids before cell division (Ebersbach and Gerdes, 2001; Ebersbach *et al.*, 2006). On the other hand, ParB consists of 91 amino acids, and binds in a site-specific manner to the partition regions, *parC1* and *parC2*. The protein binds to *parC1* and *parC2* via the C-terminal end (Ebersbach and Gerdes, 2001; Fothergill *et al.*, 2005; Ringgaard *et al.*, 2007b). *In vitro* experiments have shown that the full-length of ParB, ParBΔ19 and ParBΔ39 binds to *parC1* and *parC2* with similar affinity. However, full-length ParB binds to the partition sites and form nucleoprotein complex distinct compared to the truncated forms. ParB dimerises via the C-terminal domain, in fact it was shown that ParB dimerises even in the absence of the N-terminal ends (Ringgaard *et al.*, 2007a; Ringgaard *et al*., 2007b). Nevertheless, little information is available on the N-terminal domain, although it is believed that this domain could be involved in centromeres pairing (Ringgaard *et al.*, 2007b).

Besides its involvement in plasmid segregation, ParB also regulates *parA* and *parB* genes expression. ParB binds site-specifically to the promoter region to repress *parAB* transcription. ParB also works as a repressor of the *parMR* operon that is present in opposite orientation compared to the *parABC* cassette (Ringgaard *et al.*, 2007a). It was shown that a single mutation in the N-terminal region (R-26) leads to misdistribution of ParA as well as reduced plasmid stability (Ringgaard *et al.*, 2009).

Due to limited insights into the segregation mechanism, a clear picture of plasmid partition will be achieved with better knowledge of the three-dimensional structure of the proteins involved in the process. Structural studies revealed that CBPs of type Ib and type II belong to the Arc/MetJ transcriptional repressors superfamily (Murayama *et al.*, 2001; Golovanov *et al.*, 2003; Møller-Jensen *et al.*, 2007; Schumacher, *et al.*, 2007a; Huang *et al.*, 2011). However, the sequence similarities between CBP of type Ib and type II, as well as with the Arc/MetJ family are significantly limited. The ParB protein of pB171 plasmid is relatively longer than ParG and Omega, but shorter than ParB of pCXC100. Structural studies on ParG structure and biochemical analysis of both pCXC100 ParB and Omega revealed that the N-terminal region forms an unstructured tail. The N-terminal domain is short in Omega (20 residues) and ParG (32 residues), whereas is relatively long in ParB of pCXC100 (65 residues) (Murayama *et al.*, 2001; Golovanov *et al.*, 2003; Huang *et al.*, 2011). It seems that the length of the N-terminal domain is likely dependent on the size of the protein.

Since no structural information is available, it is unclear whether pB171 ParB adopts the same RHH C-terminal fold as other known CBPs of type Ib system or whether the N-terminal domain assumes an unstructured conformation. Secondary structure predictions using the PSIPRED server (McGuffin *et al.*, 2000) show that ParB is highly likely to have a similar RHH fold as other CBPs of type Ib (Figure 4.1), although the accuracy of the software is limited. In this chapter, a combination of NMR spectroscopy and X-ray crystallography was used to shed light on ParB 3D structure. Initially, NMR spectroscopy was implemented to determine the 3D structure of the ParB protein. Although, experiments with this technique were inconclusive due to the flexibility of some parts of the protein, they were useful to determine a preliminary secondary structure prediction through assignment of ^{1}H, ^{13}C and ^{15}N chemical shifts. The NMR studies showed that the C-terminus of the protein adopts a folded conformation, while the N-terminal domain is flexible and unstructured. Due to the limitations of NMR, X-ray crystallography was used in the attempt to obtain a high-resolution 3D structure of the protein. This technique allowed us to confirm that the protein has a homo-dimeric RHH fold, which is similar to that found in CBP of type Ib and types II systems.

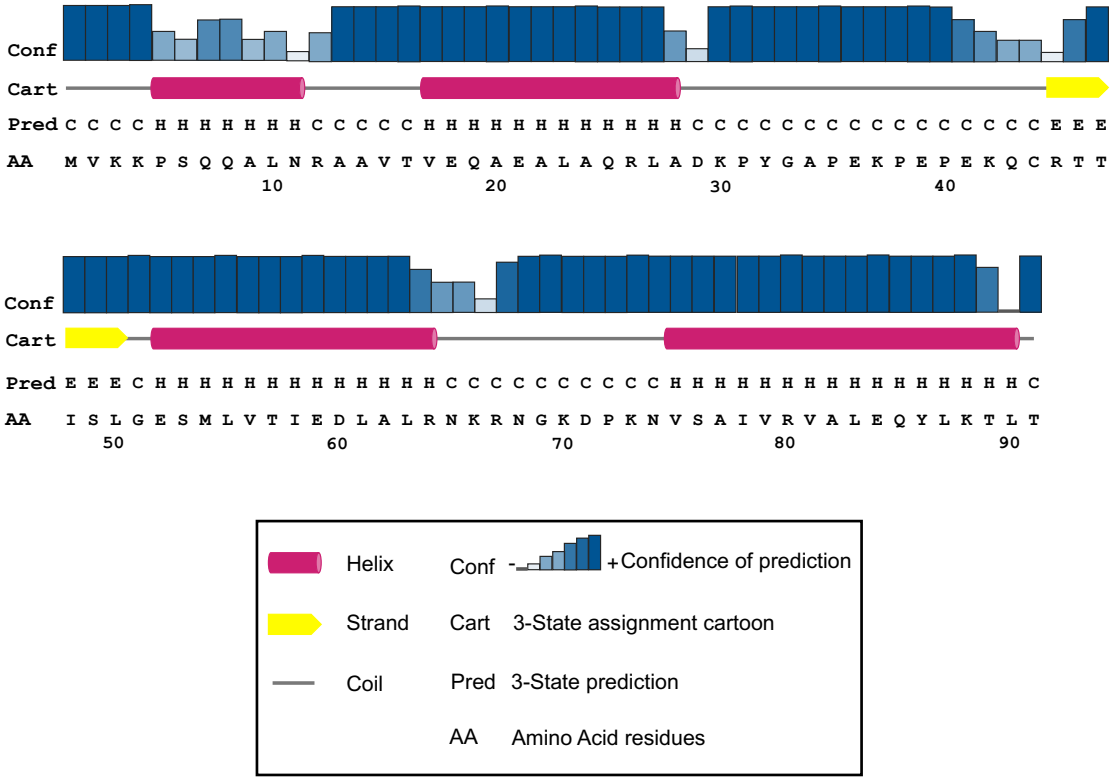


Figure 4.1 PSIPRED prediction of the secondary structure of the DNA-binding protein, ParB, from pB171

Secondary structure of ParB is predicted and shown. The cartoon structure represents the secondary structure predicted for ParB. α-helices are shown in pink colour and β-strand is shown in yellow, whereas the coils are shown with a straight grey line. Confidence intensity is shown above. The amino acids sequence is shown underneath the structure prediction, where C means coil, H helix and E strand.

4.2 Results

4.2.1 ParB overproduction and purifications

4.2.1.1 ParB overproduction, solubility and purification in LB medium

Structure determination of the protein of interest using NMR spectroscopy and X-ray crystallography requires similar sample quality. For both techniques, the protein should be soluble, folded and homogenous. The protein of interest is required to be free of contaminants with a purity of 95% or higher and needs to be highly concentrated (≥ 3 mg/ml). The pET-ParB construct was available in the laboratory collection (Fothergill *et al.*, 2005), where *parB*, was cloned in frame with codons for a N-terminal $(His)_6$-tag and downstream of the T7 promoter and *lac* operator. The target gene *parB* is expressed under control of bacteriophage T7 transcription and expression occurs in the expression host cell (BL21(DE3)) that harbour a copy of the gene encoding T7 RNA polymerase in the chromosome. Protein expression using the T7 system is induced by IPTG. The construct was sequenced before any further application to ensure no error was introduced during the cloning process.

Small-scale overexpression and solubility were carried out to ensure that the clone was effectively expressed and was capable to produce highly soluble protein. The construct was transformed into BL21(DE3) strain *E. coli* cells supplemented with 100 μg/ml of ampicillin as described in Section 2.3.2. Small-scale overexpression in LB medium was carried out as described in Section 2.4.1. 50 ml of LB culture was grown for approximately three hours until the OD_{600} reached the value 0.6. The culture was then induced with IPTG at a final concentration of 1 mM and 1 ml of the sample was collected every hour for three consecutive hours. A 15% SDS-polyacrylamide gel showed that induction of protein overproduction for three hours at 37°C was sufficient to obtain a high yield of protein, as shown in Figure 4.2*A*. Following small-scale overexpression, the solubility of the protein was tested as described in Section 2.4.2. Protein solubility was visualised by comparing the amount of protein in the soluble and insoluble fractions using SDS-PAGE. The amount of ParB protein in inclusion bodies decreased by increasing the incubation time for ParB of the overexpression, as shown in Figure 4.2*B*. These results indicate that the protein is highly soluble.

After the pilot overproduction test in a small volume, ParB overexpression was carried out at medium scale before purification. ParB appeared to be properly overexpressed in 200 ml culture as well as in 1 L culture. The expressed $(His)_6$-tagged ParB was purified using Ni^{2+} affinity chromatography as described in detail in Materials and Methods under

the Section 2.4.5. Affinity chromatography was performed at 4°C. The final concentration of ParB protein produced after the first purification was 15-30 mg/ml, depending on the volume of starting culture. The purity of the $(His)_6$-tagged ParB was verified using SDS-PAGE, as shown in Figure 4.3. Based on the SDS-PAGE, the chromatography purification yields high purity (~95%) and quantity of the $(His)_6$-tagged ParB protein. The molecular weight of ParB calculated from the primary sequence is 11.12 kDa including the $(His)_6$-tag.

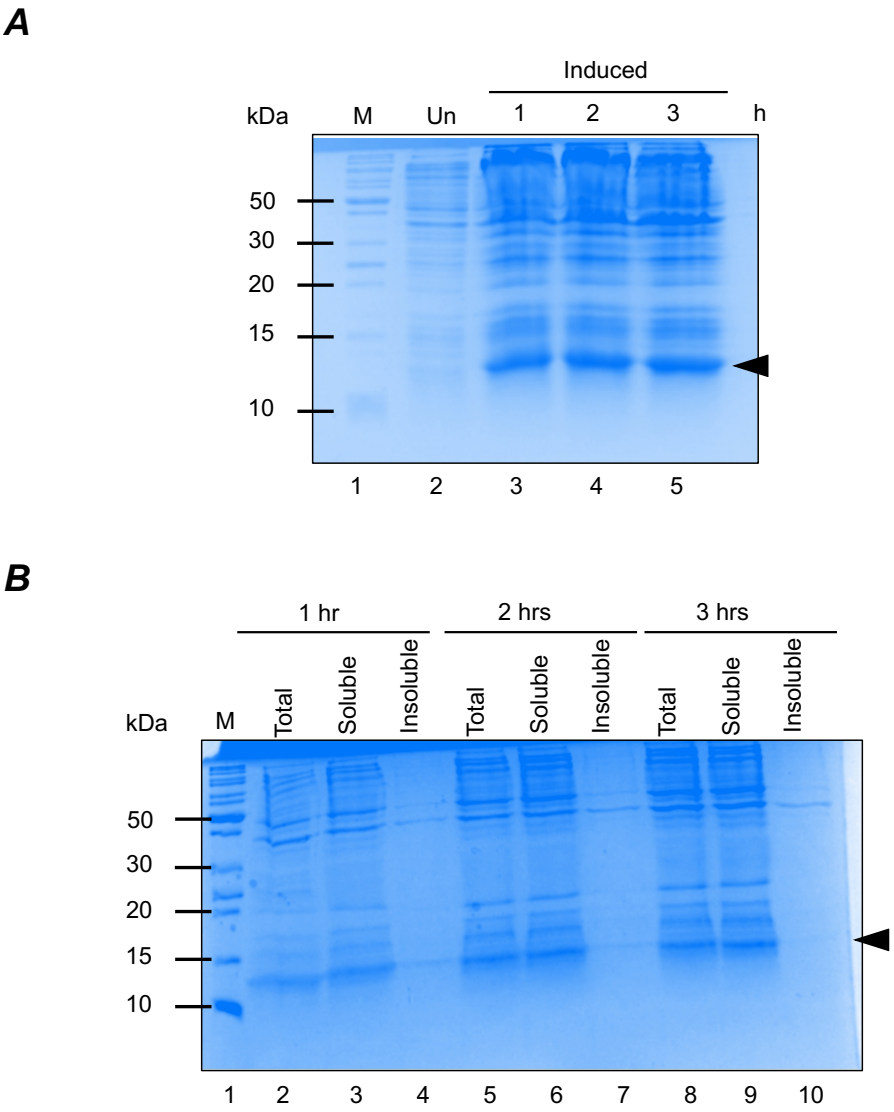


Figure 4.2 Small-scale overproduction and solubility of ParB

A) SDS-polyacrylamide gel shows overproduction of ParB in LB medium. The uninduced and induced sample are compared. The target overproduced 11.12 kDa ParB protein is indicated by the arrow. M represents the molecular weight protein marker in kDa (Pierce™ Unstained Protein MW marker). ***B)*** SDS-polyacrylamide gel of the overexpression experiment shows the effect of incubation times on the solubility of the ParB protein. The total lysate, soluble and insoluble fractions are compared after 1, 2, and 3 hours induction with 1 mM IPTG. The expressed protein indicated with an arrow. M represents the molecular weight protein marker in kDa.

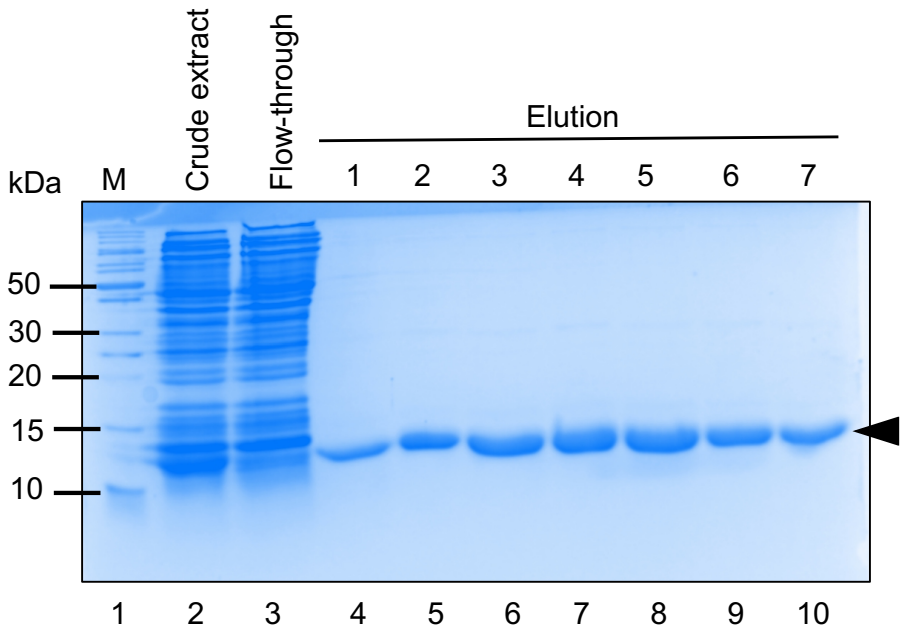


Figure 4.3 Purification of ParB using Ni^{+2} affinity chromatography

Lane 1 represents the molecular weight protein marker in kDa (Pierce™ Unstained Protein MW marker). Lane 2 shows the clear lysate of cell culture before loading on the nickel column. Lane 3 shows the flow-through. Lanes 4-10 show the elution fractions of the purified ParB protein obtained using 400 mM imidazole. The fractions were collected after two consecutive washing steps with two different imidazole concentration (15 and 85 mM). The purified 11.12 kDa (including tag) ParB protein is indicated by the arrow.

4.2.1.2 ParB overproduction and purification in minimal medium

An isotopically labelled protein sample is a prerequisite for many protein NMR experiments. However, replacing the rich medium with a minimal medium can be challenging and may affect protein production. Therefore, a small-scale overexpression in minimal medium (Section 2.4.3) was performed to determine whether the $(His)_6$-tag ParB construct was expressed at sufficient level. One colony was selected from the overnight transformation plate of BL21(DE3) *E. coli* cells transformed with the pET-ParB vector, inoculated in 10 ml of LB and grown for 6-7 hours. At this stage, the medium was switched from LB to minimal medium. The starting culture in the minimal medium had an OD_{600} = 0.05 and was grown overnight at 37°C with 180-200 rpm shaking. The overnight culture was transferred into 50 ml minimal medium, again with a starting OD_{600} = 0.05. The culture was incubated until the OD_{600} reached a value between 0.8-1, then the culture was induced with 1 mM IPTG (final concentration) and incubated either overnight at 20°C or for three hours at 37°C. Different incubation times and temperatures were used to determine which condition generates a sufficient quantity of ParB. In both cases, the expression generated high yields, as shown in Figure 4.4*A*. The amount of ParB protein produced in 50 ml culture was determined to be sufficient to process to large-scale growth.

The ParB $(His)_6$-tagged fusion construct was then produced in a 1 L volume of minimal medium, containing ^{15}N-labelled ammonium chloride as nitrogen source and ^{13}C glucose as carbon source. The amount of ParB produced after three-hour induction was sufficient to obtain a high level of protein production as shown in Figure 4.4*B*. After cell lysis, Ni^{+2} affinity chromatography was performed (as reported in Section 2.4.5) as the first step for ParB purification. The elute from this chromatography step contained contaminants that could affect the quality of the NMR spectra (Figure 4.4*B*). Therefore, SEC (Section 2.4.9) was performed to remove the contaminants. Unfortunately, this purification step did not remove the contaminants completely (Figure 4.5*C* and *D*). Several attempts were performed to improve protein purification products by growing the cells harbouring the pET-ParB plasmid in minimal medium (using the same recipe of minimal medium but replacing the isotopically labelled with ammonium chloride and glucose as sources of nitrogen and carbon, respectively (Section 2.4.4)) in which the concentration of imidazole in the washing buffer was changed to 90 mM and a different prepacked column used. However, it was impossible to eliminate all the contaminants even after the second purification step (Figure 4.5*B, C* and *D*).

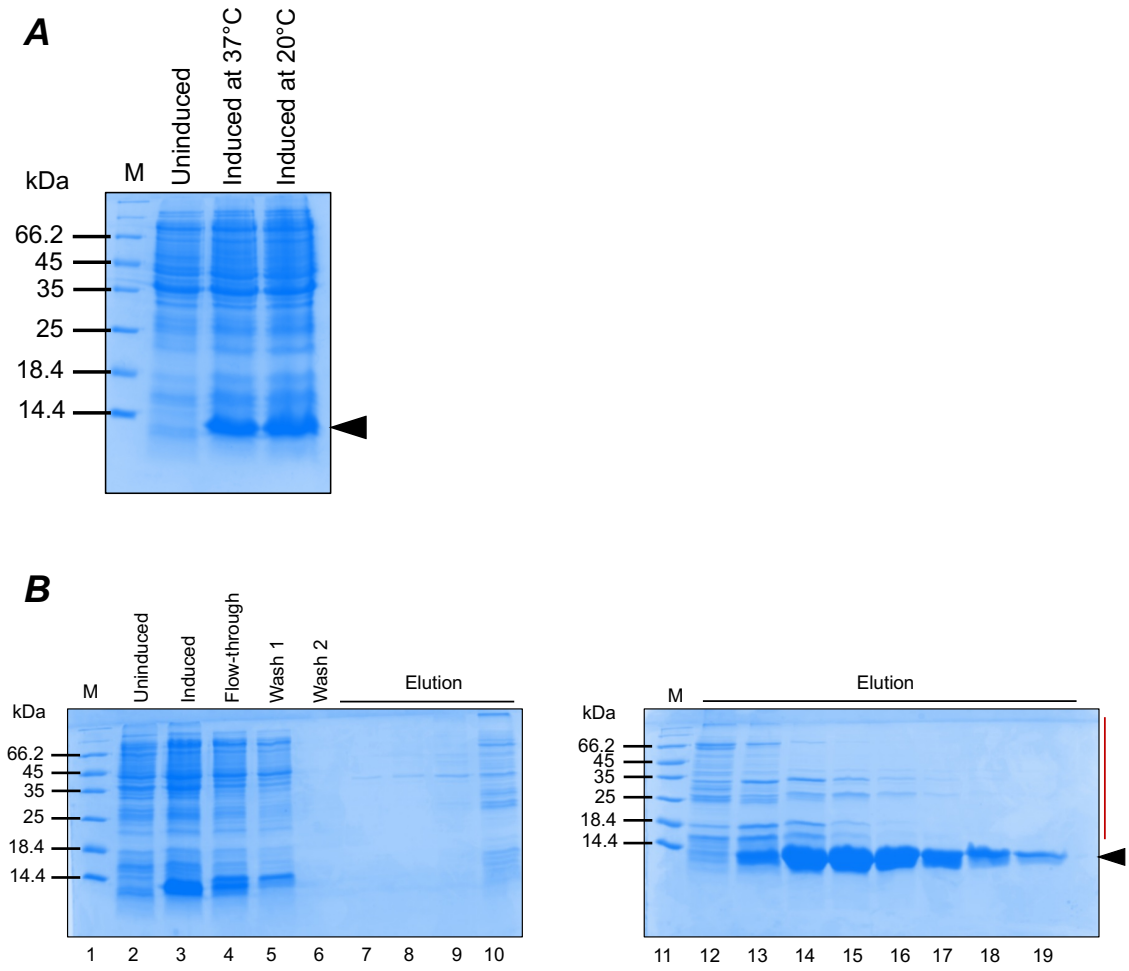


Figure 4.4 ParB overexpression and purification optimisation for NMR experiments

A) Overexpression test of the *parB* gene in *E. coli* culture grown in minimal medium at different incubation temperatures and times. M represents the unstained PageRuler protein ladder (Thermo Scientific) in kDa. Induction at 37°C and 20°C in lane 3 and 4, respectively, are compared with the uninduced culture in lane 2. ***B)*** Coomassie-stained 15% SDS-polyacrylamide gel showing fractions from different steps of the Ni^{+2} affinity chromatography ParB purification from a 1 L culture using a pre-packed 5 ml column. Left gel, lane 1 (M) is the unstained PageRuler protein ladder (Thermo Scientific) in kDa. Lane 2 shows the cell culture prior to IPTG induction. Lane 3 shows the culture after three-hour induction with 1 mM IPTG. Lane 4 represents the flow-through of the cell lysate after one-hour of circulation on the nickel column. Lanes 5 and 6 contain aliquots of two consecutive column wash steps with 15 and 80 mM imidazole, respectively. Lanes 7-10 show the protein fractions eluted with 400 mM imidazole. Right gel, SDS-polyacrylamide gel that shows the following ParB elution fractions (lanes 12-19). The level of contamination is highlighted by a vertical red line.

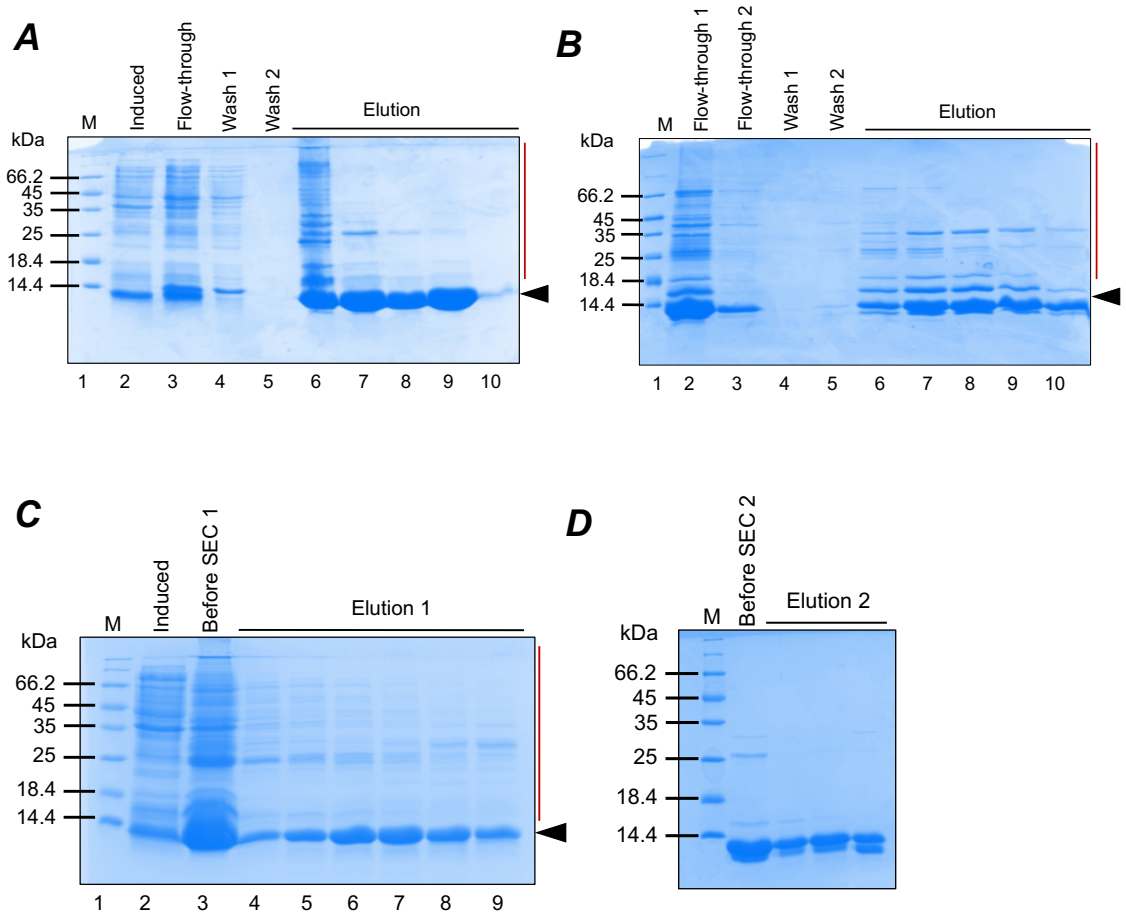


Figure 4.5 Two-step ParB purification optimisation for NMR experiments

A) and ***B)*** Coomassie-stained 15% SDS-polyacrylamide gels show two steps of Ni^{+2} affinity chromatography for ParB purification. ***A)*** the SDS-polyacrylamide gel showing the first purification step, where M in lane 1 represents the unstained PageRuler protein ladder (Thermo Scientific) in kDa. Lane 2 shows the culture overexpressing *parB* gene after three-hour induction with 1 mM IPTG at 37°C. Lane 3 represents the unbound cell lysate after one-hour of circulation on the pre-charged nickel column. Lanes 4 and 5 represent two consecutive washing steps with 15 and 80 mM imidazole, respectively. Lanes 6-10 represent the protein fractions, eluted with 400 mM imidazole. ***B)*** SDS-polyacrylamide gel showing aliquots from the second steps of Ni^{+2} affinity purification, where the elution fractions of the nickel column were collected and passed through a second nickel-charged column. The second and third lanes compare the amount of protein that did not bind to the column from the first (flow-through 1) and the second (flow-through 2) nickel column. Lanes 4 and 5 show the two washing steps to remove contaminants. Lanes 6-10 show the elution fraction collected using 400 mM imidazole. In both purification steps, the level of contamination is highlighted with a vertical red line. ***C)*** Coomassie-stained SDS-polyacrylamide gels show the fractions from the first size exclusion chromatography, which was followed by a second SEC (***D***). The level of contamination after the first step is highlighted with a vertical red line. In both gels (***C*** and ***D***), M represents the molecular weight marker.

In an attempt to obtain a higher purity for the protein of interest in minimal medium, an alternative protocol was used for ParB overproduction (Bracken *et al.*, 2001). A small-scale overexpression was carried out in which the construct was grown first in enrichment medium (LB) until the OD_{600} reached 0.6. The medium was swapped from LB to minimal medium harvesting the cells and washing the pellets several times with M9 salt (Table 2.4), prior to inducing the culture with 1 mM IPTG, as described under Section 2.4.4 in Materials and Methods chapter. The amount of $(His)_6$-tagged ParB produced after three-hour induction indicated a successful large-scale protein production (Figure 4.6*A*). The overexpression was scaled up to a 1 L culture and three-hour induction resulted in a substantial protein yield. The amount of the ParB protein produced was optimal for purification. Ni^{+2} affinity chromatography was performed as a first purification step as described in Section 2.4.5. The first step of purification resulted in a protein concentration equal to 16 mg/ml (measured by Bradford Assay, as described in Section 2.4.7) from a 1 L culture and the purity of the sample was checked using SDS-PAGE (Figure 4.6*B*). The gel shows that the change of the overexpression method was important to achieve higher ParB purity. Indeed, the alternative method had decreased the contamination level during growth and overexpression. An additional SEC purification step was performed before further downstream. SEC was performed to separate persistent contaminants from ParB as well as to remove any aggregates that could impact NMR experiments. Furthermore, SEC was used as a technique to exchange the buffer of the ParB protein to the required buffer (20 mM BIS-Tris, pH 6.0, 50 mM NaCl and 1 mM DTT) used for the NMR experiment. Again, SDS-PAGE was used to verify the purity of the sample (Figure 4.7). Following SEC, the protein was mostly free from contaminant, however, in this Figure 4.7, it is hard to conclude as different concentration were used (i.e the concentration of ParB before SEC was double the concentration obtained after SEC).

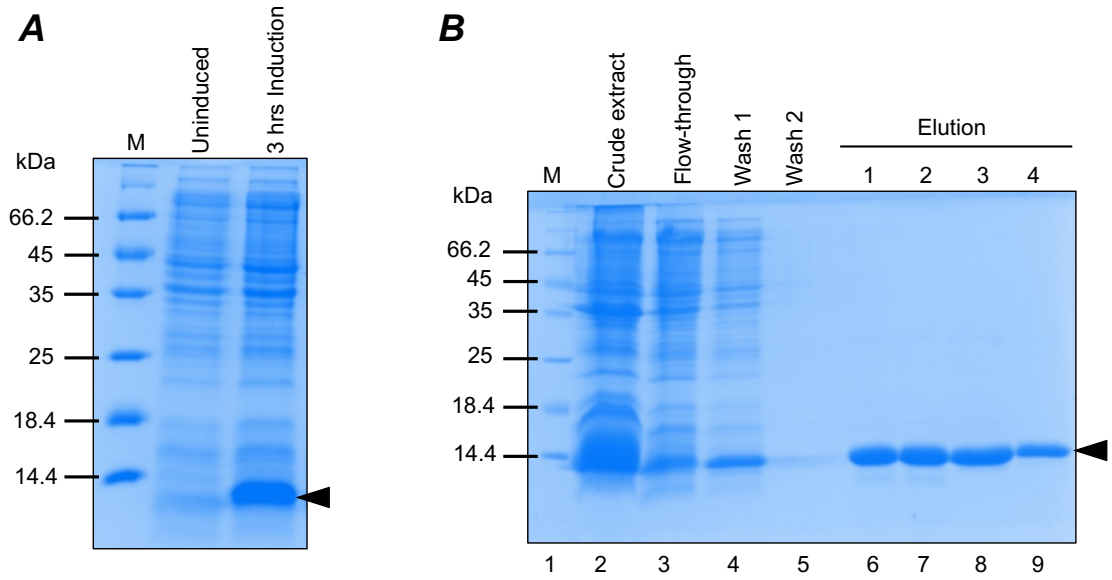


Figure 4.6 Overproduction and Ni^{+2} affinity chromatography purification of ParB from a 1L culture

A) SDS-PAGE showing overproduction of ParB in the minimal medium using the alternative protocol. Uninduced and induced *E. coli* cells are compared. The target 11.12 kDa ParB protein is indicated by an arrow. M represents the unstained PageRuler protein ladder (Thermo Scientific). ***B)*** SDS-polyacrylamide gel showing ParB purified by using Ni^{+2} affinity chromatography. Lane 1 represents the unstained PageRuler protein ladder (Thermo Scientific). Lane 2 contains an aliquot of the cell cleared lysate before chromatography. Lane 3 represents the flow-through fraction. Lanes 4 and 5 represent two consecutive washing steps using 15 and 85 mM imidazole, respectively. Lanes 6-9 represents ParB fractions eluted using 400 mM imidazole. The band representing the purified 11.12 kDa ParB protein (including the six-histidine tag) is indicated by an arrow.

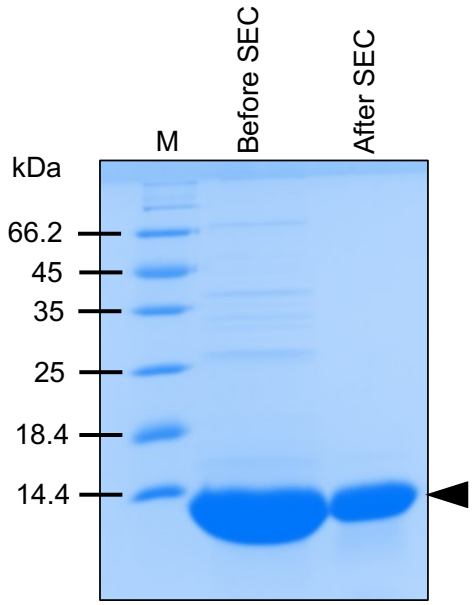


Figure 4.7 SDS-polyacrylamide gel of ParB after size exclusion chromatography

Lane 1 represents the unstained PageRuler protein ladder (Thermo Scientific). Lane 2 shows the purified ParB after Ni^{+2} affinity chromatography. All the fractions were pooled and concentrated to 8 mg/ml before being loaded onto Superdex S75 16/60 column. Lane 3 shows the concentrated elution fractions after gel filtration. The purified 11.12 kDa ParB (including the six-histidine tag) is marked by an arrow.

4.2.2 ParB is a dimer in solution as determined by using SEC-MALLS

Accurate determination of molecular weight and oligomerisation state of proteins in solution are important steps prior to perform any downstream experiments, especially for structure determination of a protein using NMR spectroscopy or X-ray crystallography. Although the purity of the protein is crucial, the oligomerisation state and homogeneity of the sample are equally important. SEC was performed to separate ParB from the aggregates and other proteins based on the size of molecules prior to multi-angel laser light scattering (MALLS) application (Figure 4.8*A*). To analyse the resulting peaks, the elution fractions of each peak were collected, and the samples checked using SDS-PAGE. This shows that the steep peak obtained during gel filtration represents ParB as shown in Figure 4.8*B*. All the samples were collected and concentrated before being subjected to MALLS. SEC coupled to MALLS is a reliable technique to determine the molar mass of proteins in solution and can determine whether the eluting peak is homogenous or heterogenous (Some *et al.*, 2019). Three different concentrations of ParB were used along with BSA that was used as a control. The concentrations were as follows: 2, 0.6 and 0.2 mg/ml for ParB and 2.5 mg/ml for BSA. The SEC-MALLS experiment confirmed the ParB protein dimerises in solution, as no monomer was observed. The molecular weight indicated by SEC-MALLS was ~ 22 kDa, which corresponds to a ParB dimer. Moreover, all three concentrations used produced similar results, as indicated in Figure 4.9. The molar masses calculated from SEC-MALLS are slightly less than the theoretically mass of the monomer.

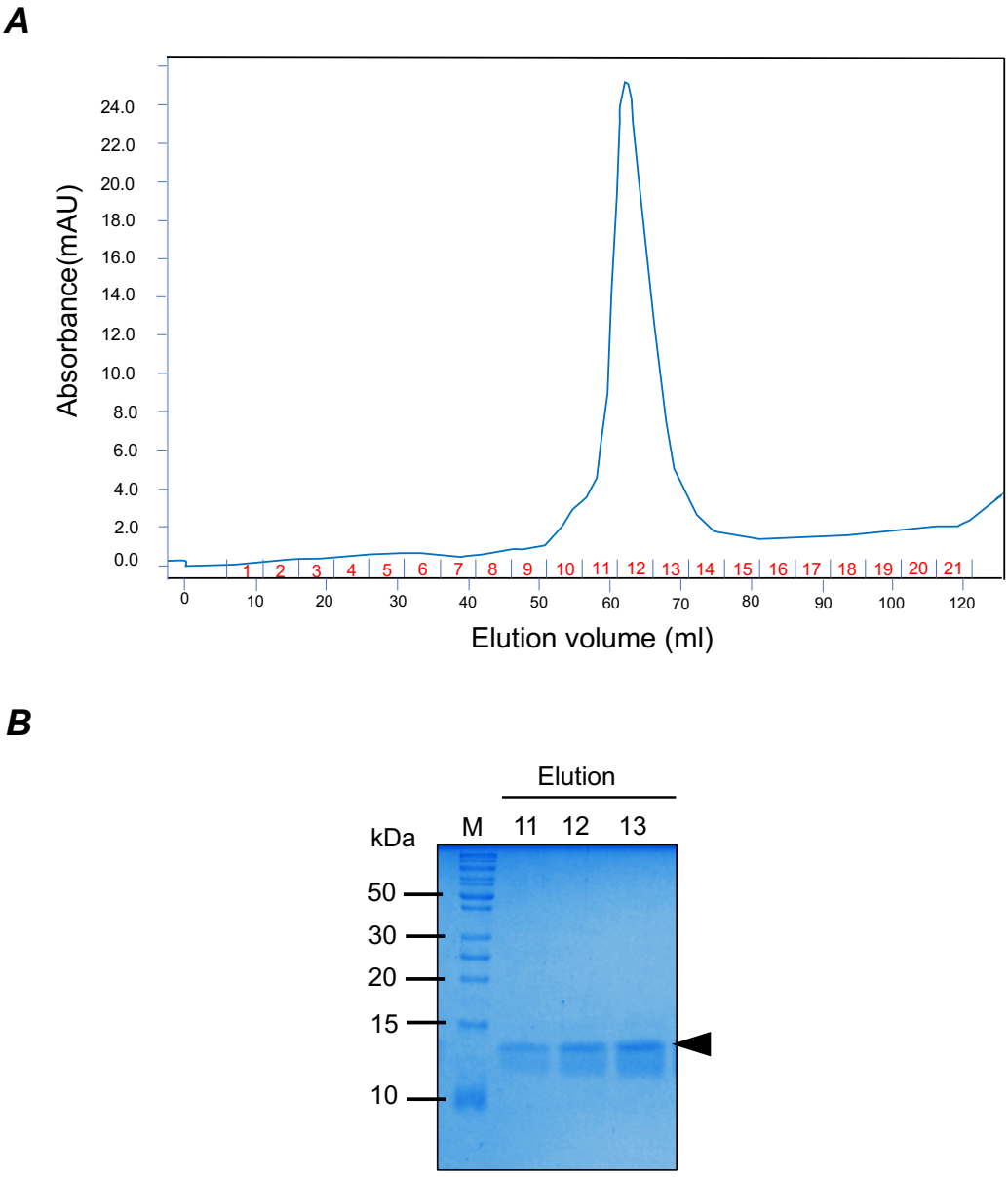
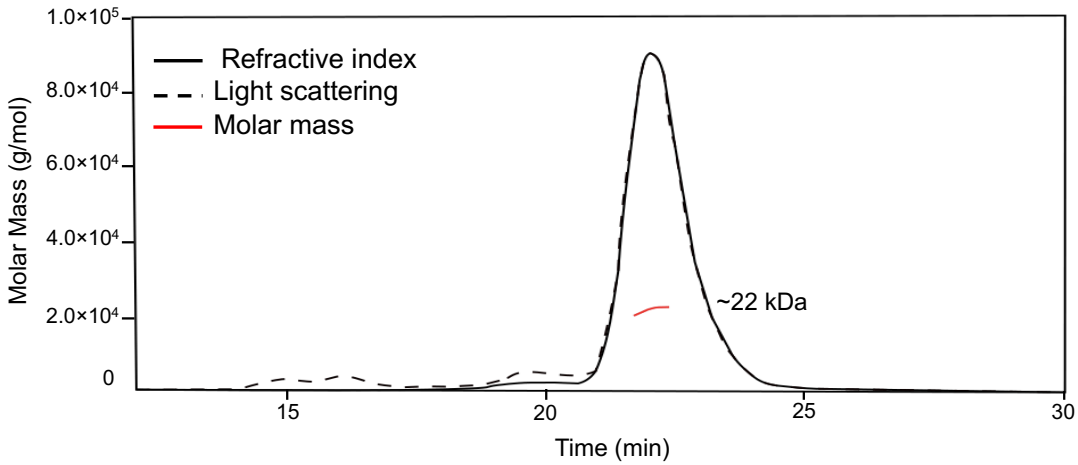


Figure 4.8 Size exclusion chromatography

A) Size-exclusion chromatography profile of ParB elution using a S75 Superdex 16/60 column (GE Healthcare). The ParB sample was concentrated down to 1 ml using a 5 kDa molecular cut-off filter concentrator before injection into the gel filtration column, pre-equilibrated with 20 mM Tris-HCl, pH 7.0, 150 mM NaCl. Purified ParB (2 mg/ml) was injected into the column. The ParB sample elutes after 50 ml. Fractions 9-14 were collected. ***B)*** Coomassie-stained 15% SDS-polyacrylamide gel showing the fractions collected during SEC. Lane 1 represents the unstained Page-Ruler molecular weight ladder (Thermo Scientific). Lanes 2-4 show ParB after the S75 16/60 Superdex purification. All the fractions were pooled and concentrated to 2 mg/ml before performing MALLS on the sample. The purified 11.12 kDa ParB protein is indicated by an arrow.



Concentration mg/ml	MALLS measured (kDa)	Expected (kDa)	Measured/ Expected (kDa)
2	21.76	11.12	1.96
0.6	21.99	11.12	1.98
0.2	21.97	11.12	1.98

Figure 4.9 ParB is a dimer in solution

SEC-MALLS data profile of the ParB protein at a concentration of 2 mg/ml. The red solid line represents the measured molar mass. The black solid line represents the refractive index of protein and the black dashed line indicates the light scattering. The table below the chromatography profile represents the average measured molar masses at three different ParB concentrations (2, 0.6 and 0.2 mg/ml). The average molar masses measured from SEC-MALLS are compared with the theoretical ones.

4.2.3 Buffer optimisations using 2D [^{1}H-^{15}N] HSQC experiment

Folding state and stability of ParB was initially assessed by 2D [^{1}H-^{15}N] HSQC NMR spectroscopy experiment with uniformly [^{15}N] labelled full-length ParB at a concentration of 150 μM. A 2D HSQC experiment, gives a "fingerprint" of the protein and provides insight into suitability of the protein sample for structure determination. From this experiment, folding and stability of a protein of interest can be assessed. In the HSQC experiment, each cross-peak represents an amino acid residue in the protein backbone, with the exception of proline, which does not have an amide proton. The first amino acid might also not be seen in the spectrum due to exchange of the terminal amine group with the solvent. The quality of the ParB samples was assessed based on the number of cross-peaks observed in the spectrum as well as the dispersion of peaks in both proton and nitrogen dimensions. The first buffer used was the storage buffer of the protein (50 mM HEPES, pH 7.5, and 150 mM KCl), when kept at -80°C post purification after the glycerol was removed from the sample. However, the spectrum showed only 52 cross-peaks including side-chain amides, which represents half the peaks that are expected from the protein sequence (Figure 4.10*A*). The reason behind missing cross-peaks was investigated further biochemically via SDS-PAGE. The SDS gel showed that the sample analysed by HSQC contained both monomers and dimers (Figure 4.10*B*). To get a single population in the sample, 1 mM DTT was added to the protein sample. Moreover, SEC-MALLS indicated that ParB is a dimer in a 20 mM Tris-HCl, pH 7.0, 150 mM NaCl buffer solution. Therefore, the buffer 20 mM Tris-HCl, pH 7.0, 150 mM NaCl, 1 mM DTT was used in a second attempt. The change of buffer and the addition of a reducing agent improved the HSQC spectrum, resulting in larger number of peaks being visible, although some cross-peaks were still missing (Figure 4.11*A*). Therefore, further buffer optimisation was needed to get an optimal spectrum for further experiments.

Several buffer conditions were tested to understand in which buffer ParB was more stable and folded. The buffers used were either 20 mM sodium phosphate, pH 5.5, or 20 mM Tris-HCl, pH 6.0 and 6.5, and with NaCl concentrations of either 50 or 150 mM. A 300 μl ParB sample was used in 3 mm NMR tube and a final concentration of 1 mM DTT was added before the samples was tested by HSQC. All the experiments were performed using a single [^{15}N] labelled ParB at a concentration of ~125 μM. All the samples were screened using 2D HSQC NMR experiment and a subset of these are shown in Figure 4.11. The results indicate that all 5 samples screened were folded. The chosen sample was judged based on the spread in the chemical shift in both dimensions. The number and linewidth of cross-peaks obtained were also considered. ParB in 20 mM Tris-HCl pH 6.0, 50 mM NaCl and 1 mM DTT produced 103 peaks, with peaks being more dispersed

compared to the other conditions. The dispersion of the peaks was observed to be between approximately 6.5 and 9.5 ppm in proton dimension, as shown in Figure 4.11*F*. The spectrum in Figure 4.11*F* shows the dispersion of the resonance of the amides of ParB residues and suggests that the protein is well folded. The undispersed region, instead, represents the un-structured region of the protein. The result also shows cross-peaks clustered in the centre of the spectrum (~8 ppm in proton dimension), which is a characteristic feature of an unfolded/flexible part of the protein. This suggests that some regions of the ParB protein are unstructured or partially folded in solution.

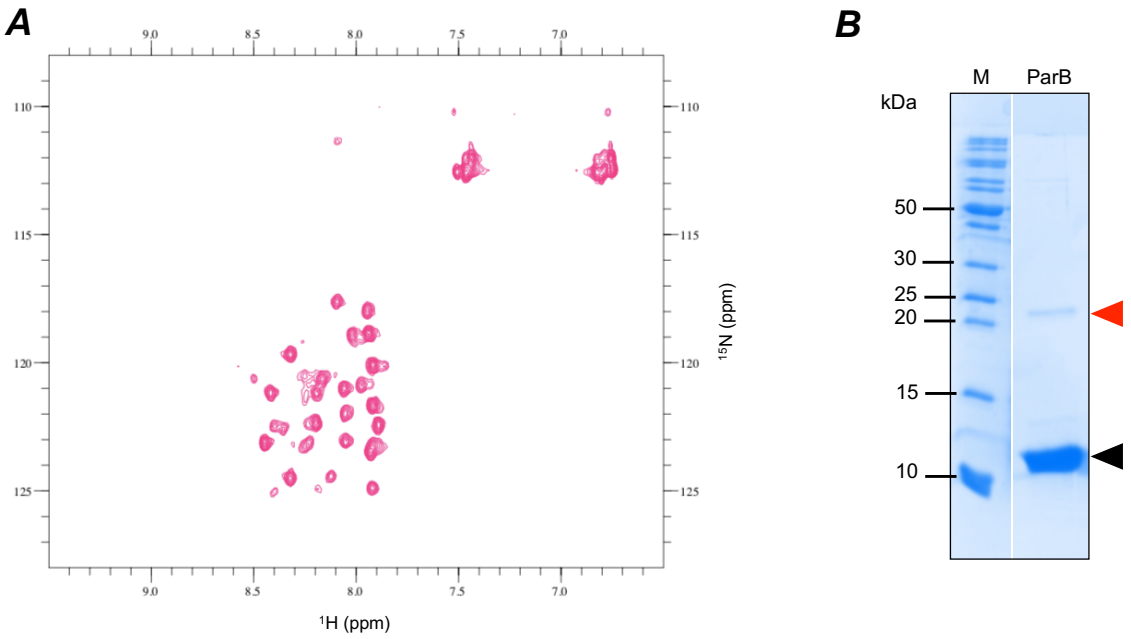


Figure 4.10 [^{1}H, ^{15}N] HSQC NMR spectrum and SDS-PAGE of ParB

A) [^{1}H, ^{15}N] HSQC NMR spectrum of ParB (125 μM dimer) in 50 mM HEPES, pH 7.5, and 150 mM KCl was recorded at 700 MHz, 25ºC. The spectrum shows 52 cross peaks. ***B)*** Coomassie-stained 15% SDS-polyacrylamide gel showing purified ParB after the NMR experiment. Monomeric ParB (11.12 kDa) is indicated by a black arrow, while dimeric ParB (22.24 kDa) is indicated by a red arrow.

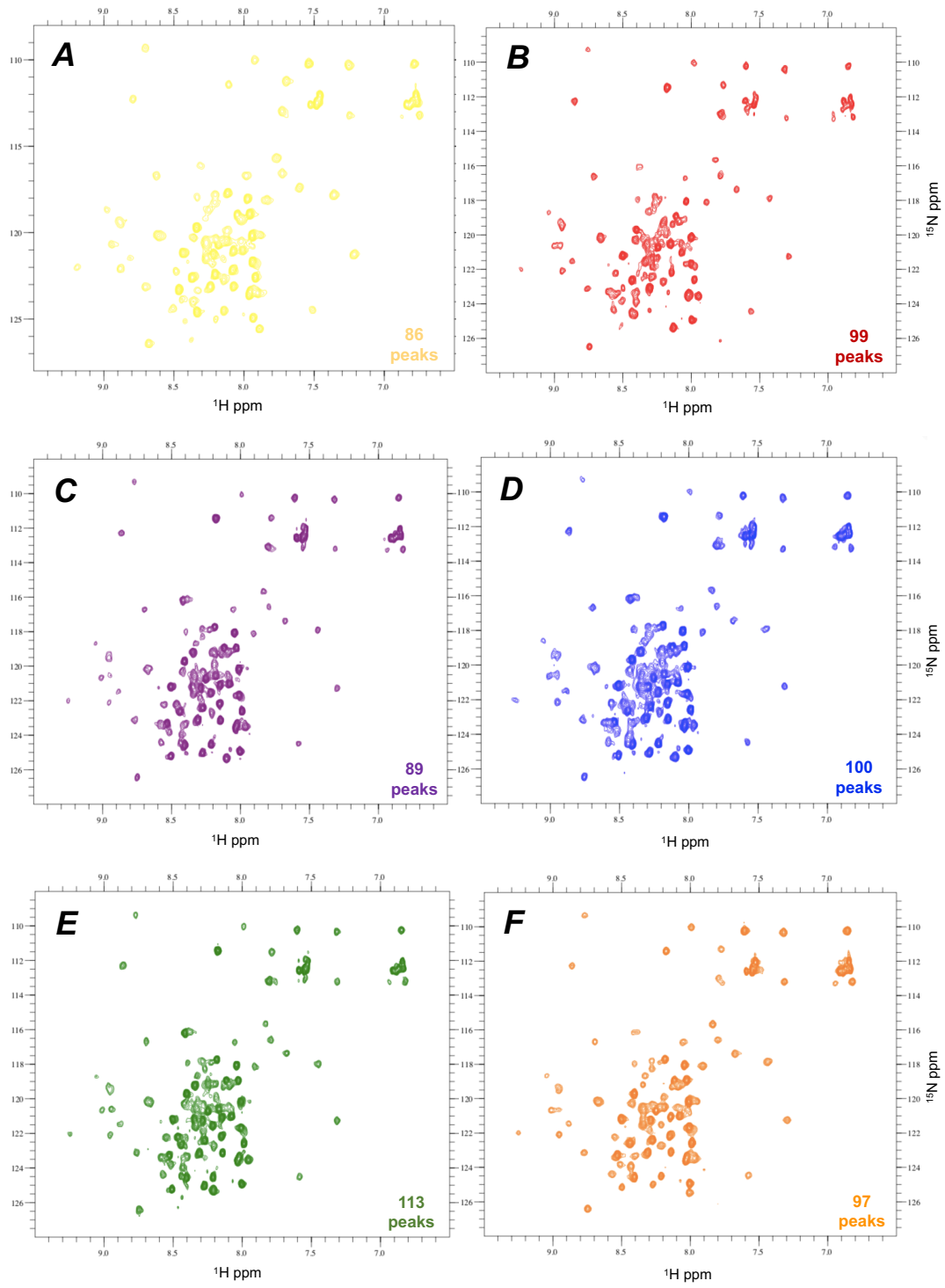


Figure 4.11 ParB buffer optimisation for NMR experiments using [^{1}H, ^{15}N] HSQC NMR

HSQC spectra of ParB (125 μM dimer) was recorded at 700 MHz, 25ºC, are shown in different buffers conditions. ***A)*** 20 mM Tris-HCl, pH 7.0, 150 mM NaCl and 1 mM DTT (yellow cross peaks). ***B)*** 50 mM phosphate, pH 6.5, 150 mM NaCl and 1 mM DTT (red cross peaks). ***C)*** 20 mM Tris-HCl, pH 6.5, 150 mM NaCl and 1 mM DTT (purple cross peaks). ***D)*** 20 mM Tris-HCl, pH 6.5, 50 mM NaCl and 1 mM DTT (blue cross peaks). ***E)*** 20 mM Tris-HCl, pH 6.0, 150 mM NaCl and 1 mM DTT (green cross peaks) and ***F)*** 20 mM Tris-HCl, pH 6.0, 50 mM NaCl and 1 mM DTT (orange cross peaks). The number of cross peaks for each spectrum are shown at the bottom right of each spectrum.

4.2.4 Backbone assignments of ParB N-terminal domain

NMR was used to determine ParB secondary structure. Following the initial buffer optimisation test, ParB was overexpressed in BL21(DE3) *E. coli* cells in uniformly [^{13}C, ^{15}N] labelled minimal medium (Bracken *et al.*, 2001) and purified as previously described. The HSQC spectrum of ParB in the buffer of choice showed 103 cross-peaks, including side-chain amides, which is slightly higher than expected from the protein sequence. Besides the 91 amino acid residues, the protein contains the six histidine residues of the tag and two extra amino acids, which corresponds to the codons of the *Xho*I restriction site at the C-terminus. ParB harbours six proline residues, which are not observed in the spectrum as well as the first methionine, therefore the number of cross-peaks should be identified are 92. Moreover, as only one cross-peak per amino acid was observed and SEC-MALLS confirmed that the protein is dimeric in solution, this result further support that the ParB protein is homo-dimeric in the buffer of choice.

The uniformly [^{13}C, ^{15}N] labelled ParB was purified using Ni^{+2} affinity chromatography followed by SEC. The NMR sample was prepared at a concentration of 0.65 mM in a 600 μl volume. Although the purity of the sample was high, it was impossible to get a higher concentration, as the protein tended to precipitate at a concentration of ~8 mg/ml. Assignment of the resonance to individual nuclei considers the first step for protein structural study by NMR spectroscopy. A 2D [^{1}H, ^{15}N] HSQC spectrum, and 3D [^{1}H, ^{13}C, ^{15}N] CBCANH and CBCA(CO)NH experiments were recorded to enable the backbone resonance assignment of ParB. 2D [^{1}H, ^{15}N] HSQC spectra were recorded following each 3D experiment to check the stability of the sample after prolonged incubation at 18°C temperature. The HSQC spectra showed that ParB is less stable after 48 hours at room temperature as the protein tends to degrade (data not shown). Thus, the sample was replaced with a fresh preparation, when necessary. Sequence-specific assignment was carried out to link the neighboring residues using Cα and Cβ chemical shifts. It was possible to assign 33% of backbone nuclei. Furthermore, the HSQC spectrum shows the NH_2 of asparagine and glutamine side chain peaks as a pair. In total, 30 out of 92 amino acids were assigned. These are located in the central region of the spectrum and produced very strong and sharp peaks. The 30 amino acids assigned belong to the N-terminal region and include residues from Glu-7 to Ala-34. This region is highly unstructured, as shown in Figure 4.12. In addition to Glu-7 to Ala-34, it was possible to assign the [^{1}H, ^{15}N] resonance of Lys-3, Lys-4 and Ser-6. However, it was challenging to sequentially link the resonance as some peaks were missing. It was not possible to record the signals from the structured part of ParB or assign the full length of the protein. This could be due to the flexibility of the N-terminal domain that slows down the overall

rotational tumbling of the protein. As a result, the signal from the structured domain relaxes quickly, resulting in very weak peaks that might not even be possible to observe in the 3D experiments.

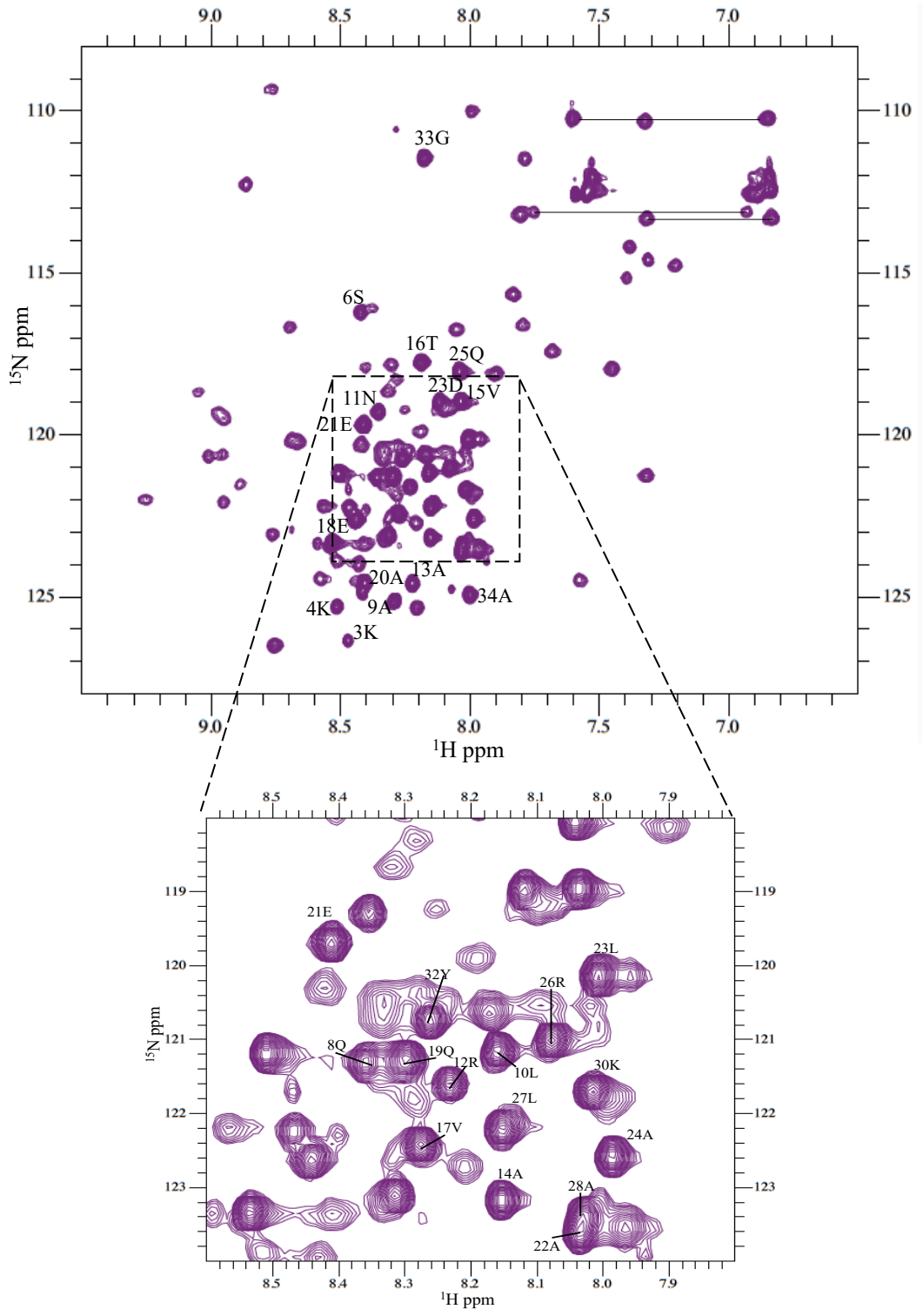


Figure 4.12 Backbone assignment of the N-terminal domain of ParB

[^{1}H, ^{15}N] HSQC spectrum of uniformly [^{13}C, ^{15}N] labelled ParB (0.6 mM monomer; 20 mM Bis-Tris, pH 6.0, 50 mM NaCl and 1 mM DTT) was recorded at 700 MHz, 25ºC. Backbone assignments were only obtained for 3, 4, 6-34 amino acids of ParB. Assignments are given with residue number and one letter for each amino acid. The congested central region of the spectrum is enlarged to show the remaining assigned peaks. Pairs of side-chain amides are connected with a horizontal line.

4.2.5 Backbone assignments of the C-terminal domain of ParB

4.2.5.1 Cloning of a truncated *parB* gene encoding ParBΔ30N

A truncated ParBΔ30N was constructed to remove the assigned residues from the flexible N-terminus. The hypothesis was that the deletion would facilitate data acquisition for the folded domain. Therefore, the first 30 amino acids were deleted from the N-terminal part and the truncated version of the *parB* gene was amplified using PCR (Figure 4.13*A* and 4.14*A*). The amplified fragment contains the *Nde*I and *Xho*I sites, upstream and downstream of the gene of interest, respectively. A 1% agarose gel in Figure 4.13*B* shows the 192 bp amplification product, which corresponds to *parBΔ30N*. Both the PCR fragment and the pET22b expression vector were digested with *Nde*I and *Xho*I restriction enzymes and were then purified using ethanol precipitation as described in Section 2.3.7.

The amplification product was then inserted into pET22b vector in frame with codons encoding a C-terminal $(His)_6$-tag. The gene was inserted downstream of T7 promoter and *lac* operator using *Nde*I and *Xho*I sites (Figure 4.14*B*). The cloning results were screened using colony PCR as described in Section 2.3.10 (Figure 4.13*C*). The successful clones were later verified using *Nde*I and *Xho*I digestion (Figure 4.13*D*). The nucleotide sequence of the construct was verified by Sanger sequencing to ensure no mutations had been introduced during the cloning procedure and the reading frame for the fusion was correct. The target protein contained eight additional amino acid residues at the C-terminus, including the six-histidine residues of the tag for purification purposes.

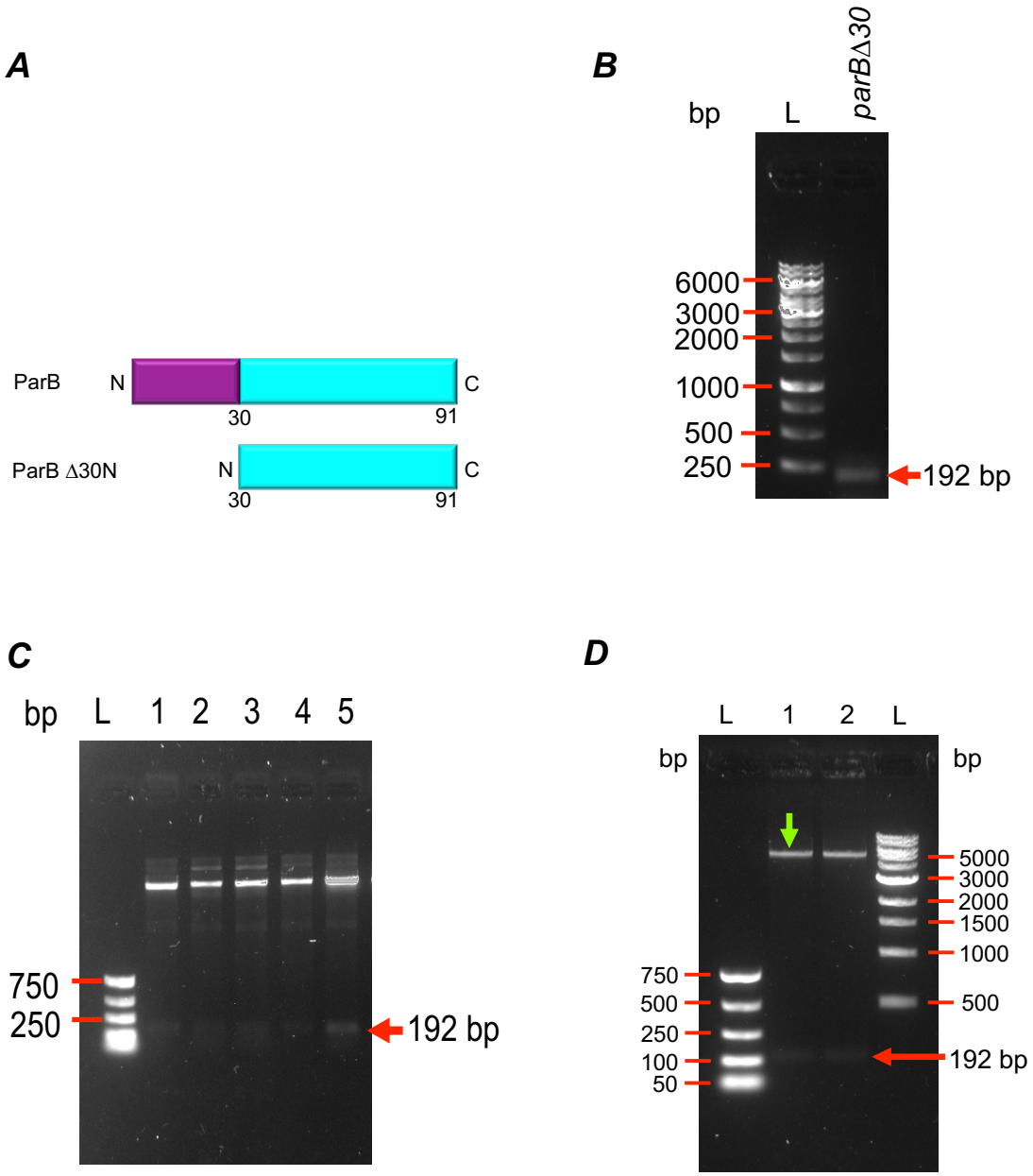


Figure 4.13 Cloning of *parBΔ30N* into pET22b (+) vector

A) Cartoon diagram comparing the wild-type ParB with ParBΔ30N. ***B)*** 1% agarose gel showing *parBΔ30N* amplified by PCR. L in lane 1 represents the 1 kb ladder (Thermo Fisher Scientific). Lane 2 shows the 192 bp PCR product that corresponds the *parBΔ30N* and is indicated by a red arrow. ***C)*** 1% agarose gel showing the colony PCR of the positive clones. L is the PCR ladder (New England Biolab) and lanes 2-6 show colony PCR products of positive clones. The amplified *parBΔ30N* gene is indicated by a red arrow. ***D)*** 1% agarose gel showing the digestion of the positive clones with *Xho*I and *Nde*I. L in lane 1 and 4 is the PCR marker (NEB) and 1 kb ladder (NEB), respectively. Lanes 2 and 3 represent recombinant clones digested with *Xho*I and *Nde*I. The two bands represent the vector and the *parBΔ30N* insert. The inserted gene (192 bp) is indicated by a red arrow. The green arrow shows the digested pET22b (+) vector.

A

M V K K P S Q Q A L N R A A V T V E Q A E A L A Q R 26
L A D K P Y G A P E K P E P E K Q C R T T I S L G E 52
S M L V T I E D L A L R N K R N G K D P K N V S A I 78
V R V A L E Q Y L K T L T 91

B

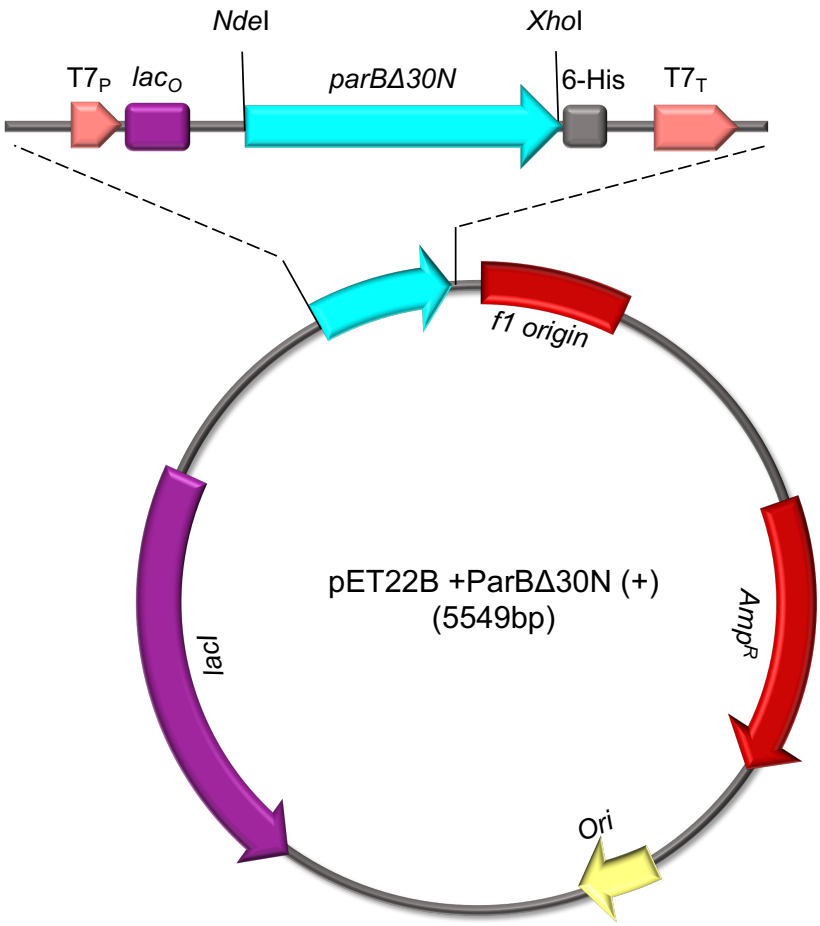


Figure 4.14 pET-ParBΔ30N plasmid map

A) ParB full-length sequence with the 30 N-terminal amino acid deleted in this construct and shown in grey. The sequence of ParBΔ30N is shown in cyan. ***B)*** Plasmid map of pET-ParBΔ30N. The truncated gene was cloned using *Nde*I and *Xho*I restriction sites in frame with a C-terminal $(His)_6$-tag. The inserted gene is located downstream of T7 promoter ($T7_P$) and the *lac* operator (lac_O) and upstream of T7 terminator ($T7_T$).

4.2.5.2 ParBΔ30N overproduction and purification

The level of expression ParBΔ30N in LB medium was tested using the same approach as per wild-type ParB. The ParBΔ30N $(His)_6$-tag fusion protein was first overexpressed in a 50 ml cultures of *E. coli* BL21 (DE3) grown in LB as described in Section 2.4.1. The expression level of the $(His)_6$-tag ParBΔ30N protein was verified by SDS-PAGE. The amount of ParBΔ30N overproduced after three-hour inductions with 1 mM IPTG at 37°C was sufficient. This result predicted a successful large-scale expression (Figure 4.15*A*). Once the overexpression was confirmed in LB medium, the construct was grown in 50 ml M9 minimal medium as detailed in Section 2.4.3. The expression level of ParBΔ30N tagged protein was verified using SDS-PAGE and it was as high as that observed in LB medium (Figure 4.15*A*). Once the expression was confirmed in 50 ml M9 minimal medium culture, ParBΔ30N $(His)_6$-tagged was overproduced in a 500 ml M9 minimal medium culture, containing ^{15}N-labelled ammonium chloride and ^{13}C-labelled glucose, following the Bracken *et al.,* (2001) approach. SDS-PAGE confirmed that the amount of protein generated after three hours of induction at 37°C was sufficient for downstream experiments.

Once overproduction was completed, cells were harvested by centrifugation and lysed by sonication as described in Section 2.4.5. Ni^{+2} affinity chromatography was then used as a first purification step to separate the target protein from the rest of the soluble *E. coli* proteins. Samples were collected during the process and examined by SDS-PAGE (Figure 4.15*B*) to determine which fractions were suitable for further purification steps. SEC was used to remove any remaining contaminants and to exchange the buffer for further NMR experiments. Following SEC, the purity of the sample was good enough and no contaminant could be seen by SDS-PAGE (Figure 4.15*B*). The total yield of ParBΔ30N purified from a 500 ml culture was 16 mg/ml. The overexpression and purification for this construct were efficient, and the amount of purified protein was two times higher than that obtained for wild-type ParB. In the ParBΔ30N construct, 30 amino acid residues are missing from the N-terminal domain and this region is believed to be unstructured or partially folded based on the chemical shift dispersion in proton dimension. It is possible that the deletion of the unfolded region increased the protein solubility. The mutant ParB protein product is 8 kDa, including the $(His)_6$-tag.

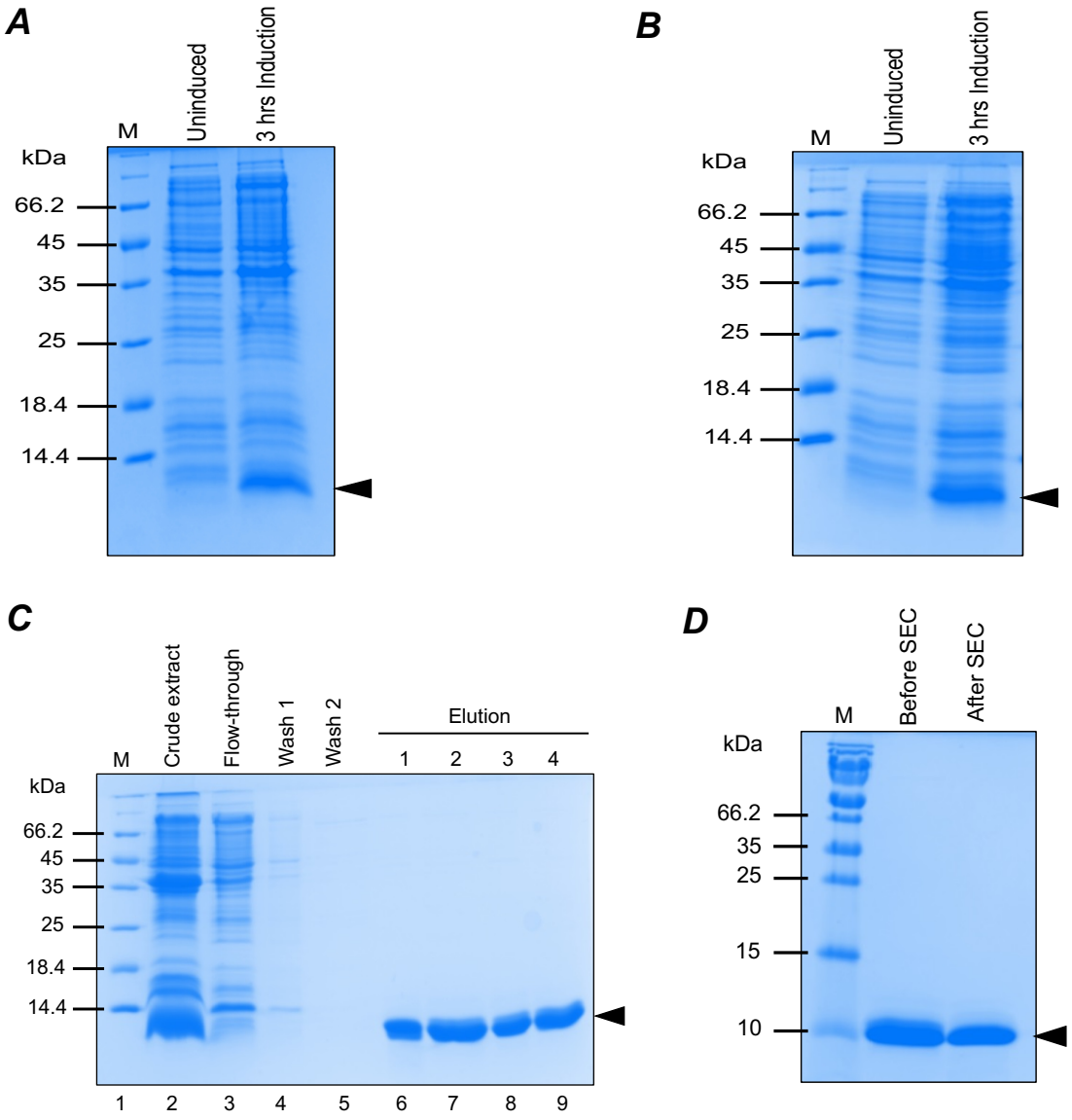


Figure 4.15 ParBΔ30N overproduction and purification from a 500 ml minimal medium culture

A) and ***B)*** SDS-PAGE showing the overproduction of ParBΔ30N in enrichment LB medium **(*A*)** and in [^{15}N and ^{13}C] labelled minimal medium **(*B*)**. In both gels, uninduced and induced *E. coli* cells are compared. The target overexpressed 8 kDa ParBΔ30N protein is indicated by an arrow. M represents the unstained PageRuler protein ladder (Thermo Scientific). ***C)*** and ***D),*** purification of ParBΔ30N using Ni^{+2} affinity chromatography and SEC, respectively. ***C)*** SDS-polyacrylamide gel showing the purification of ParB by Ni^{+2} affinity chromatography. Lane 1 is the molecular weight protein marker. Lane 2 shows the cleared lysate before loading on the pre-charged nickel column. Lane 3 shows the flow-through. Lanes 4 and 5 contain aliquots of two consecutive wash steps using 15 and 85 mM imidazole, respectively. Lanes 6-9 show ParB fractions elluted using 400 mM imidazole. The purified 8 kDa (including $(His)_6$-tag) ParB protein is indicated by an arrow. ***D)*** SDS-polyacrylamide gel showing the second purification step using SEC. The purified ParBΔ30N sample after the first purification is compared with the concentrated elution sample after SEC. The purified 8 kDa ParB protein (including $(His)_6$-tag) is indicated by an arrow.

4.2.5.3 NMR analysis shows that ParBΔ30N is folded

The amount of protein generated after the two-step purification was sufficient for downstream experiments. Although high purity and quantity of ParBΔ30N were achieved, it was necessary to check that the protein was functional and folded. Thus, 2D [^{1}H, ^{15}N] SOFAST-HMQC was performed for ^{15}N labelled ParB as well as ParBΔ30N to investigate whether the missing 30 amino acids had affected the folding state of the protein. Single ^{1}H, ^{15}N labelled ParB and ParBΔ30N NMR samples were prepared at a concentration equal to 150 μM. The HMQC experiments showed that both proteins are folded, and the cross-peaks are dispersed in both, proton and nitrogen dimensions. To assess whether truncation had affected the folding state, the two HMQC spectra were superimposed. Figure 4.16 shows that the deletion of 30 amino acids has not disrupted the structure of the protein as the two spectra overly well. Most of the peaks from the ParBΔ30N are overlaid well with the full-length ParB. This result was expected as the two proteins share the same core structure. It should be noted that a relatively small numbers of peaks are significantly shifted. These changes in chemical shifts are due to the fact the chemical environment of these residues has changed upon deletion of the first 30 amino acids. Interestingly, In the wild-type ParB, peaks with broad chemical shift distribution in the proton dimension are weak. This could be due to the fact that the unstructured part produces sharp peaks, which may overlap with the folded region. Undeniably, deleting the first 30 amino acid residues from the N-terminal region has significantly improved the signals produced by the structured C-terminal domain. This suggests that truncation of the N-terminal region, that has already been assigned, could be useful to obtain the backbone resonance assignment of the remaining peaks of the amino acids in the C-terminal region.

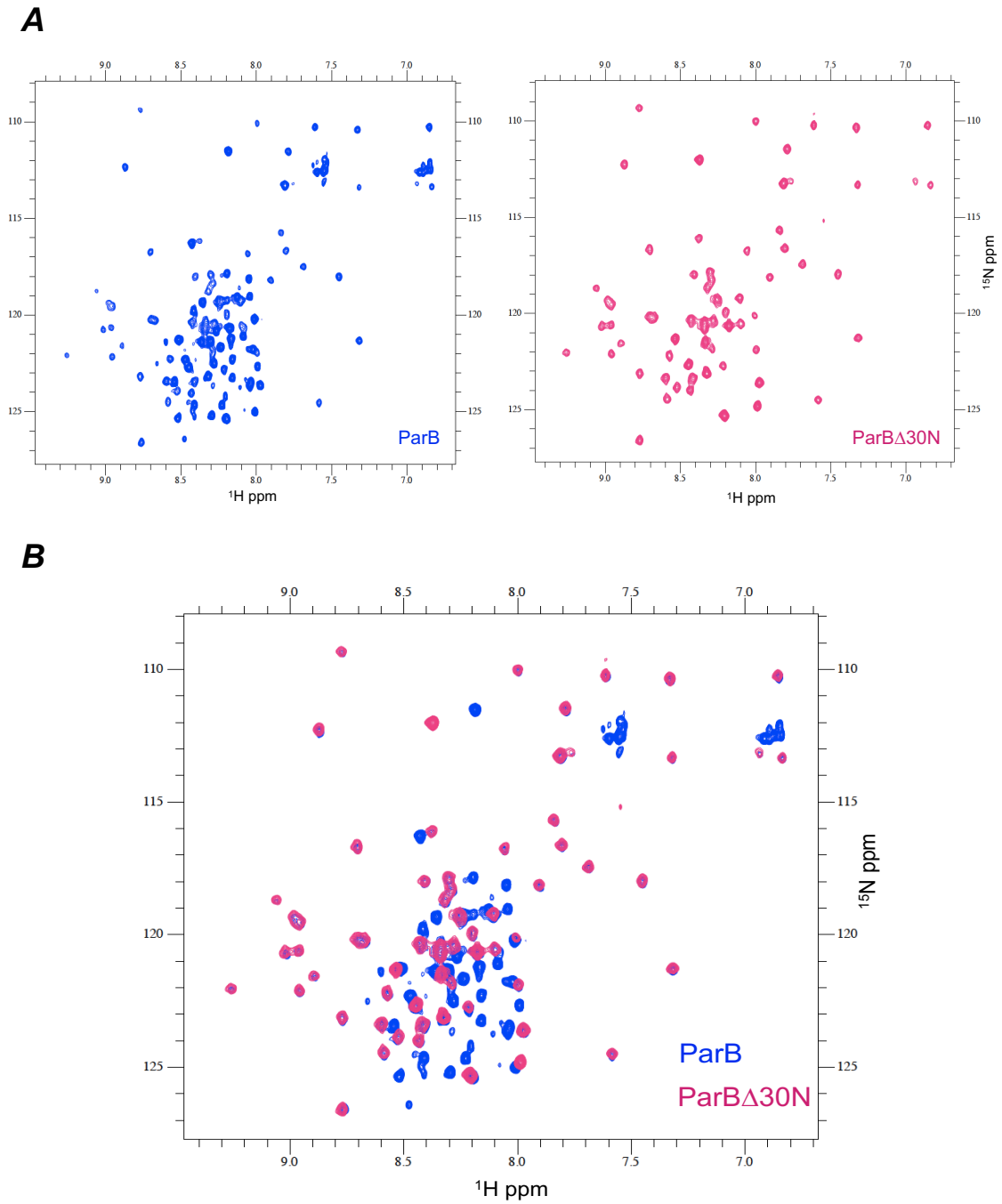


Figure 4.16 Comparison of [^{1}H, ^{15}N] HMQC NMR spectra of the wild-type ParB and ParBΔ30N

A) On the left, spectrum of wild-type ParB, represented by blue cross-peaks. On the right, spectrum of the mutant ParB (ParBΔ30N) represented by red cross-peaks. Both spectra were obtained with a final protein concentration 200 μM in the buffer 20 mM Bis-Tris pH 6.0, 50 mM NaCl and 1 mM DTT. The spectra was recorded at 700 MHz, 25ºC. ***B)*** Superposition of the spectra of ParB wild-type (blue) and ParBΔ30N (red).

4.2.5.4 Backbone assignment of ParB C-terminal domain

Following the initial test confirming that the new construct is folded, a [^{15}N, ^{13}C] double labelled sample of the ParBΔ30N protein was prepared and purified using Ni^{+2} affinity chromatography followed by SEC. The NMR sample was prepared at a concentration of ~1 mM in a 600 μl volume. Similarly, to ParB, ParBΔ30N tends to precipitate at high concentration as the sample became cloudy, when concentrated down to more than 1 mM. Overall, deletion of the flexible part from ParB, made the mutant protein more stable than the wild-type. The quality of the sample was verified using 2D HSQC before performing further experiments.

2D [^{1}H, ^{15}N] HSQC spectrum of 1.2 mM of uniformly [^{13}C, ^{15}N] labelled ParBΔ30N shows good dispersion of amide signals (Figure 4.17). A total of 69 of cross-peaks were identified, which is close to what is predicted from the protein sequence (64 peaks). The extra peaks in the spectrum are side-chain amides, which appear as a pair, as indicated in Figure 4.17. In addition to side-chain amides peaks, another four weak peaks were left unassigned. These unassigned peaks share the same Cβ chemical shift with no Cα peaks. Cross-peaks for all residues could be observed with the exception of five of the $(His)_6$-tag at the C-terminus of the ParBΔ30N. The sequential protein resonance assignment was achieved by using a standard of 3D triple resonance experiment with 1.2 mM protein concentration, namely: HNCA/HN(CO)CA, HNCO/HN(CA)CO and HNCACB, in addition to HANH and HA(CO)NH (Table 4.1). HSQC spectra were recorded following each 3D experiment to check the stability of the sample after incubation at 18°C for a prolonged time. Like ParB, the ParBΔ30N protein was also found to be less stable after 48-hour incubation at room temperature. Therefore, the sample was replaced with a fresh preparation, when required. The 3D experiments recorded were not sufficient to achieve full assignment as many Cα and Cβ peaks were missing. The reason behinds missing peaks might be that some 3D experiments, such as HNCACB, were recorded after 36-hour incubation at room temperature, and that the protein was shown to be unstable after prolonged incubation time. Thus, it was difficult to sequentially link all amino acids. Moreover, amino-acid type determination was impossible with only the presence of unique Cβ and Cα signatures for all the Ala, Gly, Ser and Thr residues. As a result, further information was necessary to connect the ambiguous regions for successful backbone resonance assignment. In addition, the presence of the same sequences of amino acids separated by proline space in between them also caused difficulties for sequentially linking fragments. Therefore, HADAMAC experiments were recorded to obtain further information on a specific amino acid type determination. HADAMAC experiment is a quick two-dimension NMR experiment that

provides significantly higher amino acid type determination with short time than any other standard ^{13}C chemical shifts (Lescop *et al.*, 2008; Lescop and Brutscher, 2009).

With all these experiments, it was possible to correctly link the pieces of fragments to achieve a complete protein backbone assignment, as shown in Figure 4.17.

Table 4.1 List of 2D and 3D NMR experiments used for backbone assignment of the ParBΔ30N

The number of cross-peaks identified in each experiment and the percentage of assignment achieved per each experiment is shown.

Experiments	No. of Peaks	% Assigned
HSQC	75	100
HNCA	127	92.7
HNCACO	111	93.1
HNCO	68	96.6
HNCOCA	66	96
HANH	69	98.6
HA(CO)NH	58	94.8
HNCACB	61	32.8

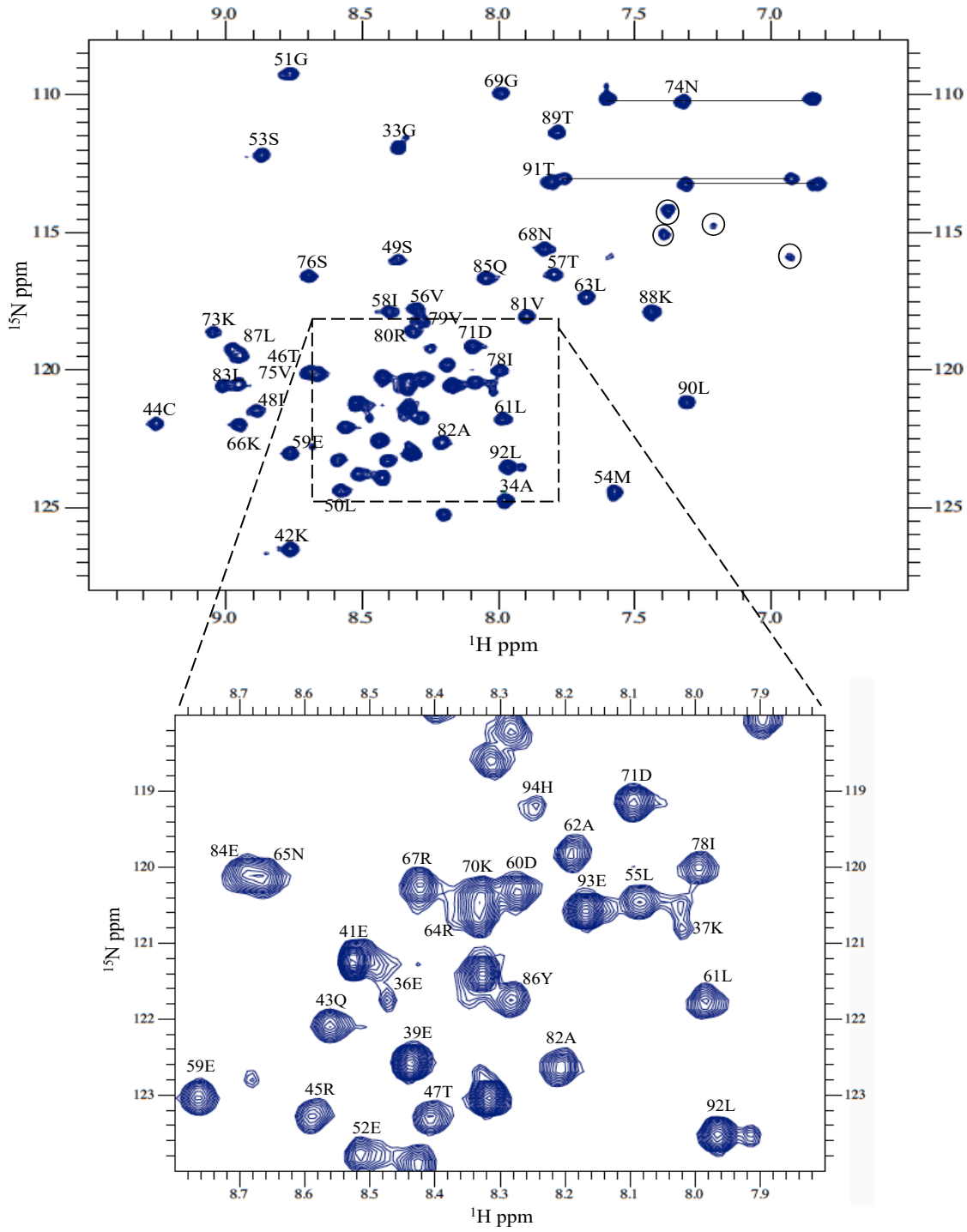


Figure 4.17 Backbone assignments for ParBΔ30N

[^{1}H, ^{15}N] HSQC spectrum of 1.2 mM monomer of uniformly [^{13}C, ^{15}N] labelled ParBΔ30N in 20 mM Bis-Tris, pH 6.0, 50 mM NaCl and 1 mM DTT was recorded at 700 MHz, 25ºC. The backbone assignments are shown through the residue number and one letter for each amino acid. The central congested region of the spectrum is enlarged to visualise the remaining assigned peaks. Pairs of side-chain amides are connected with horizontal lines. Four cross-peaks that remain unassigned are enclosed in circles.

4.2.6 Secondary structure prediction of ParB

Once the sequential backbone resonance assignment was completed for the N-terminal domain of ParB and ParBΔ30N, it was possible to predict the secondary structure of ParB. Chemical shift information obtained for both ParB N-terminal domain (residue 8-34) and ParBΔ30N were combined to predict the secondary structure of the full-length protein. Residues 3-7 were not included, as little chemical shifts information from backbone assignment were obtained. The secondary structure of ParB was predicted based on the chemical shift of HN, Hα, Cα, Cβ, C_O and N, that were obtained from backbone assignment using the TALOS-N server. The TALOS-N server is an updated version of TALOS+ that relies more on the use of artificial neural networks compared to its precursor. TALOS uses NMR chemical shifts to calculate protein backbone and side chain torsion angles. Hence, the TALOS-N server generates a highly accurate and reliable prediction of protein secondary structure (Shen and Bax, 2013, 2015). The TALOS results indicated that the ParB protein consists of two domains: N-terminal and C-terminal domain. The N-terminal domain (residue 8-43) is unstructured/flexible with a small α-helix (α1) structure in the middle region consisting of ten amino acids (Val-17-Arg-26) and with two flexible tails on both sides of the helix. In contrast, the C-terminal region adopts a RHH structure, with a β-strand, β1 (Cys-44-Leu-50), aa α-helix, α2 (Glu-52-Arg-67) and another α-helix, α3 (Val-73- Thr-89) (Figure 4.18). The two helices in the C-terminal domain, α2 and α3, are connected by a seven amino acids long loop, while the β1 and the α2 are separated by a single amino acid (Gly-51). Overall, ParB secondary structure consists of 45% loop, 47% helices and 8% β-strand.

To further investigate the conformation of ParB and its dynamics in solution, [^{1}H, ^{15}N] hetNOE experiment was performed. The hetNOE experiment provides information about the movement of N-H bond in the protein backbone. Those that undergoes movement faster than the overall tumbling of the protein (pico to nanosecond time scale) gives low NOE signal intensity. The decreased values are usually located in the N-terminal or C-terminal ends of the protein (Vögeli, 2014). In hetNOE, values >0.6 corresponds to a well-structured region and can be described as a rigid region, values between 0.4-0.6 indicate the semi-structured region or semi-rigid flexibility, and values less than 0.2 indicate flexibility. The hetNOE value in the C-terminal region of ParB was between 0.5-1, which reveals that this region is highly rigid/ structured. On the other hand, the N-terminal domain was found to be oscillating between highly flexible (residue 1- 16 and residue 26-42) to semi-rigid flexibility (residue 17-25) (Figure 4.19). This means that the N-terminal domain is highly dynamic in solution. Finally, the region between helices α2

and α3 was also shown to have a similar ratio >0.6, that suggests rigidity. However, since the secondary structure prediction showed that this region is flexible, we would assume that this value is due to the fact that the flexible loop is positioned between two helices, which slows the dynamic of this region.

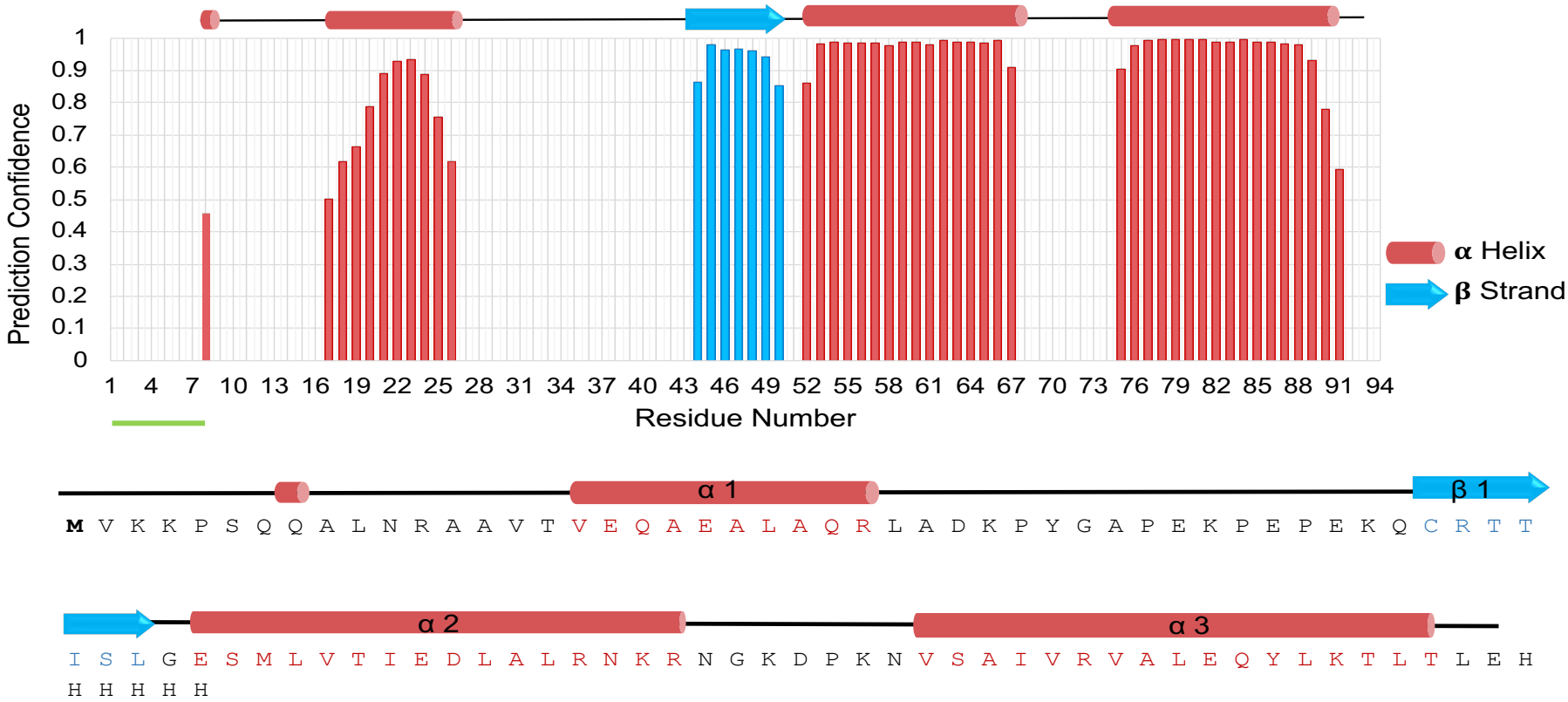


Figure 4.18 Secondary structure elements of ParB predicted by TALOS-N server from the NMR chemical shift assignment

The secondary structure prediction confidence is plotted as a function of 7-91 amino acid residues numbers. Red bars represent the α-helices, while blue bars represent the β-strand. Unpredicted residues (1-7) are highlighted with horizontal green line. The cartoon representing the secondary structure prediction of ParB is shown on top, while ParB amino acids sequences with the secondary structure prediction elements are shown at the bottom.

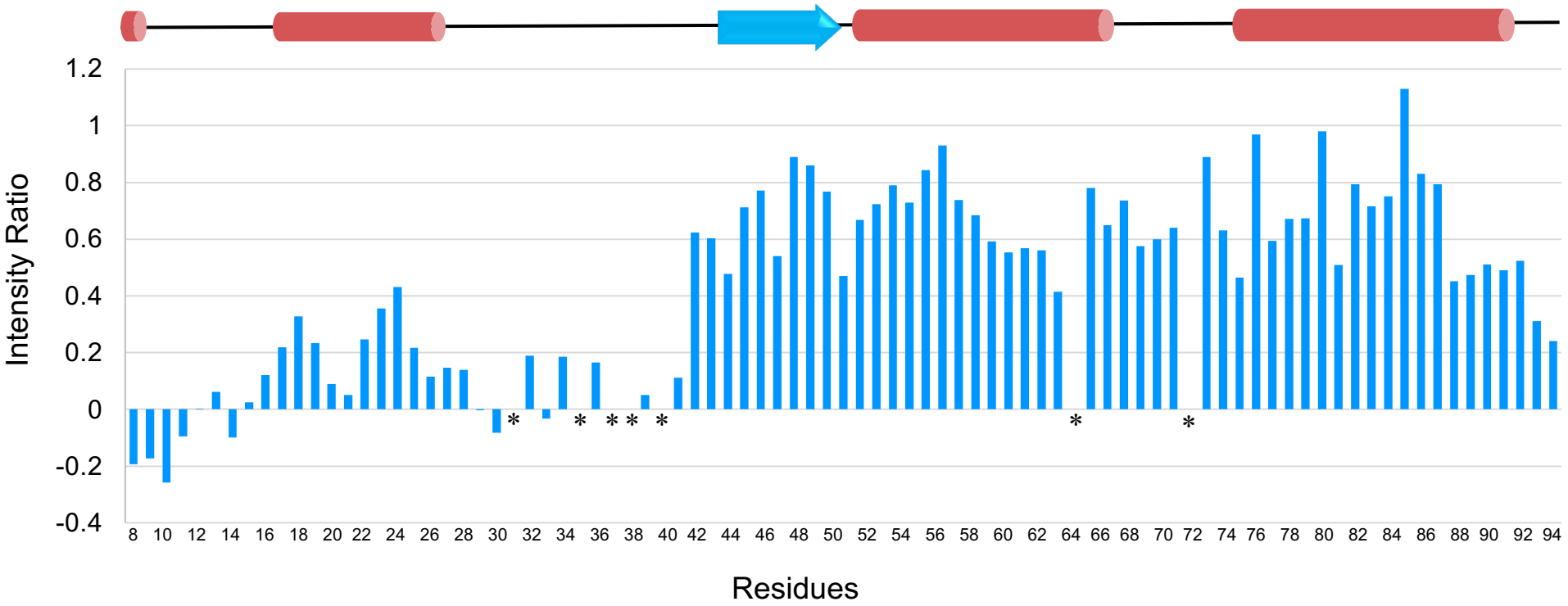


Figure 4.19 [^{1}H, ^{15}N] heteronuclear NOE

The relative intensity ratio of heteronuclearNOE is plotted as a function of residue number. The first seven residues are missing. The black asterisks represent the unassigned amino acids. The predicted secondary structure of ParB from TALOS-N is shown at the top of the figure.

4.2.7 Using X-crystallography to determine the structure of the C-terminal domain of ParBΔ30N

4.2.7.1 Crystallisation of ParBΔ30N

ParBΔ30N was used instead of the wild-type ParB due to the flexible end at the N-terminal domain that could interfere with crystal growth. ParBΔ30N was overproduced and purified as described previously from a 500 ml LB culture. Few batches of the freshly purified ParBΔ30N protein were buffer exchanged into 20 mM Bis-Tris, pH 6.0, and 50 mM NaCl, using 5 kDa MWCO dialysis tube overnight at 4°C room temperature, as described in Section 2.6.2. The protein was concentrated to approximately 6 mg/ml using 6 kDa MWCO concentrators prior to crystallisation trials. Tris (2-carboxyethyl) phosphine (TCEP) was added to 1 mM before crystal screen preparation.

Two commercially available crystal screening plates were used, PACT and X-tall-screen. ParBΔ30N was prepared at two different concentrations, 3 and 6 mg/ml. The experiment was carried out using the sitting-drop vapour-diffusion method. Each plate consisted of 96 crystallisation plate conditions, 200 nl of protein was mixed with an equivalent volume of mother liquor. Two identical plates were prepared simultaneously for incubation at 18° or 4°C. The details of crystal preparation are reported in Section 2.6.3.

Crystals were first observed after 48 hours at 18°C. Once the crystals were produced, they were harvested and flash-frozen in liquid nitrogen before diffraction experiments. Crystals of ParBΔ30N were first obtained in condition A6 (0.2 M Magnesium Chloride hexahydrate, 0.1 M Tris-HCl, pH 8.5, 30% w/v Polyethylene glycol 4,000) of the commercial crystal X-tall screen plate. To get larger crystals, the A6 condition was further optimised using the hanging drop vapour diffusion method by varying the protein concentration (2, 4 and 6 mg/ml), buffer pH (6, 6.5, 7, 7.5 and 8) and precipitant concentrations (0.1, 0.2, 0.5, 0.8 and 1 M of Tris-HCl). 2.5 μl of protein sample was mixed with an equal volume of the reservoir solution and the plate was incubated at 18°C. However, the crystals in this plate took more than two months to grow. The same crystal that was observed in condition A6 after two days, grew also in the plate incubated at 4°C. This crystal took two weeks to grow. The two crystals grown in A6 condition (X-tall screen plate) at 18°C and 4°C were sent for X-ray diffraction and the results revealed that the good crystals were those grown at 4°C.

A single long rod-like shape crystal was obtained after approximately two weeks at 4°C. The crystal was taken from 0.2 M Magnesium Chloride hexahydrate, 0.1 M Tris-HCl, pH

8.5, 30% w/v Polyethylene glycol 4,000 from X-TALL screen plate at 4°C (Figure 4.21*A*). Similar crystals were obtained from the same plate but with different incubation temperature, 18°C. However, these crystals gave lower diffraction resolution than the one grown at 4°C. More crystals were obtained, some of them are shown in Figure 4.20.

Due to rapid melting of the crystal on removal from 4°C, it was difficult to capture images. Crystal needed to be harvested quickly to avoid melting and flash-frozen in liquid nitrogen for X-ray diffraction experiments. High-resolution diffraction data were collected at beamline I03 of the Diamond Light Source synchrotron using a Pilatus 6M detector. A total of 2200 images were collected with 0.5° rotation per image and a maximum resolution of 1.93 Å (Figure 4.21).

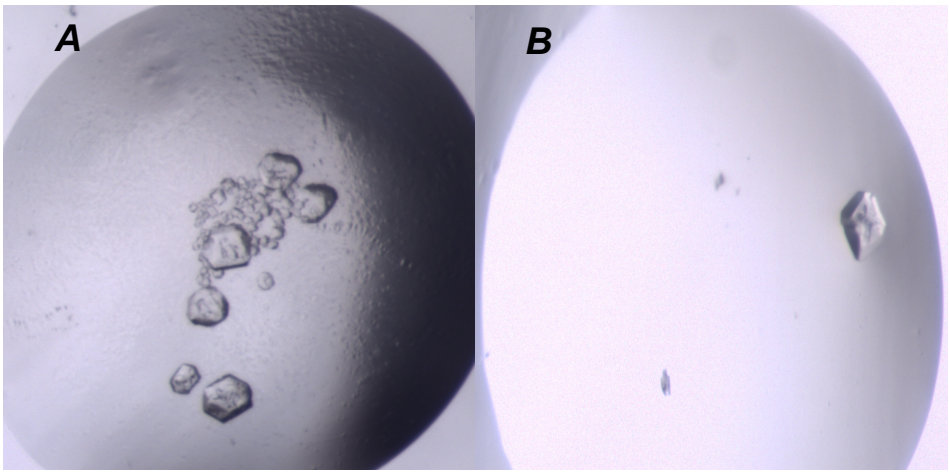
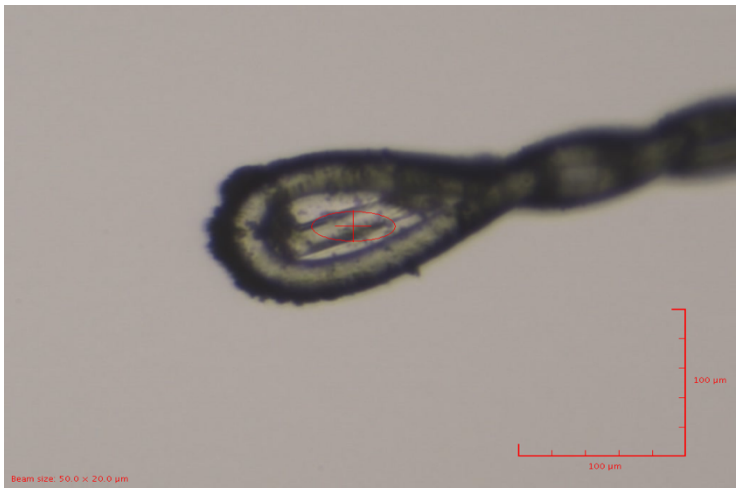


Figure 4.20 Crystals of ParBΔ30N

A) Hexagonal crystals observed after optimisation of the A6 condition from the X-TALL screening plate. 2.5 μl of 6 mg/ml protein was mixed with an equal volume of solution containing 0.2 M Magnesium Chloride hexahydrate, 0.1 M Tris-HCl, pH 8.5, 30% w/v Polyethylene glycol 4,000. ***B)*** A well-shaped single diamond crystal observed from E11 condition (0.01 M Cobalt (II) chloride hexahydrate, 0.1 M Sodium acetate trihydrate pH 4.6, 1.0 M 1,6-Hexanediol) from the PACT crystal screening plate.

A



B

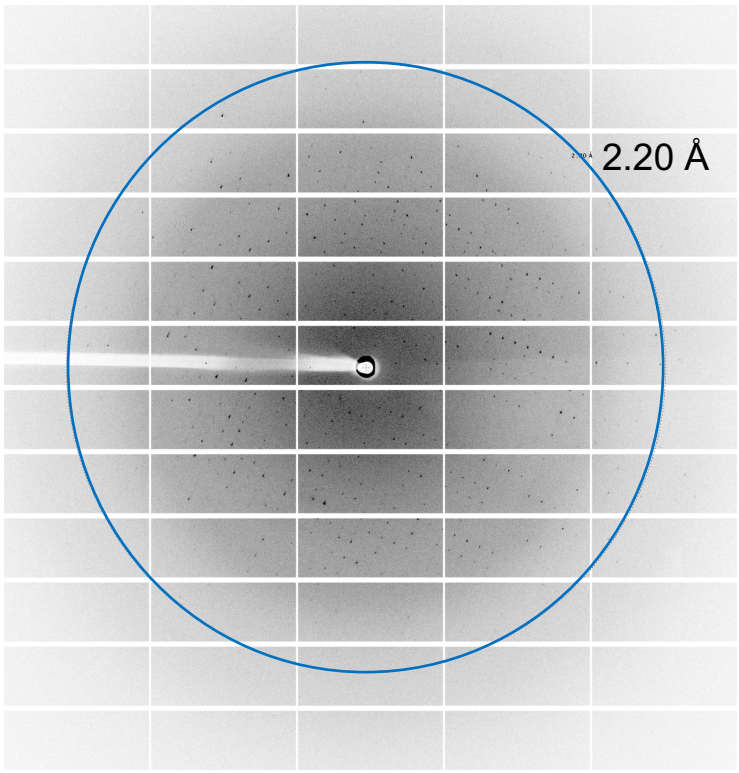


Figure 4.21 Mounted ParBΔ30N crystal and X-ray data diffraction

A) The mounted ParBΔ30N crystal fished by a fibre loop. The photograph of the crystal was shot by X-ray beam of 50 × 50 µm, indicated with red oval shape. At the bottom the size-scale bar is shown. ***B)*** Diffraction image was collected at a resolution of 1.93 Å. The blue circle corresponds to the 2.2 Å resolution, beyond which the crystal diffracted.

4.2.7.2 Structure determination of the C-terminal domain of ParB

4.2.7.2.1 Overall structure

The crystal structure of the ParB DNA-binding domain was determined. The structure obtained is limited to the C-terminal DNA-binding domain and no N-terminal domain is shown. Specifically, 11 amino acids of ParBΔ30N are missing from the crystal structure. Due to high flexibility of this region, these residues could not be identified in the electron density map. This result is consistent with the NMR result that shows that the N-terminal domain adopts an extremely flexible conformation. Therefore, data for that region was not obtained. The 3D structure of the ParB C-terminal domain was analysed by Dr Clément Dégut. The structure of the C-terminal domain of ParB was refined to a 1.93 Å resolution with an R-work of 0.2161 and R-free of 0.2577. The details of data collections and refinement are shown in Table 4.2.

Table 4.2 Crystallographic data collection and refinement statistics

Statistics for the highest-resolution shell are shown in parentheses.

Crystal data collection	**ParBΔ30N**
Beamline	I03
Resolution range	39.65 - 1.93 (1.999 - 1.93) Å
Space group	P 21 21 21
Unit cell	48.19, 51.995, 61.303, 90, 90, 90
Total reflections	24012 (2264)
Unique reflections	12008 (1132)
Multiplicity	2.0 (2.0)
Completeness (%)	99.35 (95.61)
Mean I/sigma(I)	9.27 (1.04)
Wilson B-factor	36.98
R-merge	0.02775 (0.6797)
R-meas	0.03924 (0.9613)
R-pim	0.02775 (0.6797)
CC1/2	0.999 (0.466)
CC*	1 (0.798)
Reflections used in refinement	11993 (1132)
Reflections used for R-free	603 (59)
R-work	0.2161 (0.4224)
R-free	0.2577 (0.4212)
CC (work)	0.938 (0.619)
CC (free)	0.914 (0.571)
Number of non-hydrogen atoms	893
macromolecules	864
solvent	29
Protein residues	110
RMS (bonds)	0.017
RMS (angles)	1.74
Ramachandran favored (%)	99.06
Ramachandran allowed (%)	0.94
Ramachandran outliers (%)	0.00
Rotamer outliers (%)	0.00
Clash score	6.20
Average B-factor	50.68
macromolecules	50.70
solvent	50.21

The structure revealed that the protein is a homodimer, where each chain has a RHH structure. The polypeptide chains were traceable in electron density map and covered the region of ParB starting from the amino acid 41 (which corresponds to amino acid 12 in the truncated ParBΔ30N). Therefore, the crystal structure of ParB revealed only the C-terminal domain, that adopts a dimeric RHH structure as its functional homologues ParG and Omega. The C-terminal domain of ParB is composed of two asymmetric RHH chains, intertwined to form a tight dimeric structure. The subunit I comprises of residues 41-96, while the subunit II consists of residues 41-94 (Figure 4.22). Since the (His)6-tag used for purification was not removed, in addition to the ParB 91 amino acids, the protein harbours eight additional amino acids (including Leu-92 and Glu-93 that correspond to the codons of the restriction site used for cloning). Only two amino acids, Glu-41 and Lys-42, from the N-terminal domain are observed. The C-terminal folded domain of both monomers is linked by a local two-fold rotation axis. Each monomeric subunit consists of a β-strand, followed by two α-helices, as shown in Figure 4.22. The two β-strands pair together forming an antiparallel β-sheet. Both β-strands consist of nine amino acid residues, starting from Gln-43 and ending with Gly-51. β-strands from both monomers are directly connected to the first helix. The C-terminal end of the β-strand and the N-terminal end of the α-helix 1 make a 90° turn (Figure 4.22). In both subunits, the α-helix 1 starts with Glu-52 and ends with Asn-68. The α-helix 2 in both subunits begins at Val-75 but terminates at His-95 in subunit 1 and Glu-94 in subunit 2. In each monomer, the two helices are separated by a flexible loop, which consists of six residues (Gly-69-Asn-74). The result obtained from X-ray crystallography is consistent with the data obtained through NMR spectroscopy. Figure 4.23 shows the comparisons between the results obtained from NMR and the 3D structure solved using X-ray crystallography.

Top and side view of homo dimeric ParB are shown in Figure 4.22. In the side view, the structure shows that the α-helix 1 in both monomers locate at the outlying region with the flexible loop turning towards the inside of the core structure. In contrast, α-helix 2 occupies the centre of the structure, forming the hydrophobic core structure. Thus, the rigid RHH_2 motif is held together tightly via the hydrophobic amino acid residues in the α-helices. Figure 4.22 shows the top view of the folded domain of ParB in which the twisted β-sheet occupies the centre on the top surface of the structure. The dimer interface was analysed by PISA server and it shows that 34 residues (64.8%) from subunit 1, and 35 residues from subunit 2, are involved in the dimer interface. The tight dimer is stabilised almost entirely by twelve hydrogen bonds, which are mainly located in the backbone of the β-strand and no salt bridges or covalent bonds were found in the structure. The twelve hydrogen bonds that connect the two monomers and the distance

between them are shown in Figure 4. 24 and Table 4.3. Moreover, the dimer interface area was calculated to be 1600 Å^2 (31.5%) at the total solvent-accessible area of ~5128 Å^2, and the solvation free energy average gain of dimer formation is also given by PISA server, -13.7 kcal/mol. The dimer interface is contiguous and extends from the N-terminal domain of β-strand to the C-terminal part of α2 with the total buried surface in the dimer is 3201.1Å^2.

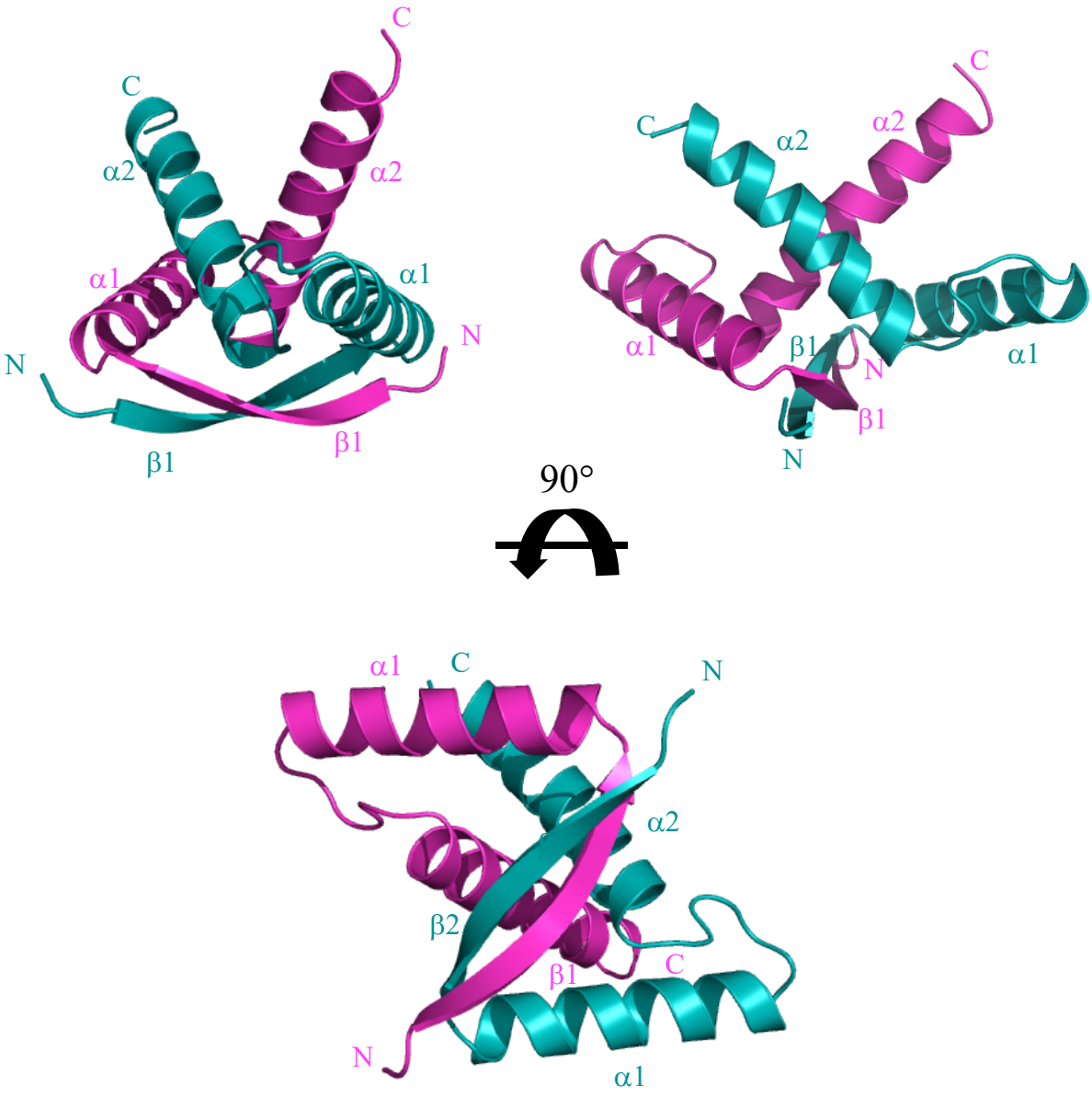


Figure 4.22 Tertiary structure of ParB C-terminal DNA-binding domain

Cartoon representing the dimeric RHH structure in three orthogonal views. The two monomers are shown in teal and magenta**.** The secondary structure elements are indicated.

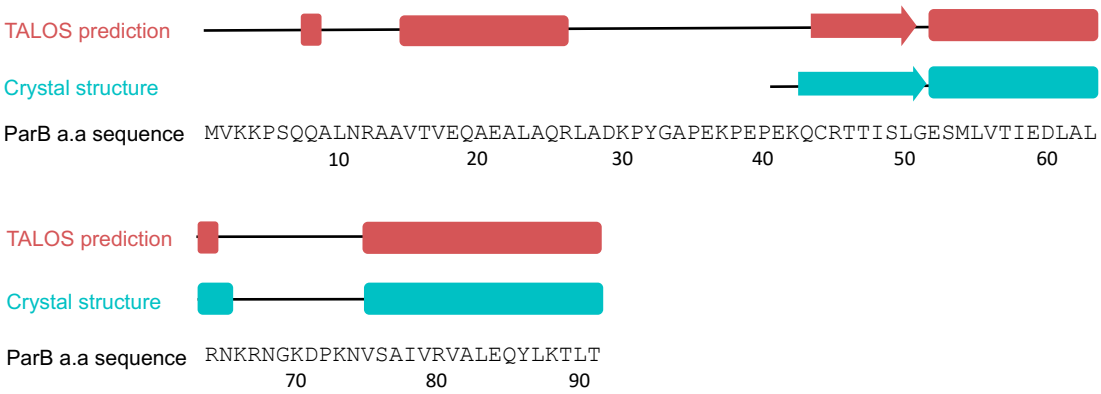
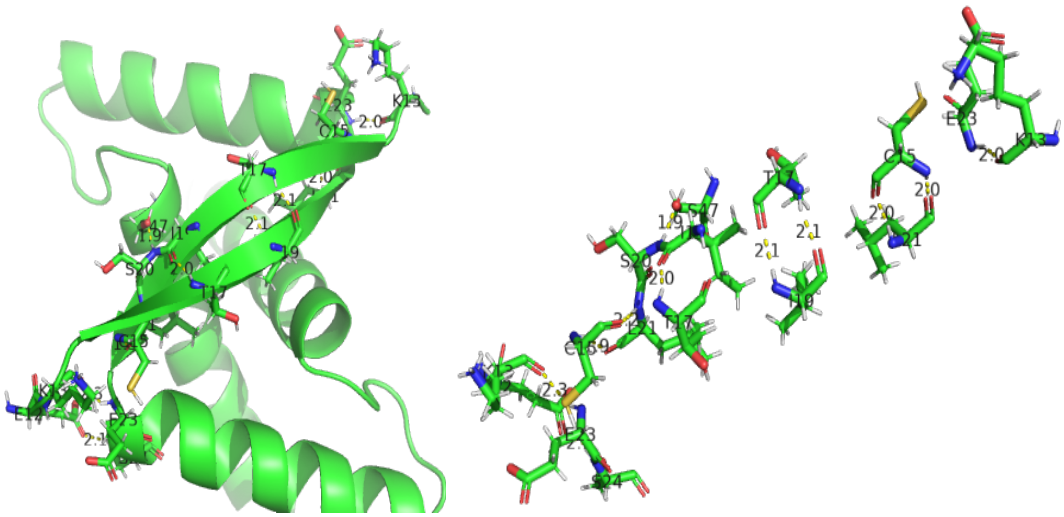


Figure 4.23 A graphical map of the secondary structure elements of ParB

A comparison of secondary structure elements of ParB obtained from NMR and X-ray crystallography. The boxes indicate the helices, while arrows indicate the β-strand. Salmon colour designates secondary structure predicted by TALOS. Teal represents secondary structure elements determined in the ParB C-terminal domain crystal structure. Both structure elements are displayed on the ParB complete sequences.



No.	Chain 1	Distance (Å)	Chain 2
1	Lys 13(O)	2.03	Glu 23(H)
2	Cys 15(O)	2.02	Leu 21(H)
3	Thr 17(O)	2.09	Ile 19(H)
4	Ile 19(O)	2.04	Thr 17(H)
5	Leu 21(O)	1.93	Cys 15(H)
6	Cys 15(H)	1.98	Leu 21(O)
7	Thr 17(H)	2.07	Ile 19(O)
8	Ile 19(H)	2.04	Thr 17(O)
9	Ser 20(H)	1.93	Ser 47(OG)
10	Lue 21(H)	2.16	Cys 15(O)
11	Glu 23(H)	2.35	Lys 13(O)
12	Ser 24(H)	2.43	Glu 12(OE1)

Figure 4.24 ParB dimer interface

Close up view of the ParB dimer interface showing the hydrogen bonds (dashed yellow lines) connecting the two subunits via backbone and side chain amides residues. The distance between connected residues are shown in Å. A list of hydrogen bonds between the two monomers is shown in the table.

4.2.7.3 ParB adopts a RHH fold and is a member of Arc/MetJ superfamily

The RHH motif is a well characterised DNA-binding domain, that is present in prokaryotic transcriptional repressors as well as in plasmid partition proteins of type Ib and type II systems (Murayama *et al.*, 2001; Golovanov *et al.*, 2003; Møller-Jensen *et al.*, 2007; Schreiter and Drennan, 2007; Huang *et al.*, 2011). Despite the fact that these proteins share a similar structure, the sequence identity among them is extremely low. This might be due to the fact that each protein recognises a specific sequence of DNA, suggesting variation in protein sequences despite maintaining a similar pattern in the interaction with the DNA. Due to the similarities with other RHH proteins, it is possible to infer that the twisted antiparallel β-strand inserts into the DNA major groove and make sequence-specific contact with base-determinants through residues in the β-strands. The interaction is firmed up through contact occurring between DNA backbone phosphates and the two α-helices (Raumann *et al.*, 1994). An attempt was made to co-crystallise wild-type ParB as well as ParBΔ30N in complex with short (harbouring two hexameric repeats) and long (harbouring four hexameric repeats) DNA sequences. However, due to time constraints, it was impossible to obtain the co-crystal structure. Although crystals were obtained, they diffracted with limited resolution and would require further optimisation.

Structural alignment of ParB monomeric subunit using DALI web server revealed that ParB is a member of the Arc/MetJ superfamily of transcriptional repressors. Within this family, sixteen proteins of known structure (including transcriptional repressors and partition proteins) were selected for structure comparison. The chosen proteins share 6-40% sequence identity in the folded RHH domain and with DALI Z-score range between 2.8-6.6. Among ParB structural homologues, ParB shares 20% sequence identity with Omega, 16% with TP228 ParG and 14% with pCXC100 ParB. The highest sequence identity recorded is between ParB and CopG (40%). Since Omega is a structural homologue of ParB and shares the highest sequence identity with ParB, the protein was used as a model to build up the ParB structure. In addition, reasons for selecting Omega were the following: first, the protein is a global transcriptional repressor as well as a partition protein. Second, the protein is a member of the Arc/MetJ transcriptional repressor family. Third, the protein is the only partition protein whose structure was solved in its apo-state as well as in complex with the operator site. Structure alignment comparison of monomer subunit of the C-terminal RHH fold of ParB generated a root means square deviation (RMSD) ranging between 1.0 Å for HicB to 5.1 Å for ParD over 42 and 46 of Cα, respectively as shown in Table 4.3. The N-terminal domains were not included in the structure alignment. The structure elements of the C-terminal fold RHH

of ParB is superimposed with structural known RHH fold of other DNA-binding protein of type Ib partition system and is shown in Figure 4.25. From structure alignment it is possible to see that the three known structure proteins of type Ib segregation seystem (ParG, Omega and pCXC100 ParB) share the same α2 fold (Figure 4.25), whereas, the β-strand and α1 are very divergent.

Table 4.3 Sequence and structure comparisons of C-terminal RHH fold of ParB with other member of the Arc/MetJ transcriptional repressor family whose structure are known

RHH protein	PDB code	Sequence identity %	No. of Cα align	RMSD (Å)
HicB	6g1nA	19	42	1.0
MetJ	1mjlA	20	49	1.2
CopG	2cpgA	40	43	1.3
Arc	1mykB	13	45	1.6
AmrZ	3qoqD	19	48	1.6
BcsZ	3gxqA	23	44	1.7
ParG	1p94	16	45	1.9
PutA	2gpeD	17	48	2.0
NikR	1q5vC	29	48	2.2
Omega	1irq	20	46	2.3
TraM	3on0A	17	53	2.4
DinJ	4q2uE	14	44	2.5
ParB	3no7	14	52	2.8
Fitab	2bsqE	20	46	2.9
Mnt	1mntA	6	48	4.0
ParD	2an7B	15	46	5.1

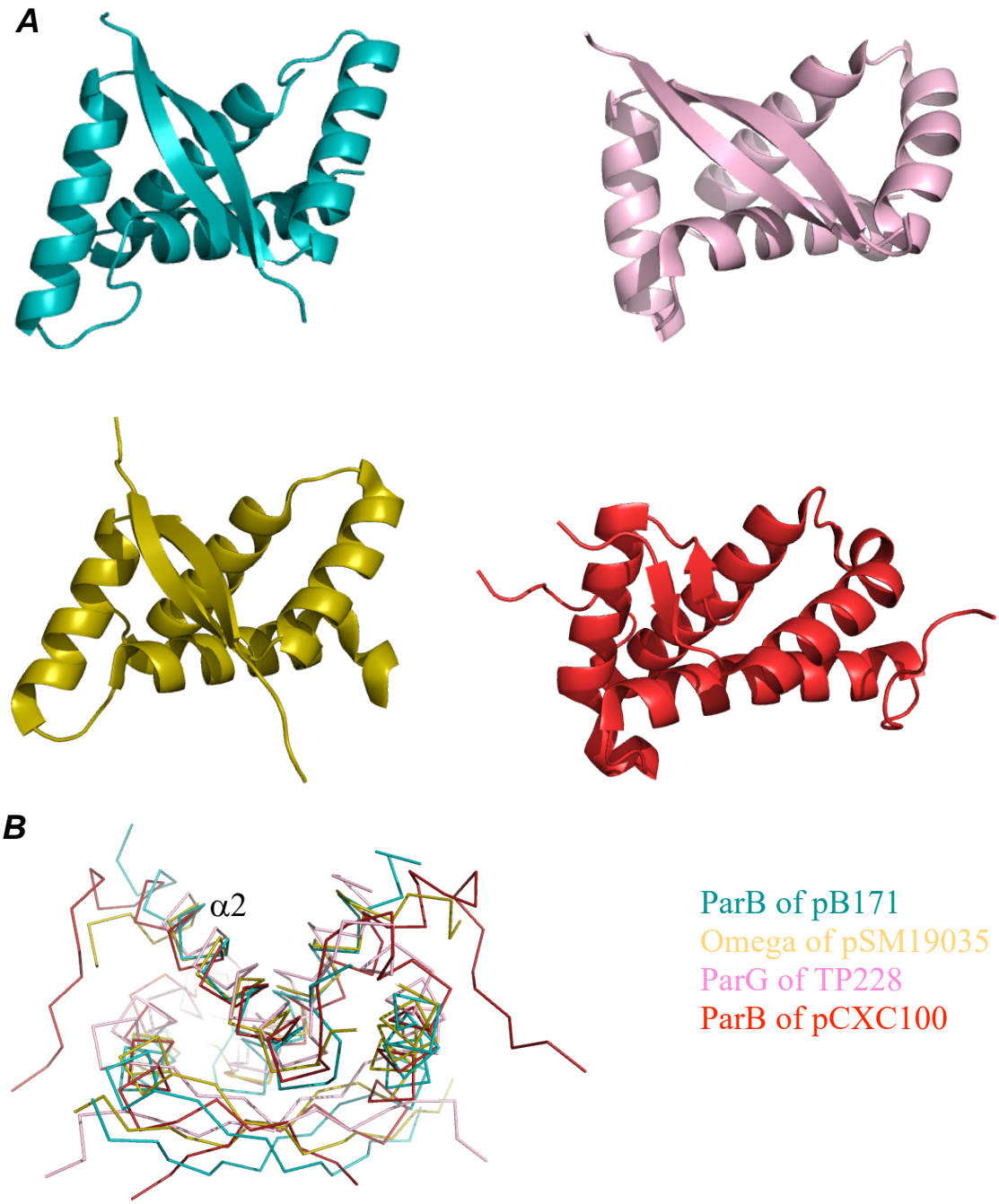


Figure 4.25 Structure superimposition of ParB with other RHH DNA-binding protein of type Ib partition systems

A) The cartoon structure of the pB171 ParB is shown in teal, TP228 ParG is shown in light purple (1p94), pSM19035 Omega is shown in olive (1irq), and structure of pCXC100 ParB is shown in red (3no7). ***B)*** The structure of all the structural homologues are superimposed onto the ParB of pB171 (teal) and shown in ribbon. These structures are aligned with focus on the structure conserved α-helix 2 (indicated). All the images were generated using PyMOL.

4.3 Conclusions and discussion

4.3.1 ParB adopts a dimeric Ribbon-Helix-Helix DNA-binding domain

A thorough investigation of molecular mechanisms that underpin DNA segregation requires us to elucidate the tertiary structure of the proteins that are involved in DNA partition. Solving the 3D structure of ParB will help to elucidate how the protein interacts with the DNA. Here, a comprehensive investigation was carried out to structurally characterise the DNA-binding protein ParB using a combination of NMR spectroscopy and X-ray crystallography. Initially, it was attempted to solve the structure of ParB in solution using NMR spectroscopy, as most of the ParB homologues have a flexible N-terminal domain, which makes it impossible to determine the structure using X-ray crystallography. The secondary structure of the full length of ParB was predicted using the chemical shift from backbone Cα, Cβ, NH, CO nuclei. Structure prediction by TALOS revealed that ParB consists of two domains: a long and highly unstructured N-terminal domain that consists of the first 43 amino acids and a structured C-terminal domain. The flexible N-terminal domain contains a loose short α-helical structure that is located at the centre of the domain. In contrast, the C-terminal domain folds to form the RHH structure. Moreover, structure prediction using chemical shift indicates that the centre region of the N-terminal domain form a helix and it proved later by hetNOE that this region is semi-rigid. The low spectral dispersion of this particular region and the hetNOE result suggest that the helix is highly dynamic. I postulate that the α-helix in the N-terminal domain will be stabilized once it interacts with either DNA or with its partner, ParA. Due to the flexible N-terminal domain, it was not possible to assign side-chain chemical shifts to allow 3D structure of ParB to be solved using NMR spectroscopy. Therefore, the structure of the truncated version of ParB was solved using X-ray crystallography. The structure shows the DNA-binding domain of ParB is located in the C-terminus and folds to form a dimeric RHH structure. The monomers of the C-terminal fold of ParB arranged symmetrically to form a tight homo-dimer with 2-fold rotation axis and the two antiparallel β-strands intertwined to form a β-sheet.

As far the N-terminal domain is concerned, it is important to note that all DNA-binding proteins identified so far from the type Ib plasmid segregation system have a C-terminal RHH fold, while the N-termini are unstructured (Murayama *et al.*, 2001; Golovanov *et al.*, 2003; Huang *et al.*, 2011). The number of amino acids residing in N-terminal domain varies according to the protein and its size. For instance, short N-terminal tails are found both in ParG (residues 1-32) and Omega (residues 1-22), whereas a long N-terminal tail is found in ParB of pCX1001(residues 1-65). It seems that the length of the N-terminal

domain depends mainly on the length of the protein, as the size of the C-terminal RHH fold in all proteins are typically very close. The N-terminal flexible tail might be considered as a unique feature of DNA-binding protein of type Ib system. In contrast, the DNA-binding protein from the type II system possess an N-terminal RHH fold and no flexible tail (Moller-Jensen *et al.*, 2007; Schumacher, Maria A, Glover, *et al.*, 2007; Popp *et al.*, 2010). Despite the fact that DNA-binding proteins of type Ib systems share the feature of a flexible N-terminal tail, no sequence identity is observed in this region among these proteins.

Structure prediction of the N-terminal domain of ParB by NMR shows that residues 17-26 could form a helix. Moreover, hetNOE shows that this region possesses a semi-rigid structure in solution. I speculate that this region might be able to fold when it interacts with the DNA or with the ParB partner, ParA. The phenomenon of changes in conformation has been seen before in the MetJ, Mnt and Arc transcriptional repressors DNA binding proteins (Knight and Sauer, 1989; Augustus *et al.*, 2006). It was shown that the N-terminal domain in the aforementioned proteins makes further base contact with their DNA cognate sites and thus, it undergoes conformational changes. To prove this hypothesis, thorough investigations need to be carried out. An attempt was made to co-crystalize ParB or ParBΔ30N with a short and long DNA fragment that corresponds to two or four repeats of the partition site to identify whether the N-terminal tail folds once it binds to DNA. Unfortunately, the trials that have been made were unsuccessful due to limited time. More investigations of the interaction of ParB with DNA are shown in Chapter 5.

Structure comparisons searches reveal that ParB belongs to the Arc/MetJ transcriptional repressors superfamily. All proteins in this family have a short RHH DNA-binding domain, which is either located at the C-terminus or the N-terminus (Schreiter and Drennan, 2007). Proteins with C-terminal RHH fold usually have a highly flexible N-terminal domain but varying in amino acid sequences and length (Schreiter and Drennan, 2007). On the other hand, the sequence comparisons in the RHH fold amongst this family have shown no significant sequence identity, especially in the β-strand and α-helix 1, but they share some sequence features. In the β-strand, despite varying in length, most of the transcriptional regulator proteins share a similar pattern with the presence of Arg or Lys either in position three, five, or seven. However, there are exceptions with short β-strands, where no positively charged amino acids are found in those positions. However, in proteins with a short β-strand, positively charged Arginines or Lysines are present just upstream of the β-strand, where the N-terminal domain is connected to the folded C-

terminal domain. Based on the position of the positively charged Arg or Lys in the β-strand, Golovanov *et al.,* categorised those located at the beginning of the β-strand as type I repressors (ParG, CopG, ParD), and those located at the end, as type II repressors (Omega, Arc, Mnt). Moreover, it is believed that the side-chain of the positively charged Arginine or Lysine makes a base contact (Golovanov *et al.*, 2003). Based on this classification, the positively charged Arg of ParB of pB171 is situated at position three, which is at the beginning of the β-strand. This observation indicates that ParB belongs to type I repressor. Moreover, RHH DNA-binding proteins make a sequence-specific contact via residues in the β-strand by inserting into DNA major grove (Schreiter and Drennan, 2007). From the structure of the CopG-DNA complexes, it was observed that side chains of three amino acids recognise bases and make sequence-specific contacts. These residues include Arg-4, Thr-6 and Thr-8 (Gomis-ruth *et al.*, 1998). As pB171 ParB share 40% sequence identity with CopG and Arg-4 and Thr-8 are conserved, I believe that Arg-45, Thr-46 and Thr-47, might make similar bases contacts. The structural alignment and superposition of C-terminal RHH fold of ParB with transcriptional regulator CopG is shown in Figure 4.27. A similar pattern of Arg and Thr in the β-strand could be seen among many of transcriptional regulators proteins, as shown in Figure 4.26. Another notable feature among transcriptional repressor proteins is that the flexible loop that connects the α-helix 1 and α-helix 2 has the G-X-S/T/N motifs consisting of 3-5 amino acid residues. However, ParB from pB171, ParB from pXC1001 and MetJ are exceptions, as they possess a longer loop with a different amino acids sequence, but with polar uncharged residues at the beginning and the end. On the other hand, no sequence identity could be seen in α-helix 1, apart from the presence of hydrophobic amino acids.

However, the aligned sequences, shown in Figure 4.26, display a distinct conserved pattern of amino acid residues that are mainly involve in DNA interaction, structure stabilisation and function. A large degree of sequence conservation is present in α-helix 2. Indeed, these conserved residues are mainly hydrophobic amino acids that form a compact central core structure and are involved mainly in dimerisation interface. All the proteins belonging to the RHH family dimerise similarly. The structural alignment and superposition of C-terminal RHH fold of ParB with transcriptional regulator CopG is shown in Figure 4.27. The two proteins, pB171 ParB and CopG share identical hydrophobic amino acids in β-strand. The highest degree of conservation between ParB of pB171 and CopG is displayed in α-helix 2 (Figure 4.27). For ParG, it was shown that two amino acids in position 68 and 71 in α-helix 2 and amino acid in position 49 in α-helix 1, are involved in the dimerisation interface and the core structure of the protein, as

substituting these residues with alanine affected dimerisation as well as the plasmid stability (Saeed *et al.*, 2015). The same pattern can be observed in the Omega protein that harbours similar amino acids at the same position (Dmowski and Jagura-Burdzy, 2011). Interestingly, similar hydrophobic residues could also be observed in ParB α-helix 2 that may share the same function as that observed in ParG and Omega.

Figure 4.28 displays the hydrophobic amino acid residues that are shown to be conserved amongst all the Arc/MetJ transcriptional repressor members in the sequence alignment. Significantly, those nine conservative substitutional hydrophobic residues and one polar residue, are those that pointed toward the centre of the core structure. Those conserved hydrophobic amino acids may make the RHH domain maintain its rigid structure and share the same dimerisation pattern.

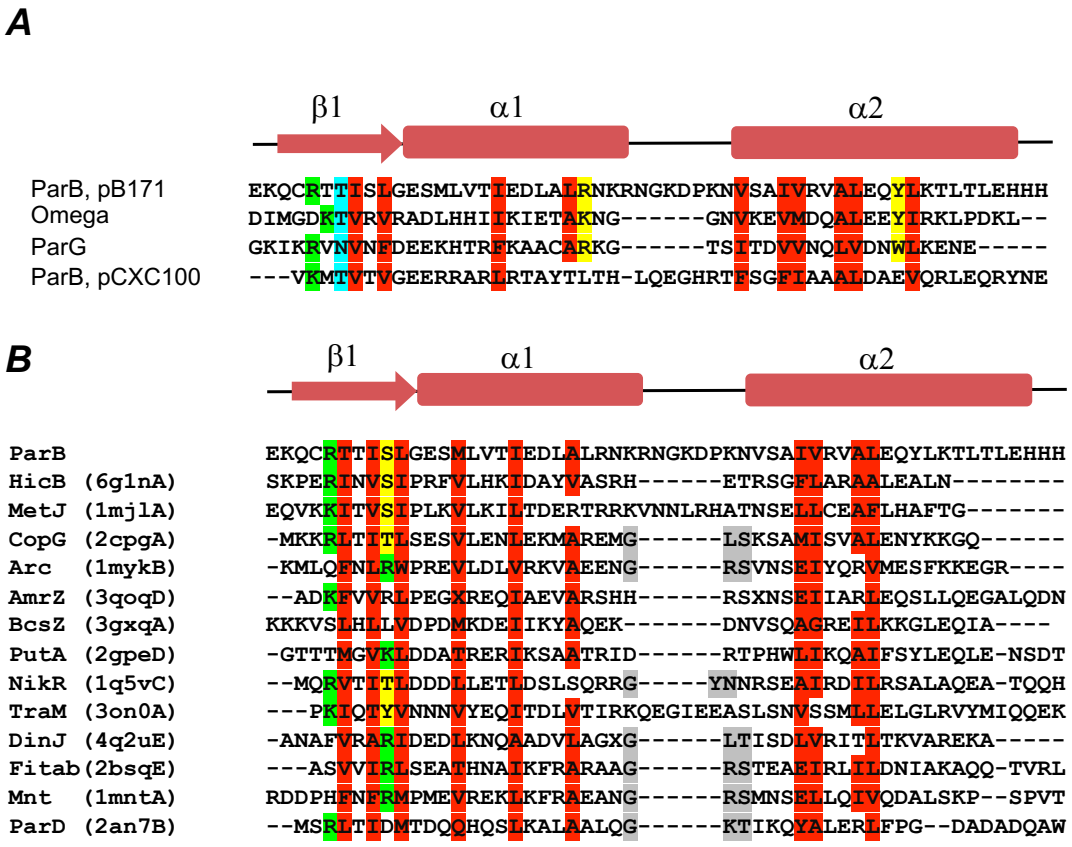


Figure 4.26 Sequence alignment of RHH of the Arc/MetJ superfamily members

A) Sequence alignment of the four RHH motif of known ParB structural homologues from type Ib partition systems (pSM19035 Omega, TP228 ParG and pCXC100 ParB). ***B)*** Sequence alignment of ParB with transcriptional repressor family members harbouring a RHH fold. ***A)*** and ***B)*** conservative substitutions hydrophobic residues are shown in red. The conserved R or K in β-strands is shown in green. The conservative substitutions polar residues are shown in blue. Partially conserved residues are shown in yellow. The G-X-T/N/S conserved sequence in loop is shown in grey.

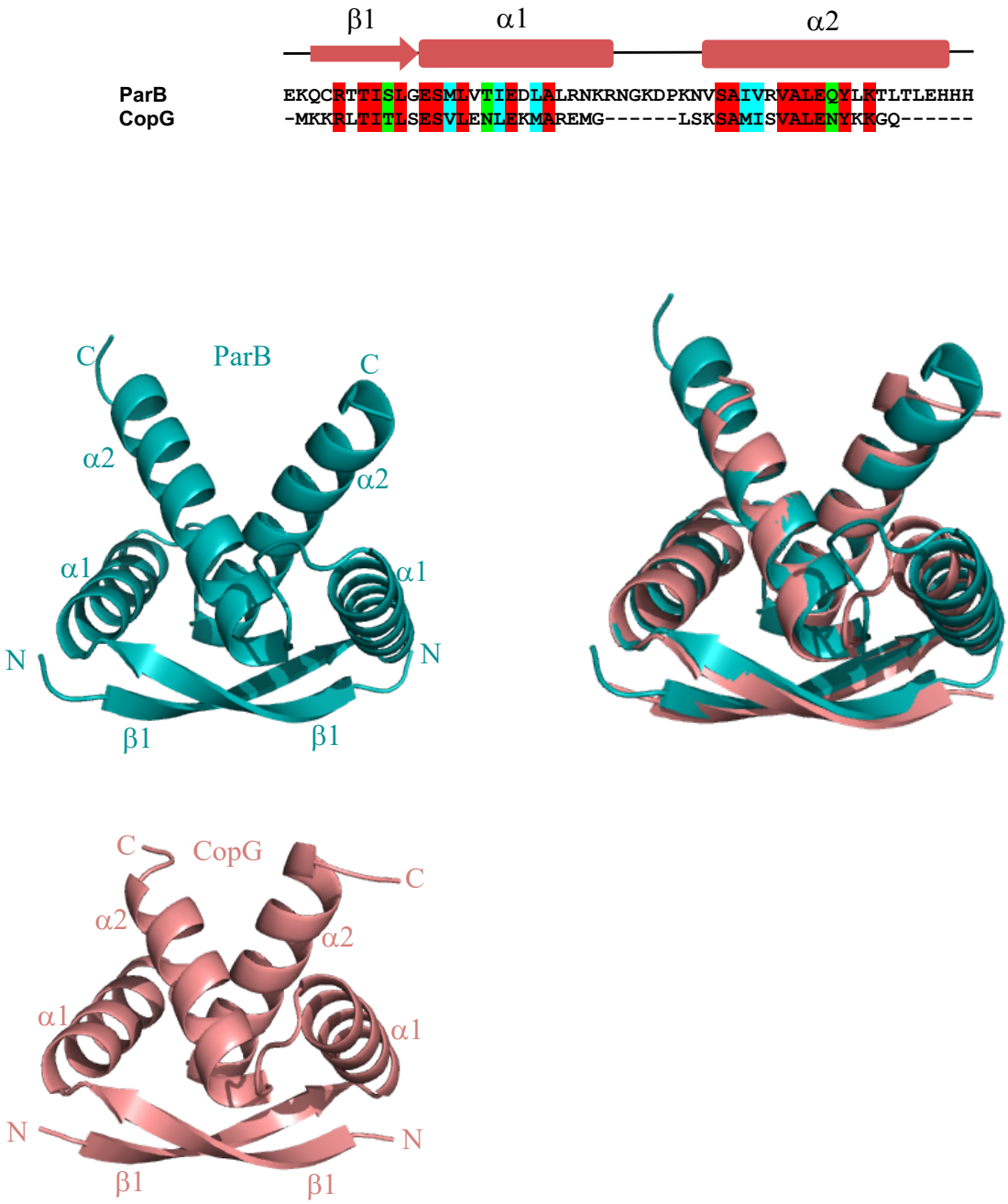


Figure 4.27 Sequence and structure alignment of ParB and CopG

The structure of the C-terminal domain of pB171 ParB is shown in teal and the structure of full length of CopG is shown in deep salmon. Superposition of ParB and CopG are shown in teal and salmon colour, respectively. The sequence alignment between ParB and CopG are displayed above. Conserved residues are in red. The conservative substitutions of polar and hydrophobic residues are shown in green and blue respectively.

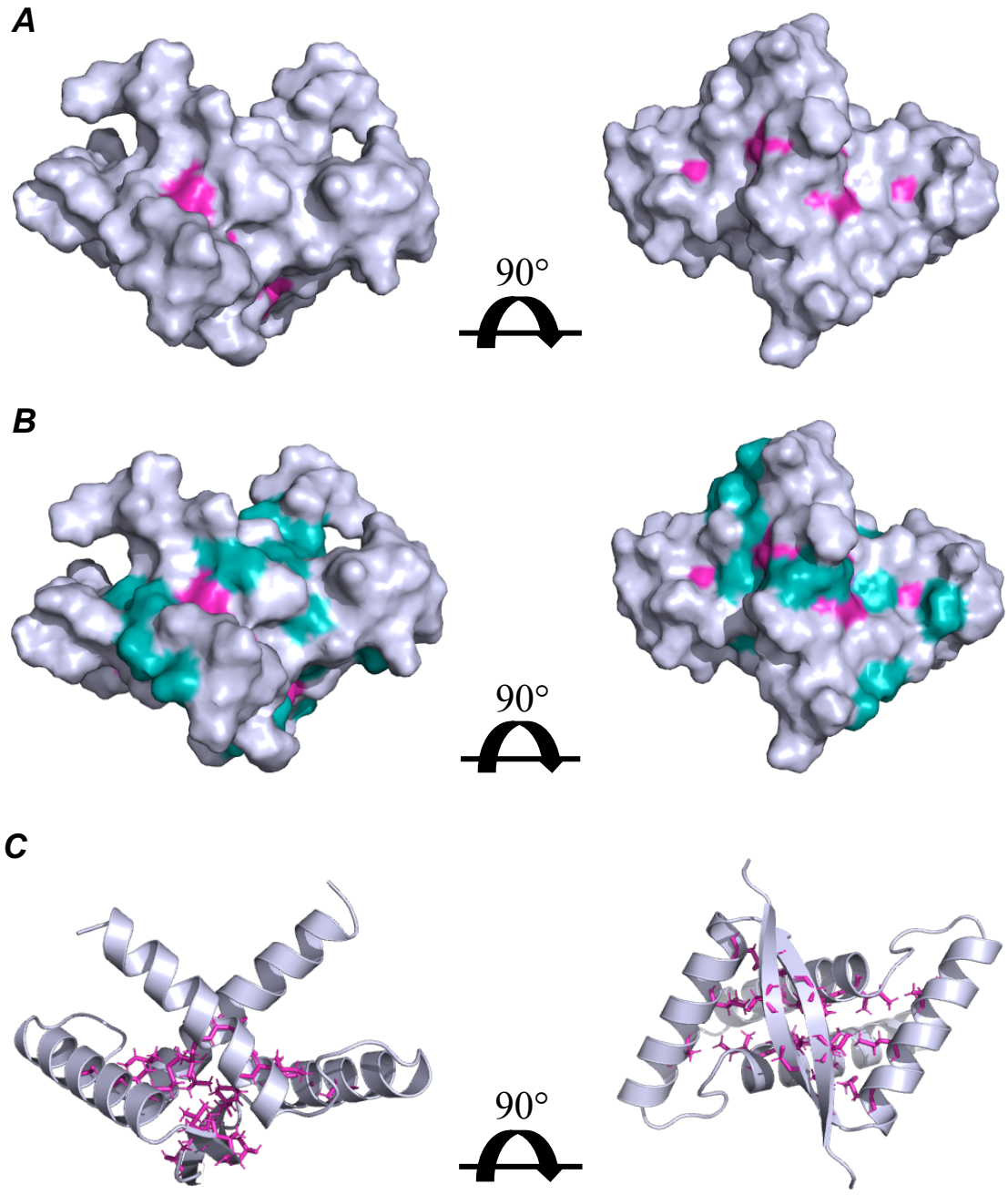


Figure 4.28 Hydrophobic amino acid residues in the ParB C-terminal domain

A) and ***B)*** A comparison of surface structure of ParB RHH fold that shows the conserved hydrophobic residues highlighted in magenta (***A***), and non-conserved hydrophobic residues highlighted in teal (***B***). ***C)*** Hydrophobic conservative substitutions shown as sticks (magenta), pointing toward the centre of core structure.

Chapter 5

Molecular interaction of the components of the ParAB-*parC* plasmid segregation system

5.1 Introduction

Plasmid partition in bacteria involves the interaction between the two segregation proteins, as well as the interaction occurring between the CBP and the partition site. These interactions are crucial for accurate assembly of the nucleoprotein and multiprotein complexes. This association is required for movement and equipartitioning of duplicated plasmids to opposite nucleoid poles before cell division (Baxter and Funnell, 2014; Funnell, 2016; Bouet and Funnell, 2019). This chapter aims to shed new light on the interactions of the segregation components using biochemical and biophysical techniques. One of the main components of pB171 plasmid segregation system is the centromere-binding protein ParB. ParB recognises short palindromic repeats located upstream (*parC1*) and downstream (*parC2*) of the *parAB* genes. Once ParB binds to its recognition sites, it forms a nucleoprotein structure that is the partition complex. The details of the loading of ParB on the centromere sites are still not fully understood. However, it was shown that ParB binds in a site-specific manner to the *parC1* and *parC2* sites via the C-terminal ends (Ebersbach and Gerdes, 2001; Fothergill *et al.*, 2005; Ringgaard *et al.*, 2007b). Therefore, the experiments reported in this chapter aimed to investigate further the features of the binding of ParB to DNA as well as to the partner ParA. Furthermore, the results described in the previous chapter (Chapter 4) were used to further characterise the site-specific interaction of ParB and cognate DNA. Type Ib ParB proteins fold into a RHH C-terminal domain, while the N-terminal domain lacks structure and the NMR results in the previous chapter have shown that this domain is flexible. The pB171 ParB C-terminal structure was solved and was shown to have the RHH fold, as detailed in Chapter 4. Moreover, ParB dimerises via the C-terminal end as it was observed that ParB dimerises even in the absence of the N-terminal domain. Based on the biochemical studies, the C-terminal domain has a dual role: DNA binding and dimerisation. On the other hand, little is known about the role of the flexible N-terminal region, but it is believed that this domain is involved in centromere pairing (Ringgaard *et al.*, 2007b). Previous studies have shown that the full length of ParB, ParBΔ19 and ParBΔ39 bind with the same affinity to the *parC1* and *parC2* partitions sites. However, full-length ParB binds to the partition sites differently as compared to truncated ParBs (Ringgaard *et al.*, 2007a; Ringgaard *et al*., 2007b). These results suggest that the N-terminal domain is somehow involved in the interaction with DNA either directly or indirectly. This observation raises some questions; for example, is the flexible N-terminal region involved in the interaction with DNA? if it interacts, does it undergo conformational changes? Moreover, what is the contribution of the N-terminus to ParB-DNA interaction? Besides its involvement in plasmid segregation, ParB regulates expression of the *parAB genes*. ParB binds site-specifically to the promoter

region to repress *parAB* transcription. ParB also represses the transcription of the *parMR* operon that is divergently transcribed compared to *parABC* (Ringgaard *et al.*, 2007b).

The second component of the partition system is the centromere site. In *parABC* of pB171, the centromere sites, *parC1* and *parC2* comprise seventeen and eighteen direct repeats, respectively. Each repeat consists of 6 bp. The repeats in *parC1* are arranged into two clusters, while in *parC2* the repeats are arranged into three clusters. The repeats are classified into two classes. The arrangement and distribution of repeats in both partition sites are different, but the repeats are related. *parC1* contains two additional identical 10 bp repeats that are located upstream of the 6 bp cluster (Ebersbach and Gerdes, 2001). The binding of ParB to the hexameric repeats is considered the initial step in the process of plasmid partition.

ParA is a Walker type ATPase protein that was previously shown to form a unique helical structure when ParB-*parC* complex is present (Ringgaard *et al.*, 2009). Although the assembly of ParA into a single, continuous helical filament has been questioned, pB171 ParA is able to form higher order structures and this behaviour is observed only when ATP is present (Ebersbach *et al.,* 2006; Machón *et al.*, 2007). Moreover, a mutation in the N-terminal region (R-26 to alanine) leads to the misdistribution of ParA as well as the plasmid stability defects (Ringgaard *et al.*, 2009). Therefore, based on these observations, we hypothesise that ParB N-terminal domain may have several roles.

In this chapter, detailed investigations were carried out to determine the binding of ParB to the centromere sites as well as identifying the region responsible for the interaction. MST and EMSA techniques were used to assess the interaction between ParB and the centromere sites. In addition, affinity level and cooperativity were established from the MST experiments. The studies showed that ParB binds with similar affinity and cooperativity to *parC1* and *parC2* sites. The data obtained from EMSA suggest that ParB binds the two different classes of repeats without any preferences. In addition, the study showed that every single repeat is bounded by a ParB dimer. In Chapter 4, the crystal structure of ParB C-terminal domain was reported, and NMR spectroscopy was used to gain insight into ParB N-terminal domain. The data obtained from NMR experiments were used as a prerequisite step to establish the recognition site of a single repeat by ParB. Based on the NMR data, both the C-terminal and N-terminal domains of ParB are involved in DNA binding. The NMR data categorised the N-terminal domain into three clusters based on the level of ParB-DNA interactions. Moreover, the study also focused on the relationship and the role played by certain components in segrosome assembly

with special emphasis on two of them, centromere site and ATP. Here, it is also reported how two amino acids could be involved in two separate roles. This chapter provides a picture of ParB recognition of the cognate DNA site and how the presence of ATP and *parC1* enhances the interaction of ParB with ParA.

5.2 Results

5.2.1 Characterisation of the *In vitro* interaction of ParB with the partition sites, *parC1* and *parC2*

5.2.1.1 Generation of *parC1* and *parC2* DNA fragments for EMSA and MST experiments

For electrophoretic mobility shift assays, the *parC1* and *parC2* DNA fragments were generated using two sets of primers. The *parC1* site was amplified using PCR. Two primers were designed to generate a 260 bp *parC1* DNA fragment. The forward primer 5′ end contained a biotin for DNA detection using the LightShift Chemiluminescent EMSA kit. The sequences of forward and backward primers are displayed in Table 2.5. Similarly, the *parC2* site was amplified using two primers with the forward primer 5′ ends modified with biotin. Additionally, the primers were used to generate the *parC2* site are listed in Table 2.5. Both fragments were amplified using PCR as detailed in Section 2.3.5. The amplified fragments were verified using a 1.2% agarose gel. The *parC1* and *parC2* PCR products were 260 and 240 bp, respectively, as shown in Figure 5.1 A and B. To obtain pure DNA pieces, gel extraction was carried out on 1.5% agarose gel. The bands were cut and purified as described in Section 2.3.12.

For microscale thermophoresis (MST) experiment, a similar strategy was used to amplify the *parC1* and *parC2* sites. Here, the DNA fragment were labelled with the fluorescent dye Cyanine (Cy5), by incorporating this dye at the 5′ end of the primers used for amplification. The *parC1* and *parC2* fragments were amplified and the PCR products were 260 and 240 bp, respectively. The size of PCR products was verified by running a 1.2% agarose gel (Figure 5.1*C* and *D*).

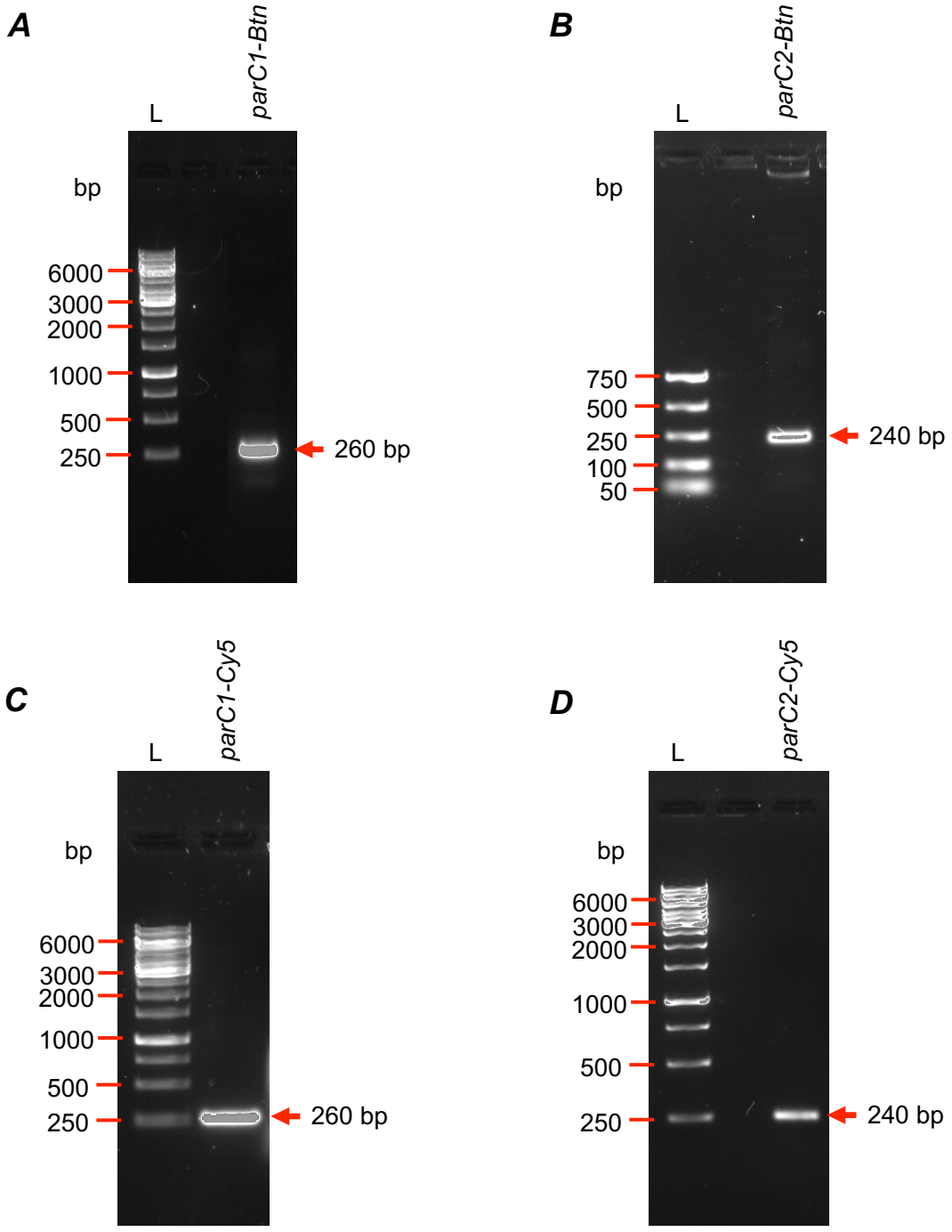


Figure 5.1 Agarose gels showing the amplified PCR products of *parC1* and *parC2* labelled with either biotin or Cy5

1.2% agarose gels showing the PCR products of *parC1* (260 bp) ***A*** and ***C***, and *parC2* (240 bp) ***B*** and ***D***, generated by amplifying the region upstream and downstream of the *parAB* genes, respectively. In all agarose gels, the first well contains the 1 kb ladder (Thermo Fisher Scientific) and the size of fragments is indicated. The second lane shows the *parC1* and *parC2* PCR products and the size of each fragment is indicated.

5.2.1.2 Analysing the binding affinity of ParB for the partition sites *parC1* and *parC2*

To examine thoroughly the interaction of ParB with its centromere sites *parC1* and *parC2* and to understand whether the ParB protein binds cooperatively, *in vitro* EMSA and MST were carried out. The gel shift assay is based on the principle that protein-DNA complexes on the native gels are not disrupted and migrate slower than the free DNA due to the size differences. The *parC1* and *parC2* DNA fragments were generated by PCR. The *parC1* DNA fragment comprised of 17 of the 6 bp repeats in addition to two identical 10 bp repeats, while the *parC2* DNA fragment consisted of 18 repeats, each spanning 6 bp. The repeats in the two DNA fragments were flanked by 25-28 bp arms to allow the ParB protein to bind efficiently to the centromeric repeats. The binding of these two DNA fragments with different concentrations of the ParB protein was analysed in EMSA and MST titration experiments. The binding of ParB with the non-specific pieces of DNA were examined in the past and were showed that ParB could not recognise random sequence of DNA (Fothergill et al., 2005).

In the EMSA experiment, a fixed concentration of biotinylated *parC1* (0.4 nM) was incubated with increasing concentration of ParB. At a very low ParB concentration (50 nM monomer), the DNA start to shift, and two low molecular weight complexes were observed (Figure 5.2*A*). Once the concentration of ParB was increased (100 nM), the intensity of the shifted bands increased, and a third distinct complex was formed. With increasing the concentration of ParB to 200 nM, the intensity of the three shifted bands increased significantly, particularly the highest molecular weight complex. When the concentration of ParB was increased to beyond 300 nM, the unbound DNA disappeared completely (Figure 5.2*A*). It should be noted that the *parC1* site comprises seventeen of 6 bp repeats that are arranged into two clusters. In addition to repeats organisation, the repeats are classified into two classes. The abundance of repeats explains the forming of multiple complexes at a very low concentration of ParB. The *parC1* fragment was shifted completely to form a high molecular weight complex with a ParB concentration of 500 nM. The same high molecular weight complex was also observed at 700 nM (Figure 5.2*A*). The free DNA contained an additional band that is visible just above the amplified *parC1* fragment (Figure 5.2*A*). This extra band might be a DNA fragment with a secondary structure (for example a hairpin) that is formed due to the presence of the biotin and results in a more retarded migration on the gel. Multiple attempts were made to remove the extra band by purifying the band of interest either from the agarose or acrylamide gel. However, it was not possible to obtain a pure band. The extra band also disappeared completely with high ParB concentration. The result was analysed using a

Gel-Doc (Bio-Rad) and the intensity of the unbound DNA was quantified using Image Lab 4.0.1 software. The binding curve was obtained by plotting the fractions of bound DNA against ParB concentration using Graph-Pad Prism software. The best fit k_D was also calculated from the binding curve using one-site binding equation. The estimated k_D value was given to be 217.8±136.6 nM (Figure 5.2*B*).

EMSA is a semi-quantitative analysis. Therefore, MST experiments were performed to characterise further the interaction of ParB with the *parC1* partition site. MST is an effective tool that can be used to calculate the dissociation constant, stoichiometry and the binding cooperativity. MST allows a quantitative analysis of molecular interactions in solution based on the direct movement of molecules in temperature gradients. For MST experiments, the *parC1* fragment was designed in a similar way as that used for EMSA experiments. MST requires a fluorescence molecule to detect the signal, therefore, the *parC1* DNA fragment was labelled with Cy5. A fixed concentration of Cy5-*parC1* (12.5 nM) fragment was added to ParB concentrations ranging from 3 nM to up to 100 µM. However, the same concentration of DNA that was used for EMSA experiment could not be used in MST experiments as it produced low fluorescence intensity that was not sufficient to detect the ParB-DNA interactions. The data obtained revealed that ParB bound to *parC1* with two folds higher affinity than was obtained from electrophoretic mobility shift assays. The dissociation constant was shown to be 119.3±27.48 nM (Figure 5.3). Moreover, the Hill coefficient of the binding of ParB to *parC1* was calculated and a value of 2.08 was obtained. This result indicate that ParB binds cooperatively to the *parC1* site.

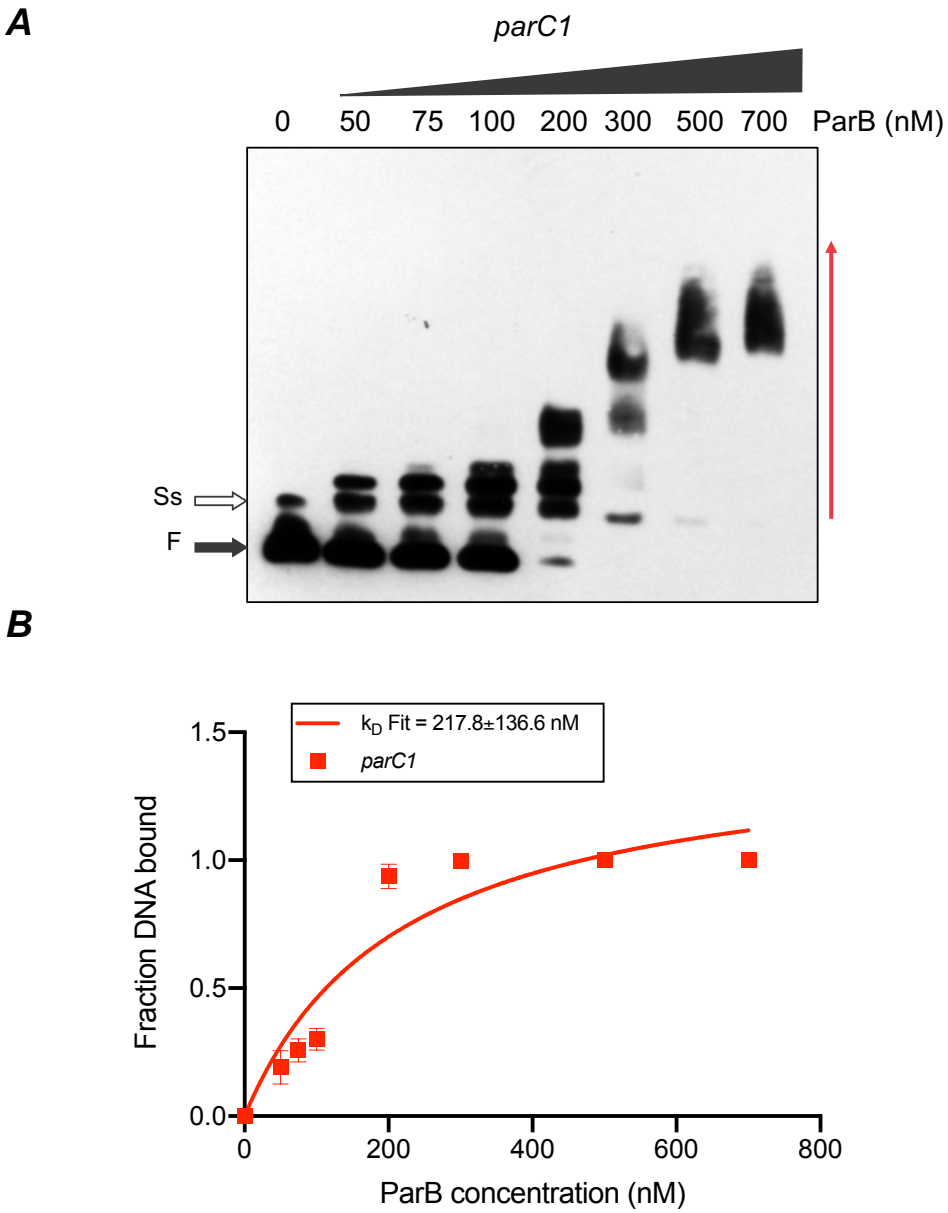


Figure 5.2 ParB binds to *parC1* forming multiple complexes

A) Representative EMSA experiment in which ParB (50 nM- 700 nM, monomer) was incubated with a fixed concentration of biotinylated labelled *parC1* (0.4 nM). F represents the unbound, free DNA (*parC1*) and is indicated by a solid black arrow. Ss represents a secondary structure band formed by biotinylated DNA and is indicated by an empty black arrow. The shifted bands are indicated by vertical red arrow. Pixel intensities of the unbound DNA were quantified using Image Lab 4.0.1 software (Bio-Rad). ***B)*** DNA binding curve was generated using Graph-Pad Prism. The fraction of bound DNA from three experiments was averaged and plotted against ParB concentration. The k_D was calculated from Graph-Pad Prism using the one site-specific binding model. The error bar represents the standard error of the mean. The k_D for *parC1* was estimated to be 217.8±136.6 nM.

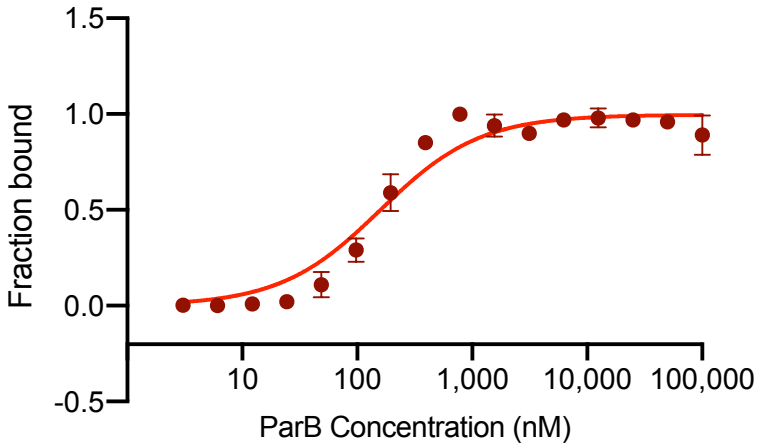


Figure 5.3 MST binding curve of ParB and *parC1* interaction

Sixteen ParB concentrations ranging from 3 nM to up to 100 μM monomer, was incubated with a fixed concentration of Cy5 labelled *parC1* (12.5 nM). DNA binding curve was generated using Graph-Pad Prism. The fraction of bound DNA from three experiments was averaged and plotted against ParB concentration. The k_D was calculated from Graph-Pad Prism using the one site-specific binding model. The error bar represents the standard error of the mean. The k_D for *parC1* was estimated to be 119.3±27.48 nM.

The interaction of ParB with its centromere site *parC2* was investigated in a similar way. In the EMSA experiments, a fixed concentration of biotinylated *parC2* was added to increasing concentrations of ParB. As observed for ParB/*parC1* interaction, the addition of the ParB protein at a very low concentration (50 nM monomer) to a 1.6 nM biotinylated labelled *parC2* fragment resulted in the formation of two low molecular weight complexes (Figure 5.4*A*). When ParB concentration was increased to 100 nM, the intensity of the shifted bands increased, and two other distinct complexes with higher molecular weights were observed (Figure 5.4*A*). Increasing ParB concentration to 200 nM, the intensity of the three shifted bands increased significantly, especially the highest molecular weight complex. All the DNA was shifted, when ParB concentration increased to above 300 nM (Figure 5.4*A*). It should be noted that the *parC2* site consists of eighteen 6 bp repeats that are arranged into three groups. Moreover, the repeats fall into two classes according to their sequence. As for *parC1*, the presence of multiple repeats explains the formation of multiple complexes at a very low concentration of ParB. Like *parC1*, the *parC2* fragment was shifted completely to form a high molecular weight complex when 300 nM of ParB was added. The same high molecular weight complex was also observed at 500 and 700 nM (Figure 5.4*A*). Similarly to what observed for *parC1*, the *parC2* free DNA showed an additional band (Figure 5.4*A*). As it was suggested previously, this extra band could be another biotinylated fragment harbouring a secondary structure that results in a different migration during electrophoresis. Multiple attempts were made to remove this band by gel extracting the DNA band of interest. However, it was challenging to obtain a single band with no secondary structure. As previously observed with *parC1*, this extra band disappeared completely at high ParB concentrations. The result was analysed using a Gel-Doc (Bio-Rad) and the intensity of the unbound (free DNA) was quantified using Image Lab 4.0.1 software. The binding curve was obtained by plotting the fraction of bound DNA against ParB concentration using Graph-Pad Prism software. The best fit k_D was also estimated from the binding curve using one-site specific binding equation from the Graph-Pad Prism software. The estimated k_D value was 165.6±112.2 nM (Figure 5.2*B*).

To quantify the interaction of ParB with *parC2* in a similar way as *parC1*, MST experiments were performed. A fixed concentration of 14.5 nM of Cy5 labelled *parC2* was incubated with increasing concentration of ParB ranging from 3 nM to 25 µM. The MST data showed that ParB binds to this centromere with lower affinity as shown by gel shift assay. The dissociation constant was calculated from the sigmoidal curve and a value of 613.38± 284.81 nM was obtained (Figure 5.5). The Hill coefficient was also

calculated from the Hill curve and a value of 1.8 was obtained. This indicates that the ParB protein binds cooperatively to *parC2*.

By comparing binding affinity of ParB to *parC1* and *parC2*, EMSA experiments have shown that ParB bound *parC2* with higher affinity than *parC1*. Furthermore, at 300 nM of ParB concentration, ParB caused a complete shift to *parC2* (Figure 5.4), while ParB caused a complete shift to *parC1* at 500 nM concentration (Figure 5.2). However, the pattern that ParB binds to *parC1* and *parC2* fragments are very similar, where both DNA fragments were shifted at low concentration of ParB and several complexes were formed when the ParB concentrations were increased above to 50 nM before it shifted completely. However, the results obtained from EMSA experiments are not conclusive due to large margin errors in the EMSA binding curve of ParB/*parC2*. In general, these results are in consistence with those obtained *in vivo* experiments, where deletion of the *parC2* site had abolished plasmid stability. On the other hand, MST experiments contradict with what have been observed in EMSA. This could be due to the labelling of *parC2* fragment with Cy5, which had shown high degree of aggregation in thermophoresis traces. This may had affected the binding of ParB to *parC2* fragment.

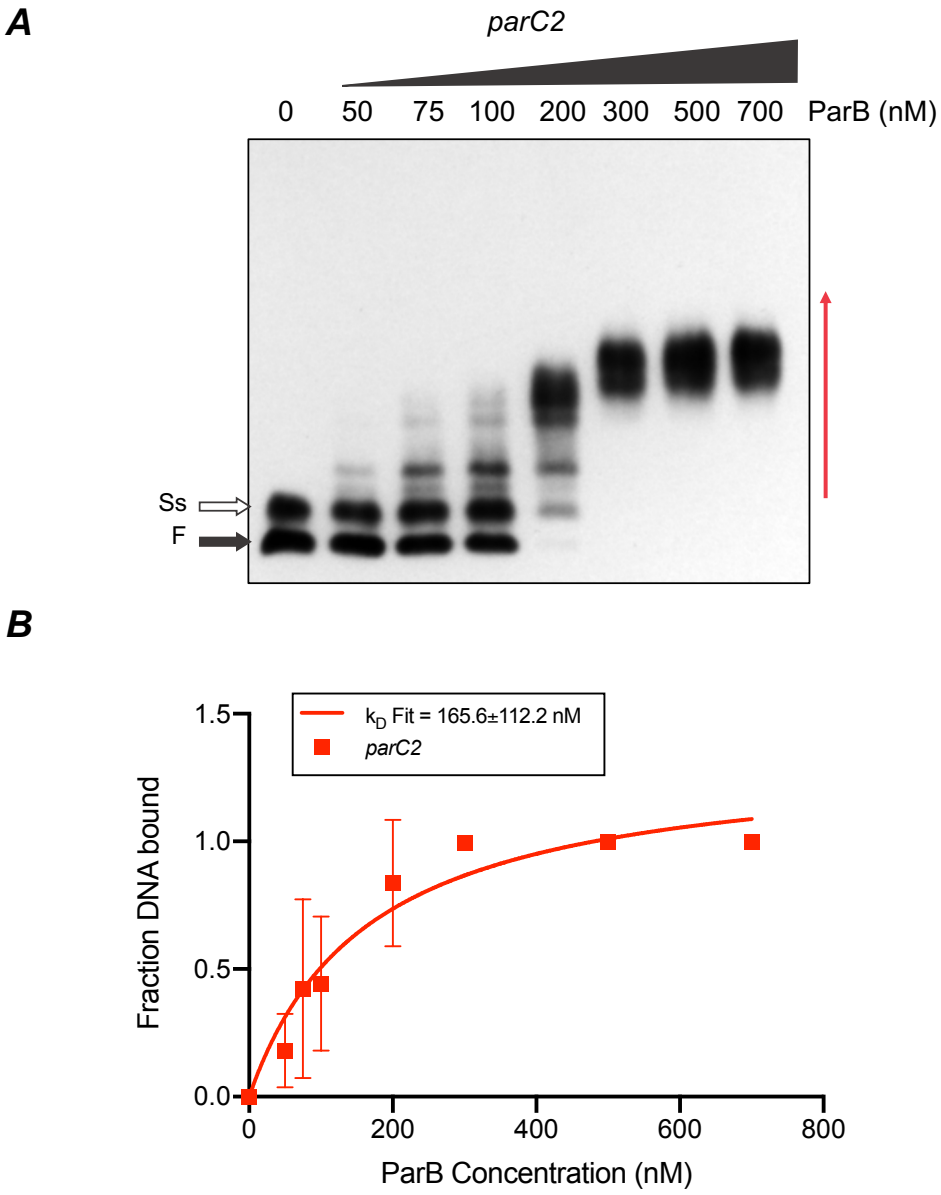


Figure 5.4 ParB binds to *parC2* forming multiple complexes

A) Representative EMSA experiment in which increasing concentrations of ParB (50 nM- 700 nM) were incubated with a fixed concentration of biotinylated labelled *parC2* (1.6 nM). F represents the unbound, free DNA (*parC2*) and is indicated by a solid black arrow. Ss represents a secondary structure band formed by the biotinylated DNA fragment and is indicated by an empty black arrow. The shifted bands are indicated by vertical red arrow. Pixel intensities of the unbound DNA were quantified using Image Lab 4.0.1 software (Bio-Rad). ***B)*** ParB-*parC2* binding curve obtained using Graph-Pad Prism software. The fraction of bound DNA was averaged from three different experiments and plotted against ParB concentration. The k_D was calculated from Graph-Pad

Prism using one site-specific binding. The error bar represents the standard error of the mean. The k_D for *parC2* was estimated to be 165.6±112.2 nM.

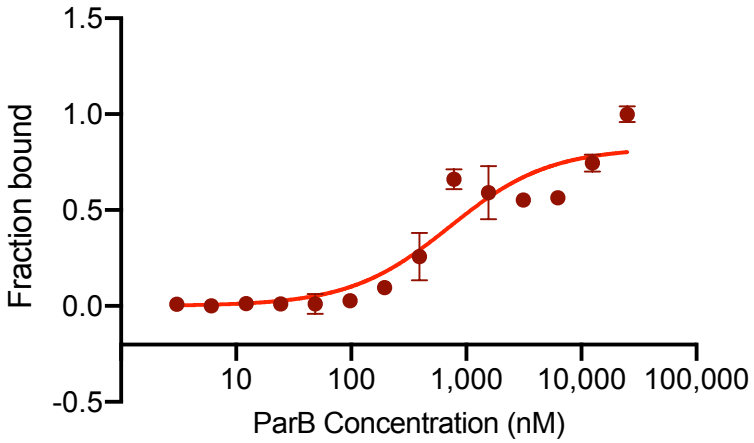


Figure 5.5 MST binding curve of ParB and *parC2* interaction

Fourteen ParB concentrations ranging from 3 nM to up to 25 μM monomer, was incubated with a fixed concentration of Cy5 labelled *parC2* (14.5 nM). DNA binding curve was generated using Graph-Pad Prism. The fraction of bound DNA from three experiments was averaged and plotted against ParB concentration. The k_D was calculated from Graph-Pad Prism using the one site-specific binding model. The error bar represents the standard error of the mean. The k_D for *parC2* was estimated to be 613.38± 284.81 nM.

5.2.2 Interaction of ParB with a single repeat in a solution using NMR spectroscopy

5.2.2.1 DNA preparation

ParB interactions with its full recognition sites, *parC1* and *parC2* was investigated using EMSA and MST techniques. However, we also wanted to characterise the interaction of ParB with a single repeat. To probe the binding of ParB to one repeat and to identify which residues of the ParB protein are involved in the DNA interaction, a [^{1}H,^{15}N] SOFAST-HMQC experiment was performed. This experiment requires an isotopic labelled ParB and a double-stranded DNA fragment. In Chapter 4, the overproduction and purification of ^{15}N labelled ParB were described. To generate a double-stranded DNA, two oligonucleotides complementary to each other were designed. The forward oligonucleotide was 18 bp long with a single 6 bp direct repeat in the middle of the sequence. This repeat was flanked by 6 bp GC-rich sequence on both sides of the repeat (Figure 5.6*A*). The reverse oligonucleotide was designed to be complementary to the forward oligonucleotide. The extra GC nucleotides were included at both ends to anneal the complementary oligonucleotides effectively. Equal molar ratios of both oligonucleotides were mixed and annealed at 95°C for 6 minutes and then the mixture was left to cool down at room temperature for an hour, as described in Section 2.3.14 (Figure 5.6*B*). To investigate whether the oligonucleotides were annealed efficiently, a 5 µl aliquot of annealed oligos was run on 12% polyacrylamide gel along with single-strand oligonucleotide used as a control and a PCR marker (NEB) (Section 2.3.13.1). Once the electrophoresis was completed, the gel was stained with SyBr Safe dye and visualized with Gel Doc (Bio-Rad). Figure 5.6*C* shows that the single-strand DNA run faster compared to the double-strand DNA and confirms that the two oligonucleotides were annealed

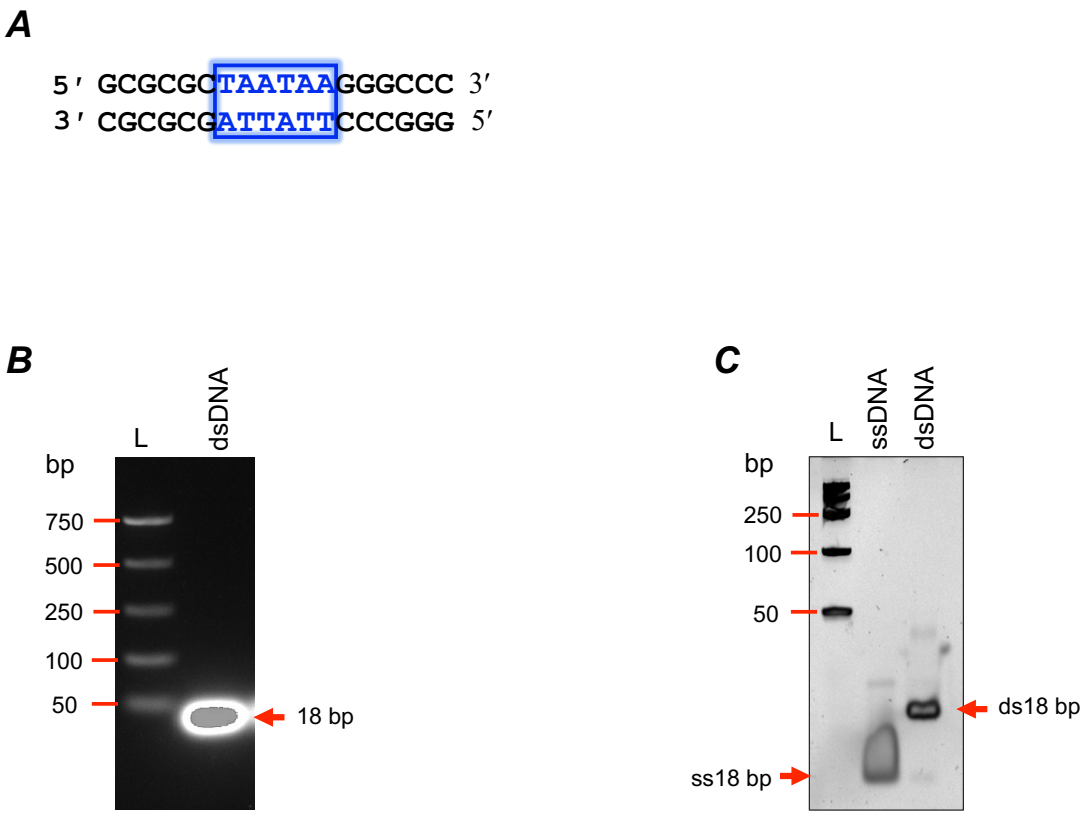


Figure 5.6 Construction of the double-stranded DNA fragment to be used in the NMR experiment

A) DNA sequence of the oligonucleotide used for the NMR study. The 6 bp repeat sequence is shown in the middle of the DNA sequence and highlighted in blue inside a blue square. ***B)*** 2% agarose gel showing 18 bp annealed oligonucleotide product. L indicate the PCR marker (NEB). Lane 2 shows the double-stranded DNA product after annealing the two complementary oligonucleotides. ***C)*** 12% acrylamide gel comparing the single-stranded (ss) and double-stranded (ds) oligonucleotides. The single-stranded DNA migrates faster and appears as a smear compared to a sharp distinct double-stranded oligonucleotides band. L indicates the PCR marker (NEB). The bands corresponding to a single-stranded and double-stranded DNA are indicated by red arrows.

5.2.2.2 Chemical shift mapping of ParB with a single DNA repeat

NMR experiments were performed to investigate the ParB-DNA interaction and the effect of DNA on the ParB conformation. Attempts to obtain a co-crystal of a ParB/DNA complex were unsuccessful due to low crystal resolution. Hence, the study of the ParB/DNA interaction in solution was carried out using NMR spectroscopy. In the Chapter 4, the backbone assignment of the full ParB protein was obtained. This information was used to probe the ParB-DNA interaction. Therefore, [^{1}H,^{15}N] SOFAST-HMQC experiments were conducted to examine which residues in the ParB protein are involved in DNA interaction. (^{15}N) labelled ParB was overproduced and purified as described in Sections 2.4.4 and 2.4.5. The [^{1}H, ^{15}N] SOFAST-HMQC experiments were performed to record the NMR chemical shift mapping of the (^{15}N) labelled ParB bound to the 18 bp double-stranded DNA that harbours a single 6 bp repeat. A fixed concentration of ^{15}N labelled ParB was added to increasing concentration of the 18 bp DNA.

First, the SOFAST-HMQC spectrum of ^{15}N labelled ParB was recorded and used as a reference. The B18 double-stranded oligonucleotides, which corresponds to a single 6 bp repeat of *parC2* site, was titrated against fixed concentration of ^{15}N labelled ParB at 0.2:1, 0.4:1, 0.5:1, 0.6:1, 0.8:1, 1:1 and 2:1 molar ratio. 2D [^{1}H,^{15}N] SOFAST-HMQC spectra were recorded at each step of the titration. The chemical shift changes between spectra were used to monitor which residues were perturbed by the binding of ParB to its recognition site. Interestingly, the binding of ParB to DNA caused major changes to the NMR spectrum. Specifically, the HMQC spectra revealed that signals for the folded domain, which corresponds to the C-terminal RHH domain, disappeared completely (Figure 5.7 and 5.8). This might be due to slow tumbling of the high molecular weight complex that causes the peaks to be widely dispersed, which makes it difficult to be observed. This indicates that this domain binds strongly to the site-specific DNA. However, signals from several residues in the C-terminal folded domain remain visible on binding DNA. Three peaks in the folded domain remain visible correspond to residues Ser-49, Arg-67 and Glu-84 (Figure 5.8). The minimal chemical shift changes seen for these residues could indicate that those residues are far from the DNA binding pocket.

Importantly, signals corresponding to residues in the N-terminal flexible domain showed significant variation in their chemical shift upon the addition of DNA. The residues that experienced significant chemical shift variations are those located in the N-terminal region of the unstructured N-terminal domain. Specifically, signals for Lys-4, Ser-6, Gln-7, Gln-8 and Ala-9 shifted progressively as the DNA concentration was increased (Figure

5.7, 5.8 and 5.9). However. A smaller change in the chemical shift was observed for Ala-9, while Leu-10 was difficult to follow due to signal overlap (Figure 5.8). Lys-4, Ser-6, Gln-7 and Gln-8 shifted at a very low concentration of DNA, but when the ratio reached 0.8:1, these peaks split into two separate cross-peaks (Figure 5.8). Notably, when the ratio was 1:2, peaks that correspond to Lys-4, Ser-6, Gln-7 and Gln-8 disappeared completely from the spectrum. These data suggest that this region of ParB may participate in the interaction with the DNA. Importantly, the peak assigned to Lys-3 disappeared from all the spectra meaning that this residue may interact directly with DNA (Figure 5.7). In contrast, most of the peaks that are located between residues Asp-11 to Lys-30 experiences slow exchanges upon binding to DNA (Figure 5.7, 5.8 and 5.9). These data may indicate that this region of the N-terminal domain of ParB undergoes conformational changes upon binding to DNA. The region that connects the N-terminal domain to the C-terminal RHH domain was not affected as the chemical shift of residues in this region did not change or slightly shifted between bound and unbound. These peaks correspond to residues Tyr-32, Gly-33, Ala-34, Glu-39 and Glu-41 (Figure 5.7 and 5.8).

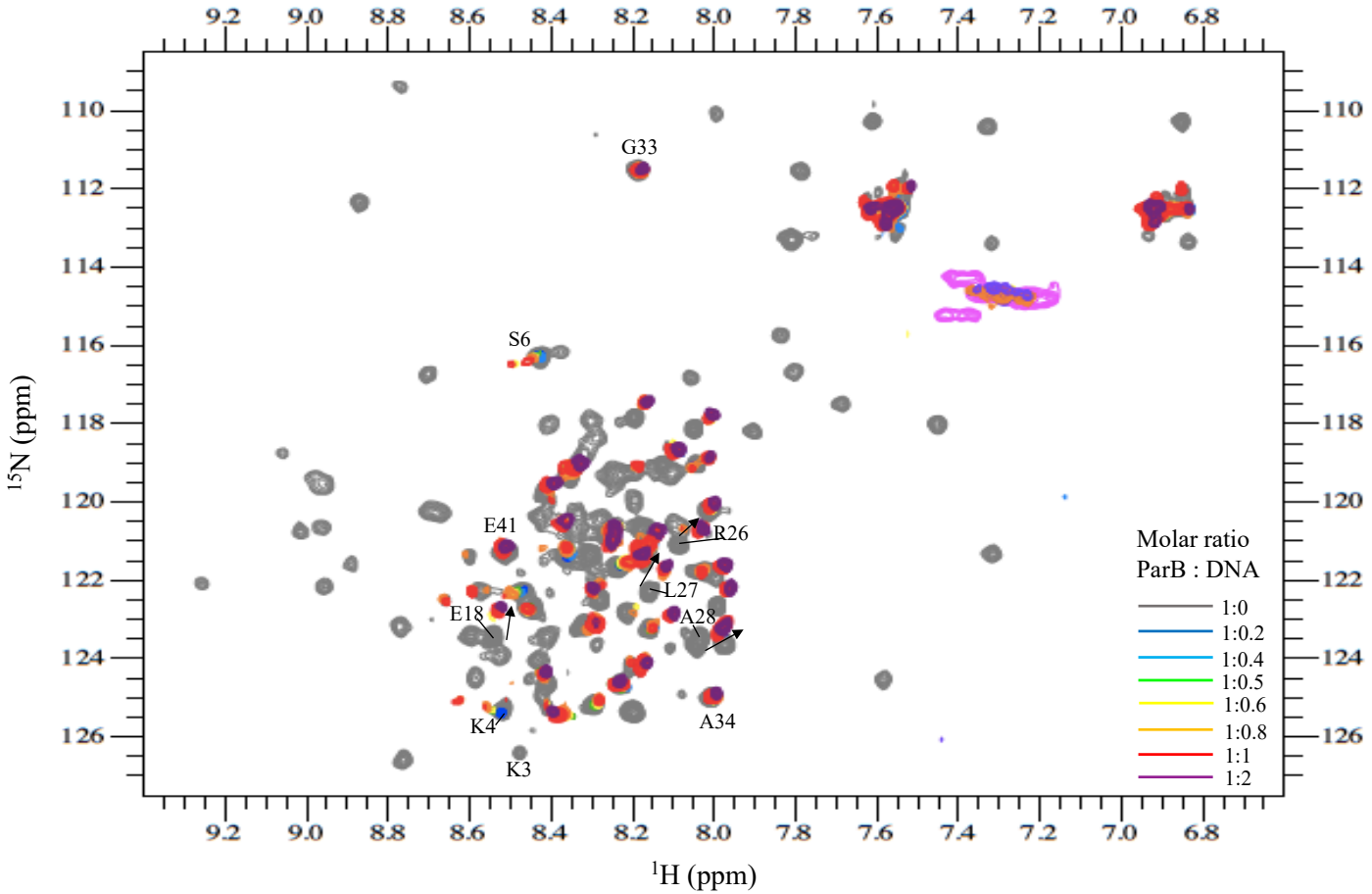


Figure 5.7 NMR chemical shift mapping of ParB-DNA interaction

2D [^{1}H, ^{15}N] SOFAST-HMQC NMR spectra of the titration of (^{15}N) labelled ParB with DNA. The grey spectrum was recorded in the absence of 18 bp dsDNA. ^{15}N labelled ParB was incubated with increasing concentration of the 18 bp dsDNA. ParB:DNA molar ratio were 1:0.2, 1:0.4, 1:0.5, 1:0.6, 1:0.8, 1:1 and 1:2. The spectra were overlaid, and the colours are indicated. All spectra were recorded at 700 MHz, 25ºC.

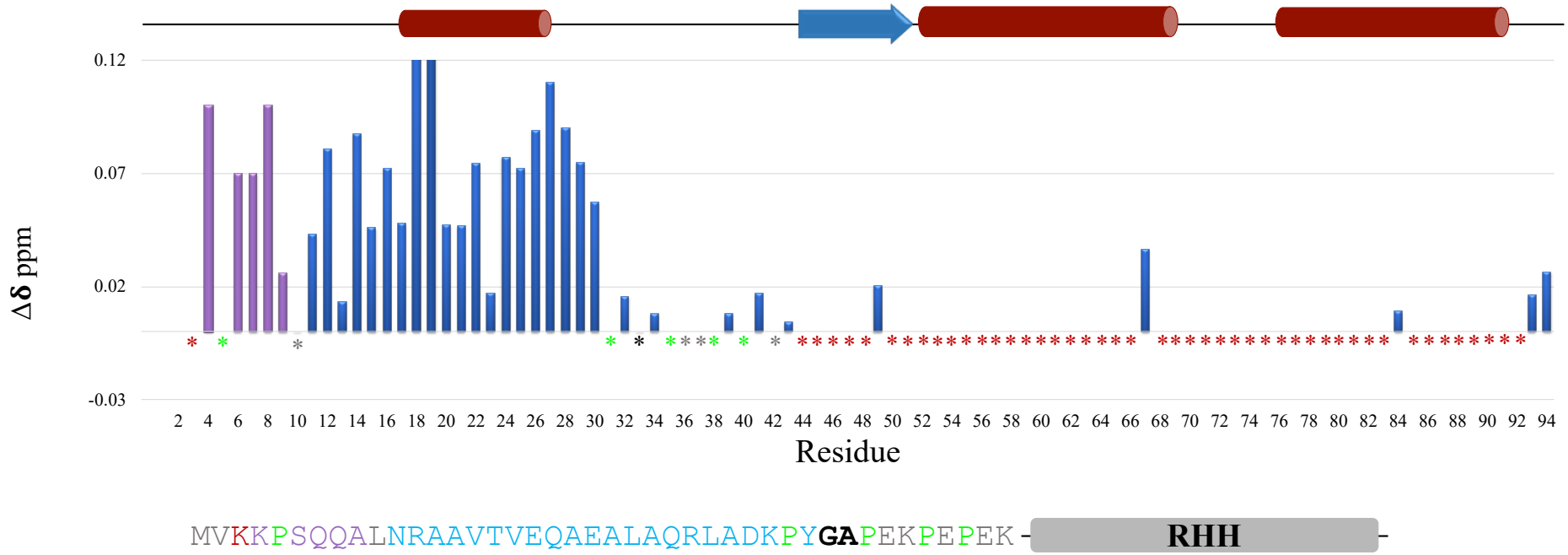


Figure 5.8 DNA-induced chemical shifted changes in ParB

The diagram shows the chemical shift perturbations at saturating DNA concentration plotted against residue numbers. The purple bars represent residues that underwent gradual increases in chemical shift and these cross-peaks disappeared from the spectra at 1:2 ParB:DNA ratio. Blue bars represent residues that show slow exchange and the most affected are positioned in the middle of the N-terminal domain. The coloured asterisks under the bar chart and the colours of bars are colour-coded to the sequence of the N-terminal domain of ParB, which is showed under the bar chart. The red asterisks represent residues that disappeared from the spectra when the DNA was added. Grey asterisks represent residues that were difficult to monitor due to multiple shifted peaks in the area. The black asterisk represents Gly-33 whose chemical shift did not change when the DNA was added. Green asterisks represent proline residues that lack N-H cross-peak. A cartoon diagram representing the ParB structure is displayed at the top of the chart.

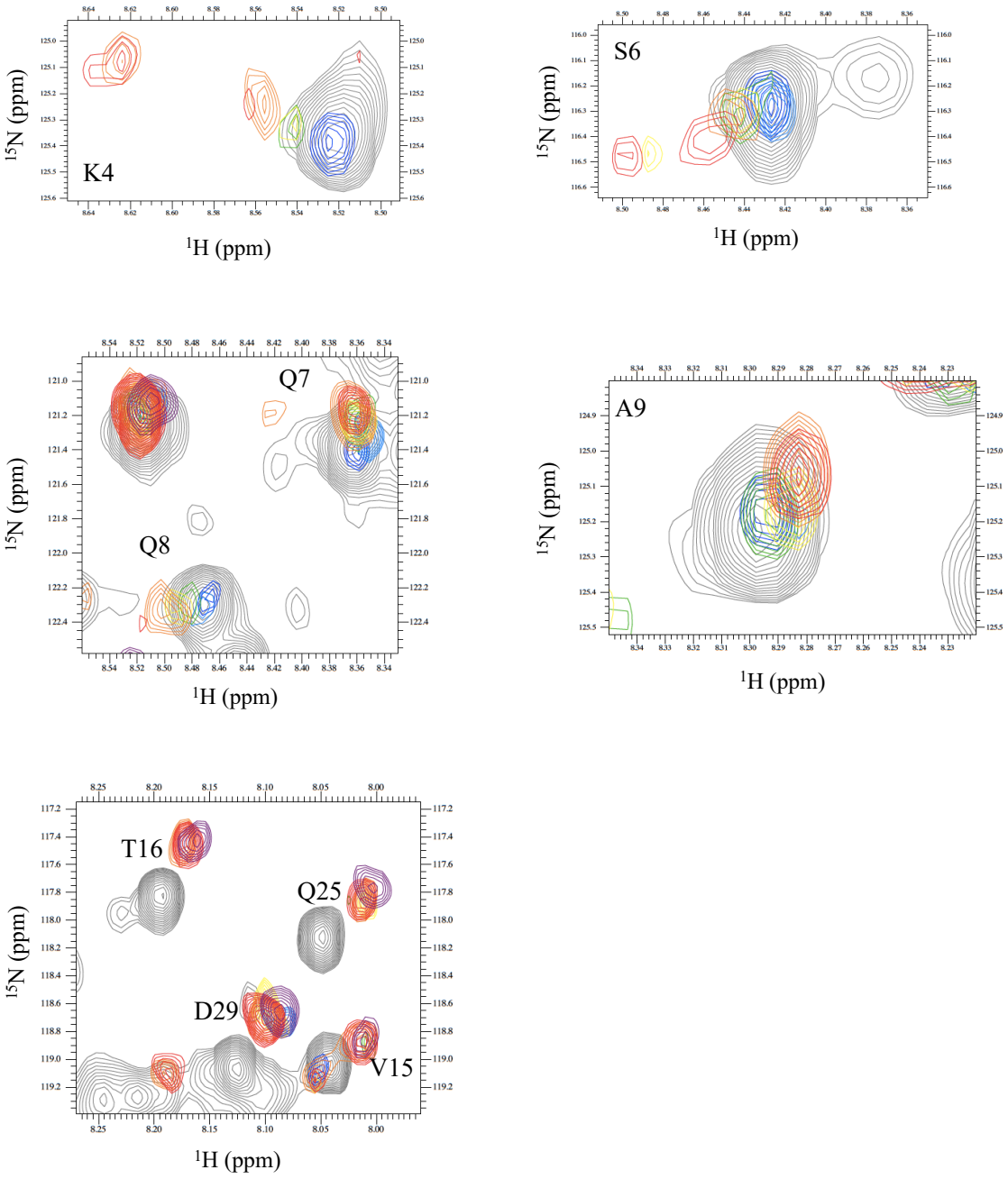


Figure 5.9 N-terminal domain ParB residues exhibiting changes mapped by NMR chemical shift perturbation

Zoomed in spectra of some of the residues located in the extreme tip and in the middle region of ParB N-terminal domain. Residues K4, S6, Q7 and 8, and A9 disappeared from the spectra, when saturation was reached at 1:2 ParB:DNA ratio. V15, T16, Q25 and D29 peaks shifted, when the DNA was present. The colour code used is same as that of Figure 5.7.

5.2.3 Alanine scanning mutagenesis of ParB residues in the N-terminal domain

The NMR experiments have shown that some residues located in the unstructured N-terminus of ParB undergo a chemical shift change upon DNA binding. The NMR results have revealed three regions of chemical shift changes. Based on these observations, the N-terminal regions could classify into three clusters: cluster I corresponds to residues 3-9, which show interaction with the DNA. Cluster II correspond to residues 11-30, which experienced slow chemical shift changes. In contrast, cluster III represent residues that are not shifted when the DNA was present. To further investigate the role played by these amino acids in plasmid segregation, site-directed mutagenesis was used to construct a number of mutants. The residues were selected based on the NMR results. These residues included Lys-3 that disappears completely from the spectrum at the lowest DNA concentration and Lys-4 that disappears when the DNA concentration was two-fold higher than that of the ParB protein. These two residues are located at the tip of the N-terminal tail. In addition, Asp-11, Arg-12 that are located at the beginning and Arg-26 located at the end of the second cluster were targeted for mutagenesis. Lys-3, Lys-4, Asp-11, Arg-12 and Arg-26 codons were changed to alanine codons. The pCC02 plasmid that harbours the wild type *parABC* segregation cassette was used as a template to generate *parB-K3A*, *parB-K4A*, *parB-N11A*, *parB-R12A* and *parB-R26A* alleles using site-directed mutagenesis by overlapping extension PCR, as described in Section 5.3.15. The pFH450 vector was used to generate all the mutant constructs by cloning the PCR fragments that harboured the mutated *parB* genes in this plasmid. Two internal overlap primers were designed, which contained the desired mutation, and two external primers were designed that contained the *Eco*RV and *Xho*I sites, respectively. The products of PCR1 and PCR2 for all mutations were around 1 kb and 600 bp, respectively (Figure 5.10*A*). Both PCR fragments were purified and annealed together before PCR3 was carried out. The final PCR3 fragment was approximately 1.6 kb with *Eco*RV and *Xho*I restriction sites at the 5′-end and 3′-end of the fragment (Figure 5.10*B*). A standard cloning protocol was followed where the pFH450 vector and the PCR3 product that harboured the mutated codon were digested with *Eco*RV and *Xho*I restriction enzymes. After digestion, the insert and the vector were purified and ligated as described in Section 2.3.9. Clones were screened via restriction enzyme digestion. Few colonies from the cloning results were picked and a double digestion was carried out using *Eco*RV and *Xho*I to verify that the clones contained the insert (Figure 5.10*C*). Once the inserts were confirmed, the clones were sent for sequencing to ensure the *parB* mutant alleles were present and no mistake had been introduced in the *parABC* cassette.

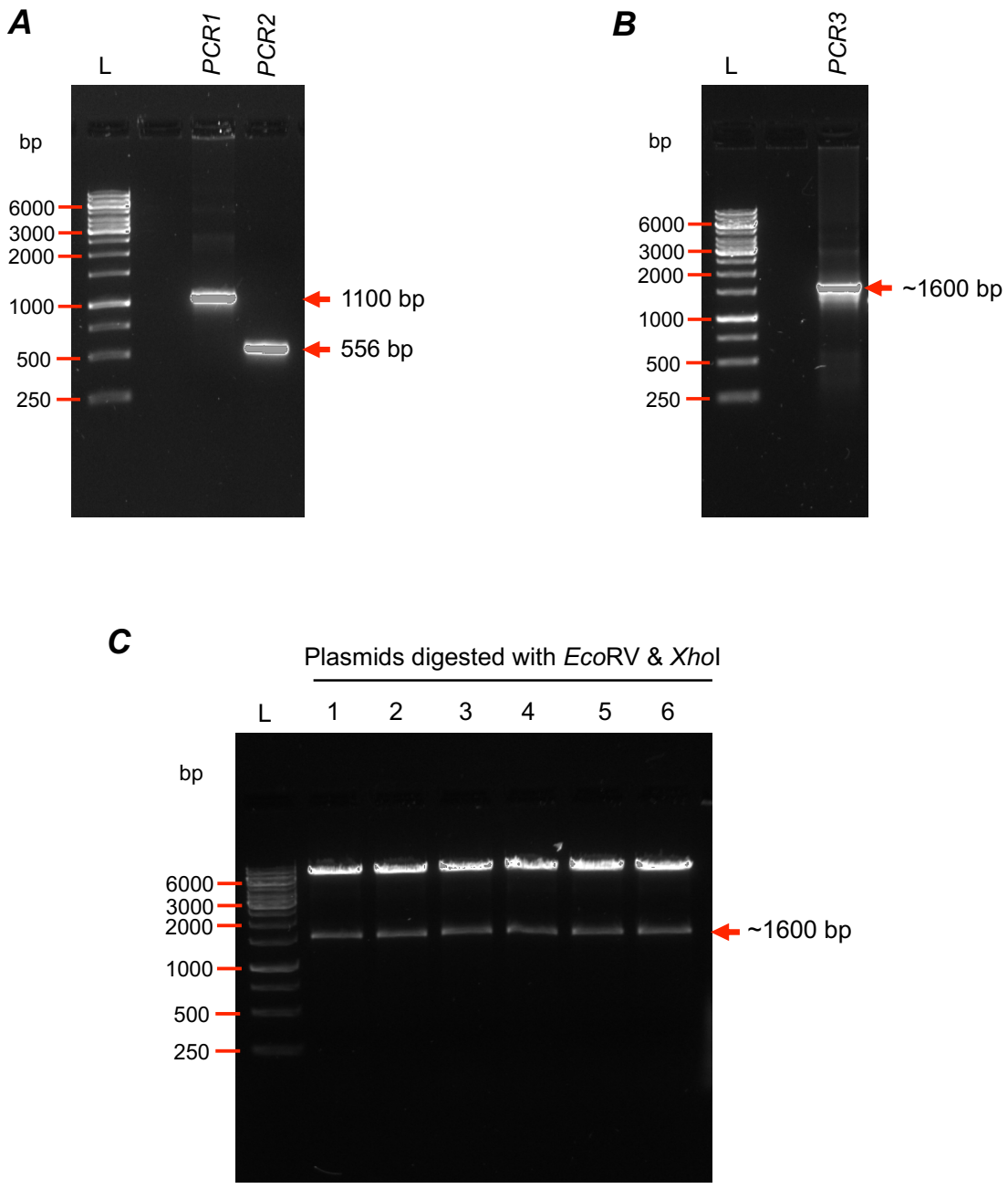


Figure 5.10 Agarose gels showing the site-directed mutagenesis steps for the construction of one of the *parB* mutant alleles

A) 1% agarose gel showing the amplification products of PCR1 (~1100 bp) and PCR2 (556 bp) harbouring the mutants *parB* allele. L indicates the 1 kb molecular weight ladder (Thermo Fisher Scientific). ***B)*** 1% agarose gel showing the PCR3 products after annealing PCR1 and PCR2 (~1600 bp). ***C)*** Representative example of the restriction digestion analysis of positive plasmid clones. The first lane contains the 1 kb ladder and the size of fragments is indicated. Lanes 2-7 show the restriction digestions analysis of some positive clones using *Eco*RV and *Xho*I restriction enzymes. The size of the insert is 1282 bp and is indicated by the red arrow.

5.2.4 *In vivo* analysis of ParB mutants

Plasmid partition assays were carried out to determine the effect of mutations on plasmid stability *in vivo*. The constructs were transformed into BR825 *E. coli* cells and grown overnight at 37°C on LB plate supplemented with chloramphenicol. The same *E. coli* strain was used to transform with the wild type pCC02 plasmid that harboured the full partition cassette and an empty plasmid pFH450 that lacks stabilising system. The last two plasmids were used as positive and negative controls, respectively. The selection was carried out to obtain individual colonies followed by growing the cells for 25 generations under non-selective pressure. This allows only the plasmid that harbour a functional partition cassette to be retained. The retention level of all clones was measured in percentage after the selection (chloramphenicol) was added to the media in the final step of partition assays. Each mutant was assessed individually along with pCC02, that contains the full *parABC* cassette and pFH450 that lacks any stabilising system. The plasmid stability assays were performed in triplicate.

The level of retention of pCC02 plasmid bearing the full partition cassette was on average of 76.3%, whereas that of the empty plasmid pFH450 was on average of 10.6% (Figure 5.11). The stability of the plasmid that carried the *parB-K3A* codon was affected. The plasmid stabilisation decreased noticeably to ~35% retention level. The plasmid carrying the *parB-K4A* mutant allele exhibited a retention level of 55.6%, which is 27% less stable than the wild type (Figure 5.11). On the other hand, the plasmid that harbours the *parB-N11A* allele showed a stability level comparable to that wild-type pCC02 with 80.7%, indicating that N11 residue is not crucial for plasmid segregation. The plasmid that carries *parB-R12A* mutation showed a very slight defect with a stability of 64.3%. The *parB-R26A* allele was affected plasmid stability conferring a plasmid retention of 56.66 (Figure 5.11). These results suggest that the two lysine residues at position 3 and 4 may play important roles in plasmid stabilisation. Moreover, the residues located in cluster I (i.e. at the N-terminus of the ParB N-terminal domain) may contribute in the interaction with the DNA as previously suggested by NMR, and thus, plasmid segregation was defective.

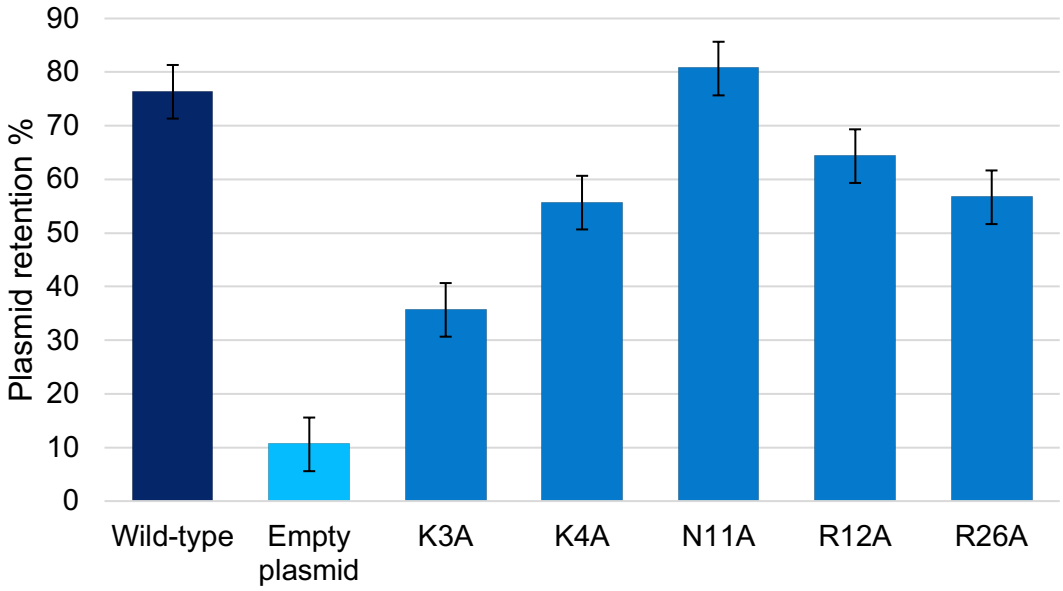


Figure 5.11 Plasmid retention conferred by the ParB mutants expressed in *E. coli* cells

Wild type (dark blue bar) represents the pCC02 plasmid that harbours the full *parABC* cassette and acts as positive control for the experiment. The empty plasmid (cyan bar) represents the pFH450 plasmid that lacks the stabilising systems and acts as negative control of the experiments. Retention of the plasmids harbouring the *parB* mutant alleles (K3A, K4A, N11A, R12A AND R26A) (blue bars) was measured by plasmid partition assays. The percentages represent the average of at least three independent experiments performed along with the two controls. Error bars represent the standard error of the mean.

5.2.5 ParA-ParB interaction in the presence of the centromere site and ATP

5.2.5.1 ParA overproduction, solubility and purification

One of the project aims was to investigate the interplay of the two pB171 segregation proteins ParA and ParB. The initial step to study the interaction of the proteins *in vitro* was overproduction and purification of proteins of interest. The overproduction and purification of ParB were described in detail in Section 4.2.1.1. The pET-ParA construct was available in the laboratory collection (Fothergill *et al.,* 2005). The *parA* gene was cloned in frame with the sequence encoding an N-terminal $(His)_6$-tag and downstream of the T7 promoter and *lac* operator. The plasmid was transformed into the expression host *E. coli* cells (BL21(DE3)) that harbours a copy of T7 RNA polymerase gene in the chromosome. Protein expression was induced with IPTG. The construct was sequenced before starting any experiment to ensure no mutation had been introduced during the cloning procedure.

Small-scale ParA overproduction and solubility were carried out to ensure that the plasmid was effectively expressing *parA* and was capable to produce soluble protein. The pET-ParA construct was transformed into BL21(DE3) strain of *E. coli* cells. Two overexpression conditions were tested, one at 30°C and the other at 37°C. The growing cultures were induced with 1 mM IPTG final concentration. The overexpression was carried out for three hours, and 1 ml of the sample was collected every hour. Induction for three hours at 37°C and 30°C was sufficient to obtain a high yield of protein and the two incubation temperatures showed a similar level of expression (Figure 5.12*A* and *B*). Following the small-scale overexpression, the solubility of the protein was tested as described in Section 2.4.2. Protein solubility was estimated by comparing the amount of protein in the cytoplasmic and the inclusion bodies fractions using SDS-PAGE. The band intensity was analysed using a Gel-Doc (Bio-Rad) and was quantified using Image Lab 4.0.1 software. At 37°C the amount of ParA protein in inclusion bodies increases progressively with time, as shown in Figure 5.13*A*. The two-hour post induction soluble fraction contains around 64% of the total overproduced ParA and at three-hour post induction 58% of ParA was obtained from soluble fraction. This means the protein is moderately soluble, however, the level of solubility improved significantly when the induction was carried out at 30°C. The amount of ParA in the soluble fraction after two hours incubation with IPTG was 15% higher than the one obtained at the same time at 37°C. Additionally, only 25% of ParA was lost in the insoluble fraction (Figure 5.13*B*). However, the figure was changed when the incubation time was continued one more

hour and a higher percentage of ParA was observed in the pellets. This indicates that ParA protein is more soluble at 30°C and the solubility was higher at two hours post induction.

Once the condition for protein overproduction was established, the ParA protein was produced in large quantity in 200 ml cultures. Once the overproduction of ParA was verified on a SDS-polyacrylamide gel, the $(His)_6$-tag protein was purified using Ni^{2+} affinity chromatography as described in Section 2.4.5. The total amount of $(His)_6$-tag ParA protein produced after purification was 5-6 mg total from a 200 ml culture. An attempt was made to purify a larger amount of ParA protein. The attempt was successful, nevertheless, after a short period of time, the protein precipitated. It was established that ParA tends to precipitate at a concentration higher than 0.8 mg/ml. The purity of the $(His)_6$-tag fusion ParA protein was verified using SDS-PAGE, as shown in Figure 5.14. Based on the SDS-polyacrylamide gel, the chromatography purification yields a good purity and reasonable quantity of the $(His)_6$-tag fusion ParA protein. The amount of purified ParA was sufficient for downstream experiments. The ParA has a molecular weight of 23.5 kDa including the hexa-histidine tag (Figure 5.14).

The gel shows an extra band in the ParA purification fractions (Figure 5.14*A*). The band size is slightly higher than the ParA protein. To assess whether this band is contamination from *E. coli* protein or another conformation of the ParA protein, a Western blot was performed as described in Section 2.4.11. After running the ParA sample onto the SDS-polyacrylamide gel, the content of gel was transferred onto a nitrocellulose membrane for protein detection. The detection was carried out using anti-his antibodies and revealed that the extra band with higher molecular weight could be another conformation or post-translational modified ParA (Figure 5.14*B*).

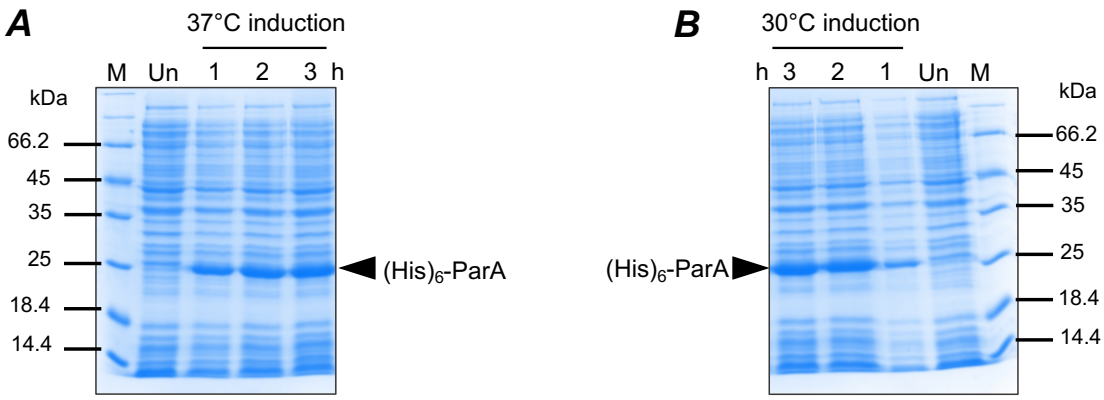


Figure 5.12 Small-scale overproduction test at 37°C and 30°C

SDS-PAGE showing the overproduction of ParA in LB medium. Total cell samples induced at 37°C (**A**) and 30°C (**B**) for one, two and three hours are compared. The overproduced 23.5 kDa ParA protein is indicated by the arrow. M represents the unstained PageRuler molecular weight protein marker (Thermo Fisher Scientific).

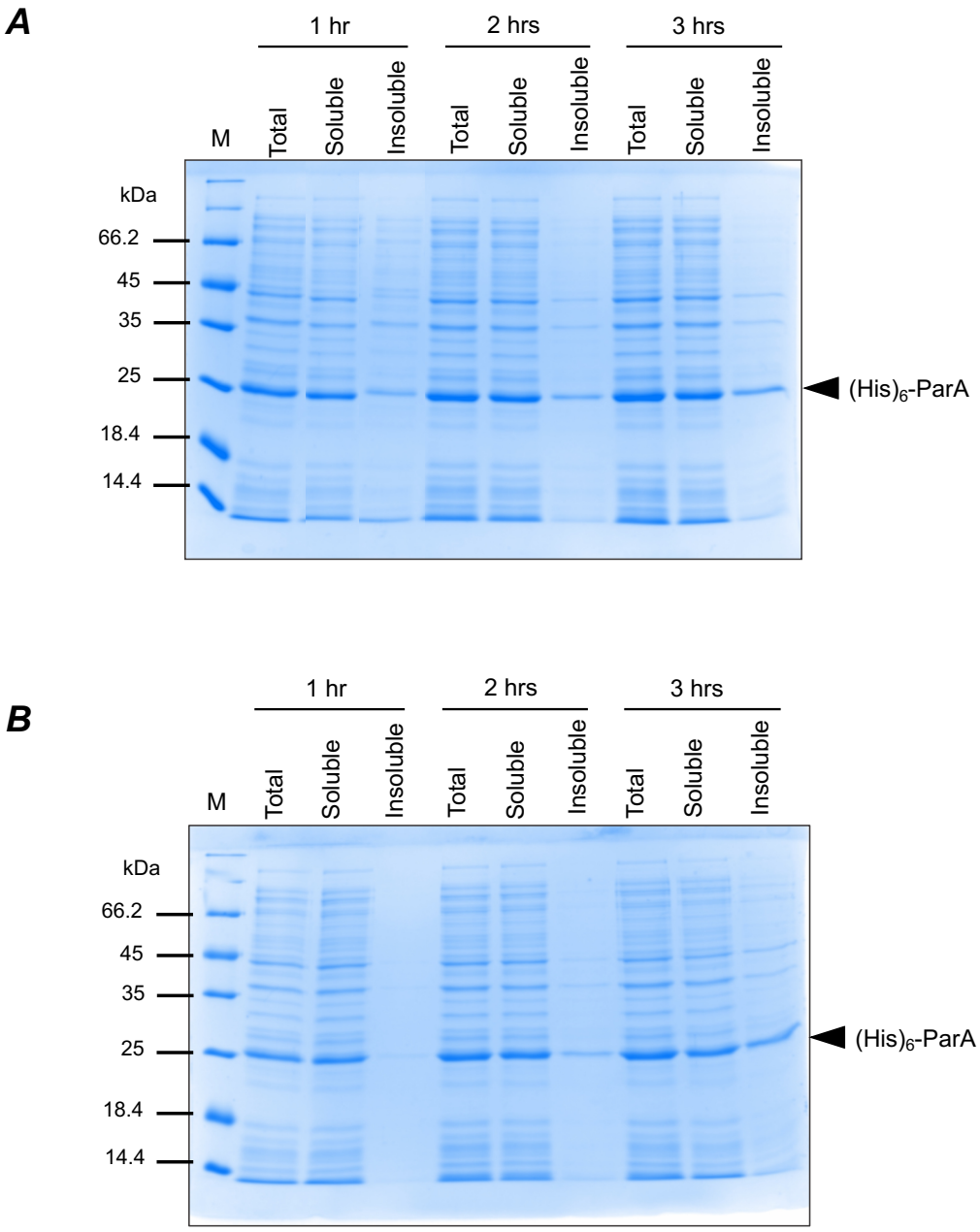


Figure 5.13 Solubility of ParA at two different induction temperatures

A) and ***B)*** SDS-PAGE showing the ParA overproduction and the effect of incubation times on the protein solubility at 37°C (***A***) and 30°C (***B***). The soluble and insoluble fractions are compared with unfractionated cell lysate (total) at 1, 2, and 3 hours post induction. The overproduced protein is indicated by the arrow. M represents the unstained PageRuler molecular weight protein marker (Thermo Fisher Scientific). The bands were quantified using a Gel-Doc (Bio-Rad) and Image Lab 4.0.1 software.

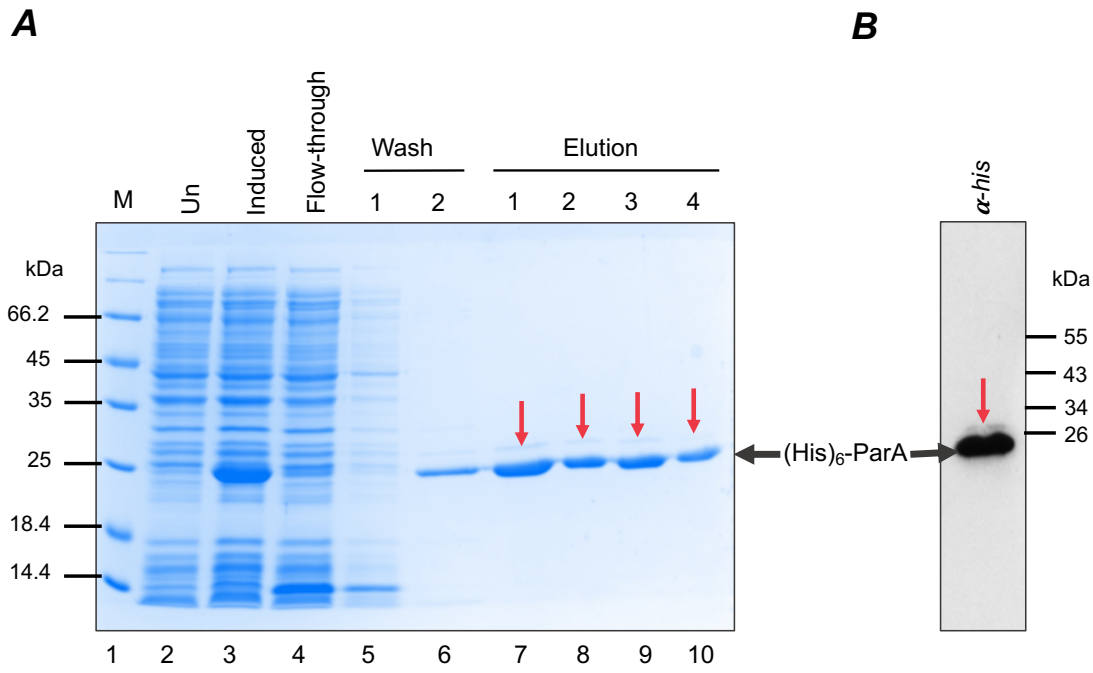


Figure 5.14 Purified ParA using Ni^{+2} affinity chromatography and Western blot

A) SDS-polyacrylamide gel showing the purified ParA protein. M in lane 1 represents the unstained PageRuler molecular weight protein marker (Thermo Fisher Scientific). Un in lane 2 contains an aliquot of the growing culture before induction with 1 mM IPTG (uninduced). Lane 3 shows a sample of the culture after two-hour-inductions with IPTG. Lane 4 shows the column flow-through after loading the clarified cell lysate. Lanes 5 and 6 contain aliquots of the two wash steps with two buffers with different imidazole concentrations (15 and 85 mM). Lanes 7-10 show the fractions of purified ParA eluted with 300 mM imidazole. The purified 23.5 kDa (including the hexa-histidine tag) ParA protein is indicated by the arrow. ***B)*** Western blot detecting the ParA protein using α-his antibodies. The ParA protein is indicated by a black arrow. Red arrows in ***(A)*** and ***(B)*** show the ParA extra band.

5.2.5.2 The *parC1* partition site and ATP enhance the binding of ParB to ParA

In plasmid segregation system, the interaction between the two partition proteins is essential to drive the newly replicated plasmids apart. In the *par2* system of pB171 plasmid, it is still unclear whether the two partition proteins, ParA and ParB are able to interact without additional factors or require specific elements to boost the interaction. Here, chemical cross-linking experiments were carried out to investigate the interaction of ParA and ParB in the presence and absence of centromere site (*parC1*) and adenosine triphosphate (ATP). Chemical cross-linking of ParA and ParB was performed using DMP (dimethyl pimelimidate) as cross-linker. The experiment was performed in triplicate to minimise the effect of experimental error.

DMP cross-linking reactions were carried out as described in detail under Section 2.8.1. The concentration of DNA used in this experiment was based on the hypothesis that a dimer of ParB binds to a single repeat. *parC1* consists of seventeen repeats, therefore a concentration equivalent to 1/17 of the ParB concentration was used, which is equal to 0.6 μM. ParA and ParB were used at 1:1 ratio (10 μM each). ATP was added at a final concentration of 1 mM final. Uncross-linked ParA and ParB were used as controls for the experiment to compare and identify the cross-linked proteins and complexes (Figure 5.15*A*). ParA and ParB were incubated together in the absence of other ligands, in the presence of the partition site *parC1* and in the presence of both *parC1* and ATP. In addition to these three reactions, ParA and ParB were cross-linked individually and used as controls to identify the monomers, dimers and multimers of the proteins (Figure 5.15*A*). The reactions were incubated at room temperature for an hour, followed by quenching the reaction with 50 mM Tris-HCl, pH 6.8, for 30 minutes. All the reactions were loaded on 12% SDS-polyacrylamide gel and run at 4°C for approximately two hours (Figure 5.15*B*). The gel was analysed, and the molecular weight of each complex was identified using Image Lab 4.0.1 software.

The results show a ParB monomer cross-linked with a ParA monomer, forming a 40 kDa complex. The molecular weight of the complex was calculated from the standard curve using the Image Lab 4.0.1 software and was slightly higher than the theoretical one (34.12 kDa). However, also the ParA and ParB monomers show a molecular weight higher than the theoretical one: ParA migrates as a 25.4 kDa protein (theoretical 23.5 kDa) and ParB migrates as a 14.4 kDa protein (theoretical 12.12 kDa) (Figure 5.15*B*). The intensity of ParA-ParB complex band increased in the presence of the cognate DNA (*parC1*), growing even further when ATP was present in the reaction. Quantification of bands intensity showed that, when *parC1* was present, the amount of ParA-ParB

complex is 22.3% higher than in the reaction containing ParA and ParB alone. Moreover, the quantity of ParA-ParB complex in the presence of both, *parC1* and ATP increased further by 28.1% as compared to the percentage observed for the complex produced by the proteins alone (Figure 5.15*C*). This observation suggests that *parC1* and ATP enhance the binding of ParB to ParA. Another complex was also observed, which could be a ParA dimer cross-linked to a ParB monomer and has a molecular weight of 64.8 kDa (Figure 5.15*C*).

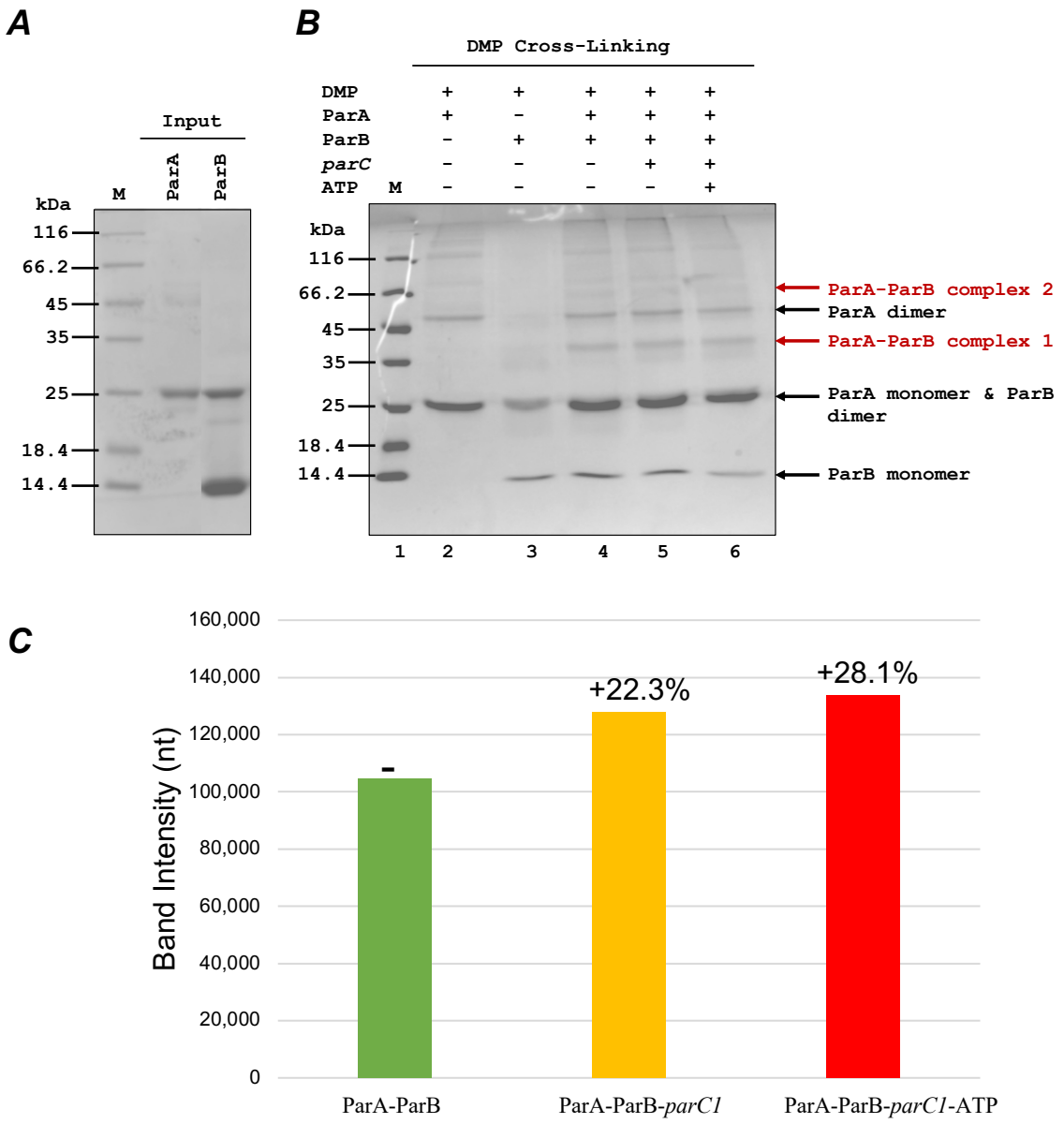


Figure 5.15 The partition site parC1 and ATP enhances the formation of the ParA-ParB complex

A) 12% SDS-polyacrylamide gel showing the input proteins, ParA and ParB that are not cross-linked. ***B)*** 12% SDS-polyacrylamide gel showing the cross-linking reactions containing ParA and ParB in the absence and presence of *parC1* and ATP. Lane 1 shows the molecular weight marker. Lanes 2 and 3 contain ParA and ParB cross-linked alone, respectively. Lanes 4, 5 and 6 show ParA and ParB crosslinked, alone, in the presence of *parC1* and the presence of *parC1* and ATP, respectively. ParA monomer, ParB monomer, ParB and ParA dimers are indicated by the black arrow. After addition of DMP cross-linker, the 39.5 kDa complex is indicated by a red arrow and represents a ParA monomer crosslinked to a ParB monomer. A second complex migrating at the level of at 64.8 kDa is also indicated by a red arrow and it is likely a ParA dimer cross-linked to a ParB monomer. ***C)*** Bar chart showing quantitation of the band intensity of ParA-ParB cross-linked complex I. The intensity of bands of three different conditions was calculated from three independent experiments using Image Lab 4.0.1 software (Bio-Rad). The increased percentages are shown on the top of the bars.

5.2.5.3 Identification of amino acid residues that are involved in the interaction of ParA and ParB using BS3 cross-linking combined with Liquid Chromatography-Mass Spectrometry (LC-MS)

DMP cross-linking experiments were able to capture a ParB monomer cross-linked to a ParA monomer. Moreover, the presence of certain ligands enhanced the interaction between the proteins. To investigate further the ParA-ParB interaction interface involved in complex formation, both proteins were chemically cross-linked using BS3 (bissulfoscuccinimidyle substrate) cross-linker followed by mass-spectrometry. Cross-linking coupled to LC-MS is a powerful tool that allows to identify protein associations and the interaction interface. BS3 is an amine specific cross-linker with an 11.4 Å spacer arm that covalently bonds lysine side chains in the functional area as well as the N-termini of the protein and brings them in close proximity.

In this experiment, the purified ParA and ParB were chemically cross-linked with BS3 in three different reactions; alone without any additions, in the presence of *parC1*, and the presence of both *parC1* and ATP. Cross-linking reactions were carried out as detailed in Section 2.8.2. The experiment was performed in triplicate and all reactions were analysed by SDS-PAGE. Addition of BS3 cross-linker to the reaction mixture led to covalently link both proteins and resulted in the formation of the 40 kDa high molecular weight complex (Figure 5.16*A*). Band intensity in the three conditions in three different experiments was quantified using Image Lab 4.0.1 software before the downstream experiment was performed. The band intensity of the ParA-ParB complex increased slightly in the presence of the partition site *parC1* and more substantially in the presence of ATP. Quantification analysis of band intensity showed that, when *parC1* was present, the average amount of ParA-ParB complex is 0.7% higher than in the reaction containing ParA and ParB alone. Additionally, the average quantity of ParA-ParB complex in the presence of both *parC1* and ATP increased significantly by 30.8% (Figure 5.16*B*). These data support the previous observation when both proteins were cross-linked using the DMP cross-linker that has a slightly different chemical structure. These results indicate that centromere *parC1* and ATP enhance the binding of ParB to ParA.

Once the bands that corresponding to ParA-ParB complex was quantified, the LC/MS was carried out. The ParA-ParB complex band was excised from the SDS-polyacrylamide gel and digested with trypsin followed by LC/MS analysis. Due to technical error, it was impossible to analyse all bands from three separate gels, therefore, only one repeat was examined and one condition from the second repeat. In total, nine

bands were obtained and digested but only four of them were analysed to identify the cross-link peptides. The results showed four cross-linked peptides. Based on the location of the cross-links, the results were categorised into two groups. The first group contains inter-protein contact, i.e. the cross-links occurred between two different proteins (i.e., when ParA is cross-linked with ParB). The second group shows the intra-protein contacts, i.e. the cross-links occurring either between or within the same protein. In this set of experiments, intra-protein cross-links were only found in ParB protein. In total four cross-linked peptides were identified, two inter-protein and two intra-protein, in which ParB-ParB were cross-linked (Figure 5.17). In the case of the ParA-ParB complex, lysine 28 from ParA was cross-linked to lysine 3 of ParB. Besides, lysine 179 of ParA was cross-linked to lysine 4 of ParB (Figure 5.17). This indicate that residues that are located in the N-terminal region of ParB, which was shown to be unstructured by the NMR results, are involved in the interaction with the partner ParA. However, it is difficult to draw definitive conclusions, as only one experiment was performed. The cross-linked peptides that were obtained between or within ParB identify lysine 66 cross-linked to lysine 30 and lysine 3 cross-linked to lysine 30 (Figure 5.17). This suggests that the ParB protein folds in a such a way to bring the N-terminal and C-terminal domains in close proximity to assemble a compact structure. This could happen either by bridging residues that are located in the tip of the N-terminal domain with residues that are located at the C-terminal end of N-terminal region (i.e. the N-terminus folds to be in close proximity to the folded C-terminal domain of the ParB tail). Or residues at the N-termini may interact with the folded C-terminal domain. This hypothesis holds true, if the interaction occurred only within one single ParB subunit. Nevertheless, the interaction may also happen between two ParB monomers, which may be the case here as the tetramer of ParB forms a slightly lower molecular weight (35 kDa) than the ParA-ParB complex. This may indicate that there are two populations in a single band.

Importantly, the number of cross-linked peptides obtained from the interaction between lysine 28 in ParA and lysine 3 in ParB was 90% higher when *parC1* and ATP were added to the reaction. The same high-level peptides were obtained in the cross-linkage between lysine 179 in ParA and lysine 4 in ParB (97%) in the presence of *parC1* and ATP (Figure 5.17). Moreover, the higher number of cross-linked peptides was not only observed in the inter-protein group but also in the intra-protein group. The number of cross-linked peptides obtained between lysine 30 and 66 within ParB, when *parC1* and ATP were presented in solution were 97% higher than the reaction contains the ParA/ParB and *parC1* (Figure 5.17). The number of cross-links between lysine 3/30 in ParB increased 4-fold, when *parC1* was added and increased further to 8.5-fold, when

both *parC1* and ATP were added to the reaction (Figure 5.17). These observations suggest that the centromere *parC1* and ATP somehow enhance the binding of ParB to ParA. Additionally, this suggests that ParB may undergo conformational changes, when *parC1* and ATP are present, which brings the two domains close together and allow ParA to bind more efficiently. However, to verify this hypothesis, additional experiments are needed. The results obtained from BS3 cross-linking are displayed in an interaction network map (Figure 5.18).

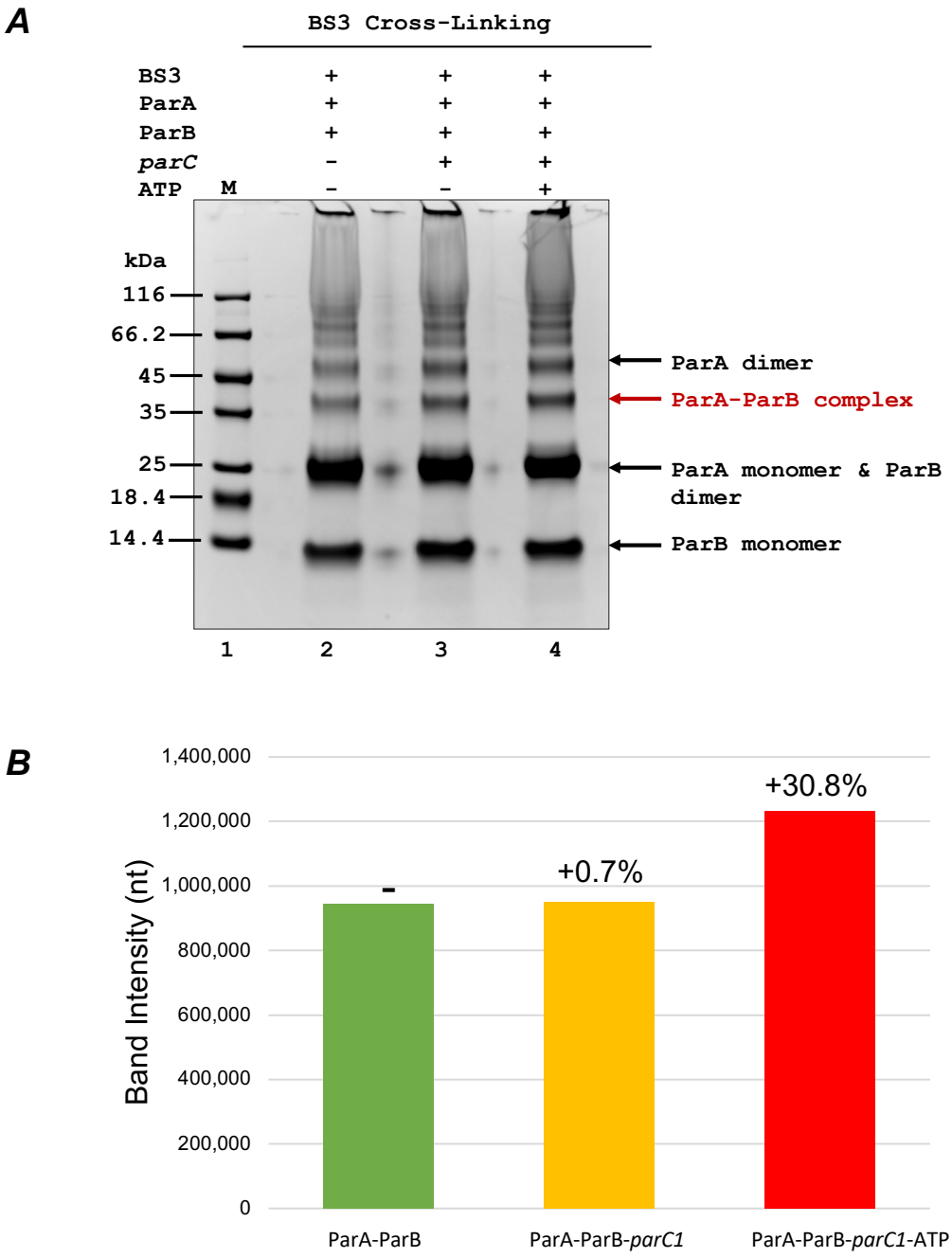
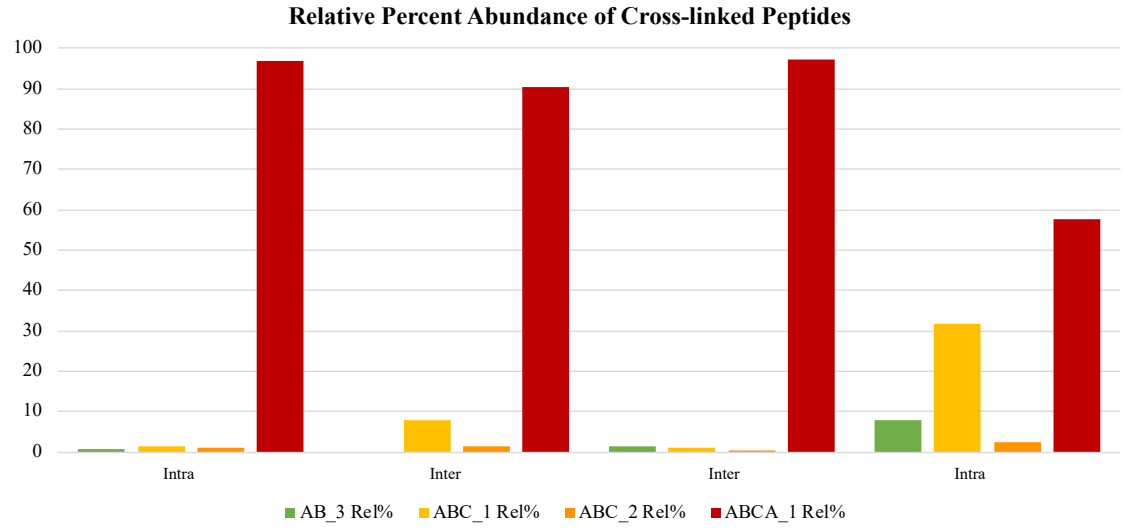


Figure 5.16 BS3 cross-linking and ParA-ParB band intensity in the presence and absence of the centromere *parC1* and ATP

A) Representative 12% SDS-polyacrylamide gel showing the cross-linking results of ParA plus ParB with and without *parC1* and ATP and the input of the LC/MS experiment. Lane 1 contains the molecular weight marker. Lanes 2, 3 and 4 show ParA and ParB cross-linked, alone, in the presence of *parC1* and in the presence of *parC1* and ATP, respectively. ParA and ParB monomers, and ParB and ParA dimer are indicated by the black arrowed. The 39.5 kDa ParA-ParB complex is indicated by a red arrow and it represents a ParA monomer cross-linked to a ParB monomer (lane 2, 3 and 4). ***B)*** Bar chart showing quantification analysis of the band intensity of the ParA-ParB complex cross-linked using BS3. The intensity of bands from three different conditions was quantified from three independent experiments using Image Lab 4.0.1 software (Bio-Rad). The percentage increases are shown on the top of the bars.



Cross-link Type	AB3	ABC1	ABC2	ABCA1	Cross-link at A position	Cross-link at B position
Intra-protein	1	1	1	97		K66-K30
Inter-protein	0	8	2	90	K28	K3
Inter-protein	1	1	0	97	K179	K4
Intra-protein	8	32	2	58		K3-K30

Figure 5.17 Relative percentages of cross-linked peptides identified through BS3 cross-linking coupled with LC/MS

The ParAB only reaction is indicated as (AB), the ParABC reaction containing *parC1* as (ABC1 and 2) and the ParABCA containing *parC1* and ATP as (ABCA). Cross-links peptides identified are classified into intra- (ParB-ParB) and inter-protein (ParA-ParB) groups. The number of cross-linking peptides obtained from each condition is shown in the table below, in which the blue colour indicates a low number of peptides identified and the red colour indicate a high level of peptides. The cross-linked peptides at position A (ParA) and position B (ParB) are shown.

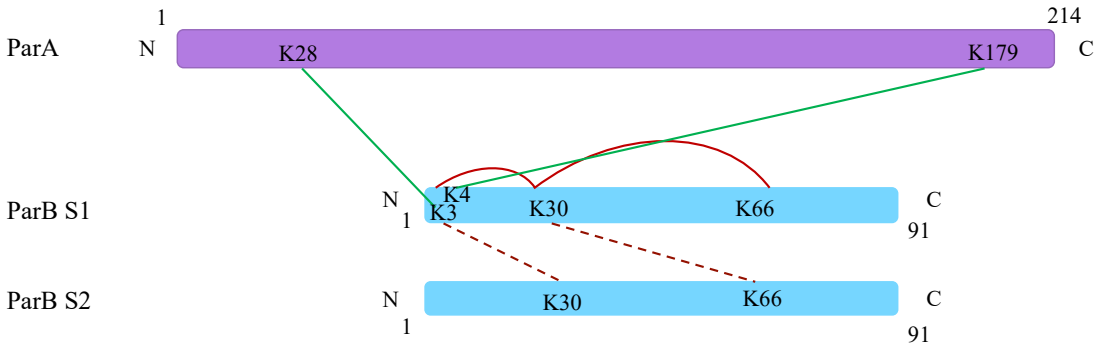


Figure 5.18 Cross-link map

The map shows the identified cross-links between ParA and ParB and within ParB. A cartoon of ParA is shown in purple and ParB subunit I (S1) and ParB subunit II (S2) are shown in cyan colour. The green lines indicate cross-links between ParA and ParB. The red solid lines indicate cross-links within the same subunit and dashed red lines indicate cross-links between two ParB subunits. The cross-linked peptides residues are indicated.

5.3 Conclusions and discussion

The molecular interactions between partition proteins are crucial in plasmid segregation. The molecular events start with the formation of the partition complex, in which ParB binds site-specifically to the partition sites, *parC1* and *parC2,* and then ParB interacts with its partner ParA. In Chapter 3, the minimal partition sites were probed *in vivo* and the results showed that *parC2* is crucial for plasmid partition. However, as part of this project, gel retardation assays have shown that ParB binds equally to *parC1* and *parC2* with similar affinity. However, MST experiments have shown than ParB binds *parC1* partition site with five-fold higher affinity than binding to *parC2*. During this study, the structure of the centromere-binding protein ParB was determined using NMR spectroscopy and solved using X-ray crystallography. The structure showed that the C-terminal domain folds to form a RHH structure, whereas in solution NMR experiments revealed that the N-terminal domain lacks structure and is highly flexible. With these insights, it is possible to thoroughly investigate the interactions among the three segregation components of the segregation complex.

In this chapter, the interactions among the pB171 segregation proteins and centromere were studied using a combination of biochemical and biophysical techniques. Moreover, the details of the ParB structure derived from the NMR study were used to determine the interactions between ParB with a single repeat of the partition site. NMR and chemical cross-linking coupled to LC/MS experiments allowed us to identify the interactions within the segrosome. EMSA and MST approaches were used to investigate the interaction and the binding affinity between ParB and the centromere sites. The biochemical results showed that ParB binds with high affinity to the *parC1* centromere. EMSA experiments showed that ParB causes a shift of *parC1* starting at a concentration of only 50 nM. *parC1* was shifted more when the ParB concentration was increased and led to the formation of multiple complexes. This might indicate that ParB binds to repeats without any preferences, forming multiple complexes with different molecular weights. A similar band shift pattern was obtained when ParB bound to the *parC2* partition site in EMSA. In contrast, MST experiments have shown that ParB binds with two-fold higher to *parC1* than that shown in EMSA, whereas, the protein binds five-fold lower to *parC2*. The dissociation constant and the cooperativity between ParB and the centromere sites were established using the MST technique. The results showed that ParB bound to *parC1* with high affinity and with lower affinity to *parC2*, the K_D values calculated from the MST binding curve were 119.3±27.48 and 613.38± 284.81 nM, respectively. Moreover, the Hill coefficient of the two binding experiments was also calculated from MST and showed that ParB binds cooperatively to *parC1* and *parC2*. These results suggest that ParB binds

as a dimer to a single repeat and once it binds, it forms a nucleation site that allows another ParB dimer to bind to the adjacent repeat.

Gerdes *et al.,* showed that the wild-type ParB and ParBΔ39 bind to *parC1* and *parC2* in a similar fashion and they suggest that the ParB C-terminal domain binds to the centromere sites (Ringgaard *et al.*, 2007b). Here, NMR studies were conducted to identify ParB DNA-binding site. Significant changes occurred in the spectrum, when site-specific DNA was added to ParB. The NMR study showed that the RHH C-terminal domain binds to the DNA oligonucleotides as the peaks corresponding this region disappeared from the spectrum. This observation supports the previous study conducted showing that the C-terminal domain is responsible for binding (Ringgaard *et al.*, 2007b). However, the N-terminal domain also showed changes in the chemical shift when the DNA was added to the solution as many residues were shifted. Based on the results obtained, the N-terminal region can be classified into three clusters. The first cluster includes amino acids located at the tip of the N-terminal region, which corresponds to residues lysine 3 to alanine 9. This cluster support an interaction occurring with DNA as many of the peaks corresponding to residues in this region disappeared, when the DNA concentration increased to 1:2 ParB: DNA ratio. The second cluster is located in the middle region of the N-terminal tail starting from asparagine 11 to lysine 30. Some residues in this region underwent chemical shift changes, when the DNA was present in solution. This might indicate that this region undergoes conformational changes or possibly it becomes structured upon addition of the DNA. The third cluster, which is located at the C-terminal end of the N-terminus was not affected as residues in this area experienced no or negligible changes in the chemical shift, when DNA was added. This region connects the N-terminus to the folded RHH C-terminus and includes residues tyrosine 32 to glutamic acid 39. However, it was difficult to prove the conformational changes hypothesis since changes in chemical shift in the second region, may also indicate changes occurring in the chemical environment and not be symptomatic of conformational changes. Moreover, we were interested in some residues that experienced a significant chemical shift change, once the DNA was added. The residues were chosen carefully for mutagenesis including lysine 3 that disappeared completely from the spectrum at the lowest DNA concentration and lysine 4 that disappeared at high DNA concentration. These two residues belong to the first cluster. On the other hand, asparagine 11, arginine 12 that are locating at the beginning of cluster II experienced a low chemical shift and arginine 26 that is located at the end of the second cluster showed a significant change in chemical shift. Those amino acids were changed to alanine and *in vivo* studies showed that changing lysine 3 to alanine significantly affected plasmid

stability. This suggests that lysine 3 may be involved in DNA binding. However, the plasmid encoding ParB with lysine 4 changed to alanine displayed a reasonable retention level, which was 27% lower than the wild type, although this amino acid disappears from the NMR spectrum when the DNA is added at 1:2 ratio. This result suggests that the adjacent lysine 3 and 4 may affect each other. In other words, the changes in the chemical shift in lysine 4 may have resulted from the changes occurred in the neighbouring lysine 3. On the other hand, changes in asparagine 11 showed no effect in plasmid retention level compared to the wild type. Changing arginine 12 to alanine had affected moderately on plasmid retention. Changing arginine at position 26 to alanine reduced the plasmid maintenance to 2.5 folds. However, these results are not conclusive as further *in vitro* study needs to be carried out to investigate the importance of residues in the N-terminal domain of ParB.

Chemically cross-linking with DMP showed that *parC1* and ATP are essential for segrosome assembly as adding these two ligands to the cross-linking reactions enhanced the interaction of ParB with ParA. In addition, the presence of the *parC1* DNA alone did not improve the interaction occurring between the two segregation proteins. To investigate further, BS3 cross-linking coupled with LC/MS was performed to identify the binding interface of ParA and ParB as well as to investigate the influence of centromere site and ATP on ParA-ParB complex formation. In agreement with the results obtained from DMP cross-linking, BS3 cross-linking with LC/MS showed that the presence of *parC1* in solution enhanced slightly the binding of ParB to ParA. The enhancement increased significantly when both *parC1* and ATP were present in solution. This suggests that *parC1* and ATP are essential elements for ParA-ParB complex formation, meaning that segrosome assembly needs the presence of all elements at once. The cross-linking results showed that lysine 3 and 4 in ParB interact with lysine 28 and 179 in ParA, respectively. Interestingly, lysine 3 and 4 were shown to interact with DNA as well, meaning that the N-terminal domain of ParB plays two roles, binding to DNA as well as binding to ParA. These two lysine residues are the only cross-linked peptides that were identified in ParA and ParB complex. However, this experiment will need repeating to allow conclusions to be made. In addition to the cross-linked peptides that joined ParB to ParA, two more peptides were retrieved between ParB molecules or within ParB molecule. The results suggest that the flexible N-terminal domain of ParB might come in close proximity to the RHH C-terminal domain. Again, it is not possible to conclusively prove the latter hypothesis, as only one repeat was analysed, and further experiments are needed. Multiple attempts of NMR experiments were carried out to identify the

interaction site between ParA and ParB, however, it proved difficult to carry on with this experiment as the ParA protein tends to precipitate quickly in the NMR buffer.

Chapter 6

Discussion and Future work

6.1 Discussion

Active partitioning systems of the low-copy number plasmids have been studied thoroughly in the last three decades. During these years of studies, many models for the molecular mechanisms have been proposed. However, there are still many aspects that need to be elucidated. The first step in the plasmid partition process that follows plasmids duplication is the site-specific binding of the centromere-binding protein (CBP) to its cognate repeats site and formation of a nucleoprotein complex. The nucleoprotein complex recruits the partner motor protein for proper positioning of the two sister plasmids prior to cell division (Schumacher, 2008; Baxter and Funnell, 2014; Hayes and Barillà, 2006b).

The pB171 plasmid is a virulence plasmid that is harboured by an enteropathogenic strain of *E. coli* (Tobe *et al.*, 1999). The plasmid carries two DNA segregation systems, *par1* and *par2*. The *par2* system, which is the focus of this thesis, comprises two proteins, ParB and ParA, and two of centromeres sites, *parC1* and *parC2* (Ebersbach and Gerdes, 2001). ParB performs a dual function, binding to the centromeres for plasmid partitioning, and to the *parAB* promoter for transcriptional repression of the genes (Ringgaard *et al.*, 2007a; Ringgaard *et al.*, 2007b). ParA is a Walker type ATPase protein. In the past, a model was proposed that ParA oscillates in a spiral-like structure from one nucleoid pole to another for moving the plasmid apart before cell division (Ebersbach *et al.*, 2006; Ringgaard *et al.*, 2009). A few years later, a model was proposed in which ParA forms gradient over the nucleoid that direct the plasmid movement towards the nucleoid pole (Hwang, *et al.*, 2013; Vecchiarelli, *et al.*, 2013a; 2013b; 2014a). More recently, a model of TP228 plasmid suggested that ParF forms a meshwork (ParA homologous) consisting of polymers and small oligomers over the nucleoid volume (McLeod *et al.*, 2017)

One of the main aims of this thesis was to understand the contribution of the centromeric repeats in plasmid partitioning *in vivo*, and the findings will be discussed in detail in the first part of the discussion. Moreover, details of understanding nucleoprotein assembly and the role played by *parC2* partition site are explained and a model is proposed. The CBP ParB was structurally characterised. The results indicate that the protein is structurally similar to other members of type Ib partition system. The protein was also characterised biochemically, and via in solution NMR. A model of the ParB-DNA interaction is discussed. The dual role played by the N-terminal domain in the interaction with *parC*s and the partner ParA is also discussed. Several recommendations for future research are provided in the future work section.

6.1.1 The *parC2* site is crucial for plasmid centromere activity

The centromere sites are widely divergent in bacteria. In plasmid partitioning systems, the centromere site consists of tandem repeats, though their sequences, lengths, numbers and orientations are extremely variable (Bouet and Funnell, 2019). The position of the centromere sites is varied, they can be located upstream or downstream of the *par* genes (Hayes and Barillà, 2006a; 2006b). Some plasmids possess multiple copies of centromeres at several positions across the plasmid, and some of them are located at a distance from the *par* genes. An example of the aforementioned plasmid segregation system is the pSM19035 plasmid that contains three sets of centromeres. Two of them are close to the segregation cassette, while the other is at a distance from the cassette. The sequences of repeats in the three centromere sites are related, but the number and orientation are varied (Dmowski and Kern-zdanowicz, 2016). On the other hand, the *par2* system of pB171 comprises two centromere sites, however, they are both adjacent to the segregation cassette, straight upstream and downstream (Ringgaard, *et al.*, 2007a). Both sites have been shown to be involved in plasmid segregation through biochemical investigation (Ringgaard *et al.,* 2007a) and in live cells (This study). The two centromeres are *parC1* and *parC2*, where *parC1* is located upstream of the *parAB* genes, whereas *parC2* is located downstream (Ebersbach and Gerdes, 2001; Ringgaard *et al.*, 2007a). The two sites are composed of seventeen and eighteen hexameric motifs, respectively. All the repeats are arranged in direct orientation only. The two centromere sites are distinctive in the arrangement and distribution of repeats. However, the motifs are related. The *parC1* repeats are arranged into two clusters of fourteen and four motifs, respectively, while the repeats in *parC2* are organised into three groups of eight, three and seven motifs, respectively. The repeats within clusters are separated by a single nucleotide (Ebersbach and Gerdes, 2001). Biochemical studies had shown that ParB recognises both centromeres and that the protein binds to both partition site, *parC1* and *parC2*, with similar affinity (Ringgaard, *et al.*, 2007a). However, our *in vivo* investigation has shown that the presence of the *parC2* site is crucial for segrosome assembly and plasmid partition. As long as *parC2* contains at least a single repeat, plasmid segregation is observed. The study showed that the presence of any ParB repeat in *parC2* is capable to mediate the centromere activity more efficiently. This result mirrors the observations related to the *sopABC* partition system from F1. In the F1 plasmid, the centromere site *sopC* comprises 12 repeats of 43 bp (Pillet *et al.*, 2011) and it was shown that one repeat in the centromere is as efficient as the full set of 12 repeats (Biek and Shi, 1994). In the P1 plasmid, the repeats in the centromere site *parS* are categorised into three regions, left, middle and right regions. The flanking regions are ParB recognition sites, while the central region is occupied by the IHF recognising site (Schumacher *et al.*, 2007b). The

centromere sites in *par2* cassette of the pB171 plasmid possess a similar repeats organisation as *parS* of P1 plasmid, however, the number of repeats is higher. The two centromere sites, which are equivalent to the two centromeric arms in *parS*, are separated by two IHF recognition sites. However, in P1 *parS*, the two regions (the left and right arms) are shorter than the two centromere sites in *par2* system of pB171 plasmid. For the P1 plasmid, it was shown that the IHF site is involved directly in plasmid partition by bending the DNA and permitting the ParB protein to bind multiple sites and form a partition complexes with a distinctive structure. However, in the absence of the IHF site, ParB can still bind the right arm with high affinity, but it cannot bind to the left arm (Schumacher *et al.*, 2007a). In this *in vivo* study, we speculate that the presence of the *parC2* site allows ParB to bind *parC1*. *parC2* could work as a nucleation site for ParB and may be an essential site to mediate segrosome assembly. As mentioned earlier, besides the two centromere sites, the *parAB* system of pB171 has two additional sites, *ihf1* and *ihf2* (Ringgaard, *et al.*, 2007a). It is still unclear whether these sites are involved in plasmid segregation. Biochemical investigations showed that the IHF protein bound to *parC2* but not *parC1*, meaning that the *ihf* site is located within *parC2* (Ringgaard, *et al.*, 2007a). The boundaries of the *ihf2* site were mapped and showed that it overlaps with the *parB* stop codon, the first B repeat and the first two bases from the second B repeat in the *parC2* site. In addition, an *in vivo* study showed that deletion of the gene encoding the IHF protein had affected moderately plasmid stability (Ringgaard *et al.*, 2007a). It is still unclear whether IHF binds to *parC2* in a similar fashion as that observed for IHF and the centromere of P1 plasmid. No other study has shown the involvement of IHF in bending DNA. However, as *ihf2* site overlaps with the first repeats and part of the second repeat in *parC2*, deleting *parC2* site may also remove part of the *ihf2* site. It remains to be elucidated whether the host factor IHF or another factor contributes to facilitate the centromere activity or, alternatively, whether the unique positioning and organisation of the repeats in *parC2* is key for the role of this centromere.

The R1 and R2 repeats in *parC1* were previously investigated *in vitro* and the result parallels with our *in vivo* observations. Gel shift assays had shown that ParB failed to associate with a fragment bearing the two boxes (Ringgaard, *et al.*, 2007a). Our *in vivo* investigations have shown that the two 10 bp repeats, R1 and R2, which are located at the upstream of *parC1*, are not ParB recognition site and the plasmid missing these repeats is maintained at the same level as the plasmid harbouring the full cassette including these repeats. This indicates that the presence or absence of these two repeats does not affect plasmid stabilisation. The importance and role of the *parC1* repeats was also investigated. Overall, the absence of most repeats from *parC1* did not affect plasmid

stabilisation. Specifically, it seems that the plasmid segregation relies for a 20% on the *parC1* site, in which the plasmid bearing four *parC1* repeats was maintained. It was difficult to draw definitive conclusions on the importance of the repeats in the *parC1* site due to the critical position of four repeats within the promoter region. We hypothesize that this partition site could work as a secondary, auxiliary centromere for optimal plasmid stabilisation. Moreover, the presence of only four repeats is sufficient to ensure proper centromere activity. However, the optimal stabilisation is reached when all the repeats are present in *parC1*. In summary, the presence of the main centromere site *parC2* and the accessory centromere site *parC1* are essential for assembly of a nucleoprotein complex with a specific structure that mediates the plasmid partitioning efficiently. The involvement of the two sites in centromere activity and the crucial role played by *parC2* suggest that the nucleoprotein complex might be formed in a fashion that brings the two sites close together in a similar way as the one observed in the P1 *parABS* system. Unlike the simple centromeres, *parC1* and *parC2* may undergo a conformational change when ParB binds forming a specific structure that could be distinct from other nucleoprotein complex in other plasmid partition systems. Nucleoprotein complexes of different plasmid partition systems exhibit distinctive structures. For example, the ParR protein of the pB171 plasmid *par1* system forms a ring-shape-like structure on *parC1* site (Moller-Jensen *et al.*, 2007). In contrast, in simple centromeres such as the *parS* site of pSM19035, the protein ω binds to its cognate site and forms a left-handed helix that wraps around *parS* without bending or distorting the DNA (Weihofen *et al.*, 2006; Pratto *et al.*, 2009).

Similarly to the repeats in *parS* of P1 plasmid, the 6 bp repeats in the *parC1* and *parC2* sites are categorised into two groups according to the DNA sequence. Class I repeats are more frequent in *parC1*, while class II are more abundant in *parC2*. The arrangement and organisation of the two classes of repeats are random in the sense that they do not follow a specific pattern. However, class II repeats are more conserved than class I. We hypothesized that the first four class II repeats in *parC2* might be essential and might play a key role in centromere activity. However, our study showed that none of the repeats were essential, and any repeat in *parC2* works to confer plasmid stability. In summary, based on the *in vivo* study results and previous *in vitro* investigations, a model of ParB-DNA nucleoprotein assembly is proposed and named as nucleation, looping and bridging model.

6.1.1.1 A nucleation, looping and bridging model for nucleoprotein assembly

Nucleoprotein complex formation is an important and prerequisite step that leads to accurate plasmid partitioning before cell division occurs. Our *in vivo* investigations have shed light on the role played by the two *parC1* and *parC2* centromeres in pB171 plasmid partitioning. Based on our observations and by comparisons with the *in vitro* studies conducted on the *par2* plasmid partitioning system, a model of ParB-*parC* complex assembly is proposed. The model has been named as nucleation, looping and bridging, based on the three steps that are proposed. We have shown *in vivo* that the centromere sites in the *par2* system are not simple, but rather complex centromeres like *parS* of P1 due to the presence of the two *ihf* sites that occupy between the two centromeres. The centromere *parC* site of *par2* of pB171 may undergo a conformational change when the CBP ParB is loaded on its recognition site. We propose that *parC2* might work as a nucleation site as deletion of this particular site abolished centromere activity and entirely abrogates plasmid retention. It could be either because of the role played by *parC2* in centromere activity or that the IHF protein may allow ParB bind to the *parC1*. Based on these hypotheses, we assumed that ParB binds to *parC2* and a nucleoprotein complex is formed. After the first complex is formed, this may allow more ParB dimers to bind to *parC2*. Once this site is saturated with ParB, the formation of a nucleoprotein complex begins on *parC1*. The building up of ParB on both centromeres might bring the sites into close proximity. A previous *in vivo* study had shown that the deletion of gene encoding IHF causes a significant decrease in plasmid stability (Ringgaard *et al.*, 2007a). Our study has shown that when *parC2* is deleted, the *ihf* site is also affected. IHF might have a role in bending DNA as observed for the centromere of the P1 plasmid. IHF or another unknown factor or the saturation of both *parC1* and *parC2* with ParB dimers might bend the DNA containing the partition locus leading to bridging of the two centromere and formation of a loop structure. In the absence of *parC2*, no looping occurs, and this results in an aberrant centromere that is unfit to mediate plasmid segregation (Figure 6.1).

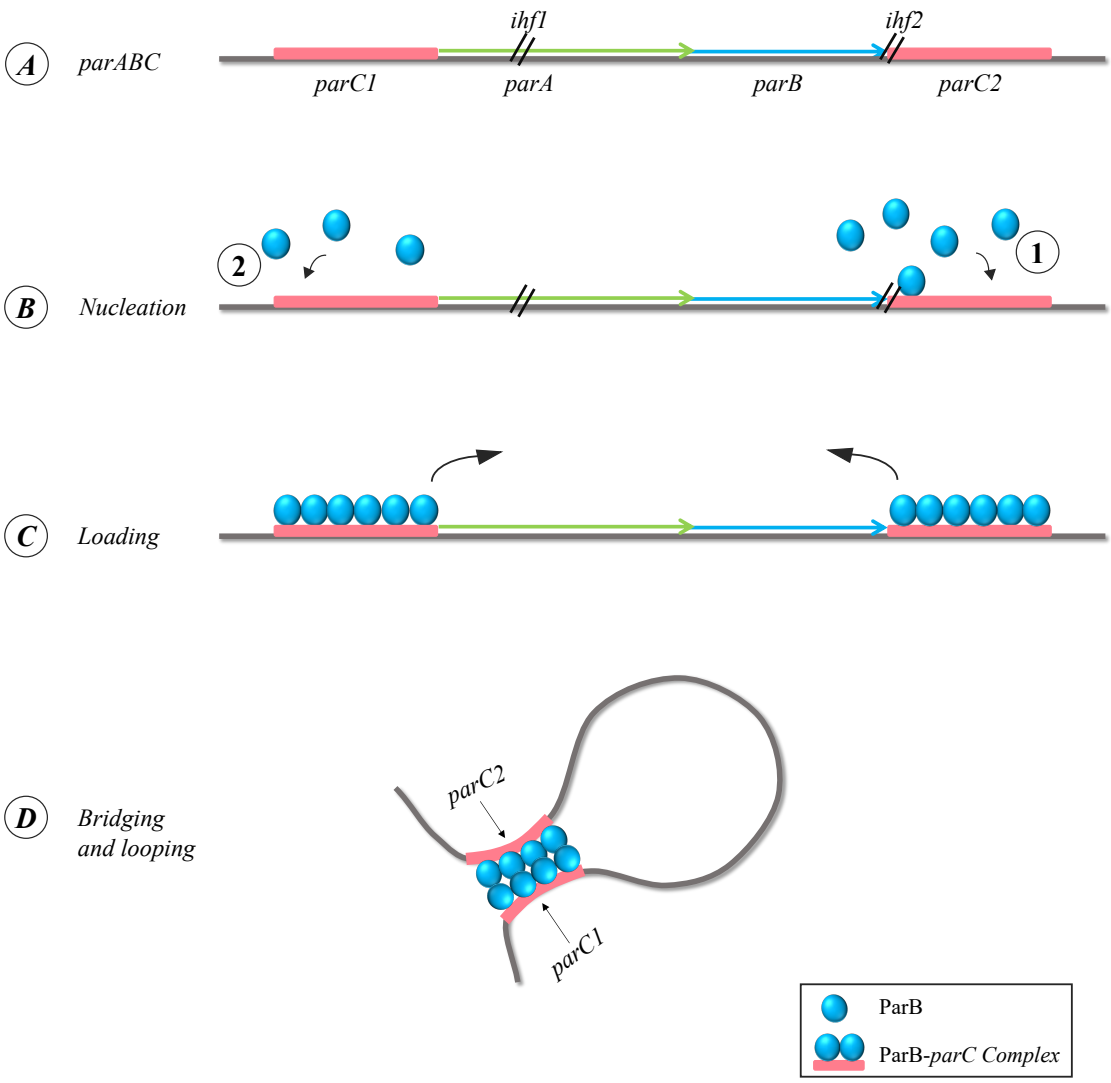


Figure 6.1 Model of ParB-*parC* nucleoprotein assembly

A) Diagram showing the *parABC* cassette. The *parA* and *parB* genes are indicated with green and blue arrows, respectively, while the *parC1* and *parC2* sites are indicated with pink bars. The *ihf1* and *ihf2* sites are indicated with double tilted lines. ***B)*** ParB (blue ball) loads on the *parC2* site first (1) followed by loading on *parC1* (2). ***C)*** The binding of the ParB proteins to both sites leads to bending the DNA. ***D)*** Bending the DNA caused by ParB or other unknown factor allows both sites to come in close proximity. A bridge is formed between the two sites leading the formation of a loop-like structure.

6.1.2 ParB is a member of Arc/MetJ superfamily with a ribbon-helix-helix C-terminal domain and an unstructured N-terminal tail.

A thorough investigation of molecular mechanisms that underpins DNA segregation requires the elucidation of the tertiary structure of the proteins involved in the process. Elucidating the three-dimensional (3D) structure of ParB will help to characterise the DNA-binding motif. Here, a combination of NMR spectroscopy and X-ray crystallography techniques was used to reveal the tertiary structure of the centromere-binding protein ParB of the pB171 plasmid. In this work, the tertiary structure of the ParB C-terminal domain was revealed using X-ray crystallography, while the N-terminal domain was predicted using the Cα, Cβ, NH and CO NMR chemical shift. The crystals of ParBΔ30N required two weeks to grow and the structure was solved to 1.9 Å. The density map does not show the first twelve residues from the truncated N-terminal domain, suggesting these residues may experience degradation during crystal growth. The 3D structure showed that the C-terminal domain folds into a RHH structure. The folded RHH structure obtained from X-ray crystallography is in agreement with the data obtained from NMR spectroscopy. The two monomer subunits of the C-terminal fold of ParB paired symmetrically to form a tight homodimer with 2-fold rotation axis in which the two antiparallel β-strands intertwined to form a β-sheet. On the other hand, the N-terminal domain structure was predicted using NMR spectroscopy that showed that residues 1-43 are highly unstructured. In addition, the NMR data revealed that the central region in the N-terminal domain is semi-structured and was shown to fold into a semi-rigid α-helix structure. Furthermore, the 3D structure prediction of ParB showed that the Glu-8 could form a helical-like structure, however, the NMR experiment showed that this region is highly dynamic in solution. Although some residues in the tip tail of the N-terminal domain were assigned including Lys-3, Lys-4, Ser-6 and Glu-8, it was difficult to obtain the secondary structure prediction for this part of the protein. Therefore, it is not possible to make conclusive statement as to whether the eight amino acids that form the tip of the tail might form a transient α-helical structure similar to the one observed in the middle of the N-terminal domain.

In plasmid segregation systems, the centromere-binding protein (CBP) falls into one of two structural classes, harbouring either a HTH or a RHH DNA-binding domain. The HTH DNA-binding domain is found in proteins belonging to type Ia partition systems, whereas, the RHH DNA-binding motifs is observed in proteins of type Ib and type II partition systems. It should be noted that all CBPs of type Ib systems characterised so far have a

C-terminal RHH fold, while the N-terminus is unstructured. In contrast, the RHH folds in type II systems proteins is located at the N-terminus and no unstructured region has been observed (Baxter and Funnell, 2014). In type Ib system CBPs, the number of residues involved in the N-terminal region varies. A short N-terminal tail is observed in TP228 ParG (residues 1-32) whose structure was solved by NMR spectroscopy (Golovanov *et al.*, 2003). The N-terminal domain of Omega (residues 1-22) was predicted to be unstructured and to have a short tail similar to that of ParG (Murayama *et al.*, 2001). In contrast, a long N-terminal tail is found only in ParB of pCX1001 (residues 1-65) (Huang *et al.*, 2011). The presence of 43 amino acids in the N-terminal domain of ParB may put this domain in the long N-terminus class like ParB of pCX1001. Therefore, based on the four known structures of type Ib systems CBPs, it seems that the number of residues involved in the N-terminal domain depends mainly on the length of the protein, since the length of the C-terminal RHH fold in all proteins are potentially very close. Furthermore, the N-terminal flexible tail can be considered a characteristic feature among known CBPs of type Ib systems. Although these CBPs share the common feature of a flexible N-terminal tail, no sequence identity is observed among these proteins.

ParG protein of TP228 plasmid and the ω protein of pSM19035 plasmid are among the RHH proteins whose tertiary structure has been solved and excessively studied. Studies of the two proteins have shown that the C-terminal domain is involved in dimerisation and binding to DNA for plasmid partitioning and autoregulation (Barillà and Hayes, 2003; Carmelo *et al.*, 2005; Pratto *et al.*, 2008; Soberón *et al.*, 2011). The same role of for the RHH domain is also observed in ParB of pB171 from type Ib system and ParR of pSK41 and pB171 plasmids from type II systems, despite the lack of sequence identity among ParB structural homologues (Moller-Jensen *et al.*, 2007; Ringgaard *et al.*, 2007a; Ringgaard *et al.*, 2007b; Popp *et al.*, 2010). In contrast, the function of the flexible N-terminal domain requires still to be elucidated. It was shown that the N-terminal flexible tail in ParG is involved in the formation of a nucleoprotein higher-order structure for transcription autoregulation. Furthermore, the N-terminal tail is also involved in the higher-order assembly and stimulation of ATPase activity of the ParG partner ParF (Carmelo, *et al.,* 2005; Barilla *et al.,* 2007). These activities were also enhanced by the N-terminal tail of the ω partner protein δ (Pratto *et al.,* 2008).

Structure comparison searches reveal that ParB belongs to Arc/MetJ transcriptional repressor superfamily. The proteins falling within this family harbour a short RHH DNA-binding domain, which is located either at the C-terminus or the N-terminus (Schreiter and Drennan, 2007). Despite having the same RHH fold, the domain shares a very poor

sequence identity. By comparing the sequences, it emerges that the conserved residues are mainly hydrophobic that form a compact central core structure and are involved mainly in the dimerisation interface All RHH family members dimerise similarly. Moreover, conserved residues are mainly involved in DNA interaction and structure stabilisation. Significantly, the ParB 3D structure shows that the nine conserved substitutional hydrophobic residues and one polar residue are pointing towards the centre of the core structure, which makes the RHH folds. The highest degree of conservation is observed between ParB of pB171 and CopG of plasmid pMV158 which share 40% sequence identity. For TP228 ParG, it was shown that Val-68 and Trp-71 in α-helix 2 and Phe-49 in α-helix 1, are involved in the dimerisation interface and the core structure of the protein (Golovanov *et al.*, 2003; Saeed *et al.*, 2015). The same conserved residues were observed in ω and the authors predicted that those amino acids are involved in the dimerisation interfaces (Dmowski and Jagura-Burdzy, 2011). Interestingly, by comparing the sequence of the folded RHH domain of ParB and ParG, similar hydrophobic residues could be observed also in ParB α-helix 2 that may share the same function as that shown for the ParG and Omega proteins. In contrast, no significant sequence identity was observed in the β-strand and α-helix 1, although they share some sequence features. Arc, MetJ, CopG and ω structures were determined in complex with the cognate DNA. In all protein-DNA structures, the antiparallel β-strands are inserted into the DNA major groove and make a sequence-specific contact through positively charged residues (Somers and Phillips, 1992; Raumann *et al.*, 1994; Gomis-ruth *et al.*, 1998; Augustus *et al.*, 2006; Weihofen *et al.*, 2006). As pB171 plasmid ParB shares the same domain organisation and shares 40% sequence identity with CopG, it is predicted that ParB may make base contact similarly to CopG, Arc, MetJ and ω. In the β-strand of the RHH domain, despite variable length, most of the transcriptional regulator proteins share a similar pattern with the presence of the positively charged residues either at beginning or end of the β-strand. Noticeably, either an arginine or lysine is present at the beginning or the ends of the β-strand. However, that is not the case in short β-strands, where no positively charged amino acids in either positions are found. In the RHH proteins with a short β-strand, the positively charged arginine or lysine is present just upstream of the β-strand, in the region where the N-terminal domain is connected with the folded C-terminal domain. These positively charged amino acids may make contact with DNA bases. Based on the position of the positively charged residues in the β-strand, Golovanov *et al.* (2003) categorised the proteins with the positively charged residue located at the beginning as type I repressors (ParG, CopG, ParD), and those located at the end as type II repressors (Omega, Arc, Mnt). Based on this classification and the position of the positively charged arginine in the β-strand of pB171

ParB, ParB belongs to the class of type I repressors. The structure of the CopG-DNA complex was also solved and showed that the side chains of Arg-4, Thr-6 and Thr-8 in the β-strand recognise bases and make a sequence-specific contact (Gomis-ruth *et al.*, 1998). We have shown that ParB of pB171 share 40% sequence identity with CopG and the conserved amino acids between the two proteins were highlighted showing that Arg-4 and Thr-8 in the β-strand are conserved. I speculate that Arg-45, Thr-46 and Thr-47, could make similar bases contacts as was shown for CopG. The presence of Arg and Thr in the β-strand has been observed among many of transcriptional regulatory proteins. Another distinctive feature among RHH transcriptional repressor proteins is that the flexible loop that connects α-helix 1 and α-helix 2 contain a G-X-S/T/N motif (Schreiter and Drennan, 2007) that has been observed only in short loops. However, this pattern is not observed in the proteins that have a longer loop, such as in ParB from pB171, ParB from pXC1001 and MetJ. The long loop in the aforementioned proteins does not exhibit a distinctive feature, and no sequence identity is observed among them. The long loop may have a distinct role or could make additional site-specific DNA contacts. However, the role played by these residues in the long loop is still unclear and further investigations need to be carried out. For HP0222, a ParB structure homologue, it was indicated that the loop separating the two α-helices is involved in the dimer-dimer interface (Popescu *et al.*, 2005)

In addition to having a similar RHH C-terminal domain, the Arc, MetJ, ω, ParG, ParB of pCX1001 and Mnt transcriptional regulators proteins share a similar flexible N-terminal tail except for CopG that possesses only a RHH domain (Schreiter and Drennan, 2007). The unstructured region was also observed in pB171 ParB by NMR spectroscopy. It was shown that Arc, MetJ and Mnt make additional DNA contacts through the unstructured N-terminal tails. Once the N-terminal tails of these proteins bind to DNA, they experience a conformational changed, which adds more specificity to the DNA interactions (Knight and Sauer, 1989; Augustus *et al.*, 2006; Schreiter and Drennan, 2007). HetNOE and structure prediction by NMR has shown that the middle region of the N-terminal domain possesses a semi-rigid structure and it is highly likely to be a transient α-helix structure. It is reasonable to speculate that this particular region may change its conformation when it binds to its cognate DNA. This observation will be discussed in further detail later in this chapter.

6.1.3 ParB binds with high affinity and cooperativity to centromeres through the C-terminal RHH folds and the unstructured N-terminal domain

Binding of the CBP to the centromere and formation of a nucleoprotein complex are essential steps to mediate plasmid partitioning (Baxter and Funnell, 2014). ParB binds site-specifically to the *parC1* and *parC2* sites, located upstream and downstream of the *parAB* genes, respectively (Ebersbach and Gerdes, 2001; Ringgaard, *et al.*, 2007a). The two sites consist of hexameric repeats arranged into two classes, based on the DNA sequence. The repeats are organised in direct orientation and arranged in clusters of two for *parC1* and three for *parC2* (Ebersbach and Gerdes, 2001). Like *parS* of the P1 plasmid, the *parC1* and *parC2* sites of pB171 have complex centromeres due to the arrangement and distribution of the two classes of hexameric repeats within the centromeres. As described previously, the importance of the repeats and minimal partition site involved in plasmid segregation were analysed *in vivo*. The study showed that *parC2* is crucial for plasmid segregation. We have also shown that the two classes of repeats are equally involved in segregation. We speculate that *parC1* could act as a support centromere. A model for the formation of the nucleoprotein complex was proposed.

In vitro studies, using EMSA and MST approaches, have shown that ParB binds avidly to the *parC1* DNA fragment, whereas, the interaction of ParB to *parC2* were different from EMSA to MST experiments. Specifically, the gel retardation assays have shown that ParB causes the two fragments of DNA that represent the two centromeres to shift at a very low ParB concentration, indicating a strong interaction. Moreover, a similar affinity between ParB and the *parC1* centromere site was observed by MST experiment. In contrast, MST experiments have shown that ParB binds *parC2* five-fold lower than that observed with *parC1* partition site. ParB was shown to bind cooperativity to its recognition site and a ParB dimer may associate with a single hexameric repeat. Cooperative interactions between ParB dimers bound to adjacent repeats may create a stable nucleoprotein complex. This suggests that the full *parC1* site requires at least seventeen ParB dimers, while at least eighteen ParB dimers are required to cover the entire *parC2* site.

As described earlier, the structure of ParB was solved and showed that the C-terminal domain folds into a RHH structure, while the N-terminus is unstructured. ParB is a member of Arc/MetJ transcriptional repressor superfamily. Moreover, NMR studies have shown that the middle region of the N-terminal domain (Val-15- Arg-26) forms a transient

α-helical structure. They also have shown that the N-terminal domain is highly dynamic in solution.

NMR investigations have shown that ParB binds the centromere sites through the C-terminal fold. The wild-type ParB and ParBΔ39 bind to *parC1* and *parC2* with similar affinity but with different pattern in which the truncated versions produced less protein-DNA complexes than compared to that of full-length of ParB (Ringgaard *et al.*, 2007a). A trial of gel shift assay was performed to investigate the binding behaviour of ParB to a single repeat. However, it proved difficult to observe the interaction with a single repeat using this technique (data not shown). As part of this project, in solution NMR experiments have shown that ParB binds to a single hexameric repeat through the C-terminal RHH domain. The region that corresponds to the folded RHH domain disappeared completely from the spectrum at 1:0.2 of protein:DNA ratio. For the RHH proteins whose tertiary structure was solved in complex with the cognate DNA site, it was shown that the antiparallel β-strands are inserted into the major groove of the DNA and make sequence-specific base contacts (Somers and Phillips, 1992; Gomis-ruth *et al.*, 1998; Weihofen *et al.*, 2006; Moller-Jensen *et al.*, 2007). However, due to the flexibility of ParB N-terminal domain, it proved difficult to co-crystallise the protein with the DNA. In eukaryotic proteins, unstructured terminal regions have been studied extensively and are widespread in the DNA-binding proteins. It was shown that flexible regions play a crucial role by increasing the affinity and specificity of DNA binding. In some proteins, the disordered region may undergo disordered-to-order transition, when it binds to DNA and may increase the affinity and specificity of the interaction. In other proteins, an unstructured tail found to support faster DNA-binding through a mechanism in which the unstructured tail binds first followed by the folded domain, however, the conformational changes process slows down (Vuzman and Levy, 2012). By contrast, the unstructured region in some of the bacterial transcriptional repressor RHH proteins has been shown to make extra nucleotide contacts (Knight and Sauer, 1989). The in-solution NMR has shown that the unstructured N-terminal region of ParB also makes base contacts, as some of the residues in this region experienced a chemical shift change.

The N-terminal region can be seen as consisting of three clusters of residues. Cluster I corresponds to the first ten residues that were shown to make DNA contacts, as most of the residues in this region disappeared from the spectrum, when the DNA concentration increased. Lysine at position 3 undergoes medium exchange in chemical shift, thus it disappeared from the spectrum when the DNA was added at low concentration. This characteristic of chemical shift changes suggest that Lys-3 is either in direct contact to

DNA, or that there was a significant change in conformation around this residue. On the other hand, Val-2 was not tracked as this residue was not assigned, therefore no information is available for this residue. The other residues in this region have experienced changes in a chemical shift at very low concentrations of DNA. However, at 1:0.8 ParB:DNA ratio, the cross-peaks corresponding to residues Lys-4 and Ser-6 were split into two peaks; one with low chemical shift change, while the other with higher chemical shift change. This may indicate that these two residues experience two modes of binding, i. e. one of the tip tails of the N-terminal domain from subunit I may interact with DNA, whereas the other tail from subunit II may bind to different nucleotide base or different phosphate backbone. As a result, two cross-peaks per residues were observed, when the DNA was added at 1:0.8 ratio. Therefore, the cluster I may be considered as an additional ParB region involved in DNA recognition besides the C-terminal RHH fold. Cluster II or middle cluster span amino acids that are experienced changes in the chemical shift upon DNA binding. There is variability in chemical shift changes in this region, some residues experienced more pronounced chemical changes than the others, such as Glu-18, Gln-19 and Leu-27. This region corresponds to residues that form the transient α-helix that may undergo disordered-to-ordered transition. An alternative explanation for the chemical changes might be the following: the changes that caused by the direct interaction of the cluster I to DNA that has affected the chemical environment of this region and become less mobile in solution. In contrast, the third cluster, which connects the N-terminal to the C-terminal domain, was not affected as residues in this region of ParB were underwent any or neglectable changes in the chemical shift, when the cognate DNA site was present. Many of the RHH Arc/MetJ family members harbour an unstructured N-terminal domain. It was shown that the DNA binding specificity of the Arc and Mnt transcriptional repressors is determined by residues located in the short unstructured N-terminal domain (Knight and Sauer, 1989). For ParG, a ParB structural homologue and whose structure is known, it was shown that the mobile N-terminal tail increases the specificity of interaction with the DNA. In addition, a transient β-strand was observed in this region (Carmelo *et al.*, 2005).

Alanine scanning mutagenesis of some residues in the N-terminal unstructured domain has shown that changing Lys-3 to alanine significantly affected plasmid stability. In solution NMR, it was noted that Lys-3 experienced medium exchange in chemical shift when DNA is added at low concentration. These results suggest that Lys-3 may have an essential role in plasmid partition system and specifically it may be involved in DNA interaction as shown by NMR. Substituting Lys-4 with alanine was also shown to affect plasmid stability, decreasing it by 27%. These results suggest that the two lysine

residues, Lys-3 and Lys-4, in the tip of the tail of the ParB unstructured N-terminal domain have an important role in plasmid segregation that involves interaction with the DNA. In contrast, changing Asp-11 to alanine had no effect on plasmid stability, whereas, changing arginine to alanine at position 12 and 26 had moderately affected plasmid stability. In NMR ParB-DNA interaction studies, it was shown that the cross-peaks corresponding to Arg-26 shifted significantly, when the DNA was present, and it was assumed that this region may undergo a conformational transition, when the protein binds to DNA. However, these results are not conclusive as further *in vitro* studies need to be carried out to investigate the importance of residues in the N-terminal domain of ParB. Due to time constraints, it was not possible to further investigate these mutants.

As ParB shares the same C-terminal RHH folds with the Arc/MetJ family, it is believed that the β-strands of C-terminal domains are inserted the major groove of the DNA in a fashion similar to that of other members of this family whose structure has been solved in complex with DNA (Schreiter and Drennan, 2007). Through NMR experiments, it was predicted that the middle region of the ParB unstructured N-terminal domain forms a transient α-helical structure. Moreover, the NMR results have shown that ParB binds a single hexameric DNA repeat through the C-terminal domain. Moreover, MST and EMSA experiments have shown high-affinity and specific interaction between ParB and the centromere sites. In addition, the NMR experiments have shown that adding the cognate DNA site to ParB causes significant chemical shift changes in the unstructured N-terminal domain. Through NMR studies, we have identified two regions involved in DNA interactions in ParB N-terminal domain: the tip of the tail including amino acids Lys3-Ala10 that directly interacts with DNA and the middle region that may either experience conformational changes due to disordered-to-order transition or become less mobile in solution due to the interaction of the two domains with DNA. Based on all the observations, a model for ParB-*parC* interaction is proposed (Figure 6.2).

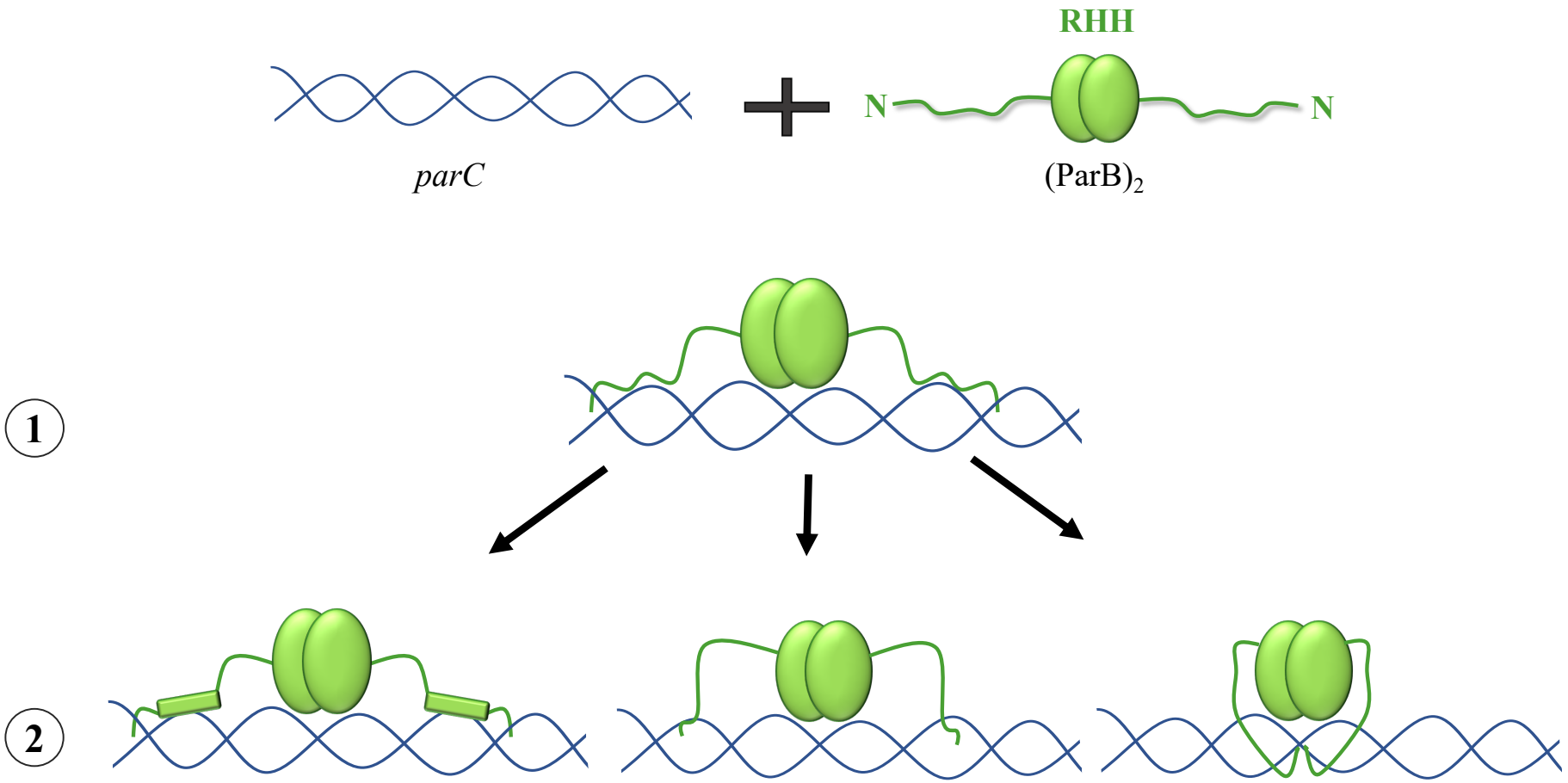


Figure 6.2 A model for segrosome assembly

ParB folds into a RHH C-terminal structure, while the N-terminal domain is unstructured. ParB binds as a dimer to a single repeat through the RHH domain and the unstructured N-terminal domain as shown in step **(1)**. Once the unstructured N-terminal domain binds through the tip of the tail, the middle region may either undergo conformational changes becoming a stable α-helix structure or become less mobile in solution when the two domains bind to DNA, as shown in step **(2)**.

6.1.4 Elucidating the role played by *parC* and ATP in ParA-ParB complex formation

Interactions among segregation system components are crucial for plasmid partitioning events. *In vivo* experiments showed that ParA oscillates from one nucleoid pole to the other in a spiral-like structure in the presence of ParB-*parC* complex (Ebersbach and Gerdes, 2004). Chemical cross-linking experiments have captured interactions between a ParB monomer and a ParA monomer. The interaction is enhanced by the presence of the centromere site and ATP. ParA was shown to polymerise in the presence of ATP and forms a bundle of filaments *in vitro* (Ebersbach *et al.*, 2006). Chemical cross-linking with DMP showed that *parC1* and ATP are essential for segrosome assembly as adding these two elements in the cross-linking solution enhanced the interaction of ParB with ParA. On the other hands, BS3 cross-linking experiments revealed that the interaction between the two segregation proteins are rather weak and enhanced, when the centromere *parC1* site was added to the reaction. However, the ParA-ParB interaction was more pronounced in the presence of both *parC1* and ATP. Thus, the formation of the ParA-ParB complex is ATP and *parC* dependent. This result suggests that binding of ParB to the centromere sites is a prerequisite step for ParA-ParB complex formation. The interface used by ParB to bind to ParA was also identified using BS3 cross-linking coupled with liquid chromatography-mass spectroscopy (LC/MS). The LC/MS experiment revealed that no cross-linked peptides were recovered, when the two proteins were cross-linked alone or even in the presence of *parC1*. However, the number of cross-linked peptides identified increased by ~90% in the presence of *parC1* and ATP. These results suggest that segrosome assembly requires the simultaneous presence in the reaction of all the components involved in plasmid partitioning systems.

As described previously, NMR studies have shown that ParB binds site-specifically to the centromere sites through the C-terminal RHH fold and the tip of the tail of the N-terminal domain. Moreover, the transient α-helix structure is predicted to undergo conformational changes and probably develop a more stable α-helix structure. LC/MS identified cross-linked peptides in the ParA-ParB complex and showed that the unstructured N-terminal domain of ParB interacts with ParA. The cross-linking results showed that Lys-3 and Lys-4 in the N-terminal domain of ParB interacted with Lys-28 and Lys-179 in ParA, respectively. Interestingly, Lys-3 and Lys-4 are predicted to interact with DNA and changing these two amino acids into alanine has also significantly affected the plasmid stability. These results suggest that the unstructured N-terminal domain of ParB plays two important roles, i.e. binding to DNA and binding to its cognate partner ParA. For ParG, a ParB structural homologue, it was shown that Arg-19 in the

unstructured N-terminal tail enhances the weak ATPase activity of its partner ParF (Barilla *et al.*, 2007). Co-crystallisation of ParF with the N-terminal flexible domain of ParG showed that Arg-19 is inserted into the ParF dimer interface and folds into a α-helix structure (Zhang and Schumacher, 2017). As reported previously, the N-terminal tail in ParG is essential in the formation of higher-order nucleoprotein structure for regulation of the *parFG* genes (Carmelo *et al.*, 2005).

Additionally, two more cross-linked peptides were observed that linked ParB molecules. However, it is unclear whether the identified amino acids are between ParB molecules or within a ParB molecule. The LC/MS result showed that Lys-3 cross-linked with Lys-30 and Lys-30 with Lys-66. This might indicate that the flexible N-terminal domain of ParB folds in close proximity to the RHH C-terminal domain. However, it is difficult to draw this as a definite conclusion, as only one repeat was analysed, and further experiments are needed. The two distinct roles of the unstructured N-terminal domain suggest that ParB might experience multiple changes depending on the function it performs. Specifically, the N-terminal domain of ParB perhaps requires binding to DNA and shaping into a defined conformation that allows ParA to bind easily and more efficiently to a fully folded structure of ParB rather than a flexible and more dynamic domain. As a result, the unstructured N-terminal domain switches between the two functions binding to DNA and stabilising its structure to facilitate ParA binding.

6.2 Future work

The pB171 plasmid is a large virulence plasmid and is maintained in *E. coli* by two types of segregation systems, a type Ib and a type II systems (Tobe *et al.*, 1999; Ebersbach and Gerdes, 2001). Type II systems have been studied comprehensively in the last two decades and the mechanisms underpinning plasmid partition have been described in detail. However, little is known about the segregation of type Ib systems. Molecular interactions among plasmid segregation components is an interesting area to explore and to shed light on the mechanism responsible for plasmid segregation.

During this project, the importance of the two pB171 centromere sites has been studied *in vivo*. It was demonstrated that the *parC2* site is crucial for plasmid stability, while the *parC1* could act as an auxiliary centromere. Two centromeres *parC1* and *parC2* are complex sites due to the presence of the two classes of repeats with different pattern of arrangement and organisation. We speculate that there are unknown factors that contribute to the plasmid partition process by bending DNA for more efficient centromere activity. Moreover, the presence of two IHF binding sites, *ihf1* and *ihf*2, within the *parABC*

cassette and our preliminary data show that the IHF protein may be involved in plasmid segregation by bending the DNA in a way similar to that observed for the partition complex of the P1 plasmid. Further investigations are necessary to confirm some of the conclusions drawn from this work. It would be useful to investigate further the involvement of the IHF protein in plasmid segregation by using electrophoretic mobility shift assays to assess the binding of IHF to ParB-*parC* complex. In addition, Atomic Force Microscopy (AFM) could be used to test whether the protein bends the DNA that harbours the two centromere sites.

The 3D structure data available for the C-terminal domain and the NMR secondary structure prediction of the N-terminal domain provide essential information for further investigation of ParB in plasmid partitioning. The new structure provides a strong base for further investigation of ParB-DNA and ParB-ParA interactions. The backbone assignment of ParB using NMR experiments allows us to identify the binding sites in ParB protein. Therefore, this information could also be used to identify in the ParB binding site for ParA as well. An initial trial has been made to recognise the binding interface between the two segregation proteins. However, due to time constraints and rapid ParA precipitation in the NMR buffer, it was not possible to continue with this experiment. It would be useful to resume what we have started and to optimise the NMR buffer to maintain ParA in solution during the experiments. The 2D [^{1}H, ^{15}N] HMQC NMR experiment is a useful tool to identify the interface between the two segregation proteins. Furthermore, chemical cross-linking coupled with LC/MS allowed the determination of the ParB region interacting with ParA. The tip end of the N-terminal unstructured domain is involved in the interaction with ParA. However, due to lack of time and technical error, this conclusion is not definite and new experiments should be performed. Alanine scanning mutagenesis showed that Lys-3 and Lys-4 at the tip end of the ParB N-terminus play a crucial role in plasmid stability in the cell. Therefore, it would be interesting to investigate further the role played by these two residues in the interaction with ParA by using the ATPase assays as well as chemical cross-linking to identify the influence of these residues on segrosome assembly.

Due to the flexibility of the ParB N-terminal domain, it was difficult to crystallise the unstructured region. Moreover, we have shown that the transient α-helical structure that is located in the middle region of the N-terminal domain may undergo conformational changes when it binds DNA. Therefore, further research is needed to confirm this interesting finding by co-crystallising ParB with its cognate DNA site. A DNA fragment that harbours two to four *parC* repeats could be used for structure calculation and might

prove an important area for future investigation. It would be interesting to use X-ray crystallography to determine the changes occurring when the protein is in complex with DNA and also identifying the map of interactions. It will be also interesting to identify the dimer of dimer interfaces, when ParB is incubated with an extended DNA fragment containing multiple repeats.

Abbreviations

2D – Two-dimensional

3D – Three-dimensional

˚C – Degrees Celsius

µg – Microgram

µl – Microlitres

µM – Micromolar

ml – Millilitre

mM – Millimolar

mg – Milligrams

ng – nanograms

ps – picoseconds

ns – nanosecond

pmol – picomole

α – Alpha

β – Beta

Å – Angstrom

A – Absorbance

ATP – Adenosine triphosphate

ADP – Adenosine diphosphate

ATPase – Adenosine triphosphate hydrolase

AP – Alkaline phosphatase

APS – Ammonium persulphate

Amp – Ampicillin

AMPPCP – phosphomethylphosphonic acid adenylate ester

AFM – Atomic Force Microscopy

Bp – Base pair

BFP – Bundle Forming Pilli

BSA – Bovine serum albumin

CBP – Centromere-binding protein

Cm – Chloramphenicol

CL-MS – Cross-linking-mass Spectrometry

CV – Column Volume

Cy5 – Cyanine dye

DMP – Dimethyl pimelimidate

DNA – Deoxyribonucleic acid

dNTP – Dixoyribonulceoside-5'-triphosphate

dsDNA – Double-stranded deoxyribonucleic acid

DTT – Dithiothreitol

EDTA – Ethylenediaminetetraacetic acid

EGTA – Ethylene Glycol Tetra-acetic Acid

EMSA – Electrophoretic Mobility Shift Assay

EPEC – Enteropathogenic *Escherichia coli*

EAF – Enteropathogenic Adherence Factor

Fnorm – normalized fluorescence

GTP – Guanosine triphosphate

GTPase – Guanosine triphosphate hydrolase

HADAMAC – HADamard-encoded Amino-Acid-type editing

His – Histidine

HSQC – Heteronuclear Single Quantum Coherence

HTH – Helix-Turn-Helix motif

IHF – Integration Host Factor

IPTG – Isopropyl beta-D-1-thiogalctopyranoside

IR – Infra-Red

Kb – kilobase pairs

K_D – Equilibrium dissociation constant

kDa – Kilodaltons

L – Litre

LB – Luria-Bertani medium

LC-MS – Liquid Chromatography-Mass Spectrometry

LC-MS/MS – Liquid Chromatography-tandem Mass Spectrometry

LED – Light emitting diode

MST – Microscale Thermophoresis

M – Molar

MW – Molecular Weight

MWCO – Molecular Weight Cut Off

NMR – Nuclear Magnetic Resonance

OD – Optical Density

P – Promoter

PBS – Phosphate Buffered Saline

PCR – Polymerase Chain Reaction

PDB – Protein Data Bank

Poly (dI-dC) – Poly(deoxyinosinic-deoxycytidylic) acid

RHH – Ribbon-Helix-Helix motif

RMSD – Root Mean Square Deviation

SDS – Sodium Dodecyl Sulphate

SDS-PAGE – Sodium-Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis

SEC – Size Exclusion Chromatography

SEC-MALLS – Size exclusion Chromatography coupled to Multi Angle Laser Light Scattering

SOFAST-HMQC – Selective optimised flip-angle short-transient-Heteronuclear Multiple Quantum Coherence

ssDNA – Single-stranded DNA

TA – Toxin-Antitoxin

TEMED – N,N,N',N'-Tetramethylethylene-1,2-diamine

U – Enzyme Unit

UV – Ultraviolet

V – Volt

w/v – Weight for volume

References

Actis, L. A., Tolmasky, M. E. and Crosa, J. H. (1999). Bacterial Plasmid: Replication of Extrachromosomal Genetic Elements Encoding Resistance to Antimicrobial Compounds. *Frontiers in Bioscience*, 4(3), pp. 43–62

Ah-seng, Y., Lopez, F., Pasta, F., Lane, D. and Bouet, J. (2009). Dual Role of DNA in Regulating ATP Hydrolysis by the SopA Partition Protein. *Journal of Biological Chemistry*, 284(44), pp. 30067–30075

Augustus, A. M., Reardon, P. N., Heller, W. T. and Spicer, L. D. (2006). Structural Basis for the Differential Regulation of DNA by the Methionine Repressor MetJ. *Journal of Biological Chemistry*, 281(45), pp. 34269–34276

Austin, S. and Abeles, A. (1983). Partition of unit-copy miniplasmids to daughter cells. I. P1 and F miniplasmids contain discrete, interchangeable sequences sufficient to promote equipartition. *Journal of Molecular Biology*, 169(2), pp. 353–372

Aylett, C. H. S., Wang, Q., Michie, K. A., Amos, L. A. and Lowe, J. (2010). Filament structure of bacterial tubulin homologue TubZ. *Proceedings of the National Academy of Sciences USA*, 107(46), pp. 19766–19771

Barillà, D., Carmelo, E. and Hayes, F. (2007). The tail of the ParG DNA segregation protein remodels ParF polymers and enhances ATP hydrolysis via an arginine finger-like motif. *Proceedings of the National Academy of Sciences USA*, 104(6), pp. 1811–1816

Barillà, D. and Hayes, F. (2003). Architecture of the ParF•ParG protein complex involved in prokaryotic DNA segregation. *Molecular Microbiology*, 49(2), pp. 487–499

Barillà, D., Rosenberg, M. F., Nobbmann, U. and Hayes, F. (2005). Bacterial DNA segregation dynamics mediated by the polymerizing protein ParF. *The EMBO Journal*, 24(7), pp. 1453–1464

Baxter, J. C. and Funnell, B. E. (2014). Plasmid Partition Mechanisms. *Microbiology Spectrum*, 2(6), pp. 1–20

Bharat, T. A. M., Murshudov, G. N., Sachse, C. and Löwe, J. (2015). Structures of actin-like ParM filaments show architecture of plasmid-segregating spindles. *Nature*, 523(7558), pp. 106–110

Bieber, D., Ramer, S. W., Wu, C. Y., Murray, W. J., Tobe, T., Fernandez, R. and Schoolnik, G. K. (1998). Type IV pili, transient bacterial aggregates, and virulence of enteropathogenic *Escherichia coli*. *Science*, 280(5372), pp. 2114–2118

Biek, D. P. and Shi, J. (1994). A single 43bp *sopC* repeat of plasmid mini-F is sufficient to allow assembly of a functional nucleoprotein partition complex. *Proceedings of the National Academy of Sciences USA*, 91(17), pp. 8027–8031

Biek, D. R. and Strings, J. (1995). Partition Functions of Mini-F Affect Plasmid DNA Topology in *Escherichia coil*. *Journal of Molecular Biology*, 246(3), pp. 388–400

Bouet, J.-Y. and Funnell, B. E. (1999). P1 ParA interacts with the P1 partition complex at *parS* and an ATP-ADP switch controls ParA activities. *The EMBO Journal*, 18(5), pp. 1415–1424

Bouet, J.-Y. and Funnell, B. E. (2019). Plasmid Localization and Partition in Enterobacteriaceae. *EcoSal Plus*, 8(2), pp. 1–23

Bouet, J., Ah-seng, Y., Benmeradi, N. and Lane, D. (2007). Polymerization of SopA partition ATPase : regulation by DNA binding and SopB. *Molecular Microbiology*, 63(2), pp. 468–481

Bouet, J. and Lane, D. (2009). Molecular Basis of the Supercoil Deficit Induced by the Mini-F Plasmid Partition Complex. *Journal of Biological Chemistry*, 284(1), pp. 165–173

Bracken, C., Marley, J. and Lu, M. (2001). A method for efficient isotopic labeling of recombinant proteins. *Journal of Biomolecular NMR*, 20(1), pp. 71–75

Breüner, A., Jensen, R. B., Dam, M., Pedersen, S. and Gerdes, K. (1996). The centromere-like *parC* locus of plasmid R1. *Molecular Microbiology*, 20(3), pp. 581–592

Carmelo, E., Barillà, D., Golovanov, A. P., Lian, L. Y., Derome, A. and Hayes, F. (2005). The unstructured N-terminal tail of ParG modulates assembly of a quaternary nucleoprotein complex in transcription repression. *Journal of Biological Chemistry*, 280(31), pp. 28683–28691

Castaing, J., Bouet, J. and Lane, D. (2008). F plasmid partition depends on interaction of SopA with non-specific DNA. *Molecular Microbiology*, 70(4), pp. 1000–1011

Choi, C. L., Claridge, S. A., Garner, E. C., Alivisatos, A. P. and Mullins, R. D. (2008). Protein-nanocrystal conjugates support a single filament polymerization model in R1 plasmid segregation. *Journal of Biological Chemistry*, 283(42), pp. 28081–28086

Clowes, R. C. (1972). Molecular Structure of Bacterial Plasmids. *Bacteriological Review*, 36(3), pp. 361–405

Dam, M. and Gerdes, K. (1994). Partitioning of plasmid R1. Ten direct repeats flanking the *parA* promoter constitute a centromere-like partition site *parC*, that expresses incompatibility. *Journal of Molecular Biology*, 236(5), pp. 1289–1298

Davis, M. A. and Austin, S. J. (1988). Recognition of the P1 plasmid centromere analog involves binding of the ParB protein and is modified by a specific host factor. *The EMBO Journal*, 7(6), pp. 1881–1888

Davis, M. A., Radnedge, L., Martin, K. A., Hayes, F., Youngren, B. and Austin, S. J. (1996). The P1 ParA protein and its ATPase activity play a direct role in the segregation of plasmid copies to daughter cells. *Molecular Microbiology*, 21(5), pp. 1029–1036

DeLano, W. (2002). Pymol: An open-source molecular graphics tool. *DeLano Scientific*

San Carlos, California USA

Delbruck, H., Ziegelin, G., Lanka, E. and Heinemann, U. (2002). An Src Homology 3-like Domain Is Responsible for Dimerization of the Repressor Protein KorB Encoded by the Promiscuous IncP Plasmid RP4. *Journal of Biological Chemistry*, 277(6), pp. 4191–4198

Díaz-orejas, R., Espinosa, M. and Yeo, C. C. (2017). The Importance of the Expendable : Toxin – Antitoxin Genes in Plasmids and Chromosomes. *Frontiers in Microbiology*, 8, pp. 1–7

Dmowski, M. and Jaguara-Burdzy, G. (2013). Active Stable Maintenance Functions in Low Copy-Number Plasmids of Gram-Positive Bacteria: I . Partition Systems. *Polish Journal of Microbiology*, 62(1), pp. 3–16

Dmowski, M. and Jagura-Burdzy, G. (2011). Mapping of the interactions between partition proteins Delta and Omega of plasmid pSM19035 from *Streptococcus pyogenes*. *Microbiology*, 157(4), pp. 1009–1020

Dmowski, M. and Kern-zdanowicz, I. (2016). Omega (ParB) binding sites together with the RNA polymerase-recognized sequence are essential for centromeric functions of the P ! region in the partition system of pSM19035. *Microbiology*, 162(7), pp. 1114–1124

Dunham, T. D., Xu, W., Funnell, B. E. and Schumacher, M. A. (2009). Structural basis for ADP-mediated transcriptional regulation by P1 and P7 ParA. *The EMBO Journal*, 28(12), pp. 1792–1802

Ebersbach, G. and Gerdes, K. (2001). The double par locus of virulence factor pB171 : DNA segregation is correlated with oscillation of ParA. *Proceedings of the National Academy of Sciences USA*, 98(26), pp. 15078–15083

Ebersbach, G. and Gerdes, K. (2004). Bacterial mitosis : partitioning protein ParA oscillates in spiral-shaped structures and positions plasmids at mid-cell. *Molecular Microbiology*, 52(2), pp. 385–398

Ebersbach, G. and Gerdes, K. (2005). Plasmid Segregation Mechanisms. *Annual Review Genetic*, 39, pp. 453–479

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

Ebersbach, G., Sherratt, D. J. and Gerdes, K. (2005). Partition-associated incompatibility caused by random assortment of pure plasmid clusters. *Molecular Microbiology*, 56(6), pp. 1430–1440

Edgar, R., Chattoraj, D. K. and Yarmolinsky, M. (2001). Pairing of P1 plasmid partition sites by ParB. *Molecular Microbiology*, 42(5), pp. 1363–1370

Eliasson, A., Bernander, R., Dasgupta, S. and Nordstrom, K. (1992). Direct visualization of plasmid DNA in bacterial cells. *Molecular Microbiology*, 6(2), pp. 165–170

van den Ent, F., Møller-Jensen, J., Amos, L. A., Gerdes, K. and Löwe, J. (2002). F-actin-like filaments formed by plasmid segregation protein ParM. *The EMBO Journal*, 21(24), pp. 6935–6943

Erdmann, N., Petroff, T. and Funnell, B. E. (1999). Intracellular localization of P1 ParB protein depends on ParA and *parS*. *Proceedings of the National Academy of Sciences USA*, 96(26), pp. 14905–14910

Ferrage, F., Reichel, A., Battacharya, S., Cowburn, D. and Ghose, R. (2010). Effects of the saturation scheme and water signal suppression. *Journal Magnetic Resonance*, 207(2), pp. 294–303

Fink, G. and Löwe, J. (2015). Reconstitution of a prokaryotic minus end-tracking system using TubRC centromeric complexes and tubulin-like protein TubZ filaments. *Proceedings of the National Academy of Sciences USA*, 112(15), pp. E1845–E1850

Fothergill, T. J. G., Barilla, D. and Hayes, F. (2005). Protein Diversity Confers Specificity in Plasmid Segregation. *Journal of Bacteriology*, 187(8), pp. 2651–2661

Fuentes-Pérez, M. E., Núñez-Ramírez, R., Martín-González, A., Juan-Rodríguez, D., Llorca, O., Moreno-Herrero, F. and Oliva, M. A. (2017). TubZ filament assembly dynamics requires the flexible C-terminal tail. *Scientific Reports*, 7, pp. 1–10

Funnell, B. E. (1988a). Mini-P1 plasmid partitioning: excess ParB protein destabilizes plasmids containing the centromere *parS*. *Journal of Bacteriology*, 170(2), pp. 954–960

Funnell, B. E. (1988b). Participation of *Escherichia coli* integration host factor in the P1 plasmid partition system. *Proceedings of the National Academy of Sciences USA*, 85(18), pp. 6657–6661

Funnell, B. E. (2016). ParB Partition Proteins: Complex Formation and Spreading at Bacterial and Plasmid Centromeres. *Frontiers in Molecular Biosciences*, 3(44), pp. 1–6

Funnell, B. E. and Gagnier, L. (1993). The P1 Plasmid Partition Complex at *parS*. *Journal of Biological Chemistry*, 268(5), pp. 3616–3624

Galkin, V., Oriova, A., Rivera, C., Mullins, R. D. and Egelman, E. H. (2009). Structural Polymorphism of the ParM Filament and Dynamic Instability. *Structure*, 17(9), pp. 1253–1264

Le Gall, A., Cattoni, D. I., Guihas, B., Mathieu-Demaziere, C., Oudjedi, L., Fiche, J.-B., Rech, J., Abrahamsson, S., Murray, H., Bouet, J. and Nollmann, M. (2016). Bacterial partition complexes segregate within the volume of the nucleoid. *Nature Communications*, 7(5), pp. 1–10

Garner, E. C., Campbell, C. S., Weibel, D. B. and Mullins, D. R. (2007). Reconstitution

of DNA Segregation Driven by Assembly of a Prokaryotic actin homolog. *Science*, 315(5816), pp. 1270–1274

Gayathri, P., Fujii, T., Moller-Jensen, J., Ent, F. Van Den, Namba, K. and Löwe, J. (2012). A Bipolar Spindle Of Antiparallel ParM Filaments Drives Bacterial Plasmid Segregation. *Science*, 338(6112), pp. 1334–1337

Gayathri, P., Fujii, T., Namba, K. and Löwe, J. (2013). Structure of the ParM filament at 8.5Å resolution. *Journal of Structural Biology*, 184(1), pp. 33–42

Gerdes, K. and Molin, S. (1986). Partitioning of plasmid R1. Structural and functional analysis of the parA locus. *Journal of Molecular Biology*, 190(3), pp. 269–279

Gerdes, K., Rasmussen, P. B. and Molin, S. (1986). Unique type of plasmid maintenance function : Postsegregational killing of plasmid-free cells. *Proceedings of the National Academy of Sciences USA*, 83(10), pp. 3116–3120

Goeders, N. and Melderen, L. Van (2014). Toxin-Antitoxin Systems as Multilevel Interaction Systems. *Toxins*, 6(1), pp. 304–324

Golovanov, A. P., Barillà, D., Golovanova, M., Hayes, F. and Lian, L. Y. (2003). ParG, a protein required for active partition of bacterial plasmids, has a dimeric ribbon-helix-helix structure. *Molecular Microbiology*, 50(4), pp. 1141–1153

Gomis-ruth, F. X., Sola, M., Acebo, P., Parraga, A., Guasch, A., Eritija, R., Gonzalez, A., Espinosa, M., el Solar, G. and Coll, M. (1998). The structure of plasmid-encoded transcriptional repressor CopG unliganded and bound to its operator. *The EMBO Journal*, 17(24), pp. 7404–7415

Hall, A. M. J., Gollan, B. and Helaine, S. (2017). Toxin – antitoxin systems : reversible toxicity. *Current opinion in Microbiology*, 36(4), pp. 102–110

Hanai, R., Liu, R., Benedetti, P., Caron, P. R., Lynch, A. S. and Wang, J. C. (1996). Molecular Dissection of a Protein SopB Essential for *Escherichia coli* F Plasmid Partition. *Journal of Biological Chemistry*, 271(29), pp. 17469–17475

Hatano, T., Yamaichi, Y. and Niki, H. (2007). Oscillating focus of SopA associated with filamentous structure guides partitioning of F plasmid. *Molecular Microbiology*, 64(5), pp. 1198–1213

Havey, J. C., Vecchiarelli, A. G. and Funnell, B. E. (2012). ATP-regulated interactions between P1 ParA, ParB and non-specific DNA that are stabilized by the plasmid partition site, *parS*. *Nucleic Acids Research*, 40(2), pp. 801–812

Hayes, F. (1998). A family of stability determinants in pathogenic bacteria. *Journal of Bacteriology*, 180(23), pp. 6415–6418

Hayes, F. (2000). The partition system of multidrug resistance plasmid TP228 includes a novel protein that epitomizes an evolutionarily distinct subgroup of the ParA superfamily. *Molecular Microbiology*, 37(3), pp. 528–541

Hayes, F. (2003). Toxins-Antitoxins : Plasmid Maintenance , Programmed Cell Death , and Cell Cycle Arrest. *Science*, 301(5639), pp. 1496–1500

Hayes, F. (2014). Regulating Toxin-Antitoxin Expression: Controlled Detonation of Intracellular Molecular Timebombs. *Toxins*, 6(1), pp. 337–358

Hayes, F. and Austin, S. (1994). Topological Scanning of the P1 Plasmid Partition Site. *Journal of Molecular Biology*, 243(2), pp. 190–198

Hayes, F. and Barilla, D. (2006a). Assembling the bacterial segrosome. *Trends in Biochemical Science*, 31(5), pp. 247–250

Hayes, F. and Barillà, D. (2006b). The bacterial segrosome : a dynamic nucleoprotein machine for DNA trafficking and segregation. *Nature Reviews Microbiology*, 4(2), pp. 133–143

Hayes, F. and Kedzierska, B. (2014). Regulating toxin-antitoxin expression: Controlled detonation of intracellular molecular timebombs. *Toxins*, 6(1), pp. 337–358

Hester, C. M. and Lutkenhaus, J. (2007). Soj (ParA) DNA binding is mediated by conserved arginines and is essential for plasmid segregation. *Proceedings of the National Academy of Sciences USA*, 104(51), pp. 20326–20331

Hiraga, S. (2000). Dynamic Localization Of Bacterial and Plasmid Chromosomes. *Annual Review Genetic*, 34, pp. 21–59

Hirano, M., Mori, H., Onogi, T., Yamazoe, M., Niki, H., Ogura, T. and Hiraga, S. (1998). Autoregulation of the partition genes of the mini-F plasmid and the intracellular localization of their products in *Escherichia coli*. *Molecular Gene Genetics*, 257(4), pp. 392–403

Hoischen, C., Bussiek, M., Langowski, J. and Diekmann, S. (2008). *Escherichia coli* low-copy-number plasmid R1 centromere *parC* forms a U-shaped complex with its binding protein ParR. *Nucleic Acids Research*, 36(2), pp. 607–615

Hu, L., Vecchiarelli, A. G., Mizuuchi, K., Neuman, K. C. and Liu, J. (2017a). Brownian Ratchet Mechanism for Faithful Segregation of Low-Copy-Number Plasmids. *Biophysical Journal*, 112(7), pp. 1489–1502

Hu, L., Vecchiarelli, A. G., Mizuuchi, K., Neuman, K. C. and Liu, J. (2017b). Brownian ratchet mechanisms of ParA-mediated partitioning. *Plasmid*, 92, pp. 12–16

Huang, L., Yin, P., Zhu, X., Zhang, Y. and Ye, K. (2011). Crystal structure and centromere binding of the plasmid segregation protein ParB from pCXC100. *Nucleic Acids Research*, 39(7), pp. 2954–2968

Hwang, L. C., Vecchiarelli, Anthony G, Han, Y., Mizuuchi, M., Harada, Y., Funnell, B. E. and Mizuuchi, K. (2013). ParA-mediated plasmid partition driven by protein pattern self-organization. *The EMBO Journal*, 32(9), pp. 1238–1249

Ietswaart, R., Szardenings, F., Gerdes, K. and Howard, M. (2014). Competing ParA

Structures Space Bacterial Plasmids Equally over the Nucleoid. *PLOS Computational Biology*, 10(12), pp. 1–20

Ikura, M., Kay, L. E. and Bax, A. (1990). A Novel Approach for Sequential Assignment of ^{1}H, ^{13}C, and ^{15}N Spectra of Larger Proteins: Heteronuclear Triple-Resonance Three-Dimensional NMR Spectroscopy. Application to Calmodulint. *Biochemistry*, 29(19), pp. 4659–4667

Jaffe, A., Ogura, T. and Hiraga, S. (1985). Effects of the ccd Function of the F Plasmid on Bacterial Growth. *Journal of Bacteriology*, 163(3), pp. 841–849

Jensen, R. B., Dam, M. and Gerdes, K. (1994). Partitioning of plasmid R1 The *parA* operon is autoregulated by ParR and its transcription is highly stimulated by a downstream activating element. *Journal of Molecular Biology*, 236(5), pp. 1299–1309

Jensen, R. B. and Gerdes, K. (1995). Programmed cell death in bacteria : proteic plasmid stabilization systems. *Molecular Microbiology*, 17(2), pp. 205–210

Jensen, R. B. and Gerdes, K. (1997). Partitioning of plasmid R1. The ParM protein exhibits ATPase activity and interacts with the centromere-like ParR-*parC* complex. *Journal of Molecular Biology*, 269(4), pp. 505–513

Jensen, R. B., Lurz, R. and Gerdes, K. (1998). Mechanism of DNA segregation in prokaryotes: Replicon pairing by *parC* of plasmid R1. *Proceedings of the National Academy of Sciences USA*, 95(15), pp. 8550–8555

Jerabek-Willemsen, M., André, T., Wanner, R., Roth, H. M., Duhr, S., Baaske, P. and Breitsprecher, D. (2014). MicroScale Thermophoresis: Interaction analysis and beyond. *Journal of Molecular Structure*, 1077, pp. 101–113

Jerabek-Willemsen, M., Wienken, C. J., Braun, D., Baaske, P. and Duhr, S. (2011). Molecular interaction studies using microscale thermophoresis. *Assay and Drug Development Technologies*, 9(4), pp. 342–353

Kay, L. E., Keifer, P. and Saarinen, T. (1992). Pure Absorption Gradient Enhanced Heteronuclear Single Quantum Correlation Spectroscopy with Improved Sensitivity. *Journal of the American Chemical Society*, 114(26), pp. 10663–10665

Kedzierska, B. and Hayes, F. (2016). Emerging Roles of Toxin-Antitoxin Modules in Bacterial Pathogenesis. *Molecules*, 21(790), pp. 1–25

Khare, D., Ziegelin, G., Lanka, E. and Heinemann, U. (2004). Sequence-specific DNA binding determined by contacts outside the helix-turn-helix motif of the ParB homolog KorB. *Nature Structural & Molecular Biology*, 11(7), pp. 656–663

Kim, S. and Shim, J. (1999). Interaction between F Plasmid Partition Proteins SopA and SopB. *Biochemical and Biophysical Research Communications*, 263(1), pp. 113–117

Knight, K. L. and Sauer, R. T. (1989). DNA binding specificity of the Arc and Mnt repressors is determined by a short region of N-terminal residues. *Proceedings of the*

National Academy of Sciences USA, 86(3), pp. 797–801

Kusukawa, N., Mori, H., Kondo, A. and Hiraga, S. (1987). Partitioning of the F plasmid: Overproduction of an essential protein for partition inhibits plasmid maintenance. *Molecular & General Genetics*, 208(3), pp. 365–372

de La Hoz, A. B., Ayora, S., Sitkiewicz, I., Fernandez, S., Pankiewicz, R., Alonso, J. C. and Ceglowski, P. (2000). Plasmid copy-number control and better-than-random segregation genes of pSM19035 share a common regulator. *Proceedings of the National Academy of Sciences USA*, 97(2), pp. 728–733

de La Hoz, A. B., Pratto, F., Misselwitz, R., Speck, C., Weihofen, W., Welfle, K., Saenger, W., Welfle, H. and Alonso, J. C. (2004). Recognition of DNA by ω protein from the broad-host range *Streptococcus pyogenes* plasmid pSM19035: Analysis of binding to operator DNA with one to four heptad repeats. *Nucleic Acids Research*, 32(10), pp. 3136–3147

Larsen, R. A., Cusumano, C., Fujioka, A., Lim-Fong, G., Patterson, P. and Pogliano, J. (2007). Treadmilling of a prokaryotic tubulin-like protein, TubZ, required for plasmid stability in *Bacillus thuringiensis*. *Genes and Development*, 21(11), pp. 1340–1352

Leonard, T. A., Butler, P. J. and Lowe, J. (2005). Bacterial chromosome segregation : structure and DNA binding of the Soj dimer — a conserved biological switch. *The EMBO Journal*, 24(2), pp. 270–282

Lescop, E. and Brutscher, B. (2009). Highly automated protein backbone resonance assignment within a few hours: The «BATCH» strategy and software package. *Journal of Biomolecular NMR*, 44(1), pp. 43–57

Lescop, E., Rasia, R. and Brutscher, B. (2008). Hadamard amino-acid-type edited NMR experiment for fast protein resonance assignment. *Journal of the American Chemical Society*, 130(15), pp. 5014–5015

Libante, V., Thion, L. and Lane, D. (2001). Role of the ATP-binding Site of SopA Protein in Partition of the F Plasmid. *Journal of Molecular Biology*, 314(3), pp. 387–399

Lim, G. E., Derman, A. I. and Pogliano, J. (2005). Bacterial DNA segregation by dynamic SopA polymers. *Proceedings of the National Academy of Sciences USA*, 102(49), pp. 17658–17663

Lim, H. C., Surovtsev, I. V., Beltran, B. G., Huang, F., Bewersdorf, J. and Jacobs-wagner, C. (2014). Evidence for a DNA-relay mechanism in ParABS-mediated chromosome segregation. *elife*, 23(5), pp. 1–32

Ludtke, D. N., Eichorn, B. G. and Austin, S. J. (1989). Plasmid-partition functions of the P7 prophage. *Journal of Molecular Biology*, 209(3), pp. 393–406

Lum, P. L. and Schildbach, J. F. (1999). Specific DNA Recognition by F Factor TraY Involves β-Sheet Residues. *Journal of Biological Chemistry*, 274(28), pp. 19644–19648

Machón, C., Fothergill, T. J. G., Barillà, D. and Hayes, F. (2007). Promiscuous Stimulation of ParF Protein Polymerization by Heterogeneous Centromere Binding Factors. *Journal of Molecular Biology*, 374(1), pp. 1–8

Marion, D. (2013). An introduction to biological NMR spectroscopy. *Molecular and Cellular Proteomics*, 12(11), pp. 3006–3025

Martin, K. A., Davis, M. A. and Austin, S. (1991). Fine-structure analysis of the P1 plasmid partition site. *Journal of Bacteriology*, 173(12), pp. 3630–3634

McGuffin, L. J., Bryson, K. and Jones, D. T. (2000). The PSIPRED protein structure prediction server. *Bioinformatics*, 16(4), pp. 404–405

McLeod, B. N., Allison-Gamble, G. E., Barge, M. T., Tonthat, N. K., Schumacher, M. A., Hayes, F. and Barillà, D. (2017). A three-dimensional ParF meshwork assembles through the nucleoid to mediate plasmid segregation. *Nucleic Acids Research*, 45(6), pp. 3158–3171

Million-Weaver, S. and Camps, M. (2014). Mechanisms of plasmid segregation: have multicopy plasmids been overlooked?. *Plasmid*, 75, pp. 27–36

Misra, H. S., Maurya, G. K. and Kota, S. (2018). Maintenance of multipartite genome system and its functional significance. *Journal of Genetics*. Springer India, 97(4), pp. 1013–1038

Misselwitz, R., De La Hoz, A. B., Ayora, S., Welfle, K., Behlke, J., Murayama, K., Saenger, W., Alonso, J. C. and Welfle, H. (2001). Stability and DNA-binding properties of the ω regulator protein from the broad-host range *Streptococcus pyogenes* plasmid pSM19035. *FEBS Letters*, 505(3), pp. 436–440

Møller-Jensen, J., Borch, J., Dam, M., Jensen, R. B., Roepstorff, P. and Gerdes, K. (2003). Bacterial mitosis: ParM of plasmid R1 moves plasmid DNA by an actin-like insertional polymerization mechanism. *Molecular cell*, 12(6), pp. 1477–1487.

Møller-Jensen, J., Jensen, R. B., Löwe, J. and Gerdes, K. (2002). Prokaryotic DNA segregation by an actin-like filament. *The EMBO Journal*, 21(12), pp. 3119–3127

Møller-Jensen, J., Ringgaard, S., Mercogliano, C. P., Gerdes, K. and Lowe, J. (2007). Structural analysis of the ParR / *parC* plasmid partition complex. *The EMBO Journal*, 26(20), pp. 4413–4422

Montabana, E. A. and Agard, D. A. (2014). Bacterial tubulin TubZ-Bt transitions between a two-stranded intermediate and a four-stranded filament upon GTP hydrolysis. *Proceedings of the National Academy of Sciences USA*, 111(9), pp. 3407–3412

Mori, H., Kondo, A., Ohshima, A., Ogura, T. and Hiraga, S. (1986). Structure and Function of the F Plasmid Genes Essential for Partitioning. *Journal of Genetics*, 192(1), pp. 1–15

Mori, H., Moris, Y., Ichinose, C., Niki, H., Ogura, T., Katoq, A. and Hiraga, S. (1989).

Purification ; and Characterization of SopA and SopB Proteins Essential for F Plasmid Partitioning. *Journal of Biological Chemistry*, 264(26), pp. 15535–15541

Murayama, K., Orth, P., De La Hoz, A. B., Alonso, J. C. and Saenger, W. (2001). Crystal structure of ω transcriptional repressor encoded by *Streptococcus pyogenes* plasmid pSM19035 at 1.5 Å resolution. *Journal of Molecular Biology*, 314(4), pp. 789–796

Ni, Lisheng, Xu, W., Kumaraswami, M. and Schumacher, M. A. (2010). Plasmid protein TubR uses a distinct mode of HTH- DNA binding and recruits the prokaryotic tubulin homolog TubZ to effect DNA partition. *Proceedings of the National Academy of Sciences USA*, 107(26), pp. 11763–11768

Ni, L., Xu, W., Kumaraswami, M. and Schumacher, M. A. (2010). Plasmid protein TubR uses a distinct mode of HTH-DNA binding and recruits the prokaryotic tubulin homolog TubZ to effect DNA partition. *Proceedings of the National Academy of Sciences USA*, 107(26), pp. 11763–11768

Nordström, K. and Gerdes, K. (2003). Clustering versus random segregation of plasmids lacking a partitioning function: a plasmid paradox?. *Plasmid*, 50(2), pp. 95–101

Ogura, T. and Hiraga, S. (1983). Partition Mechanism of F Plasmid : Two Plasmid Products and a Cis-Acting Region Are Involved in Partition. *Cell*, 32(2), pp. 351–360

Oliva, M. A. (2016). Segrosome Complex Formation during DNA Trafficking in Bacterial Cell Division. *Frontiers in Bioscience*, 3(51), pp. 1–9

Oliva, M. A., Martin-Galiano, A. J., Sakaguchi, Y. and Andreu, J. M. (2012). Tubulin homolog TubZ in a phage-encoded partition system. *Proceedings of the National Academy of Sciences USA*, 109(20), pp. 7711–7716

Page, R. and Peti, W. (2016). Toxin-antitoxin systems in bacterial growth arrest and persistence. *Nature Chemical Biology*, 12(4), pp. 208–214

Pillet, F., Sanchez, A., Lane, D., Leberre, V. A. and Bouet, J. (2011). Centromere binding specificity in assembly of the F plasmid partition complex. *Nucleic Acids Research*, 39(17), pp. 7477–7486

Popescu, A., Karpay, A., Israel, D. A., Peek, R. M. and Krezel, A. M. (2005). *Helicobacter pylori* protein HP0222 belongs to Arc/MetJ family of transcriptional regulators. *Proteins: Structure, Function and Genetics*, 59(2), pp. 303–311

Popp, D., Xu, W., Narita, A., Brzoska, A. J., Skurray, R. A., Firth, N., Goshdastider, U., Maéda, Y., Robinson, R. C. and Schumacher, M. A. (2010). Structure and filament dynamics of the pSK41 actin-like ParM protein: implications for plasmid DNA segregation. *Journal of Biological Chemistry*, 285(13), pp. 10130–10140

Pratto, F., Cicek, A., Weihofen, W. A., Lurz, R., Saenger, W. and Alonso, J. C. (2008). *Streptococcus pyogenes* pSM19035 requires dynamic assembly of ATP-bound ParA

and ParB on parS DNA during plasmid segregation. *Nucleic Acids Research*, 36(11), pp. 3676–3689

Pratto, F., Suzuki, Y., Takeyasu, K. and Alonso, J. C. (2009). Single-molecule analysis of protein·DNA complexes formed during partition of newly replicated plasmid molecules in *Streptococcus pyogenes*. *Journal of Biological Chemistry*, 284(44), pp. 30298–30306

Rafferty, J. B., Somers, W. S., Saint-girons, I. and Phillips, S. E. V (1989). Three-dimensional crystal structures of *Escherichia coli met* repressor with and without corepressor. *Nature*, 341(6244), pp. 705–710

Raumann, B. E., Rould, M. A., Pabo, C. O. and Sauer, R. T. (1994). DNA recognition by β-sheets in the Arc repressor-operator crystal structure. *Nature*, 367(6455), pp. 754–758

Ravin, N. V, Rech, J. and Lane, D. (2003). Mapping of Functional Domains in F Plasmid Partition Proteins Reveals a Bipartite SopB-recognition Domain in SopA. *Journal of Molecular Biology*, 329(5), pp. 875–889

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

Ringgaard, S., Löwe, J. and Gerdes, K. (2007b). Centromere pairing by a plasmid-encoded type I ParB protein. *Journal of Biological Chemistry*, 282(38), pp. 28216–28225

Ringgaard, S., Zon, J. Van, Howard, M. and Gerdes, K. (2009). Movement and equipositioning of plasmids by ParA filament disassembly. *Proceedings of the National Academy of Sciences USA*, 106(46), pp. 19369–19374

Rivera, C. R., Kollman, J. M., Polka, J. K., Agard, D. A. and Mullins, R. D. (2011). Architecture and Assembly of a Divergent Member of the ParM Family of Bacterial Actin-like Proteins. *Journal of Biological Chemistry*, 286(16), pp. 14282–14290

Rodionov, O., ŁObocka, M. and Yarmolinsky, M. (1999). Silencing of genes flanking the P1 plasmid centromere. *Science*, 283(5401), pp. 546–549

Saeed, S., Jowitt, T. A., Warwicker, J. and Hayes, F. (2015). Breaking and restoring the hydrophobic core of a centromere-binding protein. *Journal of Biological Chemistry*, 290(14), pp. 9273–9283

Salje, J. (2010). Plasmid segregation : how to survive as an extra piece of DNA. *Critical Reviews in Biochemistry and Molecular Biology*, 45(4), pp. 296–317

Salje, J. and Löwe, J. (2008). Bacterial actin: Architecture of the ParMRC plasmid DNA partitioning complex. *The EMBO Journal*, 27(16), pp. 2230–2238

Schanda, P. and Brutscher, B. (2005). Very fast two-dimensional NMR spectroscopy for real-time investigation of dynamic events in proteins on the time scale of seconds. *Journal of the American Chemical Society*, 127(22), pp. 8014–8015

Schildbach, J. F., Robinson, C. R. and Sauer, R. T. (1998). Biophysical Characterization of the TraY Protein of *Escherichia coli* F Factor. *Journal of Biological Chemistry*, 273(3), pp. 1329–1333

Schreiter, E. R. and Drennan, C. L. (2007). Ribbon – helix – helix transcription factors : variations on a theme. *Nature Reviews Microbiology*, 5(9), pp. 710–720

Schumacher, M. A. (2008). Structural biology of plasmid partition: uncovering the molecular mechanisms of DNA segregation. *Biochemical Journal*, 412(1), pp. 1–18

Schumacher, M. A. (2012). Bacterial plasmid partition machinery: a minimalist approach to survival. *Current opinion of Structure Biology*, 22(1), pp. 72–79

Schumacher, M. A. and Funnell, B. E. (2005). Structures of ParB bound to DNA reveal mechanism of partition complex formation. *Nature*, 438(24), pp. 516–519

Schumacher, Maria A, Glover, T. C., Brzoska, A. J., Jensen, S. O., Dunham, T. D., Skurray, R. A. and Firth, N. (2007a). Segrosome structure revealed by a complex of ParR with centromere DNA. *Nature Letters*, 450(20), pp. 1268–1271

Schumacher, Maria A., Mansoor, A. and Funnell, B. E. (2007b). Structure of a four-way bridged ParB-DNA complex provides insight into P1 segrosome assembly. *Journal of Biological Chemistry*, 282(14), pp. 10456–10464

Schumacher, M. A., Piro, K. M. and Xu, W. (2010). Insight into F plasmid DNA segregation revealed by structures of SopB and SopB – DNA complexes. *Nucleic Acids Research*, 38(13), pp. 4514–4526

Schumacher, M. A., Ye, Q., Barge, M. T., Zampini, M., Barillà, D. and Hayes, F. (2012). Structural Mechanism of ATP-induced Polymerization of the Partition Factor ParF. *Journal of Biological Chemistry*, 287(31), pp. 26146–26154

Sengupta, M. and Austin, S. (2011). Prevalence and Significance of Plasmid Maintenance Functions in the Virulence Plasmids of Pathogenic Bacteria. *Infection and Immunity*, 79(7), pp. 2502–2509

Shen, Y. and Bax, A. (2013). Protein backbone and sidechain torsion angles predicted from NMR chemical shifts using artificial neural networks. *Journal of Biomolecular NMR*, 56(3), pp. 227–241

Shen, Y. and Bax, A. (2015). Artificial neural networks: Second edition. *Methods Molecular Biology*, 1260, pp. 17–32

Simpson, A. E., Skurray, R. A. and Firth, N. (2003). A Single Gene on the Staphylococcal Multiresistance Plasmid pSK1 Encodes a Novel Partitioning System. *Journal of Bacteriology*, 185(7), pp. 2143–2152

Soberón, N. E., Lioy, V. S., Pratto, F., Volante, A. and Alonso, J. C. (2011). Molecular anatomy of the *Streptococcus pyogenes* pSM19035 partition and segrosome complexes. *Nucleic Acids Research*, 39(7), pp. 2624–2637

Some, D., Amartely, H., Tsadok, A. and Lebendiker, M. (2019). Characterization of Proteins by Size-Exclusion Chromatography Coupled to Multi-Angle Light Scattering (SEC-MALS). *Journal of Visualized Experiments*, 148, pp. 1–9

Somers, W. S. and Phillips, S. E. V (1992). Crystal structure of the *met* repressor-operator complex at 2.8 Å resolution reveals DNA recognition by β-strands. *Nature*, 359(1), pp. 387–393

Summers, D. (1998). Timing , self-control and a sense of direction are the secrets of multicopy plasmid stability. *Molecular Microbiology*, 29(5), pp. 1137–1145

Summers, D. K. and Sherratt, D. J. (1984). Multimerization of High Copy Number Plasmids Causes Instability : ColEI Encodes a Determinant Essential for Plasmid Monomerization and Stability. *Cell*, 36(4), pp. 1097–1103

Summers, D. K. and Sherratt, D. J. (1985). Bacterial Plasmid Stability. *BioEssays*, 2(5), pp. 209–211

Surovtsev, I. V, Campos, M. and Jacobs-wagner, C. (2016). DNA-relay mechanism is sufficient to explain ParA-dependent intracellular transport and patterning of single and multiple cargos. *Proceedings of the National Academy of Sciences USA*, 113(46), pp. 7268–7276

Tang, M., Bideshi, D. K., Park, H. W. and Federici, B. A. (2006). Minireplicon from pBtoxis of *Bacillus thuringiensis* subsp. israelensis. *Applied and Environmental Microbiology*, 72(11), pp. 6948–6954

Tang, M., Bideshi, D. K., Park, H. W. and Federici, B. A. (2007). Iteron-binding ORF157 and FtsZ-like ORF156 proteins encoded by pBtoxis play a role in its replication in *Bacillus thuringiensis* subsp. israelensis. *Journal of Bacteriology*, 189(22), pp. 8053–8058

Tobe, T., Hayashi, T., Han, C., Schoolnik, G. K., Ohtsubo, E. and Sasakawa, C. (1999). Complete DNA Sequence and Structural Analysis of the Enteropathogenic *Escherichia coli* Adherence Factor Plasmid. *Infection and Immunity*, 67(10), pp. 5455–5462

Tsang, J. (2017). Bacterial plasmid addiction systems and their implications for antibiotic drug development. *Postdoc Journal*, 5(5), pp. 3–9

Vecchiarelli, A. G., Han, Y. W., Tan, X., Mizuuchi, M., Ghirlando, R., Biertümpfel, C., Funnell, B. E. and Mizuuchi, K. (2010). ATP control of dynamic P1 ParA-DNA interactions: A key role for the nucleoid in plasmid partition. *Molecular Microbiology*, 78(1), pp. 78–91

Vecchiarelli, Anthony G., Havey, J. C., Ing, L. L., Wong, E. O. Y., Waples, W. G. and Funnell, B. E. (2013a). Dissection of the ATPase active site of P1 ParA reveals multiple active forms essential for plasmid partition. *Journal of Biological Chemistry*, 288(24), pp.

17823–17831

Vecchiarelli, Anthony G, Hwang, L. C. and Mizuuchi, K. (2013b). Cell-free study of F plasmid partition provides evidence for cargo transport by a diffusion-ratchet mechanism. *Proceedings of the National Academy of Sciences USA*, 110(15), pp. 1390–1397

Vecchiarelli, A. G., Neuman, K. C. and Mizuuchi, K. (2014a). A propagating ATPase gradient drives transport of surface-confined cellular cargo. *Proceedings of the National Academy of Sciences USA*, 111(13), pp. 4880–4885

Vecchiarelli, A. G., Seol, Y., Neuman, K. C. and Mizuuchi, K. (2014b). A moving ParA gradient on the nucleoid directs subcellular cargo transport via a chemophoresis force. *bioArchiteture*, 4(4–5), pp. 154–159

Vecchiarelli, A. G., Taylor, J. A. and Mizuuchi, K. (2015). Reconstituting ParA/ParB-mediated transport of DNA cargo. *Methods Cell Biology*, 128, pp. 243–269

Vögeli, B. (2014). The nuclear Overhauser effect from a quantitative perspective. *Progress in Nuclear Magnetic Resonance Spectroscopy USA*, 78, pp. 1–46

Volante, A. and Alonso, J. C. (2015). Molecular anatomy of ParA-ParA and ParA-ParB interactions during plasmid partitioning. *Journal of Biological Chemistry*, 290(30), pp. 18782–18795

Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Pajon, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J. and Laue, E. D. (2005). The CCPN data model for NMR spectroscopy: Development of a software pipeline. *Proteins: Structure, Function and Genetics*, 59(4), pp. 687–696

Vuzman, D. and Levy, Y. (2012). Intrinsically disordered regions as affinity tuners in protein-DNA interactions. *Molecular BioSystems*, 8(1), pp. 47–57

Wang, Y. (2017). Spatial distribution of high copy number plasmids in bacteria. *Plasmid*, 91, pp. 2–8

Wang, Y., Penkul, P. and Milstein, J. N. (2016). Quantitative Localization Microscopy Reveals a Novel Organization of a High-Copy Number Plasmid. *Biophysical Journal*. Biophysical Society, 111(3), pp. 467–479

Watanabe, E., Wachi, M., Yamasaki, M. and Nagai, K. (1992). ATPase activity of SopA , a protein essential for active partitioning of F plasmid. *Molecular gene genetics*, 234(3), pp. 346–352

Weihofen, W. A., Cicek, A., Pratto, F., Alonso, J. C. and Saenger, W. (2006). Structures of ω repressors bound to direct and inverted DNA repeats explain modulation of transcription. *Nucleic Acids Research*, 34(5), pp. 1450–1458

Welfle, K., Pratto, F., Misselwitz, R., Behlke, J., Alonso, J. C. and Welfle, H. (2005). Role of the N-terminal region and of β-sheet residue Thr29 on the activity of ω2 global regulator from the broad-host range *Streptococcus pyogenes* plasmid pSM19035.

Biological Chemistry, 386(9), pp. 881–894

Wider, G. (2000). Review Structure Determination of Biological Macromolecules in Solution Using Nuclear Magnetic Resonance Spectroscopy. *BioTechniques*, 29(6), pp. 1278–1294

Williamson, M. P. (2013). Progress in Nuclear Magnetic Resonance Spectroscopy Using chemical shift perturbation to characterise ligand binding. *Progress in Nuclear Magnetic Resonance Spectroscopy*, 73, pp. 1–16

Wu, M., Zampini, M., Bussiek, M., Hoischen, C., Diekmann, S. and Hayes, F. (2011). Segrosome assembly at the pliable *parH* centromere. *Nucleic Acids Research*, 39(12), pp. 5082–5097

Yang, Q. E. and Walsh, T. R. (2017). Toxin – antitoxin systems and their role in disseminating and maintaining antimicrobial resistance. *FEMS Microbioogy Reviews*, 41(3), pp. 343–353

Yao, S., Helinski, D. R. and Toukdarian, A. (2007). Localization of the Naturally Occurring Plasmid ColE1 at the Cell Pole. *Journal of Bacteriology*, 189(5), pp. 1946–1953

Yates, P., Lane, D. and Biek, D. P. (1999). The F Plasmid Centromere, *sopC*, is Required for Full Repression of the sopAB Operon. *Journal of Molecular Biology*, 290(3), pp. 627–638

Zampini, M., Derome, A., Bailey, S. E. S., Barillà, D. and Hayes, F. (2009). Recruitment of the ParG segregation protein to different affinity DNA sites. *Journal of Bacteriology*, 191(12), pp. 3832–3841

Zhang, H. and Schumacher, M. A. (2017). Structures of partition protein ParA with nonspecific DNA and ParB effector reveal molecular insights into principles governing Walker-box DNA segregation. *Genes and Development*, 31(5), pp. 481–492