

# Replication and stability of the linear plasmid pBSSB2

Submitted by

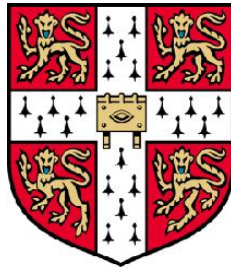
**Sunjukta Ahsan**

Lucy Cavendish College

Department of Genetics

University of Cambridge

September 2011



A dissertation submitted to the  
**University of Cambridge**  
for the degree of  
**Doctor of Philosophy**



## **Disclaimer**

---

I hereby declare that this dissertation is the result of my own work conducted between October 2007 and September 2011 under the supervision of Dr. David Summers at the Department of Genetics, University of Cambridge and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text. This work has not been submitted for any other degree at this, or any other university.

I also declare that this dissertation does not exceed 60,000 words.

.....

**Sunjukta Ahsan**

Department of Genetics

University of Cambridge

20 September 2011

## Abstract

Plasmid pBSSB1 is a 27 kb linear DNA with proteins attached at the 5' termini. It encodes the H: z66 flagellar antigen in *Salmonella enterica* serovar Typhi (*S. Typhi*) isolated from Indonesia. Together with the H: j or H: d flagellar antigen encoded by the host chromosome, pBSSB1 renders expression of the flagellar antigen biphasic in *S. Typhi*. Following the discovery of pBSSB1, initial bioinformatic analyses were carried out. However, no genetic analysis of replication and stability functions was conducted. Such studies form the basis of the present work. Plasmid pBSSB2, that contains a kanamycin cassette inserted at position 1295 bp of pBSSB1, was used in the present investigation.

The first objective of the work was to develop a method of purification for the linear plasmid. Conventional plasmid extraction methods which had been used previously were found to produce a very poor yield of plasmid DNA. It was shown in the present study that a proteinase-K treatment was essential for the removal of the linear plasmid terminal proteins to avoid loss of the plasmid in the phenol-chloroform-isoamylalcohol treatment which removes cellular proteins from the plasmid DNA.

The region containing the basic replicon of pBSSB2 was identified by screening for a region that was able to support replication in *E. coli* of a ColE1-like plasmid in a *polA* host (in which it would not normally replicate). This identified a 2831 bp fragment encompassing nucleotides 12820 to 15649 of pBSSB2. It was expected that this would encode an initiator of replication such as a Rep protein. However, mutagenesis studies showed that none of the annotated ORFs in this fragment was essential for replication. Candidate ORFs, not identified in the original annotation, have been proposed that remain to be tested as possible candidates for the *rep* encoding gene. The possibility of an alternative RNA primed initiation of replication has also been hypothesized. An adjacent region was found to exert strong incompatibility against pBSSB2, suggesting that it might encode a repressor of replication. The minimum region conferring incompatibility was 179 bp, encompassing nucleotides 10840 bp to 11018 bp of pBSSB2. A six base pair imperfect repeat, (G/T) (G/A) TGTG was found within this sequence. It is hypothesized that these imperfect repeats may function as iterons that titrate a Rep protein and regulate pBSSB2 replication.

A 1023 bp region (nucleotides 7236 to 8258 of pBSSB2) was found to confer stability in *E. coli* upon an otherwise unstable circular plasmid. Mutational analysis showed that an annotated ORF within this region (ORF09) was required for plasmid stabilisation. When expressed independently from an expression vector ORF09 killed host cells. It is proposed that the stability function acts as a toxin-antitoxin system, although the antitoxin has not yet been identified. A candidate promoter for a putative countertranscript and two potential ORFs as candidates for encoding the antitoxin have been suggested for future work to identify the antitoxin.

The preliminary functional characterization of pBSSB2 has contributed to our general understanding of the replication and stability functions of linear plasmids. However, further work will be required to achieve a complete understanding of the molecular basis of these functions in pBSSB2.

**For my parents**

## Acknowledgements

I am grateful to my supervisor, Dr. David Summers, for his continued support and advice during the course of my degree. I am also thankful to my adviser, Dr. Philip Oliver, for useful suggestions and comments throughout the length of my work.

I would like to express my gratitude to Dr. Gordon Dougan of the Wellcome Trust Sanger Institute for kindly providing for my bench fees. Thanks to Dr. Stephen Baker for identifying the linear plasmid and to Dr. Derek Pickard of the Sanger Institute for help and suggestions.

I owe sincere gratitude to Dr. Finbarr Hayes of the University of Manchester for his kind permission to use the BR825 DNA *polA* strain and the SP2 *pcnB* host.

Thanks also go out to all the members of Dr. Summers's laboratory who have made my time in the lab worthwhile. I would like to particularly thank Dr. Mathew Cahill for his kind help, particularly during the initial phases of my work. Thanks also to Azlin, Chin, Chris, Claudio, Duncan, Hannah, James, Nic and Silvia for intellectual and moral support. I would also like to express my gratitude to Mr. Mazhar Sheikh of the department of genetics for useful suggestions.

This study was supported by the Commonwealth Scholarship Commission and I express immense gratitude to the Commission. I am also grateful to all the charitable organizations that provided financial support particularly towards the end of my work. In no particular order, I would like to express my gratefulness to the Board of Graduate Studies for Lundgren Award, the Cambridge Commonwealth Trust, the Cambridge Philosophical Society, the Cambridge University Central Childcare Bursary, the Charles Wallace Bangladesh Trust, Leche Trust and to Lucy Cavendish College for various awards. Without contributions from these organizations, management of maintenance costs during the fourth year of my study would have been impossible.

I would like to express my heart-felt gratitude to my family. My parents, who have encouraged me all throughout my life, provided me with unfathomable moral support and confidence. I dedicate my thesis to them. My siblings have also been by my side all along, thank you so much. My husband and my daughter were beside me to strive through times of happiness and difficulty alike. We have worked together through our hard times with patience and endurance.

Finally, I am immensely grateful to Almighty Allah for making me so fortunate to avail such a wonderful opportunity to pursue PhD in one of the best universities in the world.

Thank you all!

Sunjukta Ahsan  
Cambridge  
September 2011

# Contents

---

	Page Number
Disclaimer .....	iii
Abstract.....	iv
Acknowledgements .....	vi
Contents .....	vii-xii
List of Figures .....	xiii-xvii
List of Tables .....	xviii
List of abbreviations .....	xix-xxi
<b>Chapter 1: Introduction</b>	<b>1-30</b>
1.1 <i>Salmonella typhi</i> : the causative agent of typhoid fever	2-4
1.2 The linear plasmid, pBSSB1	4-5
1.3 Plasmid replication and its control	5-20
1.3.1 Replication of circular plasmids	5-6
1.3.1.1 Replication and control of P1 replication	7-9
1.3.1.2 RNA mediated replication of ColE1	9-11
1.3.1.3 Rolling-circle replication of pT181	11-13
1.3.2 Replication of linear plasmids	13-20
1.3.2.1 Replication of linear plasmids with terminal proteins	14-18
1.3.2.2 Replication of linear plasmids with closed ends	18-20

1.4 General features of plasmid stability	20-29
1.4.1 Random distribution of high copy number plasmids	20
1.4.1.1 Plasmid loss and metabolic load	20-21
1.4.1.2 Plasmid clustering	21-22
1.4.1.3 Plasmid oligomer formation and resolution	22-23
1.4.2 Low copy number plasmids	23-30
1.4.2.1 Plasmid partitioning	23-26
1.4.2.2 Plasmid stabilization through the toxin antitoxin system	26-29
1.5 Aims and objectives of the present investigation	30
 <b>Chapter 2: Materials and Methods</b>	 <b>31-49</b>
2.1 Solutions	32
2.2 Reagents	33
2.3 Buffers	33
2.4 Bacterial strains	33-34
2.5 Media	34-35
2.6 Antibiotics.	35
2.7 Plasmids	35
2.8 Oligonucleotides used in this study	36-38
2.9 Plasmid Extraction	39-40
2.9.1 STET Plasmid Preparation	39
2.9.2 Single colony final sample buffer (SCFSB) technique	39
2.9.3 The Qiagen miniprep for plasmid extraction	39
2.9.4 Kado and Liu Method	40



2.10 PCR	41-43
2.10.1 Routine PCR using Taq DNA polymerase from NEB	41
2.10.2 Colony PCR	41-42
2.10.3 Large amplicon and high fidelity PCR	41-42
2.10.4 PCR using the Expand Long Template PCR system from Roche	42-43
2.11 Purification of PCR products and digested DNA	43
2.12 Quantitation of DNA	43
2.13 Restriction digestion	44
2.14 Exonuclease treatment	44
2.15 Ligation using T4 DNA ligase	44
2.16 Drop dialysis	44
2.17 Gel electrophoresis	44-45
2.18 Transformation	45-46
2.18.1 Electroporation	45
2.18.2 Chemical transformation	45-46
2.19 Phase contrast microscopy	46
2.20 Site-directed mutagenesis	46-48
2.20.1 Site-directed mutagenesis by overlap extension method	46-47
2.20.2 The QuikChange II one-day site-directed mutagenesis method	47-48
2.21 DNA sequencing	49
 <b>Chapter 3: Development of a method for purification of the linear plasmid, pBSSB2</b>	 <b>50-64</b>
3.1 Introduction	51-54
3.2 Use of conventional plasmid extraction methods for pBSSB2	54

purification	
3.3 Development and optimization of a method for the successful extraction of pBSSB2	54-56
3.3.1 Cell lysis	55
3.3.2 Removal of cell debris	56
3.3.3 Precipitation of DNA	56
3.4 Gel electrophoresis of extracted pBSSB2	58-59
3.5 Restriction digestion of purified pBSSB2	59-60
3.6 Treatment of purified pBSSB2 with $\lambda$ exonuclease	60-61
3.7 PCR amplification of pBSSB2 specific regions	61-62
3.8 Retransformation of purified pBSSB2	62
3.9 Discussion	62-64
<b>Chapter 4: Characterization of the replication functions of the linear plasmid, pBSSB2</b>	<b>66-108</b>
4.1 Introduction	66-67
4.2 Construction of a pBSSB2 library in pUC18	67-69
4.3 Identification of pBSSB2 replicon function in a circular mode	70-79
4.4 Deletion analysis of fragment RP7	80-81
4.5 Genetic analysis of fragment RP7	81-84
4.6 Analysis of RP7 for promoters, transcription terminators, ORFs and ribosome binding sites (RBS)	84-96
4.6.1. Promoters	84-87
4.6.2 Prediction of ORFs in 2083 bp fragment	88-93
4.6.3 Ribosome binding sites	93-94
4.6.4 Transcription terminators	95-96

4.6.5 Candidate replication initiator protein genes based on sequence analysis	96
4.7 Identification of pBSSB2 regions which exert plasmid incompatibility	96-98
4.8 Generation of RP2 deletion derivatives to identify the minimum region conferring incompatibility	99-101
4.9 Sequence analysis of the minimum region of incompatibility	101-102
4.10 Interaction of incompatibility and replication functions	102-104
4.11 Discussion	105-108
<b>Chapter 5: Stability functions of pBSSB2</b>	<b>110-147</b>
5.1 Introduction	110-111
5.2 Identification of regions containing stability functions of pBSSB2	111-112
5.3 Further analysis of stability functions in regions RP2 and RP4	113-114
5.4 Deletion analysis of fragment RP4 to identify minimum region necessary for stability	115-118
5.5 Mutational analysis of the pBSSB2 stability region	119-122
5.5.1 Engineering a stop codon in ORF 07	120
5.5.2 Introduction of nonsense mutation in ORF08	121
5.5.3 Introduction of stop code in ORF09	122
5.6 Role of ORFs 07-09 in pBSSB2 stability	122-123
5.7. Characterization of the RP4 stability function	124-127
5.8. The effect of ORF09 expression on cell morphology	127-128
5.9 Possible candidates for an antitoxin	129-145
5.9.1 Promoter search in the 1023 bp stability region	129
5.9.2. Prediction of ORFs in the 1023 bp region	137-142
5.9.3 BLAST analysis of nucleotide and protein sequences	142-145

5.10 Discussion	145-147
<b>Chapter 6: Discussion and future work</b>	<b>148-156</b>
6.1 Introduction	149
6.2 A method for the purification of protein-capped linear plasmids	149-150
6.3 Replication of pBSSB2	150-153
6.4 Future work necessary to characterize the pBSSB2 basic replicon	153
6.5 Stability function of pBSSB2	154-155
6.6 Conclusion	155-156
<b>References</b>	<b>157-185</b>

## List of figures

---

Figure 1.1: Site-specific recombination mediated by Hin recombinase involved in Salmonella flagellar antigenic phase variation	3
Figure 1.2. The genetic map of pBSSB1	5
Figure 1.3 Schematic diagram showing the P1 replicon	7
Figure 1.4. Schematic representation of ColE1 replication	11
Figure 1.5. Schematic representation of pT181 replication	13
Figure 1.6. Models for replication of linear plasmids containing 5' terminal proteins	16
Figure 1.7. Bidirectional replication of Borrellia linear plasmid from a central origin	19
Figure 1.8: Plasmid clustering	22
Figure 1.9. A schematic diagram of the P1 partition site	24
Figure 1.10: A schematic diagram showing the molecular assembly at the P1 <i>parS</i> site	25
Figure 1.11. Interpretation of the mechanisms of P1 partitioning	26
Figure 1.12: The <i>parB</i> post-segregational killing system of plasmid R1	29
Figure 2.1: Schematic representation of the overlap extension method for the introduction of mutation	47
Figure 3.1: Purification of pBSSB2 by an optimized method	58
Figure 3.2. Restriction digestion profile of pBSSB2	60
Figure 3.3. Lambda exonuclease treatment on pBSSB2 with pUC18 as control	61
Figure 3.4. Amplification of a 500 bp region present in pBSSB2	62
Figure 4.1 Fragments of pBSSB2 cloned into pUC18	68
Figure 4.2. Linear plasmid library	69
Figure 4.3. Validation of cloned fragments RP2-RP10 by restriction digestion	69

Figure 4.4. Amplification of a 500 bp RP7 product from strain BR825 transformed with pRP7	71
Figure 4.5. Plasmids extracted from XL1-Blue transformants following transformation with plasmid extracted from BR825 transformed with pRP7	72
Figure 4.6. Plasmids extracted from XL1-Blue transformants following transformation with plasmid extracted from BR825 transformed with pRP7	72
Figure 4.7. Extraction of plasmid from XL1-Blue colonies transformed with plasmid preparations from BR825 transformed originally with pRP7	73
Figure 4.8. Plasmids purified from XL1-Blue transformants transformed with plasmid bands cut out from gel carrying mixed plasmids	74
Figure 4.9. Digestion of four plasmids from XL1-Blue transformants	75
Figure 4.10. Flow diagram of the steps undertaken to identify the region of pBSSB2 that drives replication using the pRP2-pRP10 library	76-77
Figure 4.11. Theoretical construct RP7-Km <sup>r</sup>	79
Figure 4.12. PCR carried out to detect presence of a putative construct containing RP7 ligated to Km <sup>r</sup>	79
Figure 4.13. Deletion series of RP7 generated by PCR to identify the minimum replicon	81
Figure 4.14. Introduction of a stop codon in ORF018	82
Figure 4.15. Introduction of a stop codon in ORF019	83
Figure 4.16. Introduction of a stop codon in ORF020	84
Figure 4.17. Transcription profile from the RP7 region in pBSSB2	85
Figure 4.18. Positions of the potential promoters (arrowheads) in the two strands of the 2831 bp fragment	87
Figure 4.19. Promoters (dark blue arrowheads), transcription profile (green peaks) and ORFs (light blue arrows) present in the strand that contains ORFs 018 and 019 of the 2831 bp fragment of pBSSB2 that could support replication in a circular derivative	89
Figure 4.20. ORFs in the 2830 bp fragment in the strand complementary to that containing ORFs 018 and 019	90
Figure 4.21. Nucleotide sequence of ORF A	90

Figure 4.22. The nucleotide sequence of ORF B	90
Figure 4.23. The nucleotide sequence of ORF C	91
Figure 4.24. The nucleotide sequence of ORF D (highlighted in blue)	91
Figure 4.25. Nucleotide sequence of ORF E in blue with the putative SD in yellow	92
Figure 4.26. The sequence of ORF F highlighted in blue	92
Figure 4.27. Nucleotide sequence of ORF G in blue	93
Figure 4.28: Candidate transcription termination sequence for transcript from promoter 1	94
Figure 4.29. A potential stem-and-loop terminator structure downstream of ORF F and G	95
Figure 4.30. Incompatibility between pBSSB2 and constructs pRP2 to pRP10	98
Figure 4.31. Map showing position of fragments RP2, RP6 and RP7	
Figure 4.32. RP2 deletion series to identify minimum region expressing incompatibility against pBSSB2	98
Figure 4.33. Plasmid profile of cells transformed with linear plasmid and pRP2 deletion derivatives after growth in the absence of selection	100
Figure 4.34. 179 bp of RP2 deletion series sufficient to cause incompatibility	102
Figure 5.1 Stability of constructs pRP2 to pRP10 in a <i>pcnB</i> mutant	112
Figure 5.2 Stability of plasmid pFH450 and derivatives containing fragments RP4 and RP2	114
Figure 5.3 Schematic diagram of pRP4	115
Figure 5.4 Stability of pRP4HindIII as compared to pRP4 and pUC18	116
Figure 5.5. The location of the RP4HindIII region within pBSSB1 DNA sequence	117
Figure 5.6 Deletion derivatives of the RP4HindIII 2.2kb fragment	118
Figure 5.7. Stability of RP4HindIII deletion series	118
Figure 5.8 DNA and corresponding amino acid sequence of ORF07	120

Figure 5.9. DNA and amino acid sequences of ORF08	121
Figure 5.10. The nucleotide and amino acid sequences of ORF09 showing the introduction of mutation	122
Figure 5.11. Stability assay of constructs with ORFs 07-09 mutated	123
Figure 5.12. Growth of cultures containing pBAD18 or pBAD18-ORF09 in the presence of glucose (0.2%) or arabinose (0.2%)	126
Figure 5.13. Growth of cultures containing pBAD18 and pBAD18-ORF09 in the presence of glucose or arabinose	127
Figure 5.14. Phase contrast micrographs of <i>E. coli</i> XL1-Blue cells containing pBAD18 and pBAD18-ORF09	128
Figure 5.15. Transcription profile of fragment RP4HindIII	130
Figure 5.16. Promoters predicted for ORF09 and candidate antitoxin	131
Figure 5.17. Putative promoters in relation to the RNA-seq transcription pattern obtained for the 1023 bp RP4 derivative carrying the toxin antitoxin locus of pBSSB2	132
Figure 5.18. Promoter 2, a candidate promoter for antitoxin countertranscript formation	133
Figure 5.19. Possible transcription termination sequence (bold and underlined) for transcription initiated from promoter 2 (highlighted in blue)	134
Figure 5.20. Secondary structure of possible terminator of transcription from promoter 2	134
Figure 5.21. Predicted secondary structure for the ORF09 transcript	135
Figure 5.22. Predicted secondary structure for candidate antitoxin transcript	136
Figure 5.23. Predicted secondary structure of pBSSB2 toxin	137
Figure 5.24. Nucleotide sequence of ORF vi	138
Figure 5.25. ORFs in the same strand as ORF09	138
Figure 5.26. Nucleotide sequence of ORF I highlighted in blue	139
Figure 5.27. Nucleotide sequence of ORF ii highlighted in blue	139



Figure 5.28. Nucleotide sequence of ORF iii highlighted in blue	140
Figure 5.29. Nucleotide sequence of ORF iv highlighted in blue	140
Figure 5.30. ORFs identified within the 1023 bp fragment carrying the toxin-antitoxin locus of pBSSB2	140
Figure 5.31. Position of ORF iv relative to the toxin ORF09	142
Figure 5.32. Selected results of psi blast of the ORF09 sequence at <a href="http://www.sbg.bio.ic.ac.uk/~phyre/">http://www.sbg.bio.ic.ac.uk/~phyre/</a>	143
Figure 5.33. A proposed secondary structures of the toxin encoded by the 1023 bp RP4 derivative of pBSSB2	144
Figure 6.1 Schematic diagram showing amplicon RP7 that can drive replication of a circular pBSSB2 derivative and the region of amplicon RP2 conferring incompatibility functions in a circular mode	151
Figure 6.2. A tentative model for replication and its control in pBSSB2	153
Figure 6.3. Overview of the regions associated with replication and stability of pBSSB2 identified in this study	156

## List of tables

---

Table 2.1: Bacterial strains used in this study	33-34
Table 2.2: Antibiotics used in this study	35
Table 2.3: Plasmids used in the present investigation	35
Table 2.4: Details of oligonucleotides used in the present study	36-38
Table 2.5: Composition of PCR master mix used for routine PCRs using Taq DNA polymerase	41
Table 2.6: Reaction mix used for high fidelity PCR using MasterAmp <sup>TM</sup> DNA polymerase	42
Table 2.7: Master mix for long template PCR	43
Table 2.8: Composition of ligation reaction mix	44
Table 2.9: Composition of amplification master mix used for site-directed mutagenesis	48
Table 3.1. Examples of linear plasmids extracted through use of PFGE	53
Table 3.2. Linear plasmid purification method developed in the present study	56-57
Table 4.1 Nucleotide co-ordinates of linear plasmid fragments inserted into pUC18	68
Table 4.2. Candidate promoters identified within the 2831 bp fragment using the program available at <a href="http://linus1.softberry.com/">http://linus1.softberry.com/</a>	86
Table 4.3. ORFs present in the 2831 bp replication-competent region of RP7	88
Table 4.4 The influence of fragment RP2 on RP7 when present <i>in cis</i> or <i>in trans</i>	104
Table 5.1. Promoter prediction for the 1023 bp RP4 region	129
Table 5.2. ORFs identified within the 1023 bp fragment as identified by the program at <a href="http://www.ncbi.nlm.nih.gov/gorf/gorf.html">http://www.ncbi.nlm.nih.gov/gorf/gorf.html</a>	138

## List of abbreviations

---

A	Adenine
Amp <sup>R</sup>	Ampicillin resistant
approx.	approximately
ARS	autonomously replicating sequence
bp	base pair
C	cytosine
Cm <sup>R</sup>	chloramphenicol resistant
DNA	Deoxyribonucleic acid
DNase	Deoxyribonuclease
EDTA	Ethylenediaminetetraacetic acid
g	gram
G	guanine
hr	hour
IPTG	Isopropyl-β-D-thiogalactopyranoside
Km <sup>R</sup>	Kanamycin resistant
kb	kilo base
kDa	kilo Daltons
kV	kilo Volt
l	litre

LB	Luria Bertani
m	metre
MCS	multiple cloning site
min	minute
mRNA	messenger ribonucleic acid
nt	nucleotide
OD	optical density
ORF	open reading frame
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
RNA	Ribonucleic acid
RNase	Ribonuclease
rpm	revolutions per minute
Rep	replication
SDS	Sodium dodecyl sulphate
SDW	Sterile distilled water
sec	second
Strep <sup>R</sup>	Streptomycin resistant
T	thymine
TAE	Tris-acetate EDTA

TE	Tris-EDTA
Tet <sup>R</sup>	Tetracycline resistant
U	uracil
UV	ultra-violet
X-gal	5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactoside

# Chapter 1

---

## Introduction

---

## 1.1 *Salmonella* Typhi: the causative agent of typhoid fever

*Salmonella enterica* serovar Typhi is the causative agent of typhoid fever in humans. Unlike most other *Salmonella* serotypes, *S. Typhi* is restricted to the human host (Parry *et al.*, 2002). Typhoid fever continues to be a major public health concern, particularly in developing countries. Every year, 12-33 million cases of typhoid fever occur worldwide (Pang *et al.*, 1995; DeRoeck, 2007).

In *Salmonella* flagella play an important role in motility (Flan and Macnab, 1996; Walz and Caplan, 2000). Most *Salmonella enterica* serovars are biphasic and can express either of two flagellin antigens. These are expressed from phase-1 and phase-2 genes named *fliC* and *fljB*, located at different locations in the *Salmonella enterica* chromosome (Flan and Macnab, 1996). *fljA* is the genetic determinant of the repressor (FljA) of *fliC* and is cotranscribed with *fljB* (Bonifield and Hughes, 2003; Yamamoto and Kutsukake, 2006). Switching of flagellin antigen expression (Figure 1.1) is mediated by the product of a DNA sequence, *hin*, flanked by two inverted repeats (Johnson and Simon, 1985) and located upstream of the *fljBA* operon (Henderson *et al.*, 1999; Kelly, 2004; Kutsukake *et al.*, 2006). Binding of Hin recombinase to the inverted repeats, *hix*, leads to a site-specific recombination event that inverts the promoter for *fljBA*. In one orientation, the promoter of *fljB* faces away from it. Hence, *fliC* is expressed in the absence of FljA repressor (Figure 1.1, B) (Zieg *et al.*, 1977; Zieg and Simon, 1980; Simon *et al.*, 1980). Under conditions when *hin* reverses its direction by site-specific recombination, the promoter of *fljB* is in its right orientation, enabling expression of *fljB* as well as *fljA*, which prevents expression of *fliC* (Figure 1.1, A). The process of *hin* inversion, known as ‘phase variation’, is reversible, thereby enabling either of the flagellar antigens, FljC or FljB to be expressed at a given time.

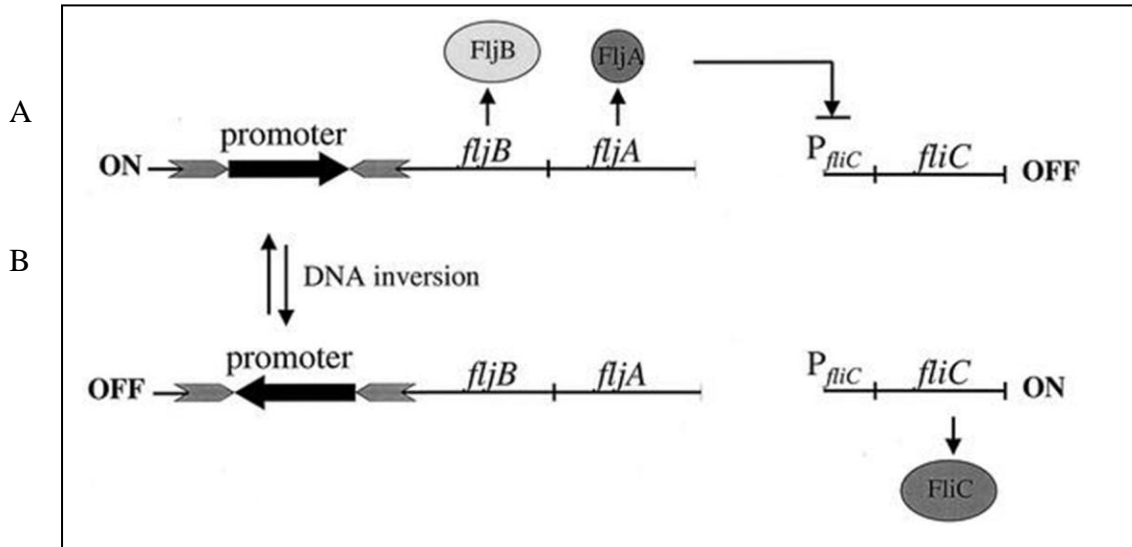


Figure 1.1: Site-specific recombination mediated by Hin recombinase involved in *Salmonella* flagellar antigenic phase variation. (Source: Bonifield *et al.*, 2003)

*Salmonella enterica* serovar Typhi is not capable of phase variation and normally expresses only one antigen, FliC (Frankel *et al.*, 1989). The chromosomal *fliC* gene codes for H:d (d) flagellin antigen. *Salmonella* Typhi is responsible for high incidence of typhoid in Indonesia, where the disease is endemic (Ivanoff, 1997). Some *S. Typhi* strains from Indonesia express an atypical flagellar antigen called H:j (j) instead of H:d (d) (Frankel *et al.*, 1989). Antigen j is coded by a truncated form of *fliC*. The truncated gene carries a 251bp deletion in the central region (Frankel *et al.*, 1989). Isolates of *Salmonella* Typhi that are motile but do not express either j or d antigen have also been reported (Guinee *et al.*, 1981). These strains were found to express a novel antigen, H:z66 (z66). Unlike other *Salmonella enterica* serovars, FliC is not the default antigen in these strains. When these strains are incubated with anti-z:66 antiserum, they express antigens j or d, depending on whether *fliC* or its deletion derivative is present in the chromosome. However, strains expressing the FljB (z66) antigen preferentially express the *fliC* gene (Baker *et al.*, 2008). It was previously assumed that these strains are biphasic like most other *S. enterica* serovars. However, the d or j positive *S. Typhi* strains do not revert to z66 expression when induced with anti-j or anti-d antiserum (Tamura *et al.*, 1988).



The gene encoding z66 was reported to be located in a linear plasmid, pBSSB1 (27kb) by Baker *et al.* (2007b). A *fljAB*-like operon was identified in the linear plasmid (Baker *et al.*, 2007b). There is 45% DNA sequence homology between the *fljAB*<sup>z66</sup> region of *S. Typhi* z66+ strains with the *fljAB* locus of *S. Typhimurium*. About 70% similarity has been reported between the amino acid sequence of the product of the *fljA*-like gene in *S. Typhi* with the FljA of *S. Typhimurium* LT2 (Huang *et al.*, 2004). No *hin*-like gene capable of site-specific recombination has been detected upstream of the *fljAB* locus in z66+ strains.

Recently, the expression of a new flagellin protein, Ind, has been reported for *S. Typhi* strains isolated from Indonesia. The sequence of Ind shows close homology to that of z66 and H:d at its amino and carboxy termini conserved sequences. On the other hand, a part of its conserved domain and the heterologous domains show homology to the FliC flagellin protein of *Serratia marcescens*. All of the isolates that were positive for Ind were also positive for H:d *fliC*. Whether Ind represents a phase 1 or a phase 2 flagellin has not yet been investigated (Hatta *et al.*, 2011).

## 1.2 The linear plasmid, pBSSB1

The plasmid, pBSSB1, was originally identified by Baker *et al.* (2007b). Linearity of pBSSB1 has been proved by restriction digestion by these authors. Linear plasmids belong to one of two structural categories: one in which the end is closed by a hairpin loop and another in which a protein is attached covalently at the 5' ends. Baker *et al.* (2007b) demonstrated that pBSSB1 was resistant to S1 nuclease and lambda exonuclease but was sensitive to degradation by 3'-5' exonuclease III. The linear plasmid, pBSSB1, therefore, belongs to the second category. Plasmid pBSSB1 is 27,037bp in length with terminal inverted repeats (TIRs) of about 1230bp. There are 33 annotated ORFs in pBSSB1, of which ORF031 codes for phase II flagellar antigen (Baker *et al.*, 2007b). The sequences of 22 ORFs have no similarity to the sequences in the public databases. The genetic map of pBSSB1 is shown in figure 1.2. The GC% of the linear plasmid is about 36.6 (Baker *et al.*, 2007b). Based on GC skew (G-C/G+C), the origin of replication lies upstream of ORF017. No direct repeats or iterons have been detected in the region predicted to contain the origin of replication. The right

terminal inverted repeat has been shown not to be essential for replication (Baker *et al.*, 2007a).

A kanamycin resistance cassette has been inserted at 1295 bp by Baker *et al.* (2007b) to give rise to pBSSB2 (indicated by the cross in Figure 1.2). The linear plasmid, pBSSB2, was transformed into *E. coli* Top10 (Invitrogen) to create a strain designated SGB33 (Baker *et al.*, 2007b). SGB33 has been used in all experimental procedures in the present investigation as a source of pBSSB2.

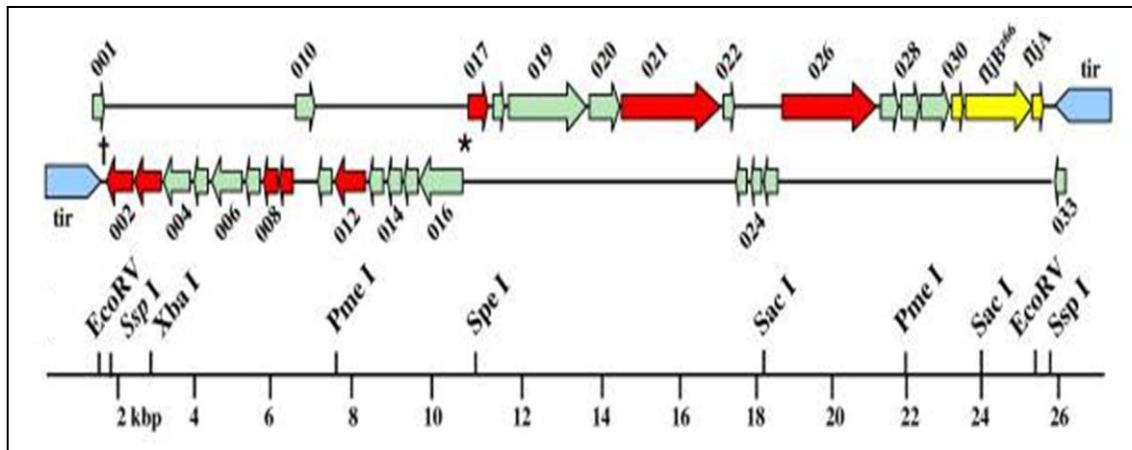


Figure 1.2. The genetic map of pBSSB1. The arrows indicate ORFs, the yellow arrows indicate ORFs with known functions, red arrows represent ORFs with sequences similar to other ORFs in the public database and blue arrows indicate ORFs with unknown function. The terminal inverted repeats, tirs, are indicated by light blue boxes with one pointed end. The restriction digestion map of pBSSB1 is shown underneath the ORF map. The cross indicates the site of insertion of the kanamycin cassette. The GC skew is indicated by the asterisk. The figure was taken from Baker *et al.*, (2007b).

## 1.3 Plasmid replication and its control

### 1.3.1 Replication of circular plasmids

Plasmids are extrachromosomal DNA molecules capable of autonomous replication. Typically, plasmids encode factors necessary for initiation of replication and use host specific elements for continuation of replication (reviewed by Actis *et al.*, 1999). The basic replicon of a plasmid is defined as the smallest region of the plasmid that is able to replicate on its own with its wild-type copy number (Kollek *et al.*, 1978). In most

plasmids it is about 4 kb in length (Actis *et al.*, 1999). The basic replicon generally includes the origin of replication (*ori*), an initiator protein (Rep) or an initiator RNA and sequences involved with regulation of replication. Primer RNAs generally act *in cis*, whereas Rep proteins can act either *in cis* or *in trans* (reviewed by Actis *et al.*, 1999).

The replication of circular plasmids has been reviewed by del Solar *et al.* (1998). Replication occurs by one of two common mechanisms. In theta type replication, replication is initiated by synthesis of a primer RNA (pRNA) at the origin. Elongation of plasmid replication is generally carried out by the host DNA polymerase III holoenzyme (HE), DnaB helicase and DnaG primase. Termination of plasmid replication generally involves DNA-protein interactions. Some plasmids replicate by rolling-circle replication in which Rep protein introduces a sequence-specific nick that generates a free 3'-OH end that acts as a primer for replication. The end-products are a double-stranded DNA and a single stranded DNA molecule. The latter DNA is converted to a double-stranded form by discontinued synthesis of Okazaki fragments using a single strand origin sequence for initiation.

Plasmids maintain themselves at a particular copy number in a given host under specific conditions. Copy number is regulated by sequences involved in control of initiation of replication. Control of plasmid replication is achieved by one of two mechanisms. In one mechanism, direct DNA repeats designated iterons, located close to *ori*, control the rate of initiation of replication through binding of Rep proteins. In some plasmids e.g. P1 (Pal and Chattoraj, 1988; Abeles and Austin, 1991; Abeles *et al.*, 1995) binding of Rep protein at low concentration allows replication. At high concentrations of Rep, due to an increased copy number of the plasmid, Rep proteins couple or handcuff origins of different plasmid molecules, hence limiting the availability of origins to initiate replication. In the other control mechanism, one or more antisense RNAs bind to pre-primer RNAs or to transcripts that encode the Rep protein and control the rate of initiation of plasmid replication (Kues and Stahl, 1989; Novick, 1987).

In this section the replication and control mechanisms of the iteron-regulated P1 replicon and the antisense RNA regulated replicons of ColE1 and pT181 will be used to illustrate copy-number control mechanisms of circular plasmids.

### 1.3.1.1 Replication and control of P1 replication

The lysogenic form of bacteriophage P1 can replicate autonomously in *E. coli* (Abeles *et al.*, 1984). The minimum region required for P1 replication is about 1.5 kb in length (Figure 1.3). A 959bp region codes for the 32 kD RepA protein (Chattoraj *et al.*, 1985). An incompatibility locus, *incC*, consisting of five 19bp iteron repeats, located about 53 bases upstream from the *repA* ORF (Chattoraj *et al.*, 1985), is essential for replication and is the binding site of RepA (Abeles *et al.*, 1984). The promoter and the operator of *repA* is present within *incC* and production of RepA is autoregulated (Chattoraj *et al.*, 1985). A region containing 67% A+T lies 50bp to the left of the RepA binding sites. A second incompatibility locus, *incA*, containing nine identical 19bp repeats is involved in P1 copy number control (Abeles *et al.*, 1985). The origin lies in a region about 245bp (Abeles *et al.*, 1984). Two DnaA boxes occur to the left of the origin and three more DnaA boxes are present about 200 bp from these, within the leader region of the *repA* gene (Abeles *et al.*, 1990). There is a group of five 7bp direct repeats, of which four contain nested GATC sequences, that are the substrates for *E. coli* DNA deoxyadenine methylase (Hattman *et al.*, 1978). These sites are separated from *incC* by a 39bp GC rich (62%) clamp sequence (Brendler *et al.*, 1991). A host protein, the product of *seqA*, recognizes the GATC methylation sites in the replicon and binds to them when hemimethylated. Hemimethylated DNA is formed immediately after replication when newly replicated DNA strands are non-methylated (Brendler *et al.*, 1995).

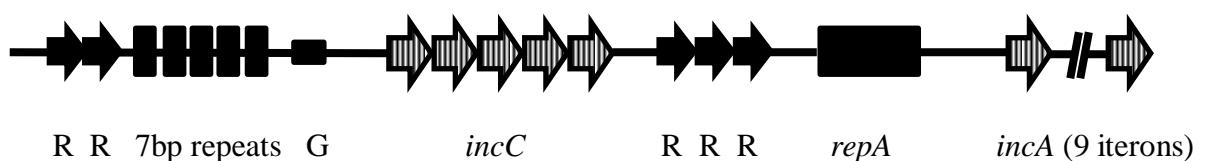


Figure 1.3 Schematic diagram showing the P1 replicon. ‘R’ represents DnaA box, the 7bp repeats contain the substrates for methylation, G represents the GC rich clamp, *incC* and *incA* contain iterons, *repA* encodes the RepA protein. The figure, not drawn to scale, is adapted from Brendler *et al.*, 1991.

RepA is essential for P1 replication (Abeles *et al.*, 1984). An amber mutation in the *repA* gene blocks replication of a mini P1 plasmid (Sternberg and Austin, 1981; Abeles

*et al.*, 1984). RepA is activated by the host-encoded heat-shock proteins, DnaJ, DnaK and GrpE, in the presence of ATP. These cause conformational changes in dimeric RepA that convert the dimer into active iteron-binding monomers (Dibbens *et al.*, 1997). RepA binds to the origin *incC* locus (Abeles *et al.*, 1989) and induces a bend in the DNA (Mukhopadhyay and Chattoraj, 1993). Melting of the DNA strands at the origin by RepA binding is stimulated by DnaA and HU (Hansen and Yarmolinsky, 1986; Ogura *et al.*, 1990). DnaA binding also aids loading of DnaB helicase (Bramhill and Kornberg, 1988; Abeles *et al.*, 1990; Wickner *et al.*, 1990). Once the DNA strands in the P1 origin of replication have separated, host-encoded DnaB helicase, RNA polymerase, DNA Polymerase III HE and gyrase are used to initiate elongation of DNA and formation of daughter strands (Wickner and Chattoraj, 1987).

The copy number of P1 is modulated by the *incA* incompatibility locus. Constructs that do not contain the *incA* locus have higher copy number than the parental plasmid (Chattoraj *et al.*, 1984). On the other hand, extra copies of *incA* on a second plasmid, or in the host chromosome, inhibit P1 replication and reduce the copy number (Austin *et al.*, 1982). When RepA binds to the *incA* repeats, the concentration of free RepA available to initiate replication is reduced, thereby limiting replication (Chattoraj *et al.*, 1984). This is known as the ‘titration model’ (Chattoraj *et al.*, 1984; Tsutsui *et al.*, 1983). The low level of initiator is maintained by autoregulation of RepA expression (binding of RepA to its own promoter prevents binding by RNA polymerase and hence prevents synthesis of more RepA).

RepA dimerization in combination with autorepression and handcuffing have been proposed to be the mechanisms underlying the ability of plasmid P1 to regulate its replication (Das *et al.*, 2005). The ability of RepA to bind simultaneously to *incC* and *incA* (Chattoraj *et al.*, 1988) enables pairing of two P1 plasmids. This is known as the ‘handcuffing model’ (Abeles and Austin, 1991; Pal and Chattoraj, 1988). The ‘handcuffing model’ results in steric hindrance whereby RepA is unable to bind to the origin to initiate replication (Park *et al.*, 2001; Pal and Chattoraj, 1988). Increase in cell volume lowers the copy number of P1 below the steady-state value (Pal and Chattoraj, 1988). Under these conditions, the concentration of active RepA is also reduced. As such, handcuffing of plasmids becomes reversible and the concentration of active RepA is increased. This allows saturation of origin iterons and replication can proceed. Replication also frees RepA bound to the RepA promoters enabling transcription of

RepA to occur (Mukhopadhyay and Chattoraj, 2000). Replication induced transcription and RepA activation by chaperones ensures a sufficient supply of RepA for replication to occur and helps to overcome handcuffing of coupled plasmids.

### 1.3.1.2 RNA mediated replication of ColE1

Replication of plasmid ColE1 is initiated by formation of an RNA primer and is independent of a plasmid-encoded Rep protein (Itoh and Tomizawa, 1980). Replication proceeds unidirectionally and can take place in extracts from cells of *E. coli* (Sakakibara and Tomizawa, 1974). The RNA transcript that is processed to form the primer of ColE1 replication, RNAIL, is about 700 nucleotides and is synthesized from approximately 555 bp upstream of the origin (Itoh and Tomizawa, 1980) (Figure 1.4, Section A step:1). It forms an RNA-DNA hybrid with the complementary DNA at the replication origin (Masukata and Tomizawa, 1984) (Figure 1.4, Section A step: 3). Formation of the stable RNAIL-DNA hybrid requires an interaction between a G-rich region in RNAIL, 265 nucleotides upstream of the 3' end and a C- rich segment of the complementary DNA, 20bp from the origin (Masukata and Tomizawa, 1990). The RNA component of this stable hybrid is processed into a primer by RNaseH (Itoh and Tomizawa, 1978) (Figure 1.4, Section A step: 5). The 3'-OH group of the processed RNA is used to prime leading-strand synthesis by DNA polymerase I (Itoh and Tomizawa, 1980) (Figure 1.4, Section A step: 6). As leading-strand synthesis continues, the template DNA unwinds exposing a primosome assembly site on the leading strand (Zavitz and Marians, 1991). About 40bp must be unwound after which DnaB helicase can be loaded. DnaB helicase and DnaG primase then work to initiate lagging-strand synthesis discontinuously with the aid of short primers (Bramhill and Kornberg, 1988). DNA polymerase I is replaced on the leading strand by DNA polymerase III gradually over about 1.3kb. DNA polymerase I processes about 20 nucleotides of the Okazaki fragments on the lagging strand (Allen *et al.*, 2011). Transition from primer RNA to DNA occurs in a region with five adenine residues. Both sides of this region are rich in G and C residues (Tomizawa *et al.*, 1977).

If RNaseH is absent, transcription continues to a point about 190 nucleotides downstream of the origin in a dA rich region (Selzer and Tomizawa, 1982). Replication

of ColE1 has however been found to occur in both RNaseH-deficient mutants (Naito *et al.*, 1984) and in DNA polymerase I mutants (Kogoma, 1984). It is possible that the formation of the RNAII-DNA hybrid unwinds the DNA (Dasgupta *et al.*, 1987; Tomizawa and Masukata, 1987) and enables loading by DnaB helicase (Bramhill and Kornberg, 1988).

Since binding of the RNAII pre-primer to the template DNA leads to initiation of leading strand synthesis, this step also is the target for control of copy number of ColE1 (Figure 1.4, section B). Formation of the active primer is inhibited by a second RNA, RNAI (Hashimoto-Gotoh and Timmis, 1981). RNAI is transcribed in a direction opposite to that of RNAII (-553 to -445bp region relative to the origin of ColE1, Figure 1.4 Section B, step: 1) (Morita and Oka, 1979). RNAI is complementary to the 5' end of the RNAII primer over 100 nucleotides (Selzer *et al.*, 1983). The RNAI-bound version fails to form a stable hybrid with the origin DNA (Tomizawa *et al.*, 1981) (Figure 1.4, Section B, step 3). Both RNAI and RNAII form three prominent stem-loop structures (Tomizawa and Itoh, 1982). Single-stranded loops of RNAI interact with complementary sequences in RNAII (Tomizawa, 1990). This initial interaction called 'kissing' becomes stronger and eventually RNAI is completely hybridized with RNAII (Tomizawa, 1984) (Figure 1.4, Section B, step: 5). However, for effective inhibition by RNAI there is a specific conformation-dependent 'window of susceptibility' (Wong and Polisky, 1985) and RNAI must bind to RNAII when the latter is between 100-360 nucleotides long (Tomizawa, 1986). The initial complex between RNAI and RNAII is unstable (Tomizawa, 1990). This unstable intermediate can be converted to a more stable complex that is resistant to RNase A in the loop region (Tomizawa, 1990) by a plasmid-encoded protein called Rom (RNA one modulator) (Tomizawa and Som, 1984). Rom is a 63 amino acid protein encoded by the *rom* gene (Tomizawa and Som, 1984), located approximately 600 nucleotides downstream of the origin of replication (Cesareni, 1982). The concerted action of RNAI and Rom negatively regulates initiation of replication and is central to copy number control of ColE1.

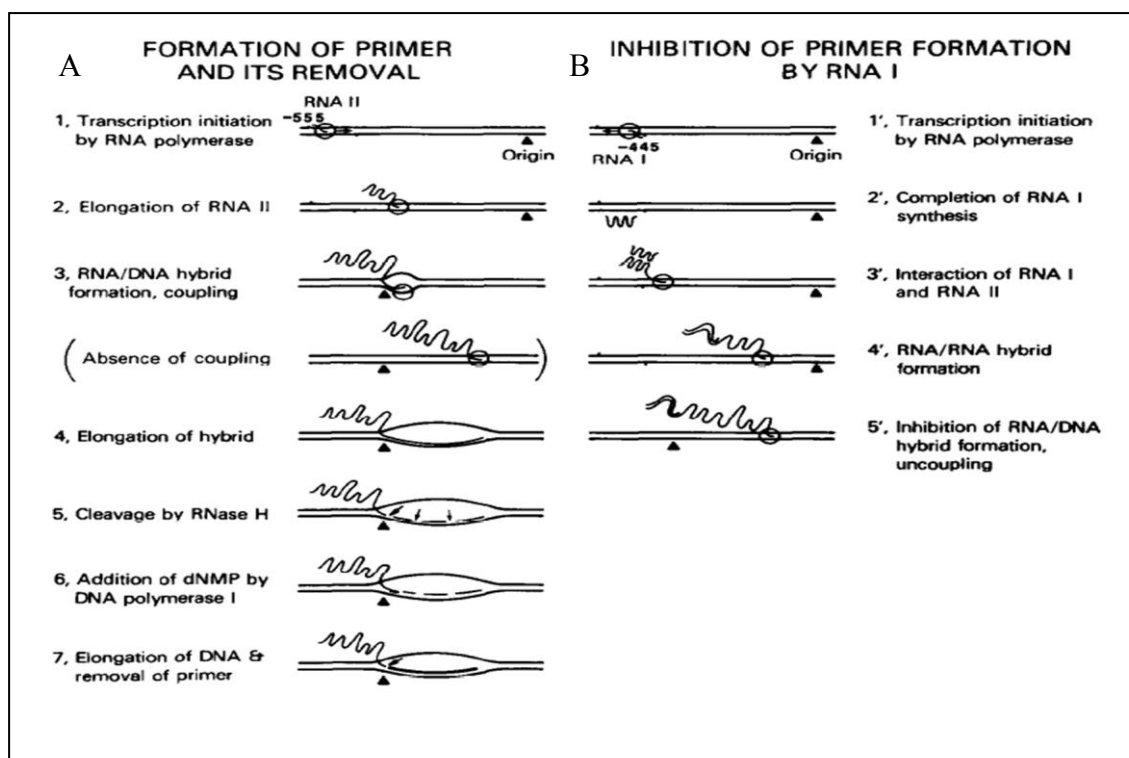


Figure 1.4. Schematic representation of ColE1 replication (adapted from Tomizawa and Itoh, 1982)

### 1.3.1.3 Rolling-circle replication of pT181

pT181 is a 4.4kb plasmid isolated from *Staphylococcus aureus* (Iordanescu and Basheer, 1991). Over many years, it has served as the paradigm for the study of Rolling Circle Replication. The basic replicon of pT181 is about 1200 bp within which a 50bp region binds the initiator protein (Khan *et al.*, 1982). Replication requires a plasmid-encoded 314 amino acid (Chang *et al.*, 2000) protein RepC (Novick *et al.*, 1982, Carleton *et al.*, 1984) that drives unidirectional (Khan *et al.*, 1982) and semi-conservative (Khan *et al.*, 1981) replication of the plasmid. Replication of pT181 is said to be asymmetric as synthesis of the leading and lagging strands are not coupled (Thomas *et al.*, 1990).

RepC exists as a monomer in solution but acts as a dimer during replication (Zhao *et al.*, 1998). This dimer is designated RepC/RepC (Rasooly *et al.*, 1994b). Replication starts with binding of RepC to the *dso* region of the plasmid (Figure 1.5). RepC has sequence-specific DNA binding, topoisomerase-like activities and endonuclease activity (Koepsel *et al.*, 1986). Residues 267-270 of RepC are involved in sequence-specific recognition



of the origin (Dempsey *et al.*, 1992; Wang *et al.*, 1992). The RepC binding site includes an inverted repeat element IRIII and the right arm of a second inverted element, IRII. Binding by RepC bends the DNA and forms a cruciform structure containing the IRII repeat element (Koepsel and Khan, 1986; Noirot *et al.*, 1990). RepC cleaves this structure between nucleotides 70 and 71 (Koepsel *et al.*, 1985) with its active tyrosine residue, Tyr-191 (Chang *et al.*, 2000) (Figure 1.5, A). A tyrosine residue of RepC exerts a nucleophilic attack on the phosphodiester bond of a dinucleotide 5'-AT-3' within the *nic* region of *dso* (Rasooly *et al.*, 1994a). One of the monomers of the RepC/RepC dimer becomes covalently linked to the 5'-end of the nick through a phosphotyrosine bond (Koepsel *et al.*, 1985; Thomas *et al.*, 1988) (Figure 1.5, B). The 3'-OH end acts as a primer for elongation of the leading strand (Koepsel *et al.*, 1985). Host factors are used for the synthesis of the leading strand including DNA polymerase III that polymerizes DNA from the 3'-OH end, single-strand DNA binding protein (SSB) that protects the displaced single stranded DNA from nucleolytic attack, and DNA helicase that unwinds the DNA. After the replisome reaches the *nic* site, elongation of the leading strand continues for about 10 nucleotides to regenerate the *dso* region (Figure 1.5, C). The RepC subunit of the dimer not bound to DNA then cleaves the newly synthesized *nic* site through its active tyrosine residue. This leads to binding of 10-12 nucleotides of the 3' half of the *nic* region to the RepC subunit (Rasooly and Novick, 1993) (Figure 1.5, D). The dimeric RepC-RepC now contains an active RepC that initiated replication and a RepC inactivated by oligonucleotide attachment, termed RepC\* (Rasooly and Novick, 1993). This heterodimer RepC/RepC\* can bind DNA but cannot relax plasmid DNA. As a result, it is inactive as an initiator of replication (Rasooly and Novick, 1993, Rasooly *et al.*, 1994b). Following completion of leading strand synthesis, the products are a single-stranded DNA (the displaced parental strand) and a double-stranded DNA that has a newly synthesized leading strand and a parental strand. The latter is supercoiled by DNA gyrase (te Riele *et al.*, 1986) (Figure 1.5, E).

The single-stranded DNA intermediate is converted to double-stranded DNA through lagging-strand synthesis. This initiation occurs from the *sso* region of the plasmid, located within a noncoding region capable of forming secondary structures (Khan, 1997). Replication of lagging strand is believed to be initiated by RNA polymerase. Elongation of lagging strand is achieved by host DNA polymerase III. Finally the host DNA polymerase I removes the primer RNA (Espinosa *et al.*, 1995) (Figure 1.5, F).

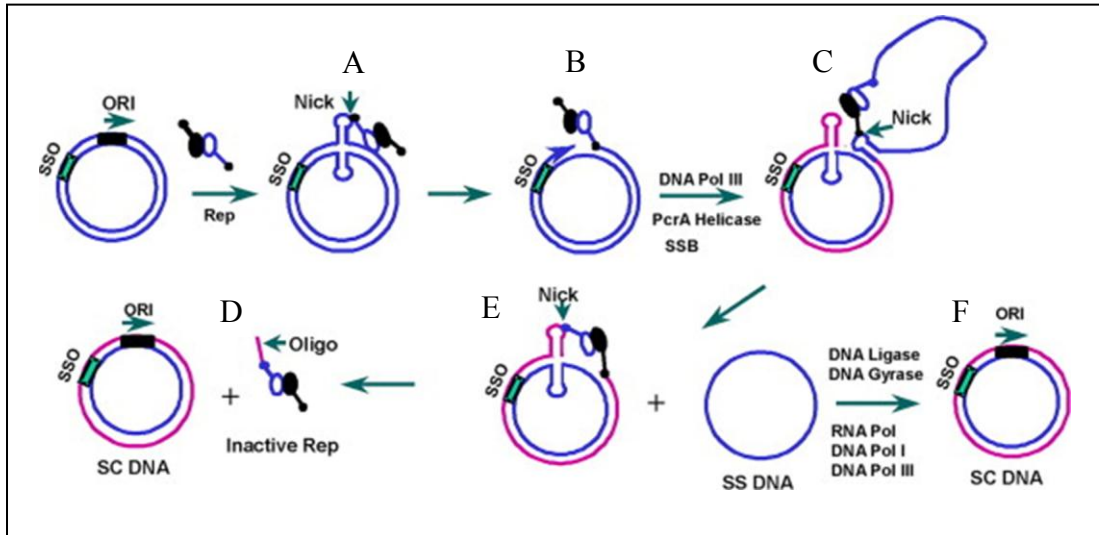


Figure 1.5. Schematic representation of pT181 replication. Source: Khan, 2005.

Replication of pT181 is regulated by anti-sense RNA that controls synthesis of RepC (Kumar and Novick, 1985). The antisense RNA or counter-transcript transcribed from the *cop* region (Rasooly *et al.*, 1994a) is complementary to the leader region of RepC-mRNA. It is capable of forming extensive secondary structures that contain hairpin regions (Novick *et al.*, 1989). It initially binds with the unpaired regions of RepC mRNA. This initial binding is converted to a linear duplex between the countertranscribed RNA and RepC mRNA (Novick *et al.*, 1989). Binding of ctRNA with RepC mRNA inhibits translation of the latter (Rasooly *et al.*, 1994a). The ctRNA is synthesized constitutively. Consequently, full-length Rep mRNA is synthesized in limited concentrations. Lower level of Rep mRNA expression reduces the concentration of the Rep initiator. This leads to lower replication rates and thereby regulates replication (Novick *et al.*, 1989).

### 1.3.2 Replication of linear plasmids

Linear plasmids belong to one of two structural categories: those with proteins attached at 5' termini (e.g. *Streptomyces* linear plasmids) and those that have covalently closed ends (*Borrelia* linear plasmids). Although the components of the replicons of several *Streptomyces* linear plasmids have been studied in the past (Section 1.3.2.1 to follow), the mechanism of replication, particularly replication at the ends, is still a mystery.

Limited studies have been conducted on replication of *Borrelia burgdorferi* linear plasmid lp17 and mechanism for telomere resolution proposed (Section 1.3.2.2).

### 1.3.2.1 Replication of linear plasmids with terminal proteins

Replication of protein-capped linear plasmids occurs bidirectionally from an internal origin (Chang and Cohen, 1994). The autonomously replicating sequence of *Streptomyces* linear plasmids generally contains an origin of replication and one or more genes required for replication (Huang *et al.*, 2003). These genes are located near the origin and may be subsidiary in some cases (Hiratsu *et al.*, 2000). Replication generates a 3' leading-strand overhang at the telomeres (Figure 1.6). The lagging strand is about 280bp short at its 5' terminus (Chang and Cohen, 1994). It is proposed that end patching occurs by one of two mechanisms: the first (Figure 1.6 A) assumes that pairing of palindromes in the 3' overhang causes it to fold back such that the terminal protein is anchored near the base of the overhang. The terminal protein (TP) acts as a primer and helps to fill the gap (Qin and Cohen, 1998). In the second model (Figure 1.6 B), TP acts as a nickase and nicks the template strand and attaches covalently to the 5' end. The gap is filled by DNA polymerase (Qin and Cohen, 1998). *Streptomyces* chromosomes are linear (Lin *et al.*, 1993) and contain telomeres similar to those present on *Streptomyces* linear plasmids. They replicate bidirectionally and it is suggested that telomeres are patched in similar manner as their linear plasmids (Chen, 1996; Huang *et al.*, 1998; Qin and Cohen, 1998). The product of the *tpg* gene is required for *Streptomyces rochei* chromosomal replication (Bao and Cohen, 2001).

In linear plasmid pSLA2, the autonomously replicating sequence lies near the centre of the plasmid. It contains two 21-mer iterons followed by a series of CT and AG residues within the *rep1* gene that codes for the Rep1 DNA binding protein which promotes plasmid replication (Huang *et al.*, 2003). Another gene downstream of *rep1*, *rep2*, encodes a protein with a helicase-like activity. Additional genes necessary for pSLA2 replication are the *tpgs*, terminal protein genes (Bao and Cohen, 2001), *tap*, the gene for a terminus associated protein (Bao and Cohen, 2003) loci required for linear replication (*rlrA/rorA*) (Qin *et al.*, 2003). The presence of *rlrA*<sup>pSLA2</sup> in a circular derivative is inhibitory to its replication. The presence of *rlrA*<sup>pSLA2</sup> is not however inhibitory for replication in the linear form (Qin *et al.*, 2003). The inhibitory effect of *rlrA*<sup>pSLA2</sup> in the

circular derivative can be overcome by another locus, *rorA*<sup>pSLA2</sup> (Kendall and Cohen, 1987; Stein *et al.*, 1989). Copurification of DNA polymerase I (PolA) and a topoisomerase I with the Tap protein suggests a role in replication. However, DNA polymerase I has not been found to be essential for *Streptomyces* linear plasmid replication (Tsai *et al.*, 2008). Tap has been demonstrated to possess reverse transcriptase activity (Bao and Cohen, 2004).

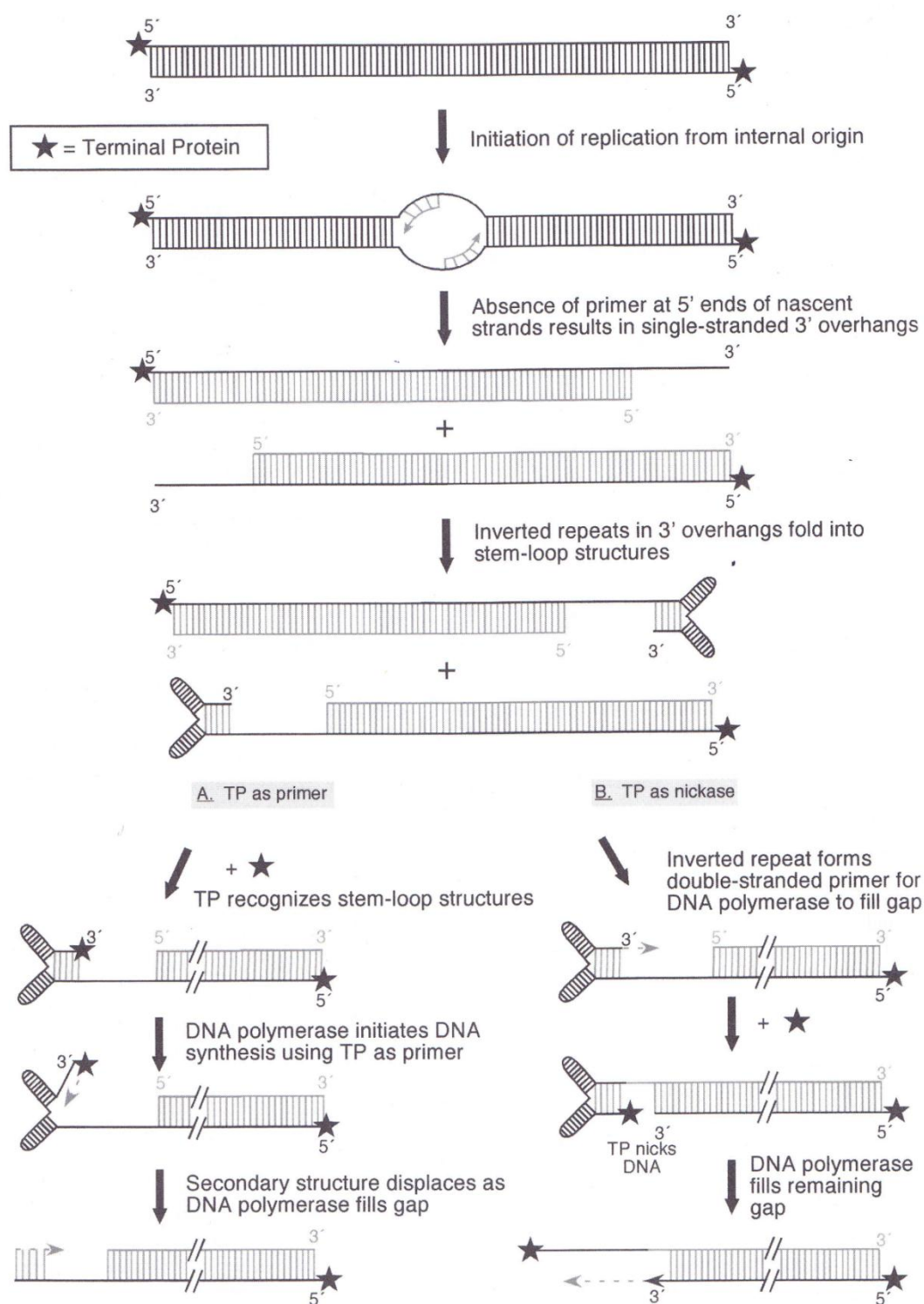


Figure 1.6. Models for replication of linear plasmids containing 5' terminal proteins (Taken from Phillips, G. and Funnell, B.G. (2004) *Plasmid Biology*. 1<sup>st</sup> Edition, ASM Press.)

In the linear plasmid SCP1, the minimum replicon is within a 5.4kb region that contains an A+T rich region and two ORFs. One of these codes for a protein that has similarity to the primase *repA* gene of *Acidianus ambivalens* and the helicase-like protein of *Sulfolobus islandicus* (Redenbach *et al.*, 1999). The A+T rich region may facilitate melting of the double-stranded DNA to initiate DNA replication. The autonomously replicating sequence (ARS) of linear plasmid SLP2 is located within approximately 5 to 10kb sequence from the left end. The ARS contains the *tpg* gene, the *dnaB*-like gene and two DnaA boxes (T/C) (T/C) (A/G/C) TCCACA (Jakimowicz *et al.*, 1998). One of these DnaA boxes is located within the *tpg* coding sequence and the other with the *dnaB*-like coding sequence. The minimal replicon of SCP1 contains an ORF (ORF SCP1.196) that codes a putative primase/helicase and a second overlapping ORF (ORF SCP1.197) that codes for a hypothetical protein and direct repeats (Redenbach *et al.*, 1999). SCP1 lacks *tpg* and *tap* homologues.

The basic replicon of plasmid pSCL1 is similar in configuration to that of pSLA2 (Wu and Roy, 1993). In the case of the linear plasmid SLP2, efficient replication as a circular derivative requires a 47bp sequence upstream of 23-mer iterons. Inclusion of these sequences increased the transformation efficiency by about 1000 fold (Xu *et al.*, 2006). This region is not transcribed as shown by RT PCR and hence may contain regulatory factor(s) that act *in cis* (Xu *et al.*, 2006). SLP2 contains two *tpg* homologues – *tpg* SLP2 and a putative pseudogene on the right arm. SLP2 requires *tpgL* gene product for replication (Lee *et al.*, 2003). SLP2 lacks a *rep* gene (Huang *et al.*, 2003). It possibly requires a host DNA polymerase for DNA replication. It has been proposed (Huang *et al.*, 2003) that the origin of replication of the plasmid lies close to the *tpg* gene, and the DnaA boxes.

A Rep-iteron structure is also found in the minimal replicon of plasmid pSHK1 of *Streptomyces* sp. HK1 (Zhang *et al.*, 2009). The gene for a putative Rep protein and adjacent noncoding sequences (*ncs*) form the basis of the basic replicon in the linear plasmids pSLA2-L from *Streptomyces rochei* (Hiratsu *et al.*, 2000), pSCL2 from *S. clavuligerus* (Wu *et al.*, 2006), pRL1 from *Streptomyces* sp. 44030, pRL2 from *Streptomyces* sp. 44414, and pFRL2 from *Streptomyces* sp. FR1 (Zhang *et al.*, 2009).

Unlike any other linear plasmid reported so far, two origins of replication have been identified in the linear plasmid pFRL1 from *Streptomyces sp.* FR1 (Zhang, 2010). One of these, *rep1A-ncs1*, lies adjacent to the telomere and the other, *rep2A-ncs2*, lies about 10kb away from it. The first origin can propagate both in linear and circular modes, but the latter requires an additional locus, *rlrA-rorA*, for propagation in the linear mode. The Rep1A protein binds to the sequence *ncs1* *in vitro*. The proteins Rep1A and Rep2A were transcribed at different levels at different times and each dominated at different points of time (Zhang *et al.*, 2010).

Rep-iteron structure is also found in *Mycobacterium celatum* linear plasmid pCLP. The basic replicon is a 2.8kb fragment comprising of a putative Rep gene and a putative origin of replication consisting of 18bp iterons and an AT-rich region (Picardeau *et al.*, 2000). Linear plasmid pRHL3 from *Rhodococcus jostii* (McLeod *et al.*, 2006) minimal replicon is also similar to that of pCLP. It contains two ORFs: RHL3.237 encoding Rep1 (Rep protein) and RHL3.235 that codes for a protein homologous to a ParA protein from *R. erythropolis* (De Mot *et al.*, 1997). In this plasmid, iterons are present as control mechanisms. An approximately 40% AT-rich 600bp region is found upstream of *rep1* (Warren *et al.*, 2004).

### 1.3.2.2 Replication of linear plasmids with closed ends

*Borrelia* linear plasmids contain covalently closed ends. Replication of *Borrelia* linear plasmids initiates from an internal origin and proceeds bidirectionally (Picardeau *et al.*, 1999) (Figure 1.7). A head-to-head dimer in which the telomere is duplicated is generated. The duplicated telomere is recognized by a telomere resolvase, ResT (Chaconas and Kobryn, 2010; Chaconas, 2005; Kobryn and Chaconas, 2005), which resolves the dimer into two covalently closed linear plasmids. ResT causes breakage of DNA, causes a conformational change by which the 5'-OH group of the DNA and the 3'-phosphate of a tyrosine residue in ResT are juxtaposed. The 5'-OH group exerts a nucleophilic attack on the 3'-phosphate group leading to the formation of a phosphodiester bond. This structure stabilizes the hairpin ends.

In *Borrelia burgdorferi* linear plasmid, lp17, the basic replicon is 1.8kb in length and contains one ORF designated BBD14 (Casjens *et al.*, 2000) as well as adjacent non-

coding sequences, 311bp and 360bp to the left and right, respectively. Expression of ORF BBD14 is required for replication (Beaurepaire and Chaconas, 2005). However, no iterons or DnaA boxes were detected in lp17 basic replicon. The exact location of the origin is not known. The gene for ResT in *Borrelia burgdorferi* is not located on a linear plasmid but on a 26-kb circular plasmid (Kobryn and Chaconas, 2002). ResT recognizes the 25kb inverted repeats for its function. It has been shown that circular plasmids of *B. burgdorferi* can replicate as linear derivatives when telomeres are added to the termini (Chaconas *et al.*, 2001). Conversely, linear plasmids have also been demonstrated to replicate as circular derivatives in circular vectors (Stewart *et al.*, 2003).

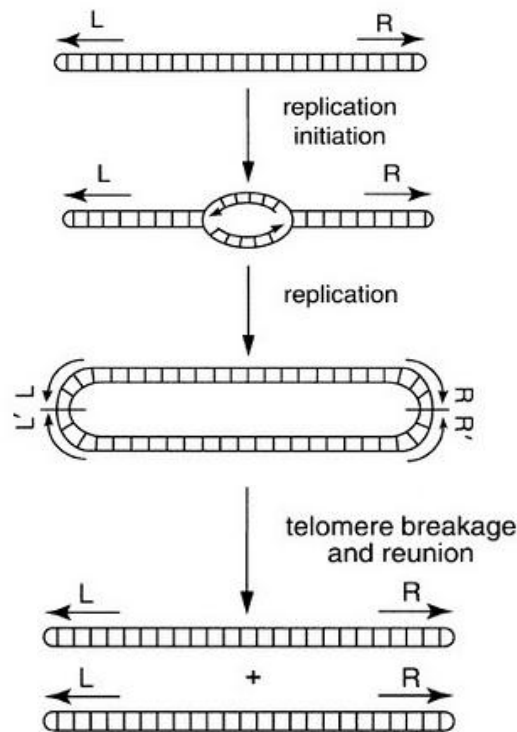


Figure 1.7. Bidirectional replication of *Borrellia* linear plasmid from a central origin. L and R indicate left and right telomeres, respectively. L' and R' are replication-generated telomeres. ResT resolves the telomeres and generates two monomer plasmids. The figure has been taken from Chaconas and Kobryn, 2010.

The basic replicon of the linear prophage N15 is 5.2kb and can drive replication of linear or circular derivatives. An ORF, *repA*, which codes for a protein with similarity to primase is required for replication (Rybchin and Svarchevsky, 1999). ResT of *E. coli*



prophage N15 is encoded by the *telN* gene located adjacent to its recognition site (*telRL*) (Deneke *et al.*, 2004, 2002, 2000).

## 1.4 General features of plasmid stability

Plasmid replication ensures plasmid duplication in the host but for efficient distribution to occur to daughter cells at cell division, plasmids also encode mechanisms for their proper partitioning to daughter cells at division. The basic replicon of a plasmid carrying the stability cassette is inherited more stably and with a much greater frequency than when the basic replicon is on its own (Gerdes *et al.*, 1985).

### 1.4.1 Random distribution of high copy number plasmids

Most high copy number plasmids are assumed to be distributed randomly to daughter cells. The probability that one daughter cell will not receive a particular plasmid is 0.5. If the copy number of the plasmid is 'n', then for all plasmids this probability becomes  $(0.5)^n$ . As two daughter cells are produced upon cell division, the probability that either one of them will not receive any plasmid is  $2(0.5)^n = 2^{(1-n)}$  (Summers, 1991). For high copy number plasmids, this probability becomes very small which means that each daughter cell is likely to receive at least one copy of the plasmid (Summers, 1996). Factors that reduce copy number, viz. plasmid metabolic load, plasmid clustering and oligomer formation, reduce the number of independent plasmids that can be segregated and hence affect plasmid stability.

#### 1.4.1.1 Plasmid loss and metabolic load

The loss of a plasmid from a cell can be due to either structural instability or segregational instability. When a plasmid is structurally unstable, loss or rearrangement of its sequence takes place (Summers, 1991). Loss of part of the plasmid DNA, or of part or the entire foreign gene inserted in the plasmid can occur owing to metabolic load (Harrington *et al.*, 1986). The cell with reduced load gets selective advantage. Metabolic load can decrease the rate of host cell growth after introduction of the plasmid (Summers, 1991; Kyslik *et al.*, 1993). Plasmid metabolic load on its host takes

place since the plasmid requires energy and resources for its maintenance. Both the size and the copy number of a plasmid affect the metabolic load imposed on the host. Metabolic load can be defined as the portion of the host cell's resources required for the maintenance of the plasmid. Resources can include energy in the form of ATP or GTP, or cellular biomolecules such as nucleotides and amino acid. The larger the size of the plasmid, the greater is the proportion of cellular energy and resources required for its maintenance (Cheah *et al.*, 1987; Ryan *et al.*, 1989; Khosravi *et al.*, 1990).

Copy number affects metabolic load in a similar way to plasmid size. The higher the copy number of the plasmid, the greater the proportion of energy and resources required to maintain the plasmid (Seo and Bailey, 1985; Birnbaum and Bailey, 1991).

#### **1.4.1.2 Plasmid clustering**

Plasmids have been reported to cluster within host cells (Eliasson *et al.*, 1992). Clustering of plasmids would be expected to affect their proper distribution to daughter cells. This is because the number of plasmid clusters has been found to be less than the copy number of the plasmid (Gordon *et al.*, 2004; Pogliano *et al.*, 2001; Weitao *et al.*, 2000). A model has been proposed to resolve this problem by Nordstrom and Gerdes (2003) (Figure 1.8). It is suggested that one plasmid is recruited at a time from a cluster to the centre of the host cell where the replication factory is present (Figure 1.8, A). Recruitment will therefore decluster plasmids. After replication (Figure 1.8, B-C), the daughter plasmids are able to diffuse freely in the cell cytoplasm, increasing the probability that each daughter cell receives a plasmid (Figure 1.8, D). Thus it is not the plasmid cluster that is the unit of segregation but it is the individual plasmid copies.

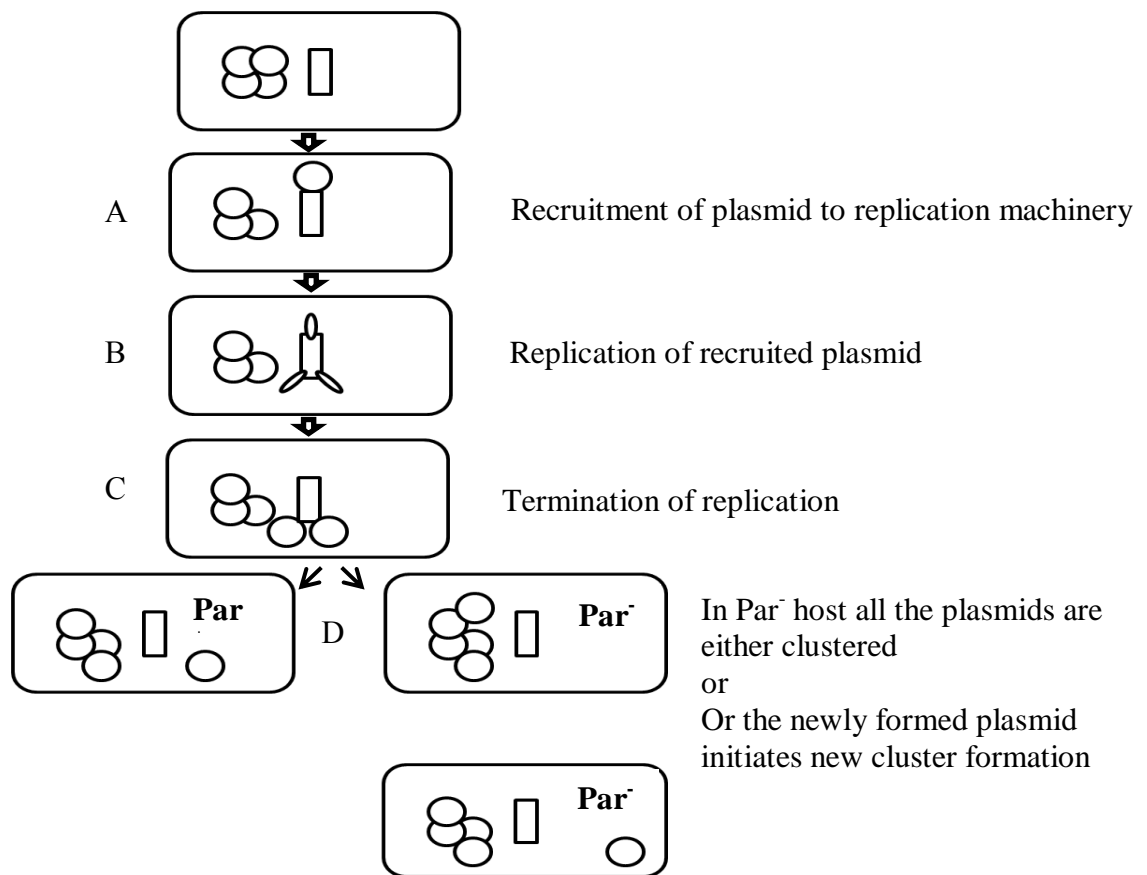


Figure 1.8: Plasmid clustering. The ultimate fate of plasmids in the cluster will vary between  $\text{Par}^+$  and  $\text{Par}^-$  cells (Adapted from Nordstrom and Gerdes, 2003)

#### 1.4.1.3 Plasmid oligomer formation and resolution

Plasmid dimers are less stable than monomers (Summers and Sherratt, 1984). Oligomers arise as a result of recombination (Patient and Summers, 1993). An oligomer of the order of  $n$  has a copy number  $1/n$  times the copy number of the monomer (Summers, 1996). This means that there are fewer plasmid entities to be distributed by random distribution. Moreover, since plasmid replication control elements count origins, dimers are considered as two plasmids instead of one (Summers and Sherratt, 1984). This means that dimers quickly out-replicate monomers in a cell. The low copy number of dimers means the rate of plasmid loss increases significantly. This phenomenon is known as ‘dimer catastrophe’ (Summers *et al.*, 1993). Multimer containing cells grow slowly, giving cells the time to overcome the issue through

multimer resolution systems. Many plasmids possess oligomer resolution sites including ColE1 (Summers and Sherratt, 1984) and pMB1 (Summers and Sherratt, 1988; Greene *et al.*, 1981). In ColE1, the oligomer resolution site is known as *cer* and is 240 bp in length. Site-specific recombination at this site converts oligomers to monomers (Summers and Sherratt, 1984). Oligomer resolution requires the products of four genes: *argR* (the repressor of arginine biosynthesis; Stirling *et al.*, 1988), *pepA* (aminopeptidase A; Stirling *et al.*, 1989), *xerC* (Colloms *et al.*, 1990) and *xerD* (Blakely *et al.*, 1993). XerC and XerD form a heterodimeric recombinase that mediates strand exchange resulting in resolution of oligomers.

Although necessary, multimer resolution is not sufficient for plasmid stability. Inactivation of the promoter of a short transcript, Rcd that lies within the *cer* site reduces ColE1 stability (Patient and Summers, 1993). Upon formation of plasmid dimers, Rcd is expressed that inhibits cell division (Patient and Summers, 1993) allowing dimers to be converted to monomers by site-specific recombination before the cell divides. The expression of Rcd is then reduced enabling cell division to occur and normal growth to resume (Summers, 1996).

## 1.4.2 Low copy number plasmids

Unlike high copy number plasmids, random distribution is not sufficient to ensure proper segregation of plasmids with fewer than approximately 20 copies in the dividing cell (Summers, 1996). Low copy number plasmids have developed dedicated functions that ensure their maintenance in the host population. One common mechanism is active (i.e. energy-consuming) partitioning of plasmids to both daughter cells at division. Another ensures that plasmid-free cells are killed so that they do not proliferate and take over the population.

### 1.4.2.1 Plasmid partitioning

The P1 plasmid which has been the subject of detailed investigation is used here as a representative low copy number plasmid that employs active partitioning for segregation to daughter cells. The idea that P1 is a low copy number plasmid may need revising, however, after the publication of a recent paper by Sengupta *et al.* (2010).

The P1 partitioning locus is known as *par*. This contains information for two proteins, ParA (598aa), ParB (333 aa) (Surtees and Funnell, 2003; Hayes and Barilla, 2006) and a centromere-like site, *parS* (Figure 1.9).

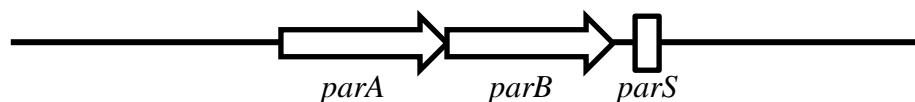


Figure 1.9. A schematic diagram of the P1 partition site. (Taken from Hao and Yarmolinsky, 2002)

The *parS* site is approximately 80 bp in length and is located downstream of the *parAB* operon (Austin and Abeles, 1983). *parS* contains two recognition sites for ParB (Davis *et al.*, 1990; Fung *et al.*, 2001). A schematic diagram of the molecular architecture that develops at the P1 *parS* site is shown in figure 1.10. The *parS* sequence motifs are arranged around an *E. coli* Integration Host Factor (IHF) binding site (Funnell, 1988, 1991). IHF increases the affinity of ParB for *parS* (Funnell, 1991). Initially IHF binds to the *parS* site and induces a bend in DNA that enables ParB binding (Funnell, 1991; Funnell and Gagnier, 1993).

ParB binds as a dimer (Funnell, 1991; Bouet, *et al.*, 2000) through recognition of the inverted repeats within *parS* (Lobocka and Yarmolinsky, 1996; Surtees and Funnell, 2001). It contains a DNA-binding region at its C-terminus that recognizes sequences within the *parS*. The region of ParB that is required for recognition of ParA lies close to the N-Terminus of ParB (Surtees and Funnell, 2001).

ParA is an ATPase (Davis *et al.*, 1992). Binding of ATP and its hydrolysis are important for partitioning to take place (Bouet and Funnell, 1999; Davis *et al.*, 1996; Fung *et al.*, 2001). It autoregulates its own expression by binding to its own promoter (Davis *et al.*, 1992). The presence of *parS* decreases expression from the *par* promoter.

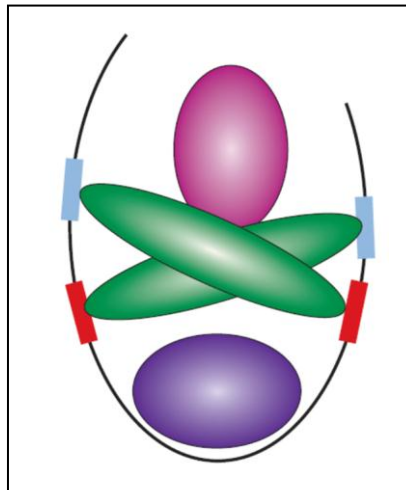


Figure 1.10: A schematic diagram showing the molecular assembly at the P1 *parS* site. IHF (purple) binds to the *parS* region and induces a bend in the DNA. ParB (Green) then binds the heptameric (red) and hexameric (blue) repeats located in the opposite arms of the *parS* region. ParA-ATP (pink) then binds to ParB. The figure is taken from Hayes and Barilla, 2006.

ParA comes into action following binding of ParB at *parS*. It interacts with ParB in the presence of ATP (Bouet and Funnell, 1999). However, ADP is a better cofactor for ParA binding than ATP (Davey and Funnell, 1994). Alternatively, it has been proposed that ParA alters conformation of ParB such that the host interaction domain of ParB is exposed for function (Philips and Funnell, 2004). ParB stabilizes the ADP-bound form of ParA by converting ParA-ATP to ParA-ADP (Philips and Funnell, 2004).

An interpretation of how P1 plasmid partitioning takes place is shown in Figure 1.11 (Sengupta *et al.*, 2010). When P1 replicates, the partitioning sites are paired through ParB (Figure 1.11, b). ParA separates two paired plasmids through filaments using the energy from ATP hydrolysis (Figure 1. 11, c). The plasmids are finally widely separated by diffusion (Figure 1. 11, e). When multiple copies of P1 are produced, pairing and separation events ensure that plasmids are evenly distributed throughout the cytoplasm (Figure 1. 11, f-k).

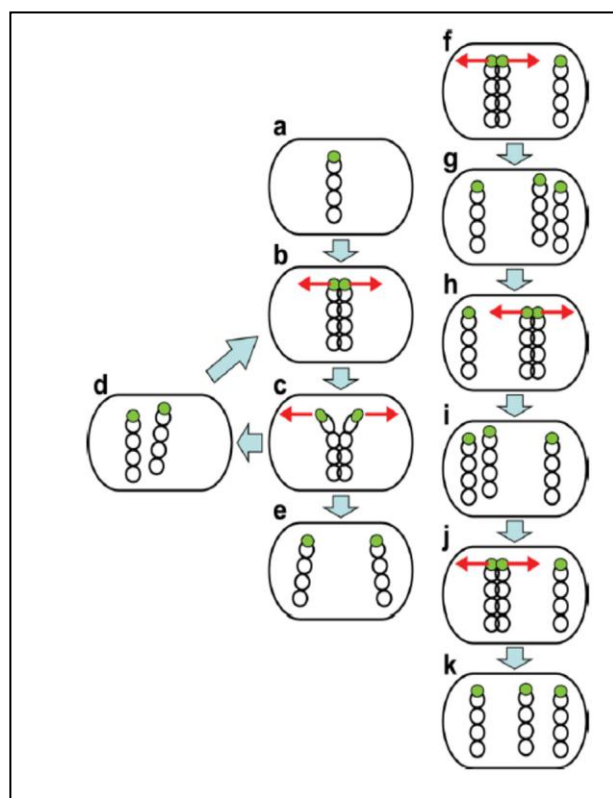


Figure 1.11. Interpretation of the mechanisms of P1 partitioning (Sengupta *et al.*, 2010). The green dots indicate parS sites with bound ParB. The red arrows represent ParA filaments.

#### 1.4.2.2 Plasmid stabilization through the toxin antitoxin system

Plasmid stability systems involving toxin and antitoxin components ensure selective killing of cells that fail to retain the plasmid following cell division (Jaffe *et al.*, 1985; Gerdes *et al.*, 1986). Unlike a plasmid partitioning system, the toxin-antitoxin system stabilizes a plasmid through the inhibition of growth or killing of cells rather than through ensuring equal distribution of plasmids to daughter cells prior to cell division (Gerdes, 1988). The term ‘Postsegregational killing’ is derived from Gerdes (1986).

Toxin-antitoxin systems kill host cells from inside rather than from the outside and are thus distinct from the operation of colicins (Nomura, 1964). The names postsegregational cell killing (PSK) and addiction systems are sometimes used interchangeably with the term toxin-antitoxin system (Hayes, 2003). Restriction-

modification systems are also considered as part of postsegregational stability systems, in which the restriction enzyme acts as the toxin (Kobayashi, 2001).

Toxin-antitoxin loci in bacteria belong to seven two-component gene families and one three-component system (reviewd by Gerdes, 2005). In some host-killing gene cassettes, the gene for the antitoxin overlaps the gene for the toxin and their synthesis is autoregulated. In others, the gene for the antitoxin alone regulates production of the toxin. In yet other systems, the product of a third gene is necessary for regulation (Engelberg-Kulka and Glaser, 1999; Gerdes, 2000). Toxins belonging to the postsegregational killing systems are usually stable proteins. The antitoxin may be either a labile protein or an untranslated counter-transcript. In Type I postsegregational killing system the countertranscript binds to the toxin mRNA and prevents translation of the toxin. In Type II systems, the protein antitoxin binds specifically with the toxin and prevents it from exerting any toxic effect on the host (reviewed by Hayes, 2003).

Toxin-antitoxin systems have been proposed to provide competitive advantage of plasmids with (TA+) and without (TA-) systems when co-infection of the same host takes place (Cooper and Heinemann, 2000, 2005). Cells that inherit the TA+ plasmid are protected from the toxin owing to the continued production of the antitoxin by the plasmid. Cells in which only the TA- plasmid is present are killed by the presence of the toxin produced by the TA+ plasmid (Cooper and Heinemann, 2005). Many plasmids have multiple toxin-antitoxin systems. A well-known example is plasmid R1 which contains the *hok-sok* system (Gerdes *et al.*, 1986) and the *kis-kid* system (Bravo *et al.*, 1987). Multiple toxin-antitoxin systems in the same plasmid have been suggested to increase the chances that one plasmid will have one system that its competitor does not. Hence, it increases fitness of the cell containing the plasmid with the multiple toxin-antitoxin systems (Cooper and Heinemann, 2010).

If a daughter cell fails to receive the plasmid owing to an error in replication of the plasmid or owing to any defect in the plasmid stability system, the toxin causes death of the host or inhibits cell growth. The toxin remains longer in the host cell because it is degraded relatively slower when compared to the antitoxin. The labile antitoxin is degraded by host enzymes. As long as the plasmid is present, more antitoxin continues to be synthesized. This leads to the maintenance of only those cells that still retain the plasmid and hence contain both the toxin and its antidote.



The toxin-antitoxin system of plasmid R1 has been studied in detail. The toxin-antitoxin locus of plasmid R1 is about 580 bp in length and contains three genes, *hok* (host killing), *mok* (modulator of killing) and *sok* (suppressor of killing) (Gerdes *et al.*, 1985, 1997; Gerdes, 1986) (Figure 1.11, A). The *hok* gene expresses a small, hydrophilic polypeptide of 52 amino acids (Rasmussen *et al.*, 1987). The *sok* gene lies upstream of *hok*, is 100bp long and is transcribed in a direction opposite to that of the *hok* toxin gene (Figure 1.11, A). The antisense RNA, *sok*, is trans-acting and its 3' end is necessary for proper function (Gerdes, 1988). It forms a single stem-loop structure (Dam and Gerdes, 1997).

Translation of the polypeptide mediator, Mok, is coupled with that of Hok (Figure 1.12, A) and the translation of *mok* mRNA is tightly regulated by *sok* RNA (Thisted and Gerdes, 1992). The single mRNA that codes for Hok and Mok is initially folded at its 3' end such that the Shine-Dalgarno sequence for *mok* (SD*mok*) is not available to initiate translation of *mok* (Figure 1.12, B). This closed structure is known as *fbi* (Fold-back inhibition). Consequently, neither *mok* nor *hok* is translated. In plasmid containing cells, this is where the suppressor of killing, *sok*, comes into play. *sok* is complementary to *sokT*, a sequence that is located close to SD*mok*. When the *sok-mok* mRNA refolds as a result of 3' end processing, *sok* can bind to *sokT* in the *mok-hok* transcript (Figure 1.12, C). The double-stranded RNA formed as a result of *sok-mok* binding is processed by RNaseIII leading to its decay. Thus, *hok* is not translated and cell death is prevented (Gerdes, 1988; Franch *et al.*, 1997). In plasmid free cells, partial processing of the 3' end changes the folding pattern that exposes the SD*mok* (Figure 1.1, D). Translation of *hok* and its mediator, *mok*, results in the production of Hok toxin (Thisted *et al.*, 1995, 1994; Franch and Gerdes, 1996) that can potentially kill cells through collapse of the cell membrane potential and inhibition of respiration (Gerdes, 1988).

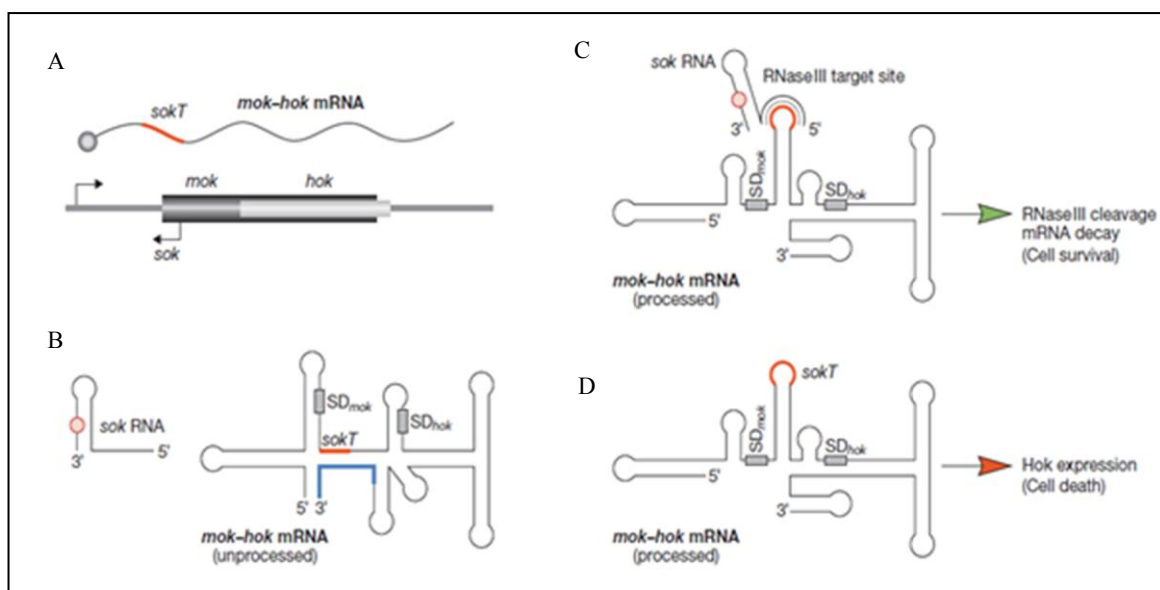


Figure 1.12: The *parB* post-segregational killing system of plasmid R1. Source: De La Cueva-Mendez and Pimentel, 2007.

**A.** The *mok-hok-sok* *parB* postsegregational system. **B.** *mok-hok* mRNA initially folds into a secondary structure in which the target of *sok* antisense RNA, *sokT*, is hidden. Processing of the 3' end of this structure exposes *sokT*. **C.** In cells containing R1, *sok* binds to *sokT* of *mok-hok* mRNA and induces RNase III decay of the resultant double-stranded RNA. **D.** In cells lacking R1, *hok* expresses toxin leading to host cell death. SD = Shine Dalgarno sequence, *mok* = mediator of killing, *hok* = host killing and *sok*= suppressor of killing.

## 1.5 Aims and objectives of the present investigation

The linear plasmid, pBSSB1, was identified from *S. Typhi* (Baker *et al.*, 2007b). Since its initial identification, only basic characterization based on bioinformatic approaches has been carried out. Based on such approaches, the putative origin of replication has been determined and ORFs identified. No laboratory based approach has yet been conducted to study the replication and stability functions of pBSSB1. The aim of the present study is to characterize regions involved with these functions in the linear plasmid.

An initial objective of the study is also to develop a method for the purification of the linear plasmid, based on the fact that no standard protocol for the purification of linear plasmids in general is available in the literature. Another objective is to identify the basic replicon (the minimum region required for replication of the linear plasmid). Finally, I will seek to identify regions necessary for stabilization of the plasmid. pBSSB1 has been shown to replicate and to be stably maintained in *E. coli*. To simplify the experimental approach I will attempt, in the first instance, to identify regions of pBSSB1 which confer competence for replication and stable maintenance upon circular *E. coli* plasmids.

Investigation of replication and stability functions of pBSSB1 can help in understanding the biology of linear plasmids in general. Results from the present work can form the basis for future work for a better understanding of specific mechanisms involved in these functions in pBSSB1. Development of a method for the purification of pBSSB1 can introduce a routine method for the extraction of linear plasmids in general.

# **Chapter 2**

---

## **Materials and methods**

---

## 2.1 Solutions

All solutions were prepared using sterile distilled water. Unless otherwise stated, chemicals were purchased from Sigma-aldrich ([www.sigmaaldrich.com/united-kingdom.html](http://www.sigmaaldrich.com/united-kingdom.html)).

### 100 mM CaCl<sub>2</sub>

Anhydrous CaCl<sub>2</sub> (1.11g) was dissolved in 100 ml sterile distilled water. The solution was filter sterilized using Millipore type VS 0.025 µm filter disc before use.

### 0.5 M EDTA

Disodium ethylenediaminetetraacetate dehydrate (186.12g) was added to 800 ml of distilled water. While stirring, the pH was adjusted to 8.0 by adding sodium hydroxide pellets. The final volume was adjusted to 1L and the solution was autoclaved at 121°C at 15 p.s.i for 20 min.

### 3 M Sodium acetate (pH 5.2)

Sodium acetate (246.09g) was dissolved in sterile 800 ml dH<sub>2</sub>O. The pH was adjusted to 5.2 by using glacial acetic acid. The final volume was adjusted to 1L using dH<sub>2</sub>O.

### Glucose (20%)

D-Glucose (2g) was dissolved in 10ml dH<sub>2</sub>O. Once prepared, the solution was sterilized by passing through a 0.2 µm filter (Millipore Type VS 0.025 µm)

### Arabinose (20%)

Arabinose (2g) was dissolved in 10ml dH<sub>2</sub>O. Once prepared, the solution was sterilized by passing through a 0.2 µm filter (Millipore Type VS 0.025 µm).

### Ethidium bromide (10mg/ml)

Ethidium bromide tablets were dissolved in sterile dH<sub>2</sub>O to prepare 10mg/ml solution. 200 µl of this was added to 2L of 1X TAE for staining gels.

## 2.2 Reagents

### dNTP (10mM)

10 mM dNTP was prepared by mixing 100  $\mu$ l of each of 100 mM dATP, 100mM dGTP, 100 mM dCTP and 100 mM dTTP and bringing the final volume to 1ml by adding sterile dH<sub>2</sub>O.

### Proteinase K (10mg/ml)

25 mg lyophilized proteinase K (Roche Applied Science) was reconstituted in 2.5 ml sterile dH<sub>2</sub>O. Aliquots were stored at 4°C for short-term storage and at -20°C for long term preservation.

## 2.3 Buffers

### Tris (0.5 M)

Tris (hydroxymethyl) aminomethane (60.57g) was dissolved in 800 ml of dH<sub>2</sub>O. The final volume was adjusted to 1 L. The pH was adjusted to 8.0 using 1 M HCl.

### TAE (50X)

TAE was stored as a 50X concentrate and diluted to 1X by adding dH<sub>2</sub>O. To prepare the concentrate, 242.2 g Tris base was dissolved in 800 ml dH<sub>2</sub>O. Glacial acetic (57.1ml) acid and 100 ml 0.5 M EDTA (pH 8.0) were added to this. The final volume was adjusted to 1L.

## 2.4 Bacterial strains

The bacterial strains used in this study are listed in Table 2.1.

**Table 2.1: Bacterial strains used in this study**

Strain	Genotype	Source
XL1-Blue	<i>recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac [F' proAB lacI<sup>q</sup>ΔM15 Tn10 (Tet<sup>r</sup>)]</i>	Agilent technologies
<i>E. coli</i> Top 10	<i>(F- mcrA Δ(mrr-hsdRMS-mcrBC) ϕ80lacZΔM15 ΔlacX74 recA1 araD139 Δ(araleu) 7697 galU galK rpsL (Str<sup>R</sup>) endA1 nupG)</i>	Invitrogen

<b>Table 2.1 (contd.)</b>		
<i>E. coli</i> DH5α	<i>F- φ80lacZΔM15 Δ(lacZYA-argF)U169 deoR recA1 endA1 hsdR17(rk<sup>-</sup>, mk<sup>+</sup>) phoA supE44 thi-1 gyrA96 relA1 λ-</i>	Invitrogen
<i>E. coli</i> BR825	<i>polA::Tn10 trp</i>	Ludtke <i>et al.</i> 1989
<i>E. coli</i> SP2	<i>pcnB, Δlac, recA</i>	Podkovyrov and Larson, 1995
BW25141	<i>lacI<sup>q</sup> rrnB<sub>T14</sub> ΔlacZ<sub>WJ16</sub> ΔphoBR580 hsdR514 ΔaraBAD<sub>AH33</sub> ΔrhaBAD<sub>LD78</sub> galU95 endA<sub>BT333</sub> uidA(ΔMluI)::<sup>+</sup> pir<sup>+</sup> recA1</i>	Datsenko and Wanner, 2000

## 2.5 Media

All media were prepared using milliQ water and sterilized at 121°C at 15p.s.i for 20 minutes. When glucose was added to any medium, it was filter sterilized separately using a 0.2 μm filter (Millipore).

### Luria Bertani (LB) broth

L-broth (Kennedy, 1971) was purchased from Merck (Merck4Biosciences, UK). The medium is sold as granules, which is reconstituted in milliQ water according to the manufacturer's instructions. The composition of the L-broth (Miller) per litre is 5 g yeast extract, 10 g peptone from casein and 10 g sodium chloride.

### Luria Bertani (LB) agar

To prepare L-agar, 12 g agar-agar was added to per liter of L-broth.

### LB agar with X-gal and IPTG

To prepare L-agar containing X-gal and IPTG, 0.3g of 5- Bromo - 4-Chloro- 3-Indoly-β-D-galactopyranoside (X-gal, Melford, UK) and 0.3g Isopropyl-β-D-thiogalactopyranoside (IPTG, Sigma Aldrich, UK) were added to 1L of molten L-agar prior to pouring plates.

### SOC medium

The recipe described in Sambrook and Russell (2001) was followed to prepare SOC. To

950 ml of dH<sub>2</sub>O, 5 g yeast extract, 20 g tryptone, 0.5 g NaCl, and 10 ml 250 mM KCl were added. The pH was adjusted to 7.0 using 10 M NaOH. The final volume was adjusted to 1 L. Before use, 5 ml of 2 M MgCl<sub>2</sub> and 20 ml of 1 M glucose were added to per litre of the medium.

## 2.6 Antibiotics

The antibiotics used in this study are listed in Table 2.2.

**Table 2.2: Antibiotics used in this study**

Antibiotic	Solvent	Stock concentration (mg ml <sup>-1</sup> )	Working concentration (µg ml <sup>-1</sup> )
Carbenicillin	Water	50	50
Chloramphenicol	Absolute ethanol	10	10
Kanamycin	Water	25	25

All antibiotics were purchased from Melford laboratories ([www.melford.co.uk](http://www.melford.co.uk)).

Following preparation, all antibiotics were filter sterilized (Millipore Type VS 0.025 µm) and stored at -20 °C.

## 2.7 Plasmids

Plasmids used in this study are in Table 2.3.

**Table 2.3: Plasmids used in the present investigation**

Plasmid	Description	Reference
pBSSB1	Linear, Kan <sup>r</sup>	Baker <i>et al.</i> 2007b
pUC18	Cloning vector, Ap <sup>r</sup>	Yanisch-Perron <i>et al.</i> 1985
pBAD18	Expression vector, P <sub>araBAD</sub> promoter, Kan <sup>r</sup>	Guzman <i>et al.</i> 1995
pFH450	P1 ori, pMB1 ori, <i>cat</i>	Hayes, 1998
pKD13	Template plasmid	Datsenko and Wanner, 2000

For short term storage, plasmids were kept at 4°C. Longer storage was done at -20°C.



## 2.8 Oligonucleotides used in this study

Oligonucleotides used in the present investigation are listed in Table 2.4.

**Table 2.4: Details of oligonucleotides used in the present study**

Oligonucleotide	Sequence 5' to 3'
<b>Oligonucleotides used for construction of linear plasmid library</b>	
EcoRIRP2	AGTCGAATTCCC GTTCTGAACAGCCTGTAA
BamHIRP2	AGTCGGATCCTCACCATCAGGGTTTATTCCA
EcoRIRP3	TAACGAATTCAGCAATTGGTACCGTTCTGG
BamHIRP3	TGCCGGATCCGAAAGCTATCCGCACGAAAG
EcoRIRP4	TTTTGAATTCGGGATTGTGTGACAGGCATT
BamHIRP4	TGATGGATCCTTTTAACGGCGGCTTATCTG
EcoRIRP5	TGTCGAATTCCTGTGGAACAATGGCAAAAA
BamHIRP5	ACATGGATCCACTGCATTTTCGAGCAGGTCT
EcoRIRP6	CAATGAATTCGCACTGTTAGCCGTAATCA
BamHIRP6	CGGCGGATCCTTCTGCCGTCTTACCTTGCT
EcoRIRP7	CTGGGAATTCGCTTGCCCCTATCGTTATCA
BamHIRP7	TATTGGATCCCCCTTTTCATTGGGTTTCCTT
EcoRIRP8	CTTGAATTCATGAACAACCGGAAGGTGAG
BamHIRP8	TGTCGGATCCTTCAGAACCCAGGAGAATG
EcoRIRP9	CTTAGAATTCAACCCAATGAAAGGGAAAGG
BamHIRP9	GTAGGGATCCCAACACGATACCCGCTACT
EcoRIRP10	CCAAGAATTCCCGTTGCGATCAAGACAGTA
BamHIRP10	AGAGGGATCCAGAGCCTTACAGACGGCAAA
<b>Oligonucleotides used for amplification of kanamycin cassette</b>	
pKD13_Fwd	ACGTGAATTCGTGTAGGCTGGAGCTGCTTC
pKD13_Rev	ACGTGGATCCCTGTCAAACATGAGAATTAA
<b>Oligonucleotides used for construction of deletion derivatives of RP7</b>	
RP7-Fwd-A	ATATATATATGAATTCATGAATGGGGAATGGCTACA
RP7-Fwd-B	ATATATATATGAATTCAGCAAGGTAAGACGGCAGAA
RP7-Fwd-C	ATATATATATGAATTCCTGCAAACTTGTTA
RP7-Fwd-D	ATATATATATGAATTCAAGACCTGCTCGAAATGCAG
RP7_del_EcoRI_2	AATTAATTGAATTCCTACGGCATGGGTGTAAT
RP7_del_BamHI_2	AATTAATTGGATCCCCCATGTGAGCTTTACCATACA
RP7_del_EcoRI_3	AATTAATTGAATTCAATGCTGCTCATGTCGTCAC
RP7_del_BamHI_3	AATTAATTGGATCCCTTGTCCAGCCACACTGAGA
RP7_delete_2_fwd	AATTAATTGAATTCATGAACAACCGGAAGGTGAG
MG6_Delete_1_Fwd	AATTAATTGAATTCTGCAAGCAAAGAAAAAGCAA

<b>Table 2.4 (contd.)</b>	
MG6_Delete_1_Rev	AATTAATTGGATCCTGTTGTCATAAGTACCCCGTTTT
MG6_Delete_2_Fwd	AATTAATTGAATTCAGCCTCTTGCTGATGTAGCAC
MG6_Delete_2_Rev	AATTAATTGGATCCTTTCTTTTATTCTGTCCTTATTGTTG
<b>Oligonucleotides used to generate deletion derivatives of RP2</b>	
RP2_Min_DelA_Rev	ATATATATATGGATCCCTTCCTGGACGACGAAGATT
RP2_Min_DelB_Rev	ATATATATATGGATCCCCACAGTTTGATGACCTGGA
RP2_Min_DelD_Rev	ATATATATATGGATCCGTGGCTCTTGCCTACACCTC
RP2_Min_DelE_Rev	ATATATATATGGATCCAAAAGGGTAAAGGGCTTCCA
RP2_Min_DelF_Rev	ATATATATATGGATCCCGTCCAGTTCGTGAACACTC
RP2_Delete_Fwd_A	ATATATATATGAATTCAATTGAATGACTGCCCCTTG
RP2_Delete_Fwd_B	ATATATATATGAATTCCAATCGCTTCCTGGTTGTTT
RP2_Delete_Fwd_C	ATATATATATGAATTCTTCAAGCTGTTGGGTGTTG
RP2_Delete_Fwd_D	ATATATATATGAATTCTTCCGTATTCCGTCTGGAAC
RP2_Delete_Fwd_1	ATATATATGAATTCTCTGCACTGCTATGGACTGG
RP2_Delete_Fwd_2	ATATATATGAATTTCGCTTGCCCCTATCGTTATCA
RP2_Delete_Fwd_3	ATATATATGAATTCATGAACAACCGGAAGGTGAG
RP2_Delete_Rev_3	ATATATATGGATCCCTCACCTTCCGGTTGTTTCAT
RP2_Delete_Rev_2	ATATATATGGATCCGCCGTAGGCACGAGAGTAAG
RP2_Delete_Rev_1	ATATATATGGATCCCCAGTCCATAGCAGTGCAGA
RP2_RevB-I	ATATATATATGGATCCTTTAACCCTGCCGTTGTAT
RP2_RevB-II	ATATATATATGGATCCCCAGTCCATAGCAGTGCAGA
RP2_RevB-III	ATATATATATGGATCCGGTCATAGCTGCTACTGGTGC
RP2_RevB-IV	ATATATATATGGATCCTGGAAGAACTTCACAATGCAA
RP2_FwdC-I	ATATATATATGAATTCTAGTTCCGAATCCCATAGGC
RP2_FwdC-II	ATATATATATGAATTCCATGTTGGCAATAGCTGACG
RP2_292_RevI	ATATATATATGGATCCCGGAATCATAAAACCCCTAA
RP2_292_RevII	ATATATATATGGATCCCCGTCAGCTATTGCCAACAT
RP2_292_RevIII	ATATATATATGGATCCAACTAAGGAGTTGCCTATGGGATT
RP2_292_FwdI	ATATATATATGAATCCTAGTTCCGAATCCCATAGGC
RP2_292_FwdII	ATATATATATGAATCCCATGTTGGCAATAGCTGACG
RP2_292_FwdIII	ATATATATATGAATCCTTCCGTATTCCGTCTGGAAC
RP4_ORF1_2_fwd	AAATTTGAATTCGGGATTTGTTGACAGGCATT
RP4_ORF1_2_rev	AAATTTGGATCCTGAAAAGCATCCTGACATGC
<b>Oligonucleotides used to generate deletion derivatives of RP4</b>	
RP4_ORF_3_4_Fwd	AAATTTGAATTCACCCAGTAAAGGGGCTTATG
RP4_ORF_3_4_Rev	AAATTTGGATCCTTAAATGGGGTTTCGTGGTTT
RP4HindIII_del_R_2	AATTAATTGGATCCATCGATGGATTGCTTTCACC
RP4HindII_del_Fwd_3	AATTAATTGAATTCTTGTGAGTTAAGCTGTGGCTATC
RP4Hind_del_Fwd_2	AATTAATTGAATTCATAGCCGTCTGTGAGCGATT
RP4_del_fwd_A	ATATATATATGAATTCCGCTGATTTGACGGCATAAC

<b>Table 2.4 (contd.)</b>	
<b>Oligonucleotides to insert stop codon in ORFs in RP4</b>	
Mutate_ORF7_RP4_Rev	TCTCTATGCCATGACCGTTT
Mutate_ORF7_RP4_Complementary	AAACGGTCATGGCATAGAGA
Mutate_ORF8_RP4_Rev	AGGGGAGACACCCACTACCTA
Mutate_ORF8_RP4_Complementary	TAGGTAGTGGGTGTCTCCCCT
Mutate_ORF9_RP4_Rev	CGGCTTTCTACAACTGGTTC
Mutate_ORF9_RP4_Complementary	GAACCAGTTGTAGAAAGCCG
New_Mutate_ORF8_Rev	CAGCTAGTGGGTGTCTCCCCT
New_Mutate_ORF8_Fwd	AGGGGAGACACCCACTAGCTG
RP4HM_DelB_Fwd	ATATATATATGAATTCAGTTCTGCAAGTCTTTCAC
BamHI_RP4HindIII	AATTAATTAAGGATCCTGGGGTTCGTGGTTTTACTT
<b>Oligonucleotides used to introduce stop codon in RP7</b>	
RP7_Mutate_ORF18_Fwd	GATTGCGGTTTAAACTGCTTTTAATGCCGGTGACTC
RP7_Mutate_ORF18_Rev	GAGTCACCGGCATTAAAAGCAGTTTAAACCGCAATC
RP7_ORF19_Mutate_Fwd	TTCACGAGCTTGATTAACAACCGGAAGGTGAGC
RP7_ORF19_mutate_Rev	GCTCACCTTCCGGTTGTTAATCAAGCTCGTGAA
RP7_mutate_ORF20_Fwd	TCTCTTCCCGATCCTTAGTTCGTGCGCAAGCTG
RP7_mutate_ORF20_Rev	CAGCTTGCCGACGAACTAAGGATCGGGAAGAGA
Mutate_RP7ORF1_Fwd	GCTTTGGCAGCGGCTGGGATATAAGCAACGGTAAC
Mutate_RP7ORF1_Rev	GTTACCGTTGCTTATATCCCAGCCGCTGCCAAAGC
<b>Oligonucleotides to amplify ORF9</b>	
EcoRI_ORF9_fwd	ATATATATATGAATTCAGGAGGTTGAGATAGAGTTAAGG GGATGAA
PstI_ORF9_rev	ATATATATATCTGCAGTCGATCAGGTCGTCAAAGAG
<b>Oligonucleotides used for sequence confirmation</b>	
P2_INLP	TTG ATG AAC AAC CGG AAG GT
P2_INRP	AAT TTC ATC TGT CGC ACC AA
P3_INT_F	ATTTCTTCCAGCCCTGCTTT
P3_INT_R	CAGCCAGCTTCCTGAATGAT
P5_INT_F	CGCACTGTTAGCCGTAATCA
P5_INT_R	AAACAACCAGGAAGCGATTG
P6_INT_F	TCTGCACTGCTATGGACTGG
P6_INT_R	GCCGTAGGCACGAGAGTAAG

## **2.9 Plasmid Extraction**

### **2.9.1 STET Plasmid Preparation**

The method described by Holmes and Quigley (1981) was followed. Five ml of DH5 $\alpha$  containing pBSSB1 was grown overnight in L-broth media containing kanamycin (25 $\mu$ g/ml). Cells were pelleted in a table-top centrifuge at 14,000  $\times$  g (Eppendorf minispin) for 2 min. The supernatant was poured off. Cells from 1.5ml culture were suspended in 110 $\mu$ l STET buffer (8% sucrose, 5% Triton X-100, 50 mM Tris pH 8.0, 50 mM EDTA pH 8.0) containing lysozyme (0.5 mg/ml final concentration). The tube containing the cells was placed in 100°C heat block for 30 sec. The tube was centrifuged for 10 min at 14,000  $\times$  g in a microfuge (Eppendorf minispin). The pellet was discarded with a sterile toothpick. Isopropanol (110  $\mu$ l) was then added and the tube centrifuged for 10 min at 14,000  $\times$  g (Eppendorf minispin). The supernatant was poured off. Pelleted DNA was washed briefly with 0.4 ml cold 70% ethanol. Next, the ethanol was poured off and the washing repeated three more times. Any remaining liquid was removed with a micropipette and the pellet was dried overnight at room temperature. Any DNA pellet was dissolved in 20 $\mu$ l TE buffer. Purified plasmid was stored at 4°C.

### **2.9.2 Single colony final sample buffer (SCFSB) technique**

This method was obtained through personal communication with Dr. David Summers. In this method, 200 $\mu$ l of culture was mixed with 200  $\mu$ l of SCFSB buffer. To prepare 100 ml SCFSB buffer, 10 g ficoll, 0.025 g of bromophenol blue, 0.25g of orange G and 0.5g of sodium dodecyl sulphate (SDS) were dissolved in 80ml dH<sub>2</sub>O water. The final volume was then adjusted to 100ml. The culture was mixed with the buffer and left at room temperature for 15 minutes. Cell debris was pelleted by centrifugation in a table-top centrifuge at 13,000 rpm for 15 min. 15-20  $\mu$ l of supernatant was loaded directly in a 1% agarose gel for electrophoresis.

### **2.9.3 The Qiagen miniprep for plasmid extraction**

The manufacturer's instructions outlined in the handbook at [www.qiagen.com](http://www.qiagen.com) was followed.

### 2.9.4 Kado and Liu Method

Fifteen ml of an overnight culture was harvested by centrifugation at  $14,000 \times g$  in a table-top centrifuge (Eppendorf minispin). Cells were washed with 5 ml of E buffer. To prepare E buffer, 4.85 g of Tris was dissolved in 800 ml dH<sub>2</sub>O. The pH was adjusted to 7.9 using glacial acetic acid. 0.74 g Disodium ethylenediaminetetraacetate dehydrate was then added to give a final concentration of 2 mM. The total volume was adjusted to 1 L.

Cells were lysed by the addition of 2ml of lysing solution. Lysing solution consisted of 3% SDS and 50 mM Tris. The pH was adjusted to 12.6 by adding 1.6 ml of 2N NaOH and the final volume was adjusted to 100ml. The lysing solution was filter sterilized by using a membrane filter (Millipore Type VS 0.2  $\mu$ m) prior to use. Cells were incubated at room temperature under gentle agitation. They were then heated at 65°C for 20 min in a heated block. NaCl (0.5 ml of 4M) and twice the volume of phenol-chloroform solution (sigma, UK) were added. The solution was mixed gently for 30 min. It was then centrifuged at  $14,000 \times g$  at 4°C for 15 min. The top aqueous phase was removed carefully using a pipette and extracted with diethyl ether. DNA in the aqueous phase was precipitated with ethanol. The DNA pellet was washed three times with 70% Ethanol and suspended in TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). Five  $\mu$ l of sample was used for electrophoresis.

## 2.10 PCR

### 2.10.1 Routine PCR using Taq DNA polymerase from NEB ([www.neb.com](http://www.neb.com)).

All PCR reaction components were thawed and assembled on ice. PCRs were done in total volumes of either 25  $\mu$ l or 50  $\mu$ l. The following master mix adopted from the NEB website was used with the modification that DMSO was included at a final concentration of 5%. The composition of the master mix used is detailed in Table 2.5.

**Table 2.5: Composition of PCR master mix used for routine PCRs using Taq DNA polymerase**

Component	25 µl reaction	50µl reaction	Final Concentration
10X Standard <i>Taq</i> Reaction Buffer	2.5 µl	5 µl	1X
10 mM dNTPs	0.5 µl	1 µl	200 µM
10 µM Forward Primer	0.5 µl	1 µl	0.2 µM
10 µM Reverse Primer	0.5 µl	1 µl	0.2 µM
DMSO	1.25 µl	2.5 µl	5%
Template DNA	variable	variable	<1,000 ng
<i>Taq</i> DNA Polymerase	0.125 µl	0.25 µl	1.25 units/50 µl PCR
Nuclease-free water	to 25 µl	to 50 µl	

The different components were mixed by pipetting up and down gently. They were then centrifuged briefly to collect everything to the bottom of the tube. Thermocycling conditions for PCR were as follows:

Initial denaturation: 95°C 3-5 min, 35 cycles: 95°C, 30 sec, 55–65°C, 30-60 sec  
72°C 1 min per kb. Final extension: 72°C 10 min. Hold: 4°C

### 2.10.2 Colony PCR

A single colony was resuspended in 20µl of sterile distilled water. The cells were lysed by heating at 100°C for 10 min. Cellular debris were removed by spinning at 14,000 × g for 5 minutes. The supernatant (2 µl) was used as template in a routine 25 µl PCR reaction as detailed in section 2.10.1.

### 2.10.3 Large amplicon and high fidelity PCR

The MasterAmp<sup>TM</sup> Extra Long DNA polymerase kit from Epicentre Biotechnologies ([www.cambio.co.uk](http://www.cambio.co.uk)) was used for all large amplicon and high fidelity PCR reactions. PCR reaction mixes were prepared and reaction conditions were maintained according to the manufacturer's instructions.

Briefly, the master mix was prepared as outlined in Table 2.6.

**Table 2.6: Reaction mix used for high fidelity PCR using MasterAmp<sup>TM</sup> DNA polymerase**

Component	Volume (µl)	Final concentration
DNA template	Variable	4-5 times more template than standard PCR
Sterile water	To a reaction volume of 25 µl	
Forward primer	5	1 µM
Reverse primer	5	1 µM
MasterAmp Extra-Long DNA Polymerase Mix	1	2.5 U
MasterAmp <sup>TM</sup> Extra-Long PCR 2X Premix 5	25	1X
Total	50	

The Thermocycling conditions were:

Initial denaturation at 98°C for 1-3 minutes followed by a three-step cycling reaction consisting of 35 cycles of denaturation at 1 minute, annealing at 2-5°C lower than the  $T_m$  of the primers for 0.5-1 min and extension at 72°C for 1min per kb of the product. A final extension was carried out at 72°C for 10 minutes. The samples were held at 4°C before further treatment.

#### 2.10.4 PCR using the Expand Long Template PCR system from Roche

High Fidelity and long PCRs were also carried out using the Expand Long Template PCR system from Roche ([www.roche-applied-science.com](http://www.roche-applied-science.com)). For PCR reaction conditions and master mix compositions, the manufacturer's instructions were followed. For a typical 50µl reaction, the master mix was as outlined in Table 2.7.

**Table 2.7: Master mix for long template PCR**

<b>Components</b>	<b>Volume (μl)</b>	<b>Final Concentration</b>
Sterile distilled water	To a final volume of 50	-
dNTPs (10mM)	1.75	350μM
Forward primer (10μM)	1.5	300nM
Reverse primer (10μM)	1.5	300nM
10X PCR buffer with MgCl <sub>2</sub>	5	1X
Template DNA	Variable	-
Expand long template enzyme mix	0.75 μl	3.75U

Thermal cycling conditions were as follows:

Initial denaturation: 94°C for 2-3 min

35 cycles consisting of: denaturation at 94°C for 10 sec, annealing at 50-60°C for 30 sec and extension at 68°C for 2 min+ 20sec for each successive cycle for up to 3 kb and 4 min+ 20 sec for each successive cycle for up to 6 kb of amplicon size.

Final extension was done at 68°C for 10 min.

## **2.11 Purification of PCR products and digested DNA**

PCR products and digested DNAs were purified using the PCR purification kit from Qiagen according to the manufacturer's instructions. The QIAquick gel extraction kit was used when a DNA sample needed to be purified from an agarose gel.

## **2.12 Quantitation of DNA**

The concentration of DNA was quantified by measuring the optical density at 260 nm using a ND-1000 spectrophotometer (NanoDrop).



### 2.13 Restriction digestion

Restriction enzymes were purchased from either NEB ([www.neb.com](http://www.neb.com)) or Fermentas ([www.fermentas.com](http://www.fermentas.com)). The manufacturers recommended reaction conditions were followed for digestion.

### 2.14 Exonuclease treatment

5' to 3' lambda exonuclease was purchased from NEB. Reactions were carried out according to the manufacturer's instructions.

### 2.15 Ligation using T4 DNA ligase

Typical ligation reaction mixes were as described in Table 2.8.

**Table 2.8: Composition of ligation reaction mix**

Component	Volume (μl)	Final concentration
DNA	Variable	Variable
10X ligase buffer	1	1X
T4 DNA ligase	1	400U
Sterile distilled water	To 10 μl	
Total	10	

Reaction mixes were incubated at 16°C for 16 hrs.

### 2.16 Drop dialysis

Ligated DNA samples were placed on a 0.2μm filter disc placed on sterile distilled water and left for about 10-15 min in order to remove excess salts prior to electroporation.

### 2.17 Gel electrophoresis

DNA samples were mixed with 5X loading dye (Bioline). Hyperladder I, supercoiled ladder or high range ladder (Bioline) was used as DNA sizing markers. Electrophoresis was done using 1% (w/v) agarose gels for routine purposes and 0.4% (w/v) agarose gels

for running linear plasmid samples. 1X TAE buffer was used for preparation of gels and for electrophoresis. Following electrophoresis gels were stained in 1X TAE containing  $0.5\mu\text{g ml}^{-1}$  ethidium bromide, and visualized using the Gel Doc XR system (Biorad).

## 2.18 Transformation

### 2.18.1 Electroporation

Ten ml of L-broth was inoculated with a single colony of the desired *E. coli* strain and incubated at 37°C overnight in an orbital shaker at 120 rpm. Cells were harvested by centrifugation (MSE Harrier) at  $3200 \times g$  for 10 min and washed with ice cold sterile distilled water. This step of harvesting and washing was repeated a total of three times. Cells were finally suspended in 200  $\mu\text{l}$  of ice-cold sterile distilled water. Fifty  $\mu\text{l}$  of competent cells were transferred to an electroporation cuvette with a gap of 0.1cm. DNA to be transformed was added to the cells and mixed briefly by pipetting up and down. Cells were then subjected to an electric pulse of 1.6 kV, 200 Ohms in a Gene Pulser unit (Bio-Rad). SOC medium (450  $\mu\text{l}$ ) was quickly added to the cells. They were allowed to recover for an hour at 37°C at 120 rpm. Cells were then plated on appropriate medium either directly or following suitable dilution. Plates were incubated at 37°C overnight in a static incubator.

### 2.18.2 Chemical transformation

Ten ml of L-broth was inoculated with a single isolated colony of the desired *E. coli* strain and incubated at 37°C at 120 rpm. Cells were harvested by centrifugation (MSE Harrier) at  $3200 \times g$  for 10 min. They were then washed with ice-cold 20mM  $\text{CaCl}_2$ . The step to harvest and wash the cells was repeated two more times before the cells were resuspended in 200  $\mu\text{l}$  of ice-cold 100mM  $\text{CaCl}_2$ . When DNA polymerase I mutant and *pcnB* mutants were used, 100  $\mu\text{l}$  of competent cells were used for transformation. DNA to be transformed was added to the competent cells and mixed gently. Cells were given a heat shock at 42°C for 45 sec-1min. They were then placed on ice for 2min. 900  $\mu\text{l}$  of pre-warmed L-broth (in case of DNA polymerase I mutant)/SOC (in case of *pcnB* mutant) was added and the cells allowed to recover *pcnB* mutants were given an hour at 120 rpm to recover before they were plated on appropriate

medium. Transformed DNA polymerase I mutants were allowed to recover for three hours at 37°C in a static incubator. They were then allowed a further overnight recovery by transferring the 1ml culture to 9 ml of L-broth. The suitable antibiotic was added to this 10ml broth. The tube was incubated overnight at 37°C once again under static conditions. Following incubation, cells were harvested by centrifugation and resuspended in 100 µl of L-broth. All cells were plated on suitable medium containing the appropriate antibiotic. Plates were incubated at 37°C overnight to allow growth of transformants.

## **2.19 Phase contrast microscopy**

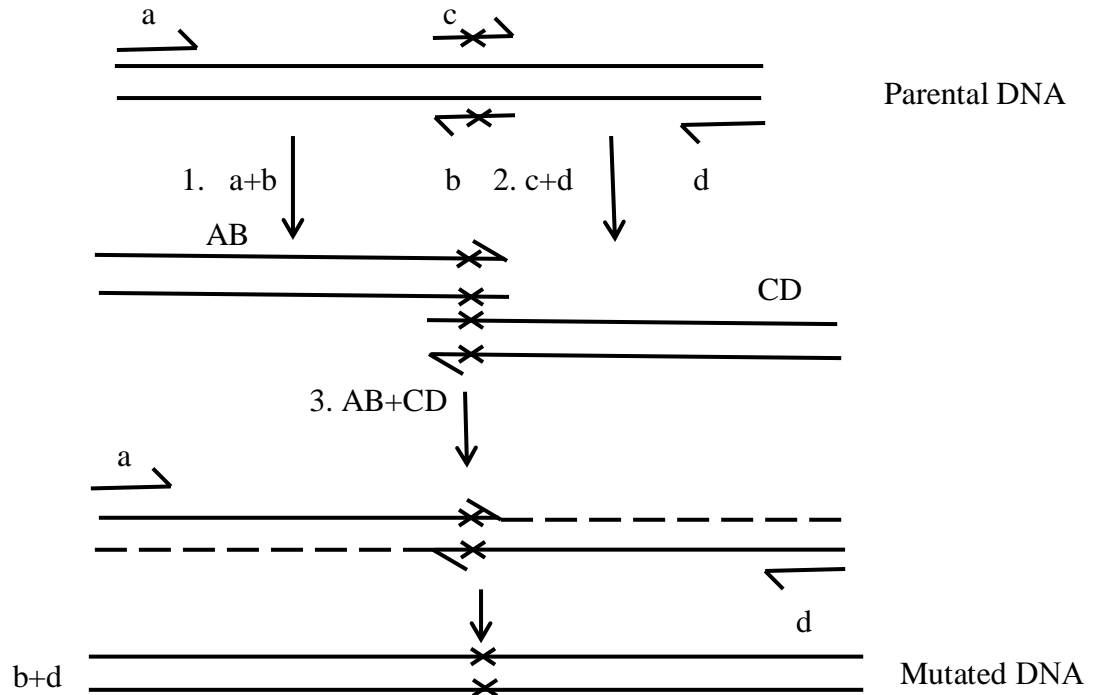
Five hundred µl of boiling 1% (w/v) agarose in SDS was quickly spread on a microscope slide and allowed to solidify. Between 5-10 µl of the bacterial culture to be observed was spread on top of the agar as soon as it had solidified. A cover slip was placed on top of the film and the slide was studied immediately under the microscope using an oil immersion objective of a phase contrast microscope (Nikon Eclipse 80i). a 100X CFI Plan Fluor objective was used. A software known as NIS-elements F 3.0 (Nikon) was used for image acquisition and the analysis of images was performed using the ImageJ software (Abramoff *et al.*, 2004).

## **2.20 Site-directed mutagenesis**

### **2.20.1 Site-directed mutagenesis by overlap extension method**

Cloned RP4 was mutated at desired sites by using mutated synthetic oligonucleotides. In separate PCR reactions, two mutated amplicon are generated. For each reaction, primer pairs were used in which one was mutated that annealed within the cloned DNA and one flanking primer that hybridized to one end of the target sequence was used. The two PCR reactions with overlapping ends were run and purified from agarose gel. They were mixed in equimolar proportions to act as template in a third PCR that used only the two flanking primers for amplification.

A schematic diagram of the method is outlined in Figure 2.1.



**Figure 2.1: Schematic representation of the overlap extension method for the introduction of mutation. Adapted from Ho *et al.*, 1989**

‘a’ and ‘d’ are non-mutated flanking primers, ‘b’ and ‘c’ are mutated complementary primers. AB and CD are products of amplifications using a+b and c+d primer pairs.

The PCR reaction mixture was the same as that when the Expand Long Template DNA polymerase (Roche) was used (section 2.10.4).

### **2.20.2 The QuikChange II one-day site-directed mutagenesis method (Smith, 2007).**

pUC18 containing the desired insert (RP7) was used as template. Synthetic primers containing the desired mutation were used to synthesize the double stranded supercoiled template using *PfuTurbo* DNA polymerase (Stratagene).

The PCR reaction mix was prepared according to the manufacturer’s recommendations with slight changes. The master mix described in Table 2.9 was adopted for PCR.

**Table 2.9: Composition of amplification master mix used for site-directed mutagenesis**

Component	Final concentration/amount	Volume ( $\mu$ l)
Reaction buffer	1X	5
Double stranded DNA template	50 ng	1
Oligonucleotide primer 1	125 ng	1.5
Oligonucleotide primer 2	125 ng	1.5
dNTP mix	300 $\mu$ M	1.5
Sterile distilled water		To 50 $\mu$ l
<i>PfuTurbo</i> DNA polymerase	0.05 U/ $\mu$ l	1

The cycling parameters used were as follows:

Denaturation: 95°C for 30 seconds

Annealing: 55°C for 1 minute

Extension: 68°C for 1 minute/kb of template plasmid length

Total number of cycles: 20

The product of the PCR was double-stranded mutated plasmid with nicks. The amplified product was purified using a QiaQuick PCR purification column (Qiagen). Parental DNA template in the amplified mix was then digested by *DpnI* (NEB) digestion following the manufacturer's recommendations. The digested product was transformed into competent XL- 1 Blue *E. coli* cells which repaired the nicks in the mutated plasmid. Transformants which grew in LA containing appropriate antibiotic were then selected for screening for the presence of the mutated plasmid. Purified plasmids were confirmed for the presence of the mutation by sequencing.

## **2.21 DNA sequencing**

DNA sequencing was carried out by Geneservice Ltd. Using BigDye Terminator chemistry and a 3730 capillary sequencer (Applied Biosystems).

# **Chapter 3**

---

## **Development of a method for the purification of the linear plasmid, pBSSB2**

---

### 3.1 Introduction

The first method for purification of plasmid DNA exploited its circular nature (Clewell and Helinski, 1970). This method depended on differential binding of the DNA intercalating dye, ethidium bromide, to covalently closed circular, nicked open circular and linear DNA fragments. During purification of plasmids from bacterial cells, chromosomal DNA usually undergoes shearing following cell lysis and is converted to smaller, linear fragments. Such linear DNA fragments and nicked open circular DNA bind more dye than covalently closed circular plasmids. Consequently, the latter undergo a smaller decrease in buoyant density and can be separated from linear DNA using high speed centrifugation in a caesium chloride - ethidium bromide solution. When the density gradient reaches equilibrium, the covalently closed circular plasmid settles below the chromosomal DNA band.

Later, more rapid methods for plasmid purification were developed that also relied on the covalently closed circular nature of plasmids. A well-known example of these was designed by Birnboim and Doly (1979). The principle of this method was the differential alkaline denaturation and strand separation of linear chromosomal and circular plasmid DNA. Alkali can break down hydrogen bonds holding the nitrogenous bases between two polynucleotide chains. During cell lysis chromosomal DNA is usually sheared into linear fragments. Linear chromosomal DNA fragments and plasmids undergo denaturation by alkaline treatment. However, because of their circular structure plasmid DNA strands are only partially separated whereas sheared linear chromosomal DNA strands are completely separated (Warner and Rush, 1969; Birnboim and Doly, 1970). Upon neutralization, chromosomal DNA strands renature randomly and become entangled and insoluble. Circular plasmid DNA strands, which remain close to each other, renature rapidly. In methods employing heat for denaturation of DNA instead of alkaline treatment, e.g. the STET method (Holmes and Quigley, 1981) a similar situation arises where chromosomal DNA strands undergo complete separation whereas circular plasmid DNA strands remain bound together. In both methods randomly reannealed chromosomal DNA strands form an insoluble aggregate in close proximity with SDS-protein complexes and most of the RNA. This insoluble aggregate can be precipitated by centrifugation (Kay *et al.*, 1952; Marko and



Butler, 1950; Crestfield *et al.*, 1955). Plasmid DNA in the supernatant can then be recovered by ethanol precipitation.

The extraction of linear plasmids using methods designed for circular plasmid purification is not appropriate. Because of their structure, linear plasmids would be expected to behave like sheared linear chromosomal DNA fragments. They are easily prone to complete denaturation of their strands and upon renaturation, become insoluble and be lost in the pellet of cellular debris and chromosomal fragments.

Methods employed for purification of some linear plasmids must also include a protease treatment prior to cellular protein removal owing to the presence of the terminal protein at the end of the linear plasmid. Without this the linear plasmid is likely to be lost during the cellular protein removal step e.g. with phenol chloroform treatment.

Caesium chloride ethidium bromide density gradient centrifugation failed to isolate the linear plasmid with 5' terminal protein, SCP1, from *Streptomyces coelicolor* A3 (2) (Schrenpf *et al.* 1975). In 1979, Hayakawa *et al.* used a neutral pH DNA extraction method for the purification of total cellular DNA which contained a linear plasmid, pSLA2, from *Streptomyces rochei*. The alkaline lysis method (Birnboim and Doly, 1979) precipitated most of the linear DNAs and was, therefore, not suitable for the extraction of the linear plasmid (Chater and Kinashi, 2007).

Conventional agarose gel electrophoresis has been used to separate the linear plasmids pSCL1 (12kb) from *S. clavuligerus* (Keen *et al.*, 1988), pSRM (43kb) from *S. rimosus* (Chardon-Lauriaux, 1986), pFS1 (16kb) from *Bacillus polymyxa* (Seldin and Rosado, 1993) and pMC3-2 (6kb) from *Morchella conica* (Meinhardt *et al.*, 1991) from the respective chromosomal DNA.

The purification of linear plasmids was made easier with the introduction of various types of pulsed-field gel electrophoresis (PFGE) (Schwartz and Cantor, 1984; Carle and Olson, 1984). In this method, cells are immobilized in agarose plugs, which are treated successively with lysozyme, protease and detergent, before being embedded in an

electrophoresis gel. In this way, shearing of linear DNA is minimized. Linear plasmids which have been purified using PFGE are shown in Table 3.1.

**Table 3.1: Examples of linear plasmids extracted through use of PFGE**

Plasmid	Size (kb)	Source organism	Reference
SCP1	363	<i>S. coelicolor</i> A3(2)	Kinashi <i>et al.</i> , 1987
pKSL	520	<i>S. lasaliensis</i>	Kinashi and Shimaji, 1987
pSLA2-M	100	<i>S. rochei</i>	Kinashi <i>et al.</i> , 1994
pSCL1, 2 and 3	11.7, 120, 460	<i>S. clavuligerus</i>	Kinashi <i>et al.</i> , 1995
pSGR1	160	<i>S. griseus</i>	Kinashi <i>et al.</i> , 1995
pSAR1 and 2	300, 200	<i>S. argenteolus</i>	Kinashi <i>et al.</i> , 1995
pZG101	387	<i>S. rimosus</i>	Gravius <i>et al.</i> , 1994
pSA1 and 2	100, 250	<i>S. avermitilis</i>	Evans <i>et al.</i> , 1994
pCLP	25	<i>Mycobacterium celatum</i>	Picardeau and Vincent, 1998
pBD2	208-212	<i>Rhodococcus erythropolis</i>	Dabrock <i>et al.</i> , 1994
pNC10, 20, 30 and 40	70, 85, 185	<i>R. corallinus</i>	Saeki <i>et al.</i> , 1999
pFiD188	200	<i>R. fascians</i>	Crespi <i>et al.</i> , 1992
pHG201, 202 and 203	270, 400, 420	<i>Nocardia opaca</i>	Kalkus <i>et al.</i> , 1990

*Borrelia* linear plasmids, which differ from most other linear plasmids in that they have covalently closed ends, were originally extracted using caesium chloride and ethidium bromide density gradient centrifugation (Barbour and Garon, 1987). Subsequently, most *Borrelia* linear plasmids have been extracted through the preparation of total cellular DNA that was either run in low-percentage agarose (Barbour, 1988, Hinnebusch *et al.*, 1996; Schwann *et al.*, 1988; Lane, 2009) or PFGE was used to separate the linear plasmids (Barbour, 1989, Baril, 1989). When necessary, the linear plasmid bands were electroeluted from the agarose gel.

The linear plasmid pBSSB2, which is the subject of the present investigation, was extracted by Baker *et al.* (2007b) using a modification (Taghavi *et al.* 1994) of a method described originally by Kado and Liu (1981) (section 2.9.4). The method employs alkaline lysis of cells followed by precipitation of insoluble chromosomal DNA and cellular debris by salt. This is followed by treatment with phenol - chloroform. Plasmid DNA in the supernatant is purified by precipitation. However, this modified method (referred to as the Taghavi method in the present study) was found to give highly variable results with poor yield of the plasmid (Stephen Baker, personal communication). An initial objective in the present study was, therefore, to optimize a rapid, reliable, reproducible and efficient method for the extraction of pBSSB2 that could be used for routine preparation of linear plasmid DNA.

### **3.2 Use of conventional plasmid extraction methods for pBSSB2 purification**

At first, various common plasmid extraction methods such as the rapid boiling lysis method of Holmes and Quigley (1981) (section 2.9.1), electrophoresis of crude cell extracts by the Single Colony Final Sample Buffer method (D. Summers, personal communication, section 2.9.2) and the Qiagen plasmid mini preparation kit (section 2.9.3) were used in an attempt to extract the linear plasmid, pBSSB2 from *E. coli* Top10 cells. However, none of these methods was found to yield a visible plasmid band following electrophoresis of the purified sample.

### **3.3 Development and optimization of a method for the successful extraction of pBSSB2**

In preliminary experiments with the Taghavi method (1994) to purify pBSSB2 no band was visible in an agarose gel. The presence of the linear plasmid, however, could be detected by PCR using primers that amplified a 500bp region of the plasmid (Data not shown). It was concluded that the Taghavi method was not an efficient option for routine purification of the plasmid. It was clearly important that an efficient and reliable method for the purification of pBSSB2 was developed. Modifications were therefore made to the Taghavi method, with each step of the procedure being considered individually for optimization as discussed in the following sections.

### 3.3.1 Cell lysis

Plasmid pBSSB2 is linear with proteins attached covalently at its 5' ends. While designing a lysis solution it was argued that the linear plasmid could be more prone to denaturation if alkaline lysis was used because of its linear structure than it would be if it had a circular structure. Linear plasmid fragments would be lost more easily with chromosomal DNA fragments that arise following cell lysis. In addition, gentle cell lysis condition can also reduce shearing of plasmid DNA and enable recovery in its original form.

In the Taghavi method, the lysis solution comprised 40mM Tris and 1% sodium dodecyl sulphate with its pH being adjusted to 12.6 with 5N NaOH. In the absence of lysozyme and sucrose this solution would cause rapid cell lysis. Rapid lysis could shear linear plasmid DNA. Additionally, the alkaline pH of the lysis solution would cause denaturation of not only chromosomal DNA but also of linear plasmid DNA strands. Subsequent renaturation would render both chromosomal and plasmid DNA insoluble and result in subsequent loss with cellular debris.

It was, therefore, decided to use a lysis solution that contained lysozyme to digest the cell wall and a non-ionic detergent, Triton-X, in place of SDS, to achieve gentle cell lysis in the presence of sucrose that provided an osmotically balanced environment in which lysis could occur. The pH of the lysis solution was also maintained near to neutral (pH 8.0) in order to prevent denaturation and possible insolubilization of pBSSB2.

In the optimized method (Table 3.2), a single colony of *E. coli* (SGB33) containing pBSSB2 was inoculated into 2L of L-broth containing kanamycin ( $30 \mu\text{g ml}^{-1}$ ) and grown overnight at 37°C and 120 rpm on an orbital shaker. Cells were harvested and resuspended in a solution containing 2% Triton X-100, 60mM EDTA and 50mM Tris, pH 8.0 (Table 3.2, step: 1). For every 250 ml of culture, 5 ml of resuspension solution was used. Cells were then incubated with gentle agitation at 37°C for 1 hour in lysis solution containing lysozyme ( $1\text{mg ml}^{-1}$ ), sucrose (10%), 50mM Tris (pH 8.0) and 0.5M EDTA (Table 3.2, step: 2).

### 3.3.2 Removal of cell debris

In the Taghavi method, cell debris was directly removed from the cell lysate using 4M NaCl and phenol-chloroform (1:1, water saturated, pH 4.8). pBSSB2 is believed to have proteins attached to the 5' termini (Baker *et al.*, 2007b). There is a danger that any protein removal step, such as salt precipitation or phenol: chloroform: isoamylalcohol treatment, would lead to the loss of the linear plasmid. A protease treatment was therefore included prior to salt precipitation with 4M NaCl and treatment with phenol: chloroform: isoamylalcohol.

Cell lysates were incubated with gentle agitation in the presence of proteinase K (200µg ml<sup>-1</sup>) for 1hr at 37°C. Salt precipitation was done using 4M NaCl and centrifugation at 21,000 rcf (SS34 rotor, Sorvall) at 4°C for 15 min. This was followed by three extractions with phenol: chloroform: isoamylalcohol to ensure complete removal of proteins (Table 3.2, step: 3).

### 3.3.3 Precipitation of DNA

No optimization was done for this step. Following a clearing spin (12,000 rcf at 4°C in a SS34 rotor, Sorvall, for 30 minutes), DNA was precipitated from the supernatant using two volumes of absolute ethanol (Table 3.2, step: 4). The precipitated DNA was washed three times with ice-cold 70% ethanol (Table 3.2, step: 5) and resuspended in TE (10mM Tris-HCl, 1mM EDTA, pH 8.0) (Table 3.2, step: 6).

**Table 3.2. Linear plasmid purification method developed in the present study**

<b>Ingredient</b>	<b>Final concentration/ value</b>
<b>Step 1: Cell resuspension solution</b>	
Triton X-100	2% (v/v)
EDTA	60mM
Tris-HCl	50mM, pH8.0
<b>Step 2: Lysis solution</b>	
Lysozyme	1 mg ml <sup>-1</sup>
sucrose	10% (v/v)
Tris (pH 8.0)	50mM
EDTA	0.5M
Incubation at 37°C for 1hr with gentle agitation (40-50 rpm on an orbital shaker)	
<b>Step 3: Removal of proteins and cell debris</b>	
Proteinase K	200µg ml <sup>-1</sup>
Incubation for 1hr at 37°C with gentle agitation (40-50rpm on an orbital shaker)	
NaCl	4M
Centrifugation at 21,000 × g at 4°C for 15 min	
3 extractions with phenol: chloroform: isoamylalcohol (25:24:1)	
1 extraction with chloroform: isoamylalcohol (24:1)	
<b>Step 4: Precipitation of DNA using absolute ethanol</b>	
<b>Step 5: Wash of precipitated plasmid with ethanol (70%)</b>	
<b>Step 6: Resuspension of plasmid DNA in TE</b>	

### 3.4 Gel electrophoresis of extracted pBSSB2

pBSSB2 purified from 2L culture of SGB33, which is *E. coli* Top10 (Invitrogen) transformed with pBSSB2 (Baker *et al.*, 2007b) was resuspended in 1-3ml Tris EDTA. Five  $\mu$ l of the extracted DNA was resolved in a 0.5% agarose gel and stained with ethidium bromide ( $100\mu\text{g ml}^{-1}$ ) before visualization (Figure 3.1). A linear ladder ( $\lambda$  DNA mono cut mix; N3019S, NEB) was run simultaneously for estimation of size. Based on the ladder, the size of the linear plasmid band was about 28kb as expected. Once the plasmid DNA was visualized in agarose gel, its concentration was measured using a ND-1000 spectrophotometer (Nanodrop). It was found to be about  $200\text{ng } \mu\text{l}^{-1}$ . The yield from the optimized method was highly reproducible and was about 200-500 $\mu\text{g}$  from a 2L culture. When this purified DNA was used as a template for PCR, the success of the PCR was variable unless the template was diluted about 1000 fold (data not shown). It was assumed that the sample was possibly contaminated with chemicals used during purification.

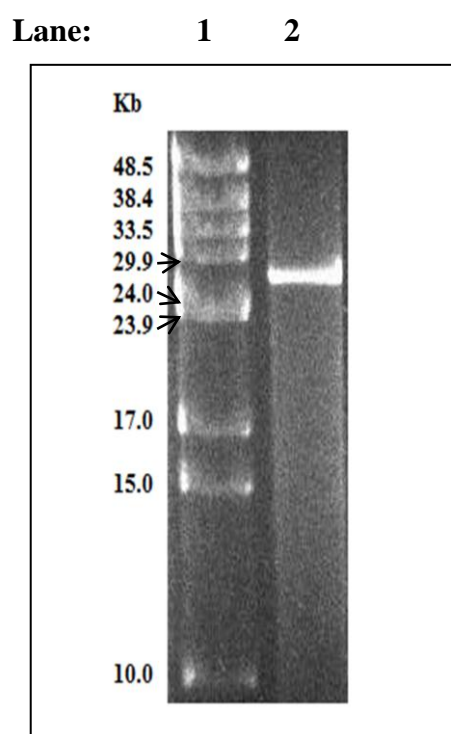


Figure 3.1: Purification of pBSSB2 by an optimized method. pBSSB2 was resolved in a 0.5% agarose gel. Lane 1:  $\lambda$  DNA mono cut mix (NEB), lane 2: Purified pBSSB2

A PCR purification column (Qiagen) was used to further purify the plasmid DNA, following the manufacturer's recommendations. Columns used for bacterial total DNA extraction in the DNeasy Blood and Tissue kit (Qiagen) were also used successfully for further purification of extracted pBSSB2. This yielded highly purified plasmid suitable for PCR and restriction digestion.

When plasmid preparation was carried out using the optimized method, but with the proteinase K step omitted to account for any effect of this on the efficiency of extraction, no plasmid DNA was obtained. This proved that the protease treatment was crucial for the extraction of pBSSB2 to be successful.

### **3.5 Restriction digestion of purified pBSSB2**

In order to confirm the restriction profile of purified pBSSB2, the extracted plasmid was digested with EcoRV and PmeI. Figure 3.2 shows the restriction pattern obtained. Alongside the gel images are predicted restriction patterns for pBSSB2 generated by the NEB cutter software (<http://tools.neb.com/NEBcutter2/>) when the linear plasmid is digested by the same enzymes. With PmeI, three bands corresponding to expected sizes of about 14kb, 9kb and 5 kb were obtained (Figure 3.2, lane 3). EcoRV yielded the expected bands of approximately 13kb, 7.5kb, 3.4kb, 1.8kb, 1.3kb, 500bp and 400bp (Figure 3.2, lane 2). The generation of bands of expected sizes provided further evidence in support of the fact that the purified DNA was pBSSB2.



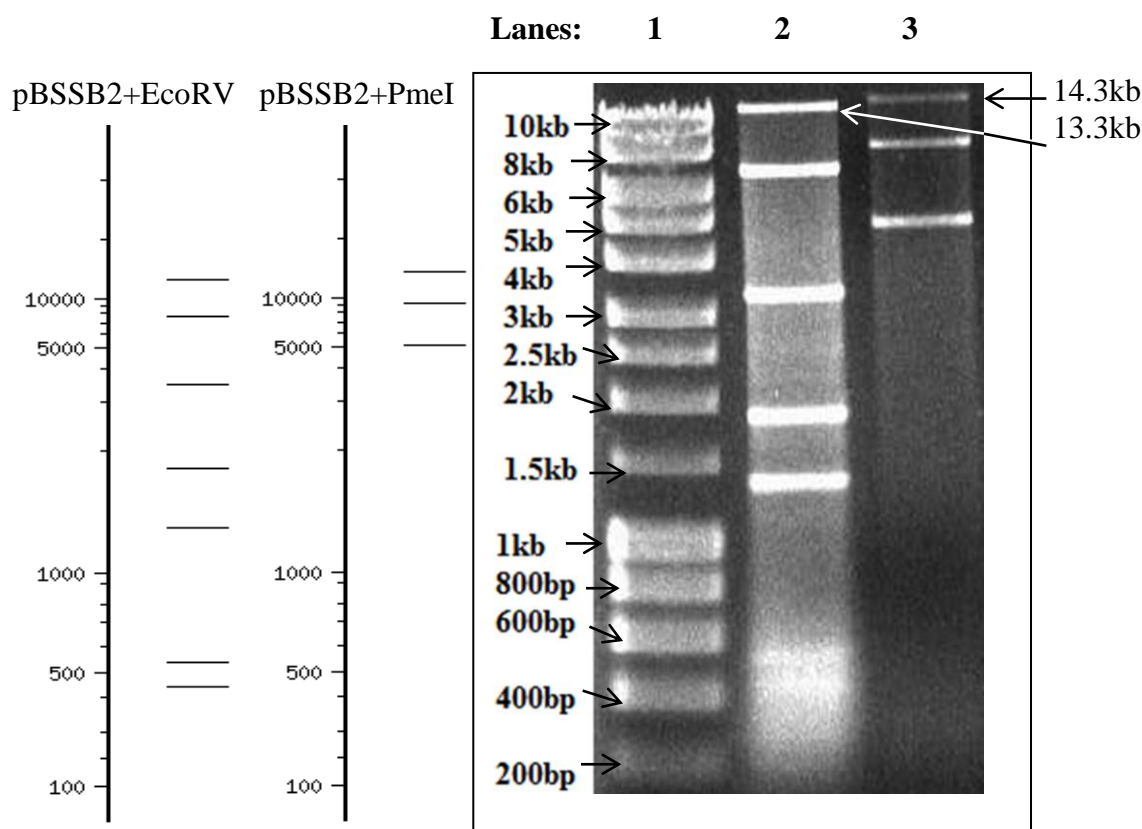


Figure 3.2. Restriction digestion profile of pBSSB2 with EcoRV and PmeI to confirm its identity. The digested DNA was resolved in a 1% agarose gel. A software (NEB cutter) generated profile is also shown in order to indicate expected sizes of bands for the digestion.

### 3.6 Treatment of purified pBSSB2 with $\lambda$ exonuclease

The linear plasmid pBBS2 has a protein attached covalently to its 5' end (Baker *et al.*, 2007b). When treated with proteinase K, the majority of the peptide bonds would be expected to be hydrolysed. However, the amino acid attached to the DNA backbone would not be removed since the bond between an amino acid and DNA would not be recognized by a peptidase. Since, the smallest peptide hydrolysed by proteinase K is a tetra-peptide (Kraus *et al.*, 1976), a protease treatment would leave at least a tetra-peptide at the linear plasmid 5' termini. Owing to the presence of this amino acid cap, the purified DNA would be expected to be resistant to treatment with  $\lambda$  exonuclease, a

5' phosphate specific exodeoxyribonuclease. However, this treatment should hydrolyse any contaminating chromosomal DNA. A linearized pUC18 was used as control for the exonuclease digestion since it should be sensitive. It was observed that the purified plasmid was still resistant to  $\lambda$  exonuclease (Figure 3.4).

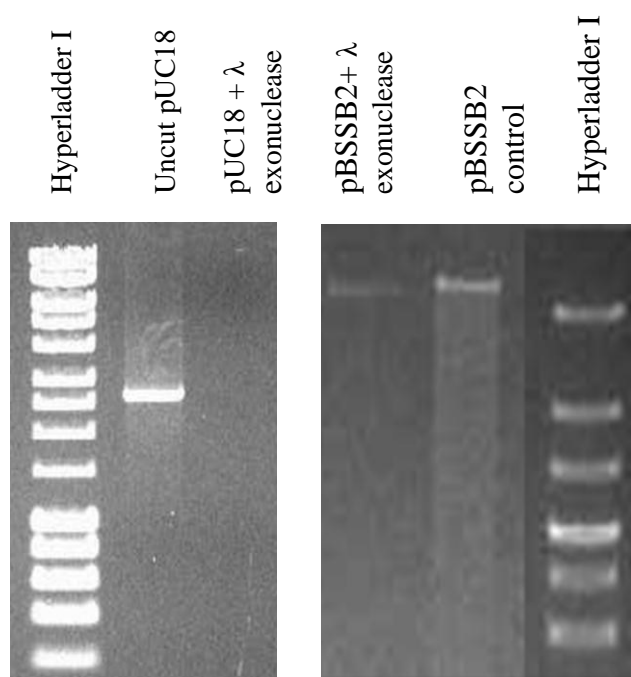


Figure 3.3. Lambda exonuclease treatment on pBSSB2 with pUC18 as control. The ladder is Hyperladder I from Bioline.

### 3.7 PCR amplification of pBSSB2 specific regions

To further confirm the identity of the linear plasmid, an amplicon specific for a 500bp region of pBSSB2 was amplified using primers (P2\_INRP and P2\_INLP, Table 2.4) that annealed to the linear plasmid template. The 500bp amplicon was successfully amplifiable. Subsequently it was sequenced which proved that the linear plasmid was present in the sample (Figure 3.4).

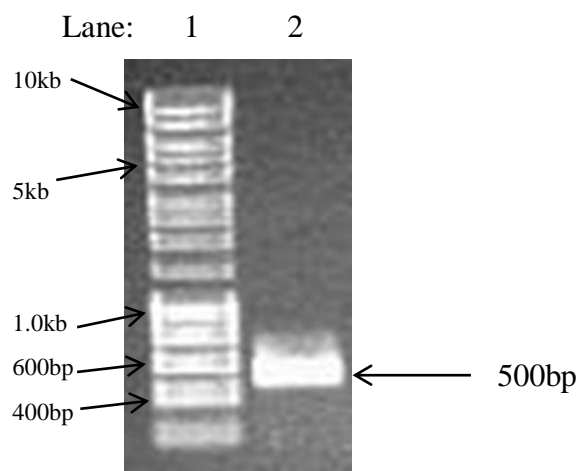


Figure 3.4. Amplification of a 500bp region present in pBSSB2. Lane 1: Hyperladder I (Bioline), Lane 2: 500bp amplicon

### 3.8 Retransformation of purified pBSSB2

Purified pBSSB2 was re-transformed into commercially available chemically competent *E. coli* (XLI-Blue, Agilent technologies) following the manufacturer's recommendations. Transformant colonies were selected on L-agar with kanamycin ( $25\mu\text{g ml}^{-1}$ ) at  $37^{\circ}\text{C}$  overnight. The transformation efficiency, at  $5\text{ cfu }\mu\text{g}^{-1}$ , was very low. To confirm the presence of pBSSB2 in the colonies, a colony PCR was carried out using primers specific for a 500bp region of pBSSB2. Generation of the expected amplicon proved that the colony contained the linear plasmid (data not shown).

### 3.9 Discussion

An initial objective of the present investigation was to develop a method for the routine purification of the linear plasmid pBSSB2. In earlier studies by Baker *et al.* (2007b) pBSSB2 was isolated using the Taghavi method (1994). However, this method was not reproducible and gave yield of the linear plasmid which was so poor that Southern blotting was necessary to reveal plasmid bands (S. Baker, personal communication). Use of the Taghavi method in the present study yielded no visible plasmid band on an agarose gel, even though the presence of the plasmid was indicated by PCR amplification of a pBSSB2-specific amplicon. This prompted literature review of methods commonly used for linear plasmid purification. As discussed previously, linear plasmids of several bacterial genera (viz. *Streptomyces*, *Borrelia*, *Rhodococcus*,

*Mycobacterium* and *Nocardia*) have been purified by using pulsed-field gel electrophoresis. The major drawback of PFGE is that it is technically demanding and time consuming (Tenover *et al.*, 1997). Since the Taghavi method used by Baker *et al.* (2007b) was successful in isolating the linear plasmid, albeit at a low concentration, attention was given to optimization of the existing protocol.

Each step of the Taghavi method was considered for further modifications. For optimization of lysis conditions, several methods were taken into consideration. It was anticipated that in methods based on alkaline lysis of cells (e.g. the Qiagen plasmid extraction kit) alkaline pH breaks the hydrogen bonds between linear plasmid strands causing complete separation. In these methods, linear plasmid strands behave similarly to sheared linear chromosomal fragments and are subsequently lost with cell debris. A similar situation arises in the boiling lysis STET protocol (Holmes and Quigley, 1981), in which linear plasmid strands are separated owing to breakdown of hydrogen bonds following the heat treatment. Finally, in the Taghavi method (1994), alkaline condition causes denaturation of linear plasmid strands and incorporation of the DNA into insoluble aggregates. Denatured strands behave like sheared, chromosomal fragments that are lost during the subsequent salt precipitation step. In all these methods, loss of linear plasmids is further enhanced by the absence of any protease treatment. In the present study, changes were made to the composition of the cell lysis solution through the incorporation of lysozyme aimed at cell wall breakdown, use of a non-ionic detergent (Triton-X) for cell lysis and inclusion of sucrose in order to maintain an osmotically balanced environment. The pH of the lysis solution was also set at 8.0 instead of an alkaline 12.0. This near-neutral pH would prevent separation of linear plasmid strands.

In the purification of a 43kb linear plasmid from *Streptomyces rimosus*, Chardon-Lauriaux *et al* (1986) found that it was necessary to carry out a ‘soft lysis’ including a protease treatment during extraction. Their method also avoided use of alkaline conditions in order to successfully purify the linear DNA. Although there was no direct evidence that gentle lysis facilitated pBSSB2 extraction in the present study, the rationale for designing the lysis solution is concordant with that favoured by Chardon-Lauriaux *et al* (1986).

Another key change was the addition of a proteinase K treatment step. It was assumed that linear plasmid would be lost with cellular proteins if the protein cap was not removed first before cellular protein removal treatment. When pBSSB2 was purified using the modified method, but with the proteinase K treatment step omitted, there was no pBSSB2-specific band visible in an agarose gel, in contrast to a bright band present for the proteinase K treated sample. It was evident that the omission of this step in the Taghavi method (Taghavi *et al.*, 1994) resulted in loss of most of the linear plasmid through the denaturation of the terminal protein attached to its 5' end during the phenol-chloroform-isoamylalcohol treatment and its precipitation with cell debris during the subsequent clearing spin.

The optimized method was similar to that used by Kemble and Baszczynski (1987) for the extraction of linear mitochondrial plasmids of sorghum and maize. As with the present work, omitting a proteinase K treatment failed to yield any linear plasmid DNA. The authors also found that conducting the complete protocol without any protease treatment but including such a treatment just prior to gel electrophoresis, did not show any plasmid band in agarose gel. From their study, it can be concluded that gentle cell lysis alone is not sufficient for purification of linear plasmids but that a protease treatment is also crucial.

The optimized method developed in the present work is rapid, reproducible and simple. It can be used routinely for purification of linear plasmids such as pBSSB2. The resulting DNA can be used for all experimental work, either as a template for PCR or for restriction digestion following an added column purification step.

# **Chapter 4**

---

## **Characterization of the replication functions of the linear plasmid, pBSSB2**

---

## 4.1 Introduction

Preliminary analysis of the linear plasmid pBSSB2 was primarily through bioinformatic approaches (Baker *et al.*, 2007b). Experimental investigation into replication and its control has not hitherto been carried out. One of the objectives of the present study was to identify regions of the linear plasmid pBSSB2 that are associated with replication.

The only indication of the location of the origin of replication is based on sequence analysis of pBSSB2 (Baker *et al.*, 2007b). According to this analysis, a GC skew lies immediately upstream of ORF017 and is close to 11000bp in the linear plasmid map. A GC skew is usually associated with the origin of replication (Picardeau *et al.* 2000). This assumption is further supported by the direction of transcription of most of the ORFs which, if the origin were located as suggested (Baker *et al.*, 2007b), are transcribed in the same direction as the movement of the replication fork. No short direct repeats that could be associated with control of replication by titrating away rep proteins were found by these authors in this region.

The central regions of *Streptomyces* linear plasmids can replicate in circular mode when their telomeres are deleted (Chang and Cohen, 1994; Hiratsu *et al.*, 2000; Redenbach *et al.*, 1999; Shiffman and Cohen, 1992). Based on this finding, and concerns that working directly with a linear plasmid or a linear derivative may pose difficulty with manipulations such as transformation, it was decided to construct a library containing 4-6 kb overlapping fragments of pBSSB2 in a circular vector (pUC18) and to screen these plasmids for their ability to replicate from the linear plasmid sequence.

Plasmid pUC18 carries a ColE1-like replicon, which requires DNA polymerase I to initiate replication (Kingsbury and Helinski, 1970). Huang and Chen (2008) showed that DNA polymerase I is not required for replication of the *Streptomyces* chromosome, the native form of which is linear (Lin *et al.*, 1993; Chen, 1996). It has been proposed that replication of the *Streptomyces* chromosome and linear plasmids occur by similar mechanisms (Qin and Cohen, 1998). The structure of pBSSB2 is similar to *Streptomyces* linear plasmids (Baker *et al.*, 2007b) so it seems reasonable to suppose that replication of pBSSB2 will also be polymerase I independent. In this study a DNA polymerase I mutant of *Escherichia coli* (BR825) is used to screen for replication of plasmids containing pBSSB2 fragments in pUC18. In this host, in

the absence of DNA polymerase I, the pUC18 replicon would be inactive. Any replication that occurs in this host must therefore be driven by a linear plasmid replicon.

## **4.2. Construction of a pBSSB2 library in pUC18**

Overlapping regions of pBSSB2 were cloned into pUC18 to generate circular derivatives which could be used to screen replication function of the linear plasmid. Fragments were amplified using EcoRI and BamHI tailed primers (Table 2.4) and the MasterAmp<sup>TM</sup> Extra Long DNA polymerase (Section 2.10.3). PCR amplicons were ligated to pUC18 using EcoRI and BamHI restriction sites. Digestion, drop dialysis and ligation were carried out according to sections 2.13, 2.16 and 2.15, respectively.

Table 4.1 shows the start and end nucleotide co-ordinates of pBSSB2 used to generate fragments to be cloned in pUC18. Fragments were named RP2 to RP10. The terminal inverted repeat (TIR) of pBSSB2 is about 1230bp (Baker *et al.*, 2007b). The fragment (RP3) amplified from the left end of the plasmid started from position 1419bp which excluded the TIR. For the right end, sequence similarity was found to the phase II flagellin gene for ORF031 and to the *fljA* repressor for ORF032 (Baker *et al.*, 2007b). The cloned fragment (RP10) closest to the right terminus began at 18230bp and ended upstream of ORF031 (start position 23,683bp). A schematic map of the overlapping fragments used in this study is shown in figure 4.1.



**Table 4.1 Nucleotide co-ordinates of linear plasmid fragments inserted into pUC18**

Construct name	Starting nucleotide	Ending nucleotide
RP2	10068	14073
RP3	1419	6015
RP4	4594	9106
RP5	6954	11514
RP6	9549	13595
RP7	11350	15933
RP8	12494	18090
RP9	15938	20493
RP10	18230	22957

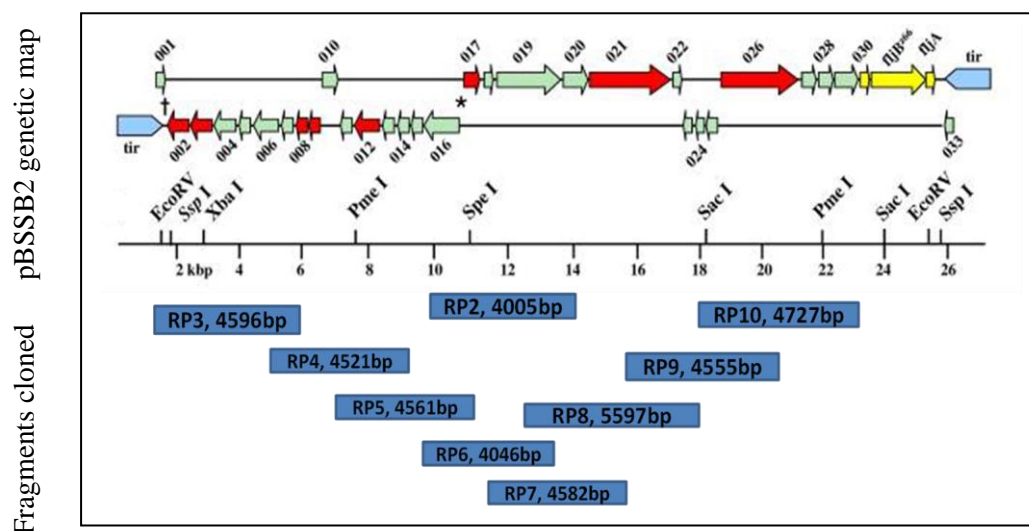


Figure 4.1 Fragments of pBSSB2 cloned into pUC18. The name and size of each fragment is shown. The pBSSB2 genetic map is adapted from Baker *et al.*, 2007b. The arrows indicate ORFs and direction of transcription. The star shows the position of the GC skew. A restriction digestion map is also shown.

After construction of the pUC18 library (Figure 4.1), the clones were validated by size confirmation following restriction digestion (Figure 4.2) and by sequencing of the ends of the inserted fragments.

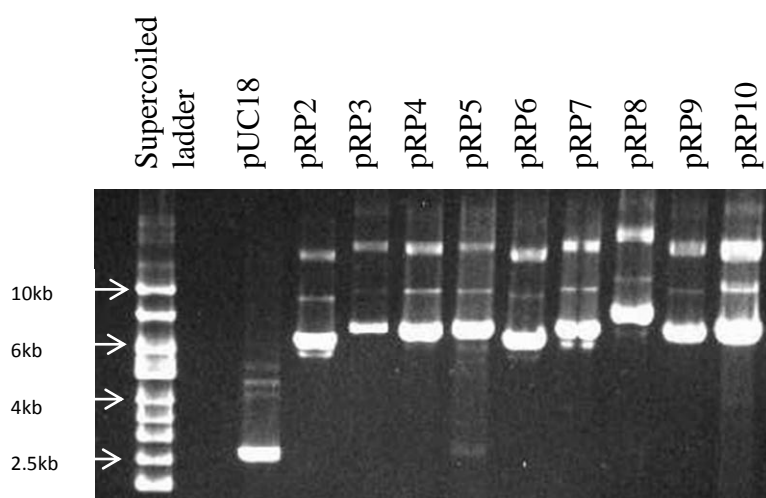


Figure 4.2. Linear plasmid library. Lane 1: Supercoiled ladder (NEB), Lane2: Blank, Lane 3: pUC18, Lanes 4-12: pRP2-pRP10

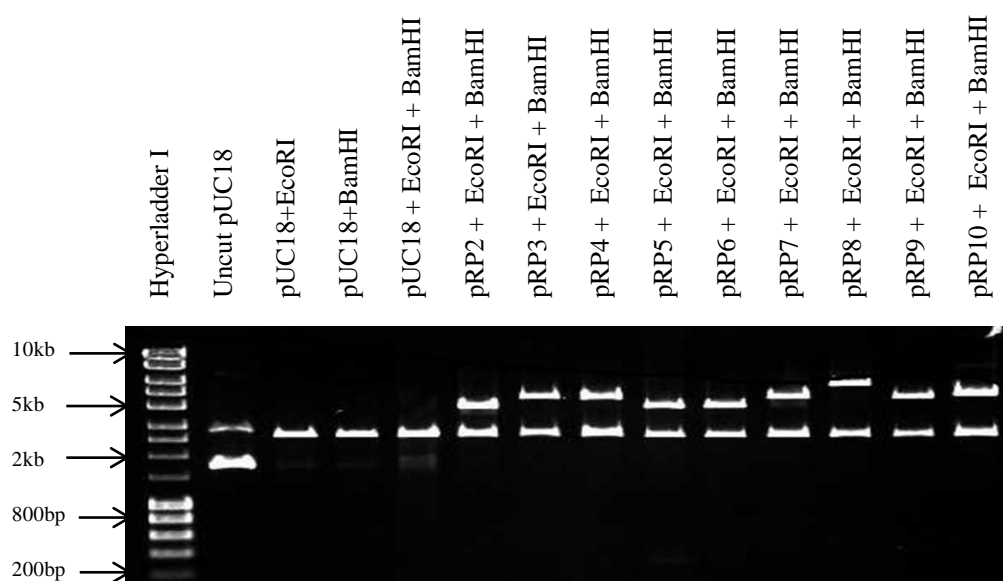


Figure 4.3. Validation of cloned fragments RP2-RP10 by restriction digestion. EcoRI and BamHI were used for confirmation of insert sizes. pUC18 was used as a control and was digested with each enzyme separately to confirm enzymatic activity of each. Uncut pUC18 contains the vector incubated in the reaction mixture minus restriction enzyme.

### **4.3 Identification of pBSSB2 replicon function in a circular mode**

In order to identify which of the pBSSB2 cloned fragments contained a replicon, pRP2-pRP10 were individually transformed into BR825, a DNA polymerase I mutant in which the pUC18 replicon is inactive (Figure 4.10, step: 1). pUC18 (no insert) was used as a negative control and pREG531 (which contains a P1 origin of replication) was used as a positive control for replication. The optimized protocol (Section 2.18.2) was used for transformation of constructs and control plasmids into BR825. Triplicate assays were done for each construct and control.

Colonies were observed only in the case of BR825 transformed with pRP7 and the positive control plasmid pREG531. There was no growth of BR825 transformed with pUC18 or with the other constructs containing amplicons of pBSSB2 (initially some colonies were obtained for pRP8 and pRP9 but upon repeated assays, this result was not reproducible).

To confirm the presence of pRP7 in the BR825 transformants, five randomly picked colonies were inoculated individually into 10ml of L-broth containing carbenicillin ( $50\mu\text{g ml}^{-1}$ ) and incubated overnight at  $37^{\circ}\text{C}$  in a static incubator. Plasmid was purified using the QIAprep Spin Miniprep Kit (Qiagen) following the manufacturer's recommendations (Figure 4.10, step: 2). About  $5\mu\text{l}$  of purified plasmid was analysed by agarose gel electrophoresis.

No plasmid band was visible in the gel (data not shown). However, the presence of pRP7 in the sample was suggested by successful PCR amplification of a 500bp fragment present within RP7 using primers described in Table 2.4 (Figure 4.4). The 500bp band was confirmed to contain linear plasmid sequence by sequencing. The amplification of the 500bp fragment, in spite of the lack of any visible plasmid band, suggested that pRP7 might replicate in BR825 with low copy number or that it had integrated into the chromosome. In order to further test for the presence of free pRP7, aliquots of the plasmid DNA preparations from the five different BR825 colonies were transformed into chemically competent XL1-Blue (Agilent technologies) cells following the manufacturer's recommendations (Figure 4.10, step: 3). XL1-Blue cells are *polA*<sup>+</sup> and hence will allow plasmid replication from the pUC18 replicon at high copy number. Transformed cells were spread on L-agar containing carbenicillin ( $50\mu\text{g ml}^{-1}$ ) and incubated at  $37^{\circ}\text{C}$  overnight.

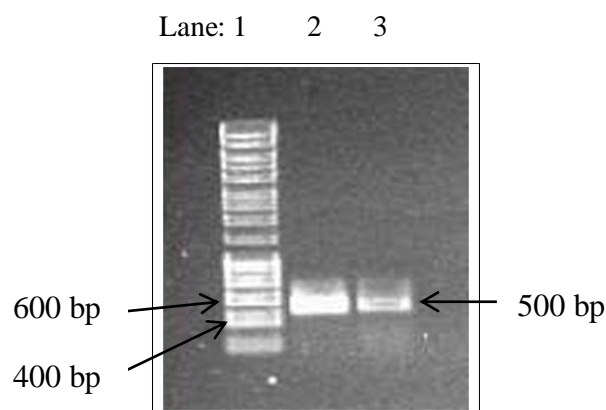


Figure 4.4. Amplification of a 500bp RP7 product from strain BR825 transformed with pRP7. Lane 1 contains Hyperladder I (Bioline) used as a molecular weight marker; lane 2 shows the amplification using pBSSB2 as template and lane 3 shows the amplicon using one of the plasmid preparations as template.

For each plasmid preparation that was transformed into XL1-Blue cells, one transformant colony was picked at random using a sterile toothpick and inoculated into an individual tube containing 10ml of L-broth and carbenicillin ( $50\mu\text{g ml}^{-1}$ ). The cultures were incubated overnight at  $37^{\circ}\text{C}$  at 120 rpm. The QIAprep Miniprep Kit (Qiagen) was used to purify plasmid DNA (Figure 4.10, step: 4) which was analysed in an agarose (1%) gel. Plasmids of two different sizes were identified from the transformants (Figure 4.5).

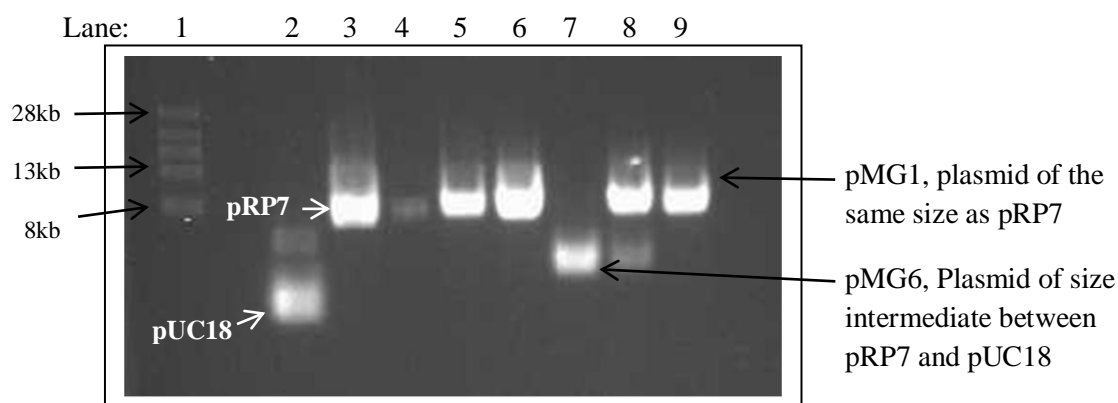


Figure 4.5. Plasmids extracted from XL1-Blue transformants following transformation with plasmid extracted from BR825 transformed with pRP7. Lane 1: Supercoiled ladder from Epicentre, Lane 2: pUC18 (2.6kb), Lane 3: pRP7 (4.5kb), Lane 4: Carry-over from lane 3, Lane 5 to 9: Plasmids extracted from transformants.

The same experiment was repeated (Figure 4.10, step: 5) and another five transformants were picked for plasmid purification (Figure 4.6). A mixture of plasmid sizes was observed in a number of XL1-Blue transformants (Figure 4.6, Lanes 5, 7, 9 and 11).

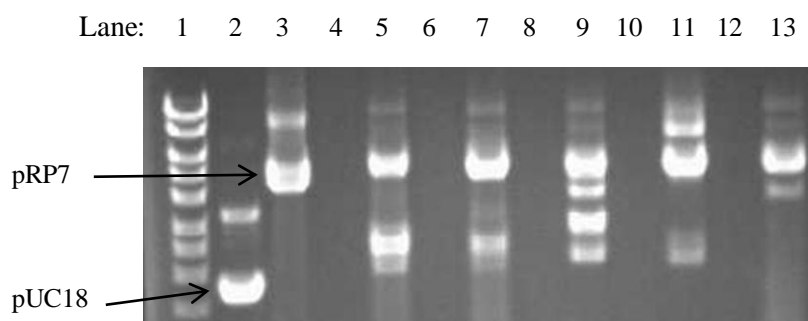


Figure 4.6. Plasmids extracted from XL1-Blue transformants following transformation with plasmid extracted from BR825 transformed with pRP7. Lane 1: Hyperladder 1 (Bioline), Lane 2: pUC18, Lane 3: pRP7, Lane 4: Blank, Lane 5, 7, 9, 11 and 13: plasmids purified from XL1-Blue.

The presence of plasmids of different sizes might be due to multiple transformation of BR825 or XL1-Blue, or picking of XL1-Blue colonies founded by more than one cell. Finally,

rearrangements of the plasmid might occur in either BR825 or XL1-Blue after transformation. Considering the possibility that smaller plasmids were natural deletion derivatives of pRP7, it was decided to test plasmids of different sizes for their ability to generate BR825 transformants. The largest plasmid from the gels was named pMG1 (Figure 4.5, lane 5) and the smallest named pMG6 (Figure 4.5, lane 7). These were purified using a QIAquick gel purification kit (Qiagen). Total plasmid DNA from the XL1-Blue transformant that contained the maximum number of plasmid bands (Figure 4.6, lane 9) was also used for transformation into BR825 to screen any of the plasmids of intermediate sizes that might contain the pBSSB2 replicon.

Plasmids pMG1, pMG6 and the plasmid mixture were transformed individually into BR825, using pRP7 and pUC18 as positive and negative controls, respectively (Figure 4.10, step: 6). BR825 transformant colonies were observed with pMG1, the plasmid mixture and the positive control, pRP7. No colonies were observed for pMG6 or with the negative control plasmid, pUC18. Plasmid pMG1 appeared to be of similar size to pRP7. With the hope that a smaller plasmid derivative of pRP7 might be recovered by screening colonies transformed with the plasmid mix, fourteen transformants were picked and plasmids purified from these (Figure 4.10, step: 7). These fourteen plasmid samples were transformed individually into XL1-Blue (Figure 4.10, step: 8). Subsequently, plasmids purified from XL1-Blue (Figure 4.10, step: 9) were resolved in agarose (1%) gel (Figure 4.7). Plasmids of several different sizes were observed in the different colonies.

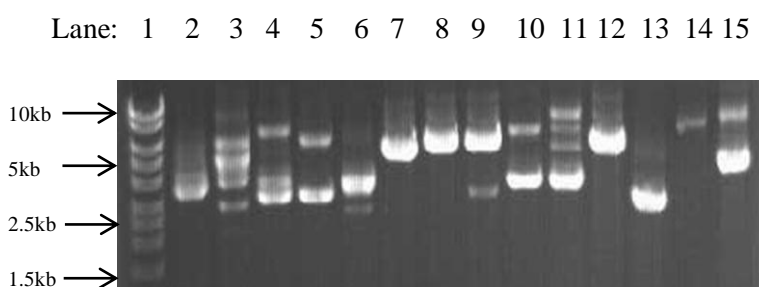


Figure 4.7. Extraction of plasmid from XL1-Blue colonies transformed with plasmid preparations from BR825 transformed originally with pRP7. Lane 1: Hyperladder I (Bioline), Lane 2-13: Plasmids purified from twelve randomly-picked XL1-Blue transformants.

Samples from lanes 2, 5 and 11 were re-run and ten plasmid bands of different sizes were purified from the gel using the QIAquick gel purification kit (Qiagen) (Figure 4.10, step: 10). Each of these plasmids was transformed again into XL1-Blue in order to obtain transformants with unique plasmid sizes (Figure 4.8; Figure 4.10, step: 11-12).

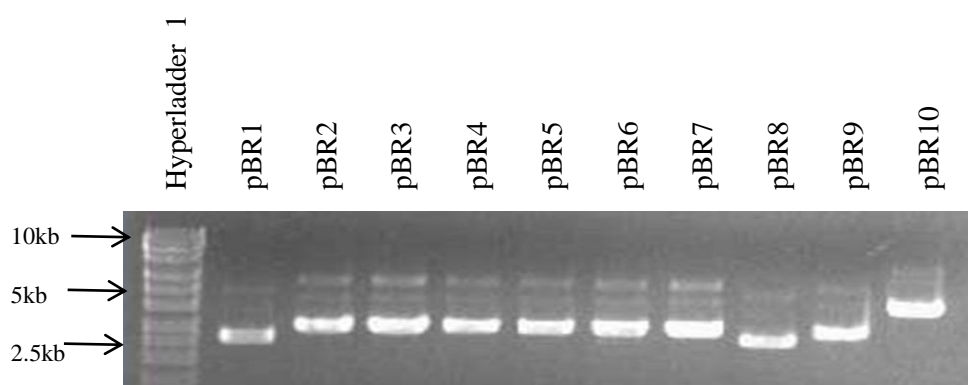


Figure 4.8. Plasmids purified from XL1-Blue transformants transformed with plasmid bands cut out from gel carrying mixed plasmids. Lane 1: Hyperladder 1 (Bioline), Lane 2-11: Purified plasmids.

Plasmids from lanes 6, 7, 9 and 10 (Figure 4.9) were selected for further study. These were designated pBR6, pBR7, pBR9 and pBR10. First they were digested with EcoRI and BamHI in order to determine the sizes of inserts (Figure 4.10, step: 13) (Figure 4.9). Upon digestion, it appeared that inserts were of two distinct sizes, about 4kb and 2kb. Three plasmids contained the smaller insert and one the larger insert. All plasmids retained the pUC18 backbone. The plasmids were transformed into BR825 to check for their ability to replicate in the *polA* mutant. Colonies were obtained for plasmids pBR6 and pBR9 only (Figure 4.8). Both of these plasmids contained inserts of approximately 2.0kb.

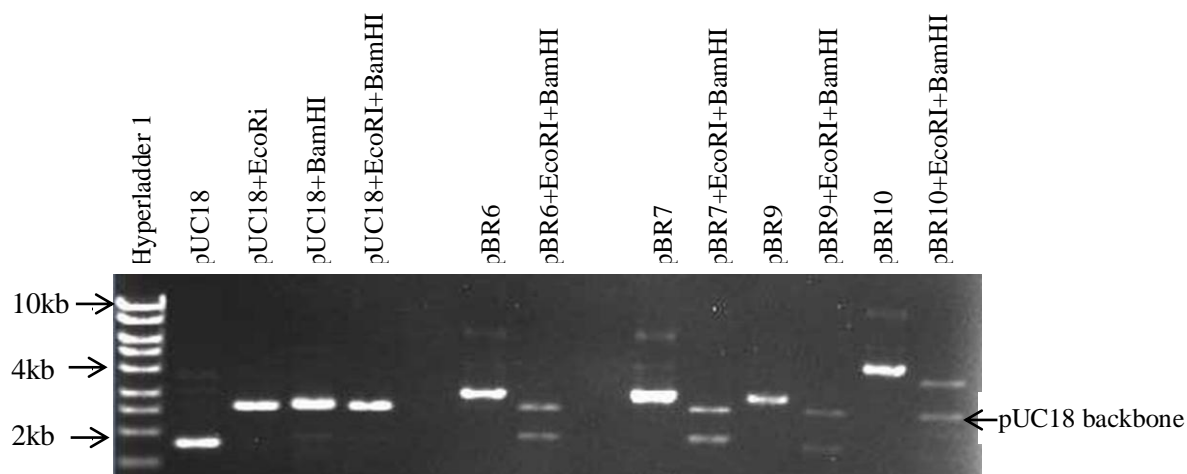


Figure 4.9. Digestion of four plasmids from XL1-Blue transformants. Lane 1: Hyperladder 1 (Bioline), Lane 2: undigested pUC18, Lane 3: pUC18 digested with EcoRI, Lane 4: pUC18 digested with BamHI, Lane 5: pUC18 digested with EcoRI and BamHI, Lanes 6 and 9: Blank, Lane 7: pBR6, Lane 8: Digested pBR6, Lane 10: pBR7, Lane 11: Digested pBR7, Lane 12: pBR9, Lane 13: Digested pBR9, Lane 14: pBR10, Lane 15: Digested pBR10.

M13/pUC18 sequencing primers were used in order to determine the exact nature of the 2kb inserts of plasmids pBR6 and pBR9 (Figure 4.10, step: 14). The insert retained a total of 1.1kb of RP7 sequences in two fragments separated by a sequence of approximately 530bp with similarity to *E. coli* insertion sequence (IS1). It contained a 815bp sized region covering nucleotides 15119-15933 (nucleotides 3770-4584 of RP7) and a 285bp sized fragment spanning nucleotides 11350-11698 (nucleotides 1-349 of RP7) of pBSSB2. The 815bp pBSSB2 fragment was amplified by PCR from pBR6 plasmid and cloned in pUC18 using EcoRI and BamHI restriction sites (Figure 4.10, step: 15). The ability of the construct to replicate in BR825 was tested (Figure 4.10, step: 16). Unfortunately, the plasmid failed to replicate in BR825.



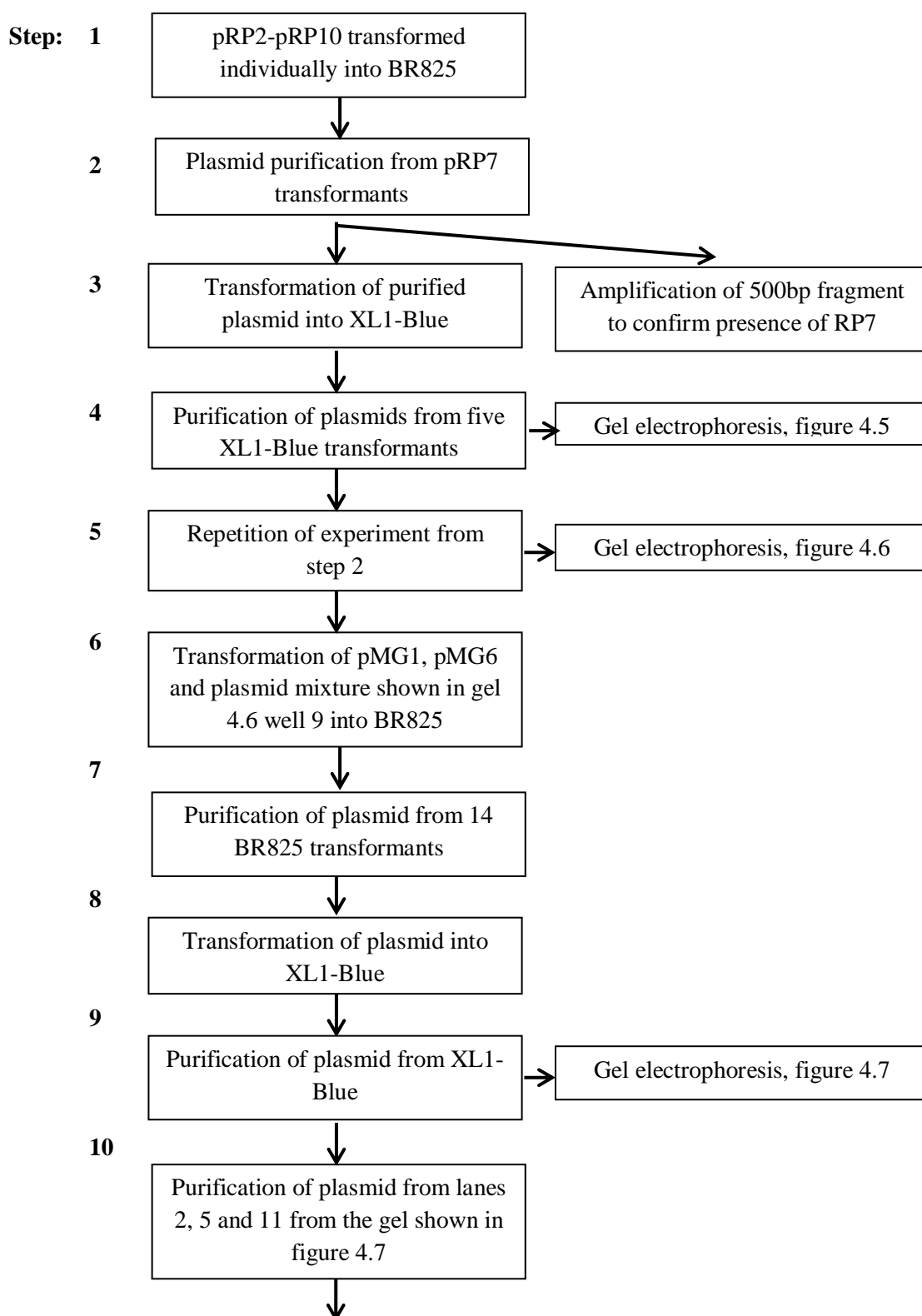


Figure 4.10. Flow diagram of the steps undertaken to identify the region of pBSSB2 that drives replication using the pRP2-pRP10 library (continued in next page)

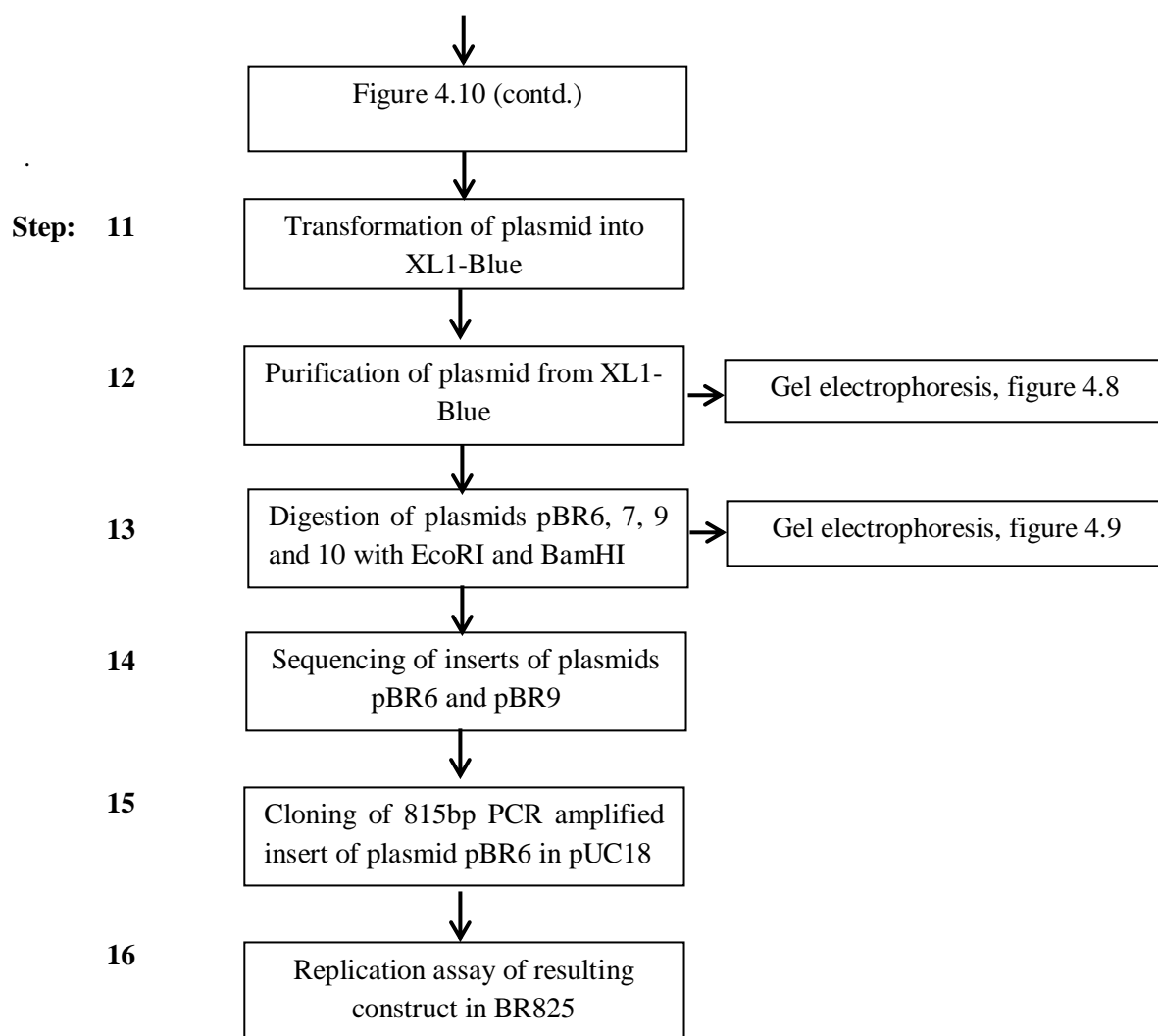


Figure 4.10. Flow diagram of the steps undertaken to identify the region of pBSSB2 that drives replication using the pRP2-pRP10 library (continuation from previous page)

In the absence of autonomous replication driven by the 815bp sequence, attention was re-directed towards pRP7. In order to further confirm the presence of an autonomous linear plasmid replicon in RP7 attempts were made to ligate fragment RP7 to a kanamycin cassette, with the assumption that this construct should be able to replicate independently. Fragment RP7 in pRP7 and a  $Km^r$  cassette cloned in pUC18 were individually digested out with EcoRI and BamHI. The digestion mix was treated with ligase and dialyzed.

About 5µl of the dialyzed DNA was transformed directly into commercially prepared chemically competent XL1-Blue cells (Agilent technologies) following the manufacturer's instructions. Transformed cells were spread on L-agar containing kanamycin ( $25\mu\text{g ml}^{-1}$ ) and incubated at  $37^\circ\text{C}$  overnight. A total of 300 randomly picked colonies were patched onto L-

agar containing kanamycin ( $25\mu\text{g ml}^{-1}$ ) and carbenicillin ( $50\mu\text{g ml}^{-1}$ ) in separate plates. Those colonies that grew in the presence of kanamycin, but not in the presence of carbenicillin, were candidates to contain RP7 ligated to  $\text{Km}^r$ , since cells with pUC18 religated to  $\text{Km}^r$  would grow in the presence of both antibiotics, whereas those containing pUC18 ligated to RP7 would grow in the presence of carbenicillin only. Three colonies were found to grow in the presence of kanamycin only. These were individually inoculated into 10ml of L-broth containing kanamycin ( $25\mu\text{g ml}^{-1}$ ). Plasmid DNA was purified from each culture using the QIAprep Miniprep Kit (Qiagen) and the sample was analysed on agarose gel. Unfortunately, no plasmid bands were visible in the gel (data not shown). It was presumed that since pRP7 replicated in BR825 with low copy number, the RP7- $\text{Km}^r$  constructs might also replicate at a copy number too low for them to be visible on the gel.

PCR was used to probe for the presence of the RP7- $\text{Km}^r$  construct by amplifying the RP7 (4.5kb) region, as well as junction regions between RP7 and  $\text{Km}^r$ , including the antibiotic cassette (Figure 4.10). From the gel of the PCR products (Figure 4.11) it was observed that even though a single PCR product (4.5kb) was detectable for RP7, multiple PCR amplicons were detected when gradient PCR was used to amplify across the junctions of RP7 and the  $\text{Km}^r$  cassette. PCR was carried out at annealing temperatures ranging from  $50^\circ\text{C}$  to  $65^\circ\text{C}$ . The same PCR primers used to generate the  $\text{Km}^r$  for the ligation were used in this reaction. Even though a single band for  $\text{Km}^r$  has been obtained consistently in earlier reactions, multiple bands were obtained in the present PCR. Consequently, in the absence of a single band for  $\text{Km}^r$  and junction regions it was not possible to confirm the presence of the construct RP7- $\text{Km}^r$ .

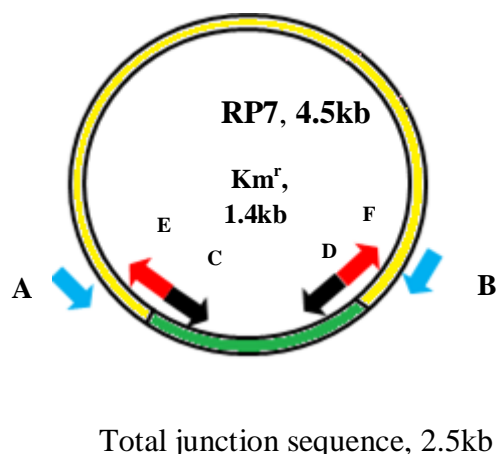


Figure 4.11. Theoretical construct RP7-Km<sup>r</sup>. The red arrows (E and F) indicate forward and reverse primers used to amplify fragment RP7 (region in yellow), black arrows (C and D) represent oligonucleotides used to amplify Km<sup>r</sup> (region shown in green) and blue arrows (A and B) indicate right and left primers used to amplify across the junction of RP7 and Km<sup>r</sup>.

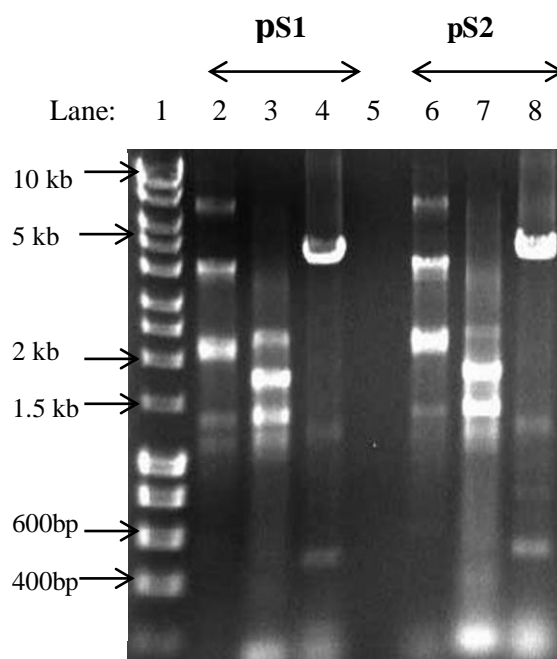


Figure 4.12. PCR carried out to detect presence of a putative construct containing RP7 ligated to Km<sup>r</sup>. Plasmid DNA from two transformants (pS1 and pS2) believed to contain RP7-Km<sup>r</sup> were used as templates. Lane 1: Hyperladder I (Bioline), lane 2: 2.5kb junction amplicon (A+B), lane 3: Km<sup>r</sup> amplicon (C+D), lane 4: RP7 amplicon (E+F). Lanes 2-4 use pS1 as template. Lanes 5, 6 and 7 use pS2 as template. Capital letters refer to primers shown in Figure 4.10.

#### **4.4 Deletion analysis of fragment RP7**

In the absence of clear evidence for the construction of RP7-Km<sup>r</sup> and consequently the inability to test it for autonomous replication, it was decided to use the DNA polymerase I mutant BR825 as the system to identify the basic replicon of the linear plasmid when inserted into a pUC18 vector.

pRP7 had been shown to replicate in BR825 (*polA*). The next step was subcloning of RP7 to identify the minimal replicon. A deletion series of RP7 was generated by PCR using the high fidelity enzyme (Section 2.10.3). Deletions were generated from either end. After amplification of each fragment, it was cloned in pUC18 using EcoRI and BamHI restriction sites. Every construct was validated by restriction digestion and size validation in agarose gel. A schematic diagram showing the deletion series is shown in figure 4.12. Each construct was individually transformed into BR825 using the optimized protocol. Construct pRP7 and pUC18 were used as positive and negative controls for replication in the BR825 host, respectively.

All deletion derivatives except one were unable to replicate. This deletion derivative was 2831bp (nucleotides 12820 to 15649 of pBSSB2) and contained the left end of RP7 with 1751bp deleted from the right end. The exclusion of even the first 144bp from the left end of RP7 prevented replication.

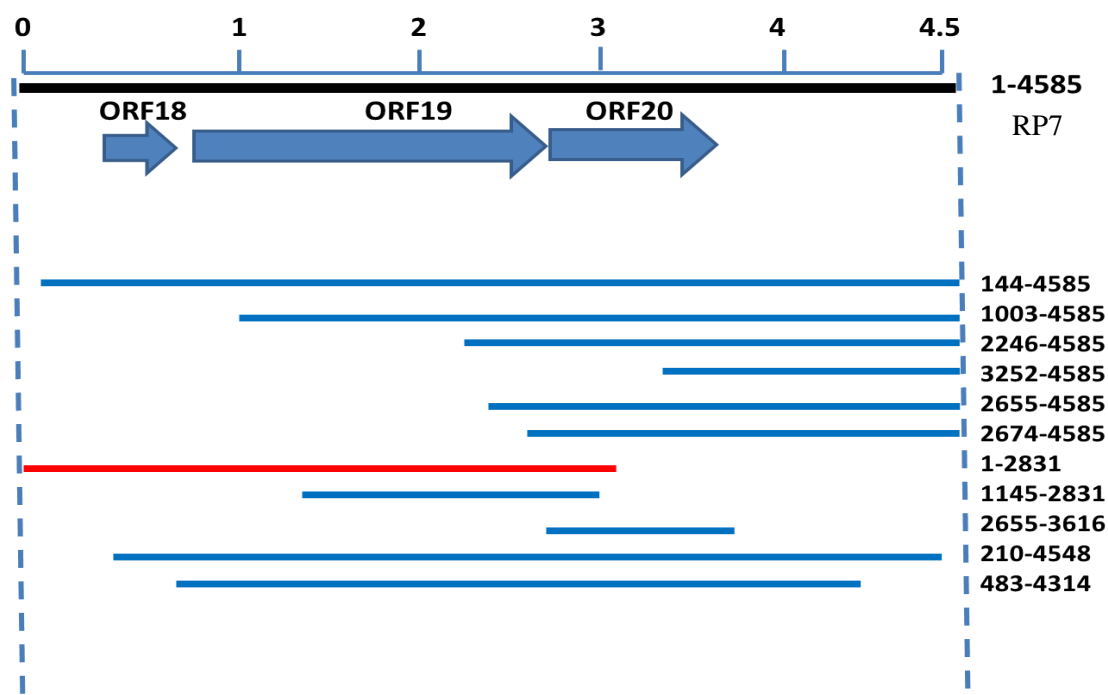


Figure 4.13. Deletion series of RP7 generated by PCR to identify the minimum replicon. The numbers represent nucleotide co-ordinates of RP7. The thick dark line represents RP7. All the other fragments are subclones of RP7 used to locate the basic replicon. Plasmids containing blue fragments failed to replicate in BR825. Only a plasmid containing the red fragment replicated successfully. Broad arrows represent open reading frames within RP7.

## 4.5 Genetic analysis of fragment RP7

According to annotations carried out by Baker *et al.* (2007b), three ORFs (ORFs 018-020), are present in fragment RP7. In order to determine whether any of these ORFs is required for replication, a premature stop codon was introduced into each ORF in turn. Suitable primers were designed using the software Primer 3 available at <http://frodo.wi.mit.edu/primer3/> so that they were present within the first one third of the total sequence. The primer sequence was designed to introduce a stop codon in the correct reading frame.

In order to introduce nonsense mutations, the QuikChange II one-day site-directed mutagenesis protocol was used (Smith, 2007). The manufacturer's instructions were followed, details of which can be found in section 2.20.2. The complete pRP7 construct was regenerated with a nonsense mutation in the reading frame of the targeted ORF. The amplified plasmids were transformed into chemically competent, commercially prepared

XL1-Blue cells (Agilent technologies). Transformant colonies were spread on L-agar containing carbenicillin ( $50\mu\text{g ml}^{-1}$ ) and incubated at  $37^{\circ}\text{C}$  overnight. A single colony was picked and inoculated into 10ml L-broth containing carbenicillin ( $50\mu\text{g ml}^{-1}$ ). The culture was incubated at  $37^{\circ}\text{C}$  at 120 rpm overnight. The QIAprep Miniprep Kit (Qiagen) was used to purify plasmid from each culture. The size of purified construct was validated by agarose gel electrophoresis. The introduction of the desired mutation was confirmed by sequencing in each case.

In the case of ORF018 a GAA triplet coding for glutamic acid at position 10 was changed to TAA, a non-sense mutation (Ochre) (Figure 4.14).

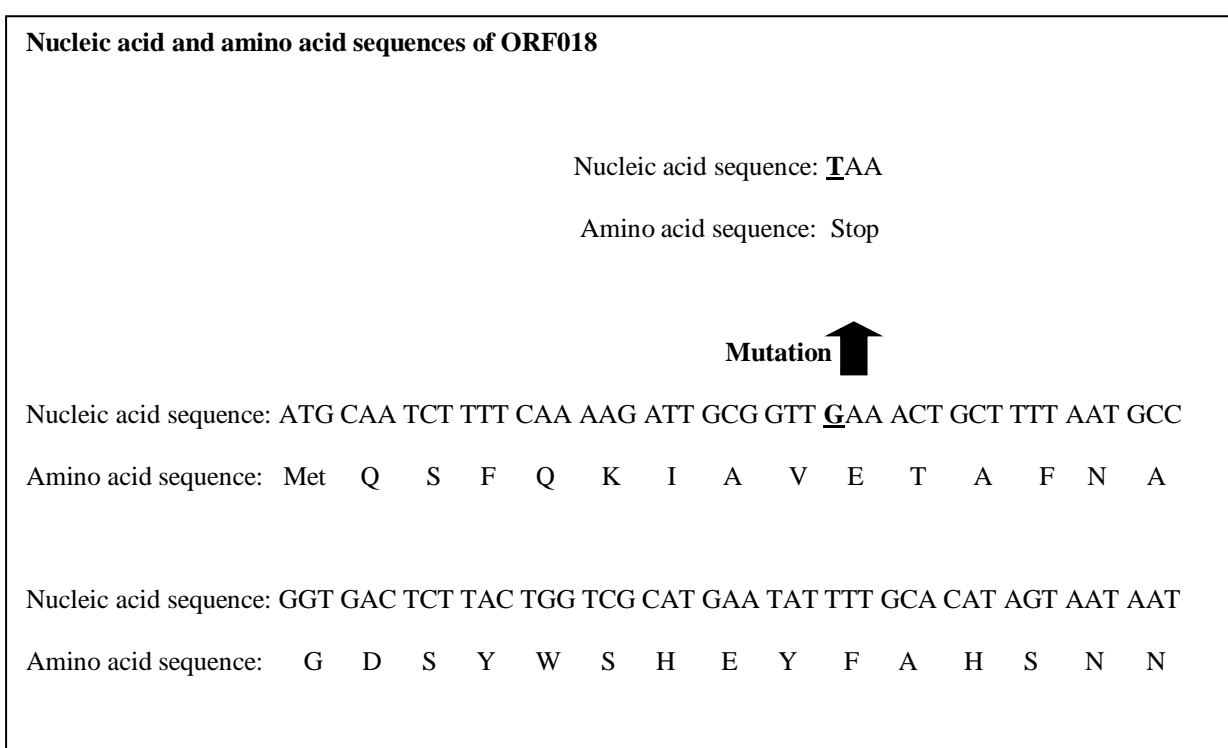


Figure 4.14. Introduction of a stop codon in ORF018. An ochre mutation was introduced at the position indicated.

For ORF019 a glutamic acid codon (GAA) was changed to a stop codon (TAA) (Figure 4.15).

Insertion of a stop codon in ORF020 was achieved by changing a leucine codon, TTG, to a TAG amber codon (Figure 4.16). (The mutagenesis was carried out before the deletion analysis (Figure 4.10) and showed that ORF020 was not required for replication.)


<b>Nucleotide sequence of ORF019</b>	
Nucleic acid sequence:	ATG GCA AAA GTT AAT GCA AAA GGT TTA AAA ACA TAT AAC CAT
Amino acid sequence:	<b>Met</b> A K V N A K G L K T Y N H
Nucleic acid sequence:	CAG GGT TTT TTA ACT TCA AAC TCA GGA GCA AAC AGC AGT GAT
Amino acid sequence:	Q G F L T S N S G A N S S D
Nucleic acid sequence:	AGT CCT CTA TAC CCA ACA CAA GAA GAT GCT TAT TAT ATC GAT
Amino acid sequence:	S P L Y P T Q E D A Y Y I D
Nucleic acid sequence:	GCC TGG TCA GGT ATA GAG ATT ACA GAT CCA AAT CCT GCA AAA
Amino acid sequence:	A W S G I E I T D P N P A K
Nucleic acid sequence:	CTT GTT AAA GAT ATA GAG GAA GGT GTA TAC GAT AAA GTT CAT
Amino acid sequence:	L V K D I E E G V Y D K V H
Nucleic acid sequence:	TTC CCT ACA AAA AGA TTT CTT AAC GAT TTA AAA GTG AAA GAT
Amino acid sequence:	F P T K R F L N D L K V K D
Nucleic acid sequence:	CTT GCA GGT GGT TTA AAG AAA TTT AAT AAG <b>CTT CAC GAG CTT</b>
Amino acid sequence:	L A G G L K K F N K L H E L
Nucleic acid sequence:	GAT <u><b>G</b></u> A A CAA CCG GAA GGT GAG CAC .....
Amino acid sequence:	D E Q P E G E H .....
<b>Mutation</b> 	
Nucleic acid sequence:	<u><b>T</b></u> A A
Amino acid sequence:	<b>Stop</b>

Figure 4.15. Introduction of a stop codon in ORF019. An ochre mutation was introduced at the position indicated.




<b>Nucleotide sequence of ORF020</b>	
Nucleic acid sequence:	ATG AAC ATA AAC AAC AAT AAG GAC AGA ATA AAA GAA ATA
Amino acid sequence:	<b>Met</b> N I N N N K D R I K E I
Nucleic acid sequence:	GAG AGT CTT GGT GTA AGC GGG GAG GAA ACT CCC CTT ACT CCA
Amino acid sequence:	E S L G V S G E E T P L T P
Nucleic acid sequence:	CCT GAT GGT ATT ATA GGA GCG ATT AAC GAG TTA AGG GAA
Amino acid sequence:	P D G I I G A I N E L R E
Nucleic acid sequence:	GAA GAG TGT TCT CTT CCC GAT CCT <u>TTG</u> .....
Amino acid sequence:	E E C S L P D P L .....
Mutation 	
Nucleic acid sequence:	<u>TAG</u>
Amino acid sequence:	<b>Stop</b>

Figure 4.16. Introduction of a stop codon in ORF020. An amber mutation was introduced at the position indicated.

Each plasmid with a mutation in ORF018, ORF019 or ORF020 was transformed into BR825 to test its ability to replicate. pRP7 was used as positive control and pUC18 as negative control for replication. Transformant colonies were obtained for all three mutant plasmids, indicating that none of these ORFs encodes a protein product essential for replication.

## 4.6 Analysis of RP7 for promoters, transcription terminators, ORFs and ribosome binding sites (RBS).

### 4.6.1 Promoters

When ORFs 018-020 were not found to play any role in replication of RP7, the 2831 bp sequence was analysed in order to identify regions that could be associated with replication.

A transcriptional analysis of pBSSB1 has been carried out at the Wellcome Trust Sanger Institute as part of PhD work of Chinyere Okoro. This analysis showed that most of the transcription was taking place from the strand in which ORFs 018-020 were present (Figure 4.17). The sequence analysis described here is complementary to this RNA-seq study.

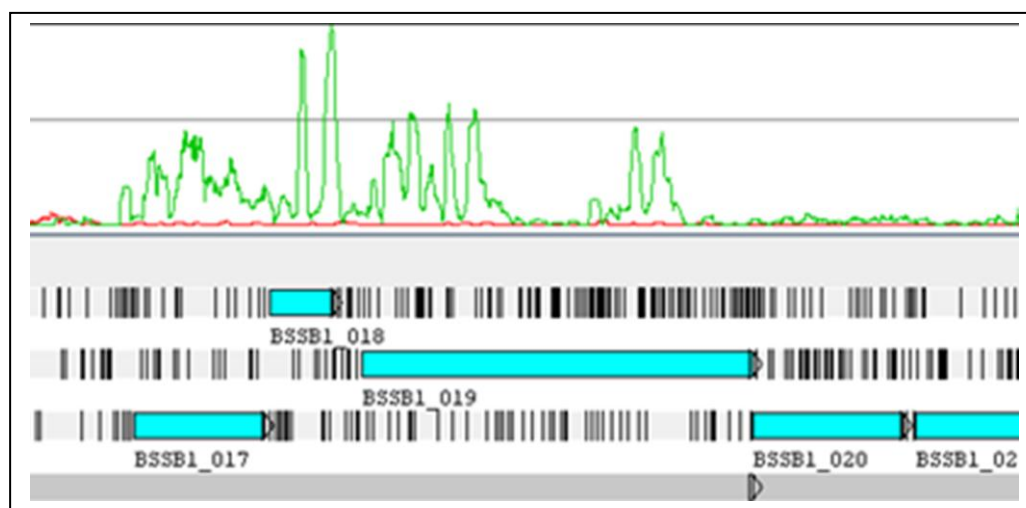


Figure 4.17. Transcription profile from the RP7 region in pBSSB2. The green trace shows transcription from the region containing ORFs 017-019. The blue bars represent ORFs with the arrows pointing in the direction in which they are transcribed. The black bars show start and stop codons. The figure was constructed from data provided by Chinyere Okoro, a graduate student at the Wellcome Trust Sanger Institute, UK.

Although the transcription profile indicates transcription of only the strand containing ORFs 018 and 019, both strands of the 2831 bp sequence were scanned for potential promoters. The scan was carried out using software available at <http://linux1.softberry.com/>. Twelve promoters were predicted by this software; seven (promoters 1-7) were in the same strand as ORFs 018 and 019 and five (promoters 8-12) were in the complementary strand. Their sequences are shown in Table 4.2 and their positions within the 2831 bp fragment are shown in figure 4.18.

Each promoter sequence was assessed for its similarity to the consensus sequences of TATAAT at the -10 Pribnow box and TTGACA at the -35 box, as well as the length of the intervening spacer. Invariants were particularly considered during analysis (Harley and Reynolds, 1987). In terms of these parameters, promoters 3 and 6 are good potential

promoters. All the other potential promoters in this strand have relatively poor matches to the consensus. In the 2083 bp strand complementary to that containing ORFs 018 and 019, promoters 8 and 10 are good potential promoters. Although the -10 and -35 sequences of promoter 9 conform closely to the consensus, the intervening spacer length of 22 bases is considered too long for this sequence to function effectively as a promoter.

**Table 4.2. Candidate promoters identified within the 2831 bp fragment using the program available at <http://linus1.softberry.com/>.**

Promoter number	-35 consensus <sup>1</sup> (TTGACA)	-35 sequence quality	-10 consensus <sup>2</sup> (TATAAT)	-10 sequence quality	Length of spacer (bp) <sup>3</sup>	Transcription start base
<b>Promoters in the same strand as ORF018 and ORF019</b>						
1	TGGACG	Poor	TATCGT	Moderate	16	C
2	TTGAAA	Good	GACTCT	Poor	17	G
3	GTGACA	Good	TAGAAT	Good	18	T
4	TTTCCA	Moderate	TATCAC	Poor	21	A
5	TTGATG	Moderate	TTTAAT	Good	19	C
6	TTGAAA	Good	TAATCT	Good	16	G
7	ATGCAA	Moderate	GAATAT	Moderate	17	T
<b>Promoters in the complementary strand</b>						
8	TTGAAA	Good	TTTAAT	Good	16	T
9	TTGTAA	Good	TATATT	Good	22	T
10	TTATTA	Moderate	TTTAAT	Good	18	C
11	TTGACT	Good	TATTCT	Moderate	22	C
12	TTTTTA	Poor	TATATT	Good	15	T

<sup>1</sup>In column 2, the bases that match with the expected -35 sequence are shown. <sup>2</sup>In column 3, the bases of the putative promoter sequences are matched with the consensus TATAAT. The invariants for the -35 and -10 sequences are indicated by light blue and bold letters. <sup>3</sup>The optimal spacer length between the -35 and -10 consensus is  $17 \pm 1$  bp.

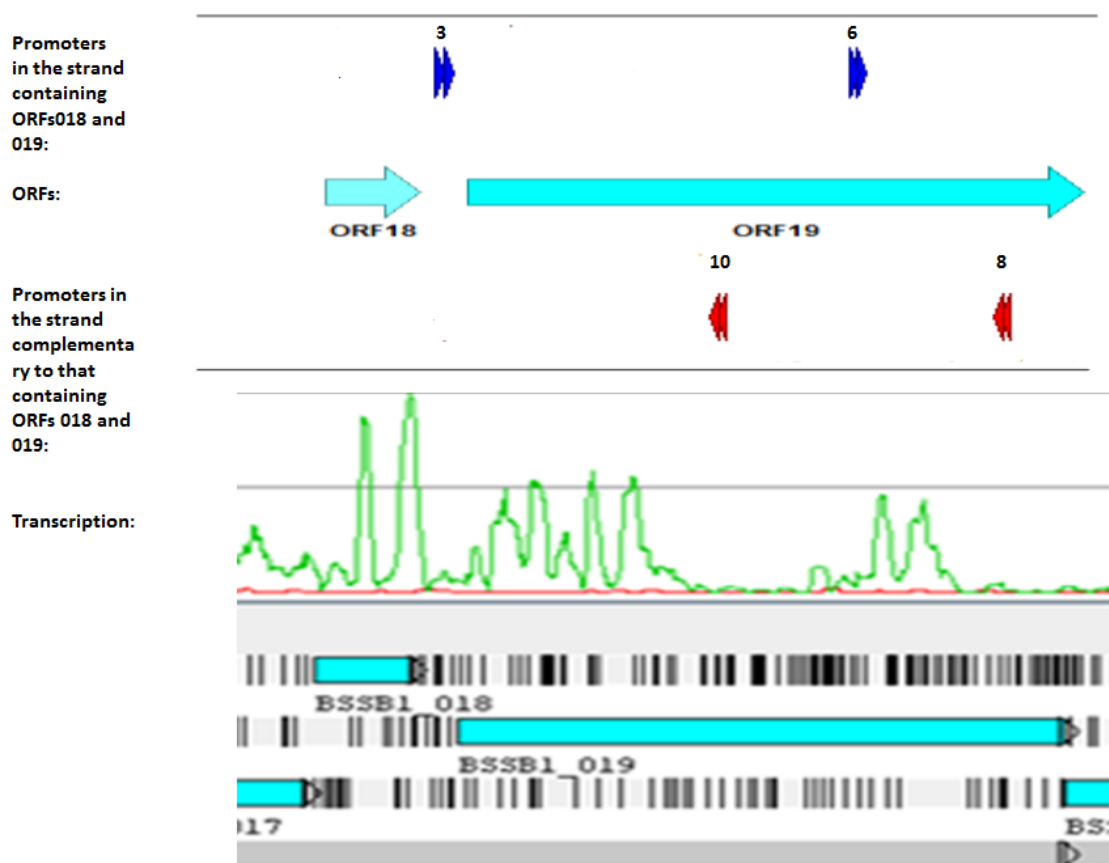


Figure 4.18. Positions of the potential promoters (arrowheads) in the two strands of the 2831 bp fragment. The dark blue arrowheads indicate promoters in the same strand as ORFs 018 and 019. The red arrowheads represent promoters in the complementary strand. Annotated ORFs are shown by light blue arrows and transcription profile by green and red traces. ORFs 018 and 019 are in the same strand from which the transcription pattern indicated by the green trace occurs.

Based on the promoter search, promoters 3 and 6 in the same strand as ORFs 018 and 019, and promoters 8 and 10 in the complementary strand were considered as potential candidates for further investigation.

#### 4.6.2 Prediction of ORFs in 2083 bp fragment

Having identified potential promoters, the sequence was next analysed for ORFs over and above those already identified, using the ORF finder tool available at [www.ncbi.nlm.nih.gov/projects/gorf/](http://www.ncbi.nlm.nih.gov/projects/gorf/), with default parameters. Six additional ORFs were found: five (ORFs A, B, C, D and E) in the same strand as ORFs 18 and 19, and two (ORFs F and G) in the complementary strand (Table 4.3; Figures 4.19 and 4.20). The sequences of ORFs A-G are given in figures 21-27. ORFs C, D and E are within ORF019 but have alternative start codons. The mutation that was introduced into ORF019 in fragment RP7 would not have affected ORFs C, D and E because the start codons for these ORFs are downstream of the introduced mutation.

**Table 4.3. ORFs present in the 2831bp replication-competent region of RP7**

Frame	Start and end nucleotide <sup>2</sup>	Length (bp)	ORF <sup>1</sup>
<b>ORFs in the same strand as ORFs 018 and 019</b>			
+3	396, 698	313	18
+1	850, 2784	1935	19
+2	158, 358	201	A
+2	182-358	177	B
+1	1561, 2784	1223	C
+1	1756, 2784	1029	D
+1	1873-2784	912	E
<b>ORFs in the complementary strand as ORFs 018 and 019</b>			
+3	2409-1207	1203	F
+1	2195-1425	771	G

<sup>1</sup>ORFs A-G have been identified using ORF finder tool available at [www.ncbi.nlm.nih.gov/projects/gorf/](http://www.ncbi.nlm.nih.gov/projects/gorf/) using default parameters. <sup>2</sup>The numbers represent nucleotide co-ordinates of the 2831 bp derivative of RP7 (see Figure 4.19 and 4.20).

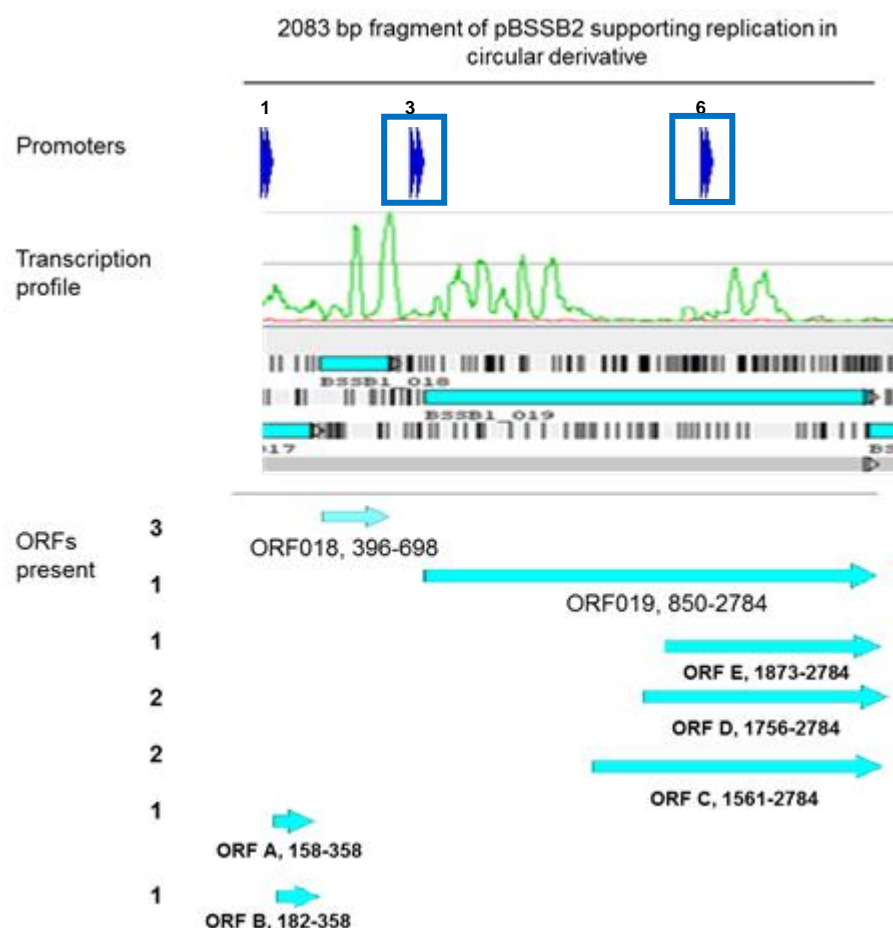


Figure 4.19. Promoters (dark blue arrowheads), transcription profile (green peaks) and ORFs (light blue arrows) present in the strand that contains ORFs 018 and 019 of the 2831 bp fragment of pBSSB2 that could support replication in a circular derivative. The numbers on the left indicate reading frames.

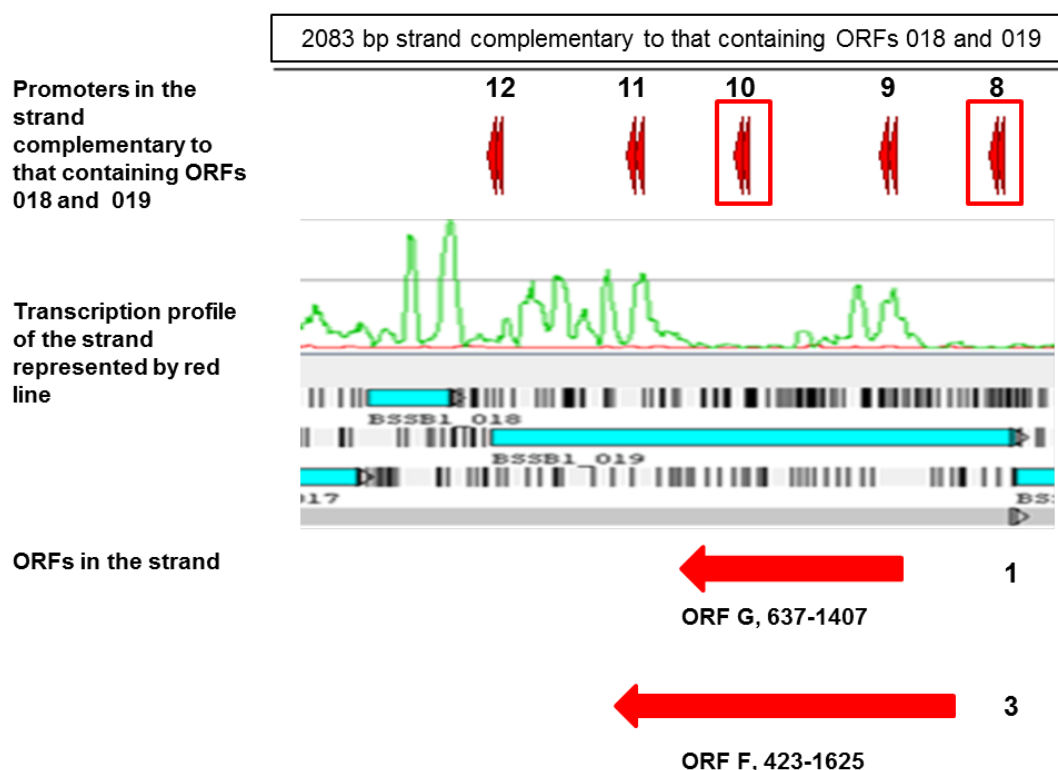


Figure 4.20. ORFs in the 2830 bp fragment in the strand complementary to that containing ORFs 018 and 019. The red arrowhead pairs indicate possible promoters. ORFs F and G are represented by solid red arrows. Transcription from this strand is indicated by the red curved line underneath the green peaks; numbers on the right indicate reading frames.

```

TGTTACGAAC TGGACG GAACGGACAAGCTCTG TATCGT CATTCCCTAATCGTGTAGAACTGCAAAACAAAGAG
CAT AAAGAC CTGCTCGAA ATGCAGTTTGAAAATGTCCCCTACATGAAGAAGCGAGTCTCTTACTCTCGTGCCT
ACGGCATGGGTGTAACCTATCTATTCGGAAGGTACAGGCTTACCGAACCTGGGGATCACTCGCCAGGAGTTTCGA
CGGTGTAGCTAAATACATTCTATAAACGAATCGAAAAACACATTGCTGATATGAAAGCAAAAGAGAAGGCATAA
CTTATGAGCAAAGAAAAAATTCATGGCTTGATGATATGCAATCTTTTCAAAGATTGCGGTTGAAACTGCT
TTTAATGCCGGTGACTCTTACTGGTCGCATG
    
```

Figure 4.21. Nucleotide sequence of ORF A (highlighted in blue). The putative Shine-Dalgarno sequence is highlighted in yellow and promoter 1 is shown in green.

```

TGTTACGAAC TGGACG GAACGGACAAGCTCTG TATCGT CATTCCCTAATCGTGTAGAACTGCAAAACAAAGAGC
ATAAAGACCTGCTCGAAATGCAGTTTGAAAATGTCCCCTAC ATGAAGAAGCGAGTCTCTTACTCTCGTGCCTAC
GGCATGGGTGTAACCTATCTATTCGGAAGGTACAGGCTTACCGAACCTGGGGATCACTCGCCAGGAGTTTCGACGG
TGTAGCTAAATACATTCTATAAACGAATCGAAAAACACATTGCTGATATGAAAGCAAAAGAGAAGGCATAA CTTA
TGAGCAAAGAAAAAATTCATGGCTTGATGATATGCAATCTTTTCAAAGATTGCGGTTGAAACTGCTTTTAA
TGCCGGTGACTCTTACTGGTCGCATG
    
```

Figure 4.22. The nucleotide sequence of ORF B (highlighted in blue). Promoter 1 is shown highlighted in green.

```

AAAGATAACAAAGTAAGAACGATTGTTGAAAAACAGATAAATATGGCTTTGGCAGCGGCTGGGATAGAAGCAACGGTAA
CCCTTGTTCTTTTACAAGAGGAAATTGGTGCGACAGATGAAATTAATCAGAAAAATTAAAAAAGAAGCAGAAGAAAA
TAATAAGCCTCTTGCTGATGTAGCACAGAACATATCAATGACGAATCTCTTGATGATGAAGAAATTTCCGACTCTTTA
ATGCAACATGGAAATCTTTCAGAACAATCCATAGAAAACTTGAGTCTATAATTAATAAAAAGCATTGACAGTAATG
AACAGTAAGAAGAAAAAGGCTCAGATAAAAGAGCTAATGGCTGAACTTGTCGAAAGAGAAGCAAGGAATTACAAGCTAT
TAATAAAGGTATTCTTTATGAAGCAAAATATATTCAAACCTAAACCTGTTGAATGAACTTCCAGAAAGCTTGAATCT
GTAAATAAAGAACTGGAAGAAGCAAAATAAACTAAATGACGATTTCAATGAAAAATAATTGAAACACTAATACTATTT
CTAATCTTAATGATGAATTAGCGAGAGAAAGAAAATATAACGAAGAACAAGAATTACAACCTAACCAGTTAGAAAAAGA
ACTTGAACCGGAAAAAATATAAATGCGGAAACCTCAGCACTCATTACAACCCCTTACCGACAAGAATACTGTTTTAGCA
GCTAACCTTGAACCTGAGCAAGGTAAGACGGCAGAACTGACAGCCGAAAAATGCAGAACTTAAAGAGAAGCATAAAGTTG
AGATCGCCACTCTTCGTAGTGAGTTCGCTCAGAACTGGTTAAATACCGTAAAACTTTGTGGCTCGTGTTAAAAACAA
AGTGGCAACGGCTGTTGAACATGCAATACAGGCATTTAGAGATAATGAATTACCTTCATTACTTAAAGATGCAACAACC
AAAGCTGTTAGTGAATATAAATTAAAGAATTGCCTGGAAAACTTAGAAAAAGTAAGAGAGGAAGAAGCCAAAAAATCAG
AGAAAGAAATTAAAGCATTAACAAACGATGTTTCAACATTGAAGACTAATGTAAATACTTTAACGGCTGATAAGATAC
ATTGCAGAAAGATTTTGACGAATTGTTTAAATGCAAGCAAAGAAAAAGCAACGGAAATGAATAATGAAATTGCTCGTCTT
AAAGAAATCTTAAAGCAATGGAATAAACCCCTGATGGTGATAACTCACCAAAAAAGCCTAAAAATGACAATAAGATAT
AAGGGGTGACTTATGAACATAAACACAATAAGGACAGAATAAAAGAAA

```

Figure 4.23. The nucleotide sequence of ORF C (highlighted in blue). The SD sequence is shown in yellow.

```

GACGAATCTCTTGATGATGAAGAAATTTCCGACTCTTTAATGCAACATGGAAATCTTTCAGAACAATCCATAGAAAACT
TGAGTCTATAATTAATAAAAAGCATTGACAGTAATGAACAAGTAAGAAGAAAAAGGCTCAGATAAAAGAGCTAATGG
CTGAACTTGTCGAAAGAGAAGCAAGGAATTACAAGCTATTAATAAAGGTATCTTTATGAAGCAAAATATATTCAAACCTAA
AACCTGTTGAATGAACTTCCAGAAAGCTTGAATCTGTAATAAAGAAGCTTGAAGAAGCAAAATAAACTAAATGACGATTT
CAATGAAAAATAATTGAAACACTAATACTATTTCTAATCTTAATGATGAATTAGCGAGAGAAAGAAAATATAACGAAG
AACAGAATTACAACCTAACCAGTTAGAAAAAGAACTTGAAACCGAAAAAATATAAATGCGGAAACCTCAGCACTCATT
ACAACCCCTTACCGACAAGAATACTGTTTTAGCAGCTAACCTTGAACCTGAGCAAGGTAAGACGGCAGAACTGACAGCCGA
AAATGCAGAACTTAAAGAGAAGCATAAAGTTGAGATCGCCACTCTTCGTAGTGAGTTCGCTCAGAACTGGTTAAATACC
GTAAAACTTTGTGGCTCGTGTTAAAAACAAAGTGGCAACGGCTGTTGAACATGCAATACAGGCATTTAGAGATAATGAA
TTACCTTCATTACTTAAAGATGCAACAACCAAGCTGTAGTGAATATAAATTAAAGAATTGCCTGGAAAACTTAGAAA
AGTAAGAGAGGAAGAAGCCAAAAAATCAGAGAAAGAAATTAAAGCATTAACAAACGATGTTTCAACATTGAAGACTAATG
TAAATACTTTAACGGCTGATAAGATACATTGCAGAAAGATTTTGACGAATTGTTTAAATGCAAGCAAAGAAAAAGCAACG
GAAATGAATAATGAAATTGCTCGTCTTAAAGAAATCTTAAAGCAATGGAATAAACCCCTGATGGTGATAACTCACCAAA
AAAGCCTAAAAATGACAATAAGATATAAGGGGTGACTTATGAACATAAACACAATAAGGACAGAATAAAAGAAA

```

Figure 4.24. The nucleotide sequence of ORF D (highlighted in blue).



```

AACAGGCTCAGAT AAAAGA GCTA ATGGCTGAACTTGTGCGAAAGAGAACAGGAATTACAAGCTATTAATAAAGG
TATTCTTTATGAAGCAAAATATATTCAAACATAAAACCTGTTGAATGAACTTCCAGAAAGCTTGAATCTGTAA
ATAAAGAAGCTTGAAGAAGCAAAATAAACTAAATGACGATTTCAATGAAAAATAATTGAAAACACTAATACTATT
TCTAATCTTAATGATGAATTAGCGAGAGAAAGAAAAATATAACGAAGAACAAGAATTACAACCTTAACCAGTTAGA
AAAAGAAGCTTGAACCGAAAAAAATATAAATGCGGAAAACCTCAGCACTCATTACAACCCCTTACCGACAAGAATA
CTGTTTTAGCAGCTAACCTTGAACTTGAGCAAGGTAAGACGGCAGAACTGACAGCCGAAAATGCAGAACTTAAA
GAGAAGCATAAAGTTGAGATCGCCACTCTTCGTAGTGAGTTCGCTCAGAACTGGTTAAATACCGTAAAAACTT
TGTGGCTCGTGTAAAAACAAAGTGGCAACGGCTGTTGAACATGCAATACAGGCATTTAGAGATAATGAATTAC
CTTCATTACTTAAAGATGCAACAACCAAGCTGTTAGTGAATATAAATTAAAAGAATTGCCTGGAAAACCTTAGA
AAAGTAAGAGAGGAAGAAGCCAAAAATCAGAGAAAGAAATTAAAGCATTACAAACGATGTTTCAACATTGAA
GACTAATGTAAATACTTTAACGGCTGATAAAGATACATTGCAGAAAGATTTTGACGAATTGTTTAAATGCAAGCA
AAGAAAAAGCAACGGAAATGAATAATGAAATTGCTCGTCTTAAAGAAATTCTTAAAGCAATGGAATAAACCCCT
GATGGTGATAACTCACCAAAAAAGCCTAAAAATGACAATAAGATATAA GGGGTGACTTATGAACATAAACACA
ATAAGGACAGAATAAAAGAAA

```

Figure 4.25. Nucleotide sequence of ORF E in blue with the putative SD in yellow.

```

TTCTGCAATGTATCTTTATCAGCCGTTAAAGTATTTACATTAGTCTTCAATG TTGAAA CATCGTTTGTTAATG
CTTTAAATTTCTTTCTCTGATTTTTTGGCTTCTTCCTCTCTTACTTTTCTAAGTTTCCAGGCAATCTTTTAA
TTTATATTCTACTAACAGCTTTGGTTGTTGCATCTTTAAGTAATGAAGGTAATTCATTATCTCT AAATGC CTGT
ATTGCATGTTTCAACAGCCGTTGCCACTTTGTTTTTAACACGAGCCACAAAGTTTTTACGGTATTTAACCAGTT
TCTGAGCGAACTCACTACGAAGAGTGGCGATCTCAACTTTATGCTTCTCTTTAAGTTCTGCATTTTCGGCTGT
CAGTTCTGCCGTCTTACCTTGCTCAAGTTCAAGGTTAGCTGCTAAAACAGTATTCTTGTGCGTAAAGGTTGTA
ATGAGTGTGAGTTTTCCGCATTTATATTTTTTTCGGTTTTCAAGTTCTTTTTCTAACTGGTTAAGTTGTAATT
CTTGTTCTTCGTTATATTTCTTTCTCTCGCTAATTCATCATTAAAGATTAGAAATAGTATTAGTGTTTTCAAT
TATTTTTTCATTGAAATCGTCATTTAGTTTATTTGCTTCTTCAAGTTCTTTATTTACAGATTCAAGCTTTCTG
GAAGTTTCATTCAACAGGTTTTTAGTTTGAATATATTTTGCTTCATAAAGAATACCTTTATTAATAGCTTGTA
ATTCCTGTCTCTTTTCGACAAGTTTCAGCCATTAGCTCTTTTATCTGAGCCTGTTTTCTTCTTACTTGTTCATT
ACTGTCAAATGCTGTTTTATTAATTATAGACTCAAGTTTTCTATGGATTGTTCTGAAAGATTCCATGTTGC
ATTAAAGAGTCGGAAATTTCTTCATCATCAAGAGATTGCTCATTGATATGTTCTTGTGCTACATCAGCAAGAG
GCTTATTATTTTCTCTGCTTCTTTTTTAATTTTTCTGATTTAATTTTCATCTGTGCGACCAATTTCTCTTG
TAAAGAACCAAGGGTACCCTTGCTTCTATCCCAGCCGCTGCCAAAGCCATATTATCTGGTTTTCAACAATC
GTTCTTACTTTGTTATCTTTTAGTAATAAAGCAGCATTGACAATCTTTTATTAGGATTGCTGCATTTATAA
ATGGAATAAACTCTCCTTTATCATTTCGTTGAAACACCATAATCTTATCAGTGATACAATGAGCAGAACGGAA
GAGGTGGAATGGTCTGGAATTACATTTCCCAATGATCTGTCTGGTTGCGGTGAATCCCCAGGCAAGCACA
CTTTGCAGACCTTTATTGACTATAATCTGTGGTTTTCTACCGCTATTCTGTTCTTTCTGAATATAAAATCGG
GGTGAGAGATTTCTCTTATGATAATTTCCCTGGTTTTTGA TTGTAGGTGCTTAGGAACACCATCAAGATAAAA
GACAAAGTGCTCACCTTCCGTTGTTTCATCAAGCTCGTGAAGCTTATTAAATTTCTTT

```

Figure 4.26. The sequence of ORF F highlighted in blue. The ribosome binding site is indicated in yellow, promoter sequences in green and the sequences involved in termination of transcription in pink.

```

TTCTGCAATGTATCTTTATCAGCCGTTAAAGTATTTACATTAGTCTTCAATGTTGAAACATCGTTTGTTAATGCT
TTAATTTCTTTCTCTGATTTTTTTGGCTTCTTCCTCTCTTACTTTTCTAAGTTTTCCAGGCAATCTTTTAATTTA
TATTCATAACAGCTTTGGTTGTTGCATCTTTAAGTAATGAAGGTAATTCATTATCTCTAAATGCCTGTATTGCA
TGTTCAACAGCCGTTGCCACTTTGTTTTTAACACGAGCCACAAAGTTTTTACGGTATTTAACCAGTTTCTGAGCG
AACTCACTACGAAGAGTGCGGATCTCAACTTTATGCTTCTCTTAAAGTTCTGCATTTTTCGGCTGTCAGTTCTGCC
GTCTTACCTTGCTCAAGTTCAAGGTTAGCTGCTAAAACAGTATTCTTGTCGGTAAGGGTTGTAATGAGTGCTGAG
TTTTCCGCATTTATATTTTTTTTCGGTTTCAAGTTCTTTTTCTAACTGGTTAAGTTGTAATCTTGTCTTCGTTA
TATTTTCTTTCTCTCGCTAATTCATCATTAAAGATTAGAAATAGTATTAGTGTTTTCAATTATTTTTTCATTGAAA
TCGTCATTTAGTTTATTTGCTTCTTCAAGTTCTTTATTTACAGATTCAAGCTTCTGGAAGTTTCATTCAACAGG
TTTTTAGTTTGAATATATTTTGCTTCATAAAGAATACCTTTATTAATAGCTTGTAAATCCTGTTCTCTTCGACA
AGTTCAGCCATTAGCTCTTTTATCTGAGCCTGTTTTCTTCTTACTTGTTCACTACTGTCAAATGCTGTTTTATTA
ATTATAGACTCAAGTTTTTCTATGGATTGTTCTGAAAGATTTCATGTTGCATTAAAGAGTCGGAATTTCTTCA
TCATCAAGAGATTTCGTCATTGATATGTTCTTGCTACATCAGCAAGAGGCTTATTATTTTCTTCTGCTTCTTTT
TTTAATTTTCTGATTTAATTTTCATCTGTCGACCAATTTCCCTCTTGTAAGAACCAAGGGTTACCGTTGCTTCT
ATCCCAGCCGCTGCCAAAGCCATATTTATCTGGTTTTTCAACAATCGTTCTTACTTTGTTATCTTTTAGTAATAAA
GCAGCATTGACAATTCCTTTATTAGGATTTGCTGCATTTATAAATGGAATAAACTCTCCTTTATCATTTCGTTGAA
ACACCATAATTCTTATCAGTGATACAATGAGCAGAACGGAAGAGGTGGAATGTTCTGGAATTACATTTCCCAAA
TGATCTGTCTGGTTGCGGTGAATCCCCAGGCAAGCACACTTTGCAGACCTTTATTGACTATAATCTGTGGTTTT
CTACCGCTATTCTGTTCTTTTCTGAATATAAAATCGGGGTGAGAGATTTCTCTTATGATAATTTCCCTGGTTTTT
GATTGTAGGTGCTTAGGAACACCATCAAGATAAAAGACAAAGTGCTCACCTTCCGGTTGTTTCATCAAGCTCGTGA
AGCTTATTAAATTTCTTT

```

Figure 4.27. Nucleotide sequence of ORF G in blue. The SD sequence is in yellow.

Based on the relative positions of the candidate promoters (3 and 6) and the ORFs in the strand containing ORFs 018 and 019, it appeared that promoter 3, which is 67 bases upstream of the start codon of ORF019, could be a potential promoter for ORFs 019, C, D and E.

In the complementary strand, promoter 8 could potentially transcribe ORFs F and G. Promoter 8 is 144 bases upstream of the start codon of ORF F (Figure 4.26) and 358 bases upstream of ORF G (Figure 4.27). Thus Promoter 8 can potentially direct transcription of an mRNA containing ORF F and ORF G. On the other hand, promoter 9 is contained within ORF G and cannot prime its transcription. At this stage both ORFs F and G are potential candidates for encoding rep protein, with ORF F being a stronger candidate because of its proximity to promoter 8.

#### 4.6.3 Ribosome binding sites

In order to predict which of the ORFs were likely to be expressed, the sequences were analysed further to identify putative Shine-Dalgarno sequences upstream (<10 bp) of the

ORFs and similar to the conserved *E. coli* SD sequence (AGGAGG). A candidate SD sequence, AAAGAC, was identified adjacent to the start of ORF A (Figure 4.21). No putative SD could be identified for ORF B (Figure 4.22) or D (Figure 4.24). A candidate SD sequence of GAAAAC was found adjacent to ORF C (Figure 4.23). The sequence AAAAGA could act as a SD sequence for translation of ORF E (Figure 4.25). In case of ORF F a candidate Shine-Dalgarno of AAATGC was found (Figure 4.26). Scanning the sequences upstream of ORF G also indicated a potential SD sequence of GTAAGG (Figure 4.27). On the basis of the finding of candidate SD sequences, it appears that ORFs A, C, E-G could be potentially translated.

#### 4.6.4 Transcription terminators

Intrinsic terminators of transcription were searched by looking for GC-rich sequences that were followed by T-rich segments (Farnham and Platt, 1981; reviewed by Platt, 1986; Wilson and Hippel, 1995). Alternatively, rho-dependent terminators were searched by looking for C-rich regions (reviewed by Richardson and Greenblatt, 1996). The sequence immediately adjacent to the termination codon of ORF A/B was not significantly GC rich to allow conventional termination of transcription. However, a relatively GC rich sequence (GCTGCTCATGTCGTCAC) that was followed by a region containing 13 A+T was found 380 bases downstream of promoter 1. This GC rich sequence can potentially form a stem-and-loop structure consisting of two GC bonds that could help in termination of transcription of any transcript from promoter 1 (Figure 4.28).

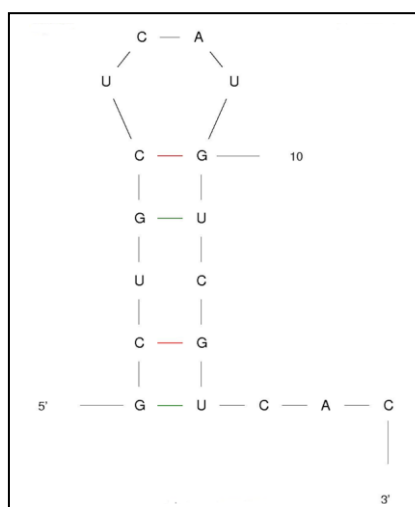


Figure 4.28: Candidate transcription termination sequence for transcript from promoter 1.

In case of the complementary strand to that containing ORFs 018 and 019, a sequence (TTGTAGGTGCTTAGGAACACCATCAAG) immediately adjacent of the termination codon of ORF F, followed by a string of 6 pyrimidines (mostly A's), could act as a potential transcription terminator (highlighted in pink in figure 4.23). The mfold software (www.mfold.rna.albany.edu) predicted a stem-and-loop structure with four GC- bonds and four AU bonds (Figure 4.29). This could act as a potential transcription terminator for both ORFs F and G.

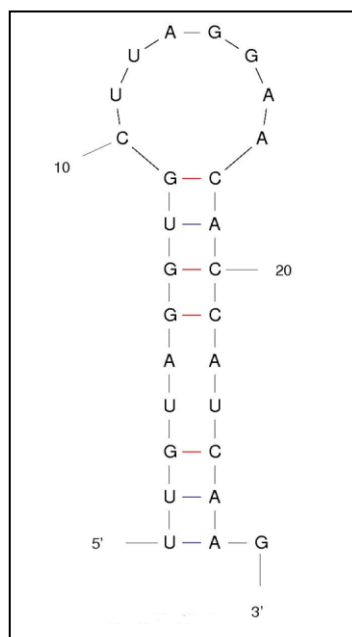


Figure 4.29. A potential stem-and-loop terminator structure downstream of ORF F and G. GC bonds are shown in red.

#### 4.6.5 Candidate replication initiator protein genes based on sequence analysis

In section 4.4, it was found that deletion of the first 144 bp from the left end abolished replication driven by the 2831 bp sequence. Deletion of 144 bases deletes promoter 1. Promoter 1 which is located 99 bp downstream of the end at which the deletion was made may actually play a role in transcription from this end. Further analysis indicates that ORF A starts 52 bases and ORF B starts 76 bases downstream of the -10 box of promoter 1 (Figure 4.18 and 4.19, respectively). Based on the deletion analysis and in the absence of a putative SD sequence for ORF B, ORF A is proposed to be a more likely candidate for translation of a putative rep protein than ORF B in the strand that contains ORF 019.

In the complementary strand, considering the position of promoter 8, the presence of a candidate SD sequences and the presence of a candidate terminator of transcription, it seems possible that either ORF F or ORF G might be a suitable candidate for further investigation as an initiator of replication or rep protein.

#### **4.6.6 Alternative hypothesis based on possible candidates for RNA initiators of replication**

ORF019 has already been experimentally demonstrated not to be involved in replication. The other alternative is priming of replication by RNA which would not require translation to take place. However, this would indicate transcription of a potential anti-transcript (repressor of replication) from the RP2 region (the region conferring incompatibility) that would bind to the transcript to inhibit replication from the RNA primer. When the RP7 sequence was aligned at <http://blast.ncbi.nlm.nih.gov/> with the RP2 region, no alignment was found. Also, no candidate promoter and transcription terminator sequences could be identified in RP2. In the present sequence analysis promoter 3 in the same strand as ORF 019 and promoter 10 in the complementary strand are both potential promoters. Transcription terminator sequences have also been identified in both strands. It is possible that promoter 3 may transcribe a replication initiator RNA as in the case of ColE1. The transcript from promoter 10 could potentially function as a regulator of initiation of replication by binding to the initiator. Northern blot for RNA profile or alternatively qPCR could be done to identify the presence of the candidate RNAs. The hypothesis that replication might be initiated by RNA however needs to be tested in future studies.

#### **4.7 Identification of pBSSB2 regions which exert plasmid incompatibility**

The phenomenon in which one plasmid is lost owing to the presence of a second plasmid with similar replicon or similar stability functions in the same host is termed plasmid incompatibility. It was expected that by identifying a fragment in the pBSSB2 library that exerts incompatibility to the parental plasmid, a region that carries the origin of replication or replication control elements or alternatively a region that carries information for stability function might be found. When two plasmids carrying similar replicons co-exist in a host

cell, they are regulated as a single pool instead of as separate plasmids. With time, one of the plasmids may be lost from the population as its copy number may drift downwards and not be corrected.

Vectorial incompatibility may occur when a cloned fragment from a test plasmid is present in an unrelated (and normally compatible) vector and is present at a high copy number in a cell also containing the test plasmid (Novick, 1987). Specific regions of the test plasmid when cloned cause destabilization of the test plasmid owing to incompatibility (Novick, 1987). Typically these regions will be those involved in the regulation of replication (e.g. producing a *trans*-acting repressor or a site which titrates a Rep protein). Cross-reaction between stability (e.g. *par*) functions can also cause incompatibility.

In order to identify incompatibility determinants, the pRP2-pRP10 library of pBSSB2 fragments cloned in pUC18 was used in incompatibility assays. Cells harbouring the linear plasmid were made electrocompetent following the procedure outlined in section 2.18.1. Constructs pRP2-pRP10 were individually introduced by electroporation. Transformed cells were spread on L-agar containing kanamycin ( $25\mu\text{g ml}^{-1}$ ) and carbenicillin ( $50\mu\text{g ml}^{-1}$ ) to select for both plasmids. As a negative control the vector backbone of pUC18 was also tested for incompatibility to pBSSB2.

A total of 50 transformants was picked for each construct and streaked onto L-agar containing carbenicillin only to select for the pUC18 derivative plasmids. Plates were incubated at  $37^{\circ}\text{C}$  overnight. Fifty colonies from each plate were streaked the next day on L-agar containing carbenicillin and L-agar with kanamycin. Plates were incubated at  $37^{\circ}\text{C}$  overnight. The percentage of  $\text{Km}^{\text{r}}$  colonies (i.e. those that contained the linear plasmid) compared to the number of carbenicillin resistant colonies was then scored. Since selection was maintained for only the circular constructs, and pBSSB2 is stable over the length of the experiment in the presence of pUC18 (Baker *et al.*, 2007b), the percentage loss of pBSSB2 provides a measure of incompatibility exerted by the insert in the circular plasmid. The assay was done in triplicate for each construct and control (Figure 4.30).

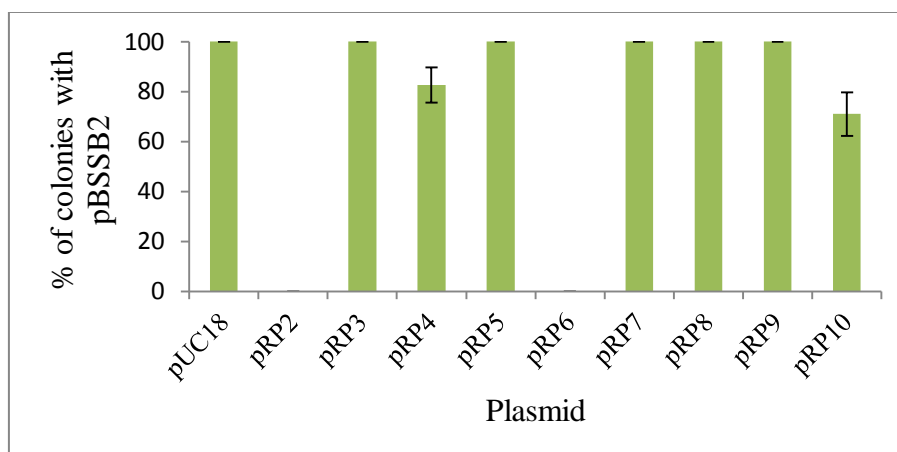


Figure 4.30. Incompatibility between pBSSB2 and constructs pRP2 to pRP10. pUC18 was used as a control to check its compatibility with pBSSB2. The results are the averages of at least three assays  $\pm$  S.D. The % for pRP2 and pRP6 is zero.

The result of the assay showed that pRP3-pRP5 and pRP7-pRP9 were compatible with the linear plasmid. With the exception of pRP4, 100% of the transformed cells retained the linear plasmid. With pRP4 and pRP10, 76% and 71.1% of the colonies retained the linear plasmid, respectively, whereas pRP2 and pRP6 showed strong incompatibility. Fragment RP2 includes most of RP6. In the linear plasmid map, RP2 (10068bp-14073bp) and RP7 (11350bp-15933bp) overlap (Figure 4.33). It was decided to use pRP2 as the starting point for identification of the incompatibility region in a deletion analysis.

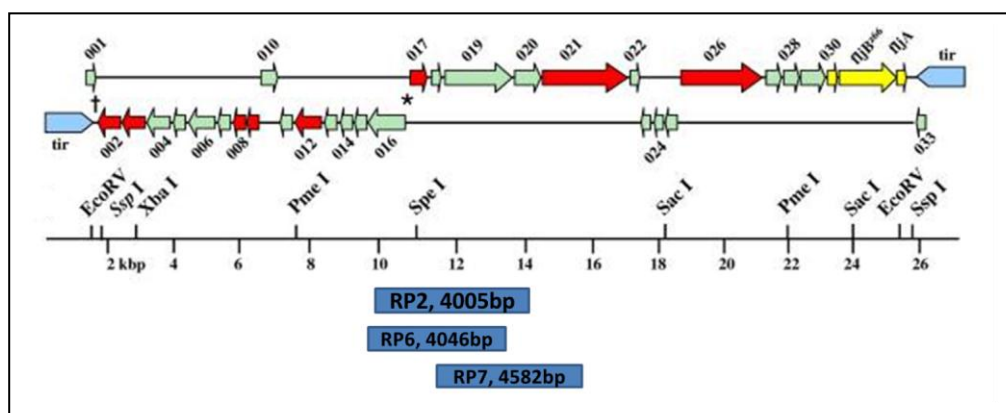


Figure 4.31. Map showing position of fragments RP2, RP6 and RP7. Fragment RP2 includes most of RP6 and overlaps with fragment RP7.

## **4.8 Generation of RP2 deletion derivatives to identify the minimum region conferring incompatibility**

In the initial analysis of RP2 deletion derivatives a rapid test for incompatibility was adopted. The circular derivative was transformed into cells containing the linear plasmid and incompatibility was scored based on the presence or absence of transformant colonies in plates (L-agar) containing antibiotic selecting for both the linear and the circular plasmids. The justification for the rapid assay was the fact that no transformant colonies had been found on plates containing carbenicillin and kanamycin when a pBSSB2-containing strain was transformed with pRP2 or pRP6.

Sub-sections of RP2 were amplified and cloned into pUC18 to identify the minimum region conferring incompatibility. PCR (MasterAmp PCR kit, Section 2.10.3) was used to generate a set of deletions extending from both ends of RP2 (Figure 4.34). Oligonucleotides used for PCR are detailed in section 2.8. PCR primers were tailed with EcoRI and BamHI recognition sites. Using these restriction sites, each PCR product was cloned in pUC18. Each construct was checked by agarose gel electrophoresis following digestion with EcoRI and BamHI (data not shown). Further confirmation was obtained by sequencing of the insert. Following validation of a construct, it was transformed into electrocompetent DH5 $\alpha$  cells containing pBSSB2. Transformant colonies were spread on L-agar containing both carbenicillin and kanamycin. pUC18 was used as a negative control and pRP2 as positive control for the incompatibility assay. Following transformation the absence of colonies on agar containing both carbenicillin and kanamycin was taken as evidence of incompatibility. The pattern of incompatibility exhibited by the deletion series is shown in Figure 4.32 in which incompatible fragments are indicated by blue lines and compatible fragments by black lines.



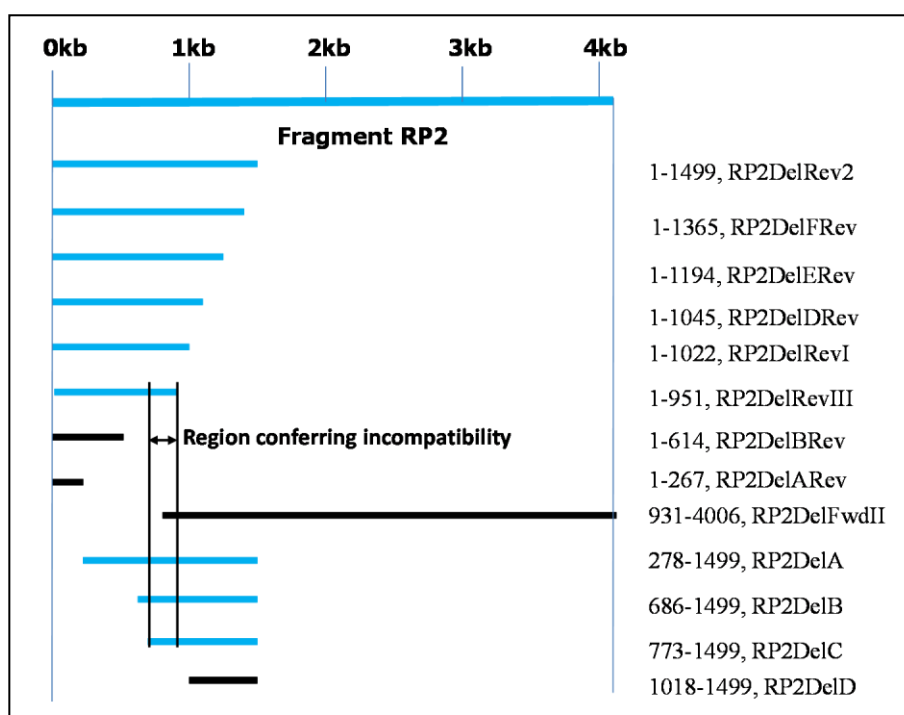


Figure 4.32. RP2 deletion series to identify minimum region expressing incompatibility against pBSSB2. The blue lines indicate fragments that exerted incompatibility, whereas the black lines represent fragments that were compatible with pBSSB2. The numbers indicate the nucleotide positions within RP2.

Subsequently a more thorough incompatibility analysis was performed with selected constructs. Each construct (pRP2, its deletion derivatives and pUC18 control) was transformed into pBSSB2-containing cells and transformant colonies were selected on L-agar with carbenicillin ( $50\mu\text{g ml}^{-1}$ ). A total of ten colonies were picked randomly from this plate and streaked onto L-agar in the absence of selection. The plate was incubated at  $37^\circ\text{C}$  overnight. Ten colonies were picked and transferred to L-agar without selection. This process was repeated once again. Finally, a total of 100 colonies were picked and replica-plated using sterile toothpicks onto L-agar containing carbenicillin and L-agar with kanamycin. The pattern of antibiotic resistance indicated which plasmids(s) had been retained (Figure 4.33).

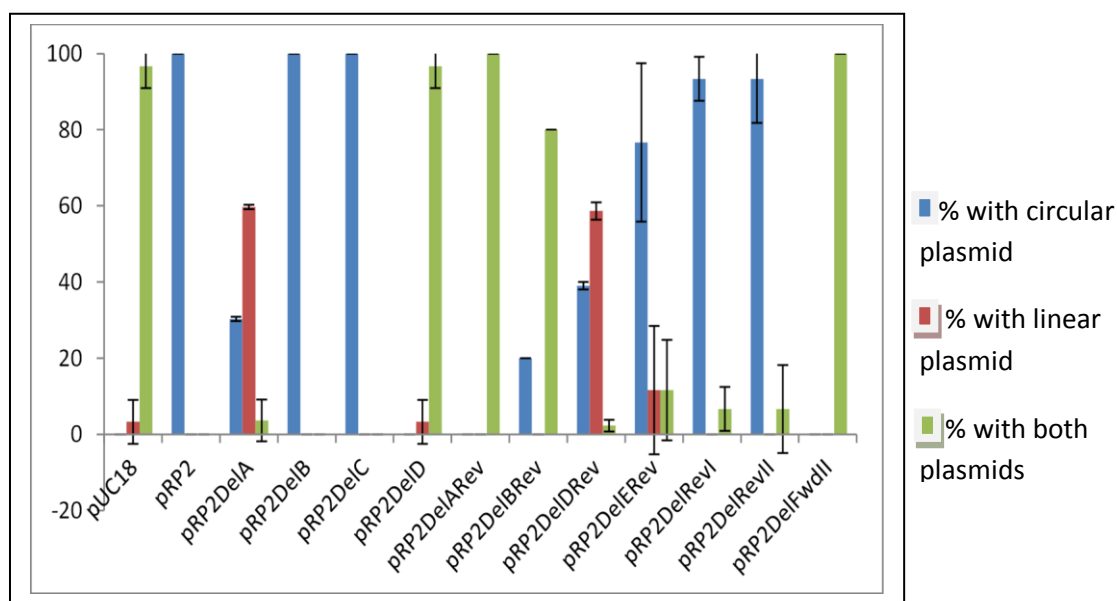


Figure 4.33. Plasmid profile of cells transformed with linear plasmid and pRP2 deletion derivatives after growth in the absence of selection. Blue bars indicate proportion of colonies containing only the circular plasmid; pink bars indicate proportion of cells containing pBSSB2 and green bars represent colonies with both plasmids. Results are averages of three assays  $\pm$  S.D.

It was observed that in most cases when incompatibility occurred, the circular pUC18-based plasmid was retained preferentially. This is expected since, in this experiment, incompatibility is directional, i.e. it is possible for the pUC18 derivative to repress replication of pBSSB2, but not *vice versa*. Exceptions were pRP2DelA and pRP2DelDRev where significant numbers of cells containing only the linear plasmid were recovered.

#### 4.9 Sequence analysis of the minimum region of incompatibility

Control of plasmid replication can be mediated by antisense RNA, protein and iterons (Summers, 1996). When the 179bp sequence responsible for RP2-mediated incompatibility was analysed using an ORF finder tool available at <http://www.ncbi.nlm.nih.gov/projects/gorf/>, no complete ORF was found. The only possible candidate as element for incompatibility therefore was an RNA or iterons. However, an insignificant level of transcription was observed from the region containing RP2 (Figure 4.31, RP2 is immediately to the left of ORF017). The 179bp sequence was analysed by eye. A six base pair imperfect repeat, (G/T) (G/A) TGTT was found within this sequence (Figure 4.34).

It is possible that these repeats might act as iterons that bind a hypothetical replication initiator protein, titrating the latter away from events leading to initiation of replication.

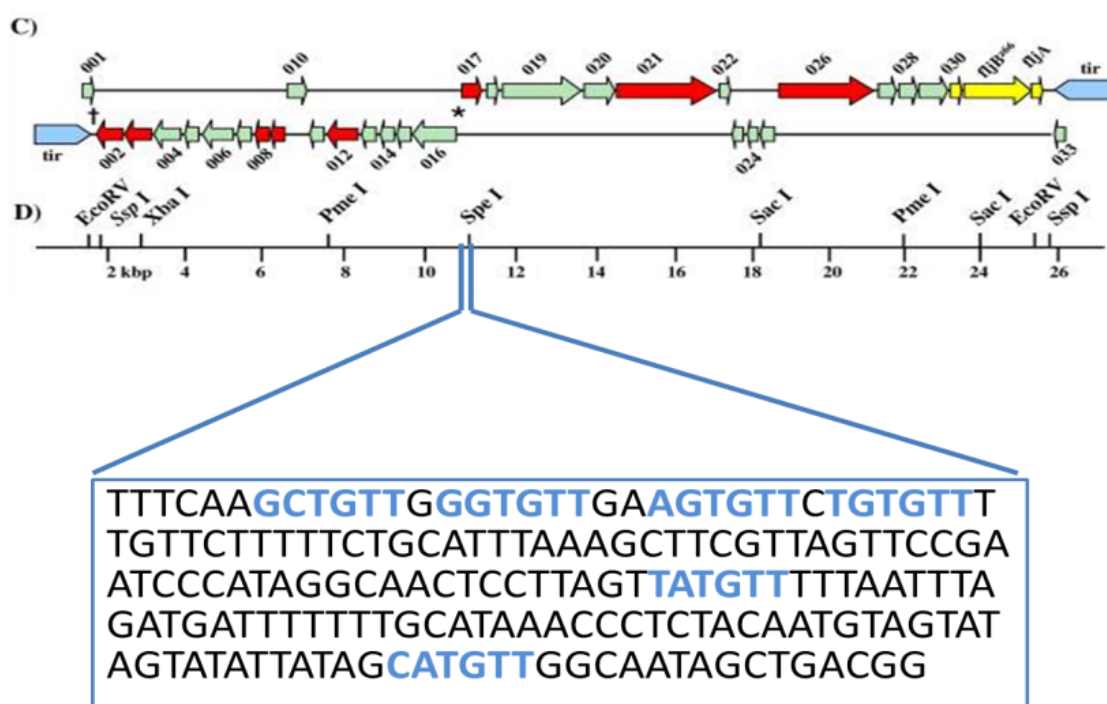


Figure 4.34. 179bp of RP2 deletion series sufficient to cause incompatibility. Location of this sequence in linear plasmid map is shown.

#### 4.10 Interaction of incompatibility and replication functions

Replication of pRP7 in a *polA* host indicated that fragment RP7 carries information for replication of the linear plasmid in a circular mode. But the very low copy number and the inability to isolate a colony containing fragment RP7 ligated to Km<sup>r</sup> cassette provided evidence that replication from fragment RP7 is inefficient. It was speculated that if the region of RP2 containing the incompatibility function of the linear plasmid was joined to RP7, replication might be better regulated.

Using PCR the region of the linear plasmid including both RP2 and RP7 was amplified using the left EcoRI tailed primer used originally to generate RP2 and the right BamHI tailed primer used originally to generate RP7 (Table 4, Section 2.8). This single fragment, designated RP27, was cloned in pUC18 using EcoRI and BamHI. The construct was validated by restriction digestion and sequencing. The plasmid (pRP27) was transformed into

the *E. coli* DNA polymerase I mutant, BR825. pRP7 and pUC18 were used as positive and negative controls, respectively. The optimized transformation method designed for BR825 was followed. Transformant colonies were selected at 37°C on L-agar containing carbenicillin. No colonies were obtained from the pRP27 transformation. The control plasmids behaved as expected with pRP7 resulting in a lawn of growth and no colonies with pUC18. This experiment was repeated with pRP27 purified from ten different, randomly-picked transformant colonies from the initial transformation plate and repeating the transformation using each plasmid on its own as well as combining the ten plasmids in a mixture and using an aliquot for transformation. This was done to exclude the possibility of a chance mutation in the pRP27 construct initially tested. However, the same failure to obtain transformants was observed in every case. Clearly the presence of the incompatibility region was inhibiting replication from the RP7 origin.

In order to investigate whether the inhibitory effect of the RP2 *inc* determinant also occurs *in trans*, the RP2 region was cloned separately in a low copy number vector, pFH450. The resulting construct was designated pFHRP2. pFH450 contains a high copy number, DNA polymerase I-dependent ColEI origin, a low copy number, DNA polymerase I-independent P1 origin and a chloramphenicol resistance cassette. Hence pFH450 is able to replicate in the *polA* mutant strain, BR825. In BR825 pFH450 replicates from the low copy number P1 origin (the ColEI origin being non-functional). The plasmids used and the results of transformation into BR825 in this experiment are shown in Table 4.4.

**Table 4.4 The influence of fragment RP2 on RP7 when present *in cis* or *in trans***

Name of plasmid(s) transformed into BR825	Purpose of the experiment	Presence of transformants in BR825	
		LA+Carbenicillin	LA+Chloramphenicol
pRP7	Positive control for pBSSB2 replication in circular mode	>300 <sup>1</sup>	ND <sup>2</sup>
pFH450	Positive control for pFH450 replication	ND	>300
pUC18	Negative control for pUC18 replication in BR825	0	ND
pFHRP2 and pRP7	Effect of RP2 on RP7 when present <i>in trans</i>	0	>300 <sup>1</sup>
pRP27	Investigate effect on replication from RP7 by RP2 <i>in cis</i>	0	ND

<sup>1</sup> >300 refers to the number of colonies in a plate spread with 100µl of transformants. <sup>2</sup>Not done

It was seen that pRP7 and pFHRP2 replicated in BR825, whereas pUC18 did not. When pRP7 and pFHRP2 were transformed sequentially in either order, pRP7 did not replicate when pFHRP2 was present in the same host (BR825). pRP27 failed to replicate. Thus, absence of transformants was observed when RP2 and RP7 were present together in the same cell, either in the same plasmid or in different plasmids. From this observation it was concluded that the inhibitory effect of RP2 on replication from RP7 occurred both *in cis* and *in trans*.

## 4.11 Discussion

Since its identification and isolation by Baker *et al.* (2007b), the linear plasmid pBSSB2 has previously only been characterized by sequence analysis. The GC skew (Picardeau *et al.*, 2000) (G-C/G+C) lies upstream of ORF017 and this region has been proposed to contain the origin of replication (Picardeau *et al.*, 2000). No direct repeats or iterons were found in this region (Baker *et al.*, 2007b).

One of the major objectives of the present study was a functional analysis of pBSSB2 replication. To map the basic replicon it was decided to work with circular derivatives of pUC18 containing inserts from the linear plasmid. Working with circular plasmids was expected to be technically easier. For example transformation efficiency would be higher for circular than for linear plasmids. A search of the pertinent literature revealed that there are both advantages and disadvantages of using circular derivatives of a linear plasmid or chromosome. For example, it has been reported that linear plasmids pSLA2, pSLA2-L, SCP1 and pSCL from the well characterized *Streptomyces* group can replicate as a circle from a central origin when their telomeres are deleted (Chang and Cohen, 1994; Hiratsu *et al.*, 2000; Redenbach *et al.*, 1999; Shiffman and Cohen, 1992). In the *Streptomyces* linear plasmid pSCL mutations that inhibited replication in a linear mode also rendered replication of the circular mode non-functional (Shiffman and Cohen, 1992). Such an observation indicated that these two modes require the same gene products for replication. Linear plasmids from *Rhodococcus* and *Mycobacterium* also replicate from centrally located loci and have similar structure to pBSSB2 (Warren *et al.*, 2004; Picardeau *et al.*, 2000). However, inefficient replication of circular derivatives of plasmids that are linear in their native state has been reported previously (Zakrzewska-Czerwinska and Schrempf, 1992; Chang *et al.*, 1996; Qin and Cohen, 2003; Beaurepaire and Chaconas, 2005; Xu *et al.*, 2006). Linear plasmid lp17 of *Borrelia burgdorferi* has also been found to replicate more efficiently in a linear mode than in a circular mode. The circular derivative exhibited reduced stability with no detectable plasmid following 50 generations (Beaurepaire and Chaconas, 2005). The reason for the effect of topology on replication is not always clear but differences in supercoiling between circular and linear replicons can potentially influence melting of the replication origin, formation of essential secondary structure or transcription (Alverson *et al.*, 2003).

Since most previous investigation of linear plasmid replicons have been based on experiments with circular derivatives, it was decided to adopt this approach in order to look for pBSSB2 replicon.

Upon working with circular derivatives, only one region of pBSSB2, RP7, was able to rescue replication of pUC18 in a *polA* mutant. Formally this could be due to:

1. Presence of sufficient pBSSB2 information in pRP7 to enable replication in a circular mode.
2. Integration of the construct into the chromosome.
3. Activation of the pUC18 origin by information contained in RP7 that enabled replication in BR825.

Purification of pRP7 from XL1-Blue transformed with plasmid preparation purified from BR825 made it unlikely that pRP7 had integrated into the chromosome. It also seems unlikely (although not impossible) that only one of the pRP2-pRP10 plasmid series would have been capable of integration.

Some of the BR825 transformants were found to contain a pRP7 derivative that had undergone rearrangement and contained 815bp of linear plasmid sequence. This 815bp cloned in pUC18 could not drive replication of the resultant construct in BR825. This suggested that presence of the insertion sequence exerted some influence on replication of the construct. It is possible that the insertion sequence enabled integration of the plasmid into the host chromosome, where it replicated with the chromosome but this was not investigated further.

From the transformation of pRP7 into BR825 it could not be concluded whether RP7 contained functions required for pBSSB2 replication in a circular mode or whether it somehow activated the pUC18 origin. An attempt was made to resolve the question by ligating RP7 to a kanamycin cassette ( $Km^r$ ). If the resultant construct was replicated from RP7 this construct should be viable. Unfortunately, exhaustive attempts to obtain an autonomously replicating construct proved futile. The possibility that information contained in RP7 activated the pUC18 origin cannot therefore be excluded.

Deletion analysis of RP7 identified a 2830bp fragment which supported replication of a pUC18 derivative in a *polA* mutant. A 144bp deletion from the left end rendered RP7 non-replicating, indicating that this region is essential for RP7 function. The smallest replicating

fragment carries ORF018 and ORF 019 of pBSSB2. Mutagenesis studies showed that neither of these ORFs is essential for replication. However, analysis of transcription data generated at the Wellcome Trust Sanger Institute (Hinxton, UK) indicated significant transcription of the region containing ORFs 017, 018 and 019 (Figure 4.14). Upon sequence analysis, three ORFs in the same strand as that containing ORF 019 (ORFs A, C and E) and two in the complementary strand (ORFs F and G) are potential candidates for rep-protein encoding ORFs in future investigations. An alternative hypothesis of RNA primed replication has also been proposed. This hypothesis also needs to be tested in future studies.

Incompatibility assays and deletion walking on fragment RP2 identified a 179bp region exerting incompatibility, just 331bp from the boundary of RP7. Fragment RP2 contained the GC skew (Baker *et al.*, 2007b), which had previously been cited as evidence that this region contained the origin of replication. On the other hand, the incompatibility phenotype would also be consistent with being (or encoding) a repressor of replication. This can be by the production of a repressor (antisense RNA or protein) or by DNA direct repeats or iterons titrating a Rep protein. The absence of a complete ORF sequence made the production of a repressor protein unlikely. The four 6bp imperfect repeats in RP2 may be acting as incompatibility determinants possibly by titrating a Rep protein. In any case these data show that additional sequences outside of RP7 play a role in regulation of replication in the native linear plasmid. Requirement for sequences in addition to the Rep gene and iterons has been demonstrated for other linear plasmids viz. pSLA2 (Chang *et al.*, 1996). In case of SLP2 efficient replication as a circular derivative requires additional 47bp sequences upstream of 23-mer iterons. Inclusion of these sequences increased the transformation efficiency by about 1000 fold (Xu *et al.*, 2006). The ability of RP2 to inhibit replication from RP7 also provided evidence against the hypothesis that RP7 simply activates the pUC18 origin. The presence of iterons, if confirmed, would favour a Rep protein as the initiator of replication rather than pre-primer RNA.

Based on comparison with other linear plasmids with similar structural features to pBSSB2, the encoding of an initiator protein in region RP7 is plausible. The origin of replication seems likely to lie in fragment RP2. Control of replication is possibly regulated by iterons in RP2. In the event that the origin of replication is contained in fragment RP2, it is possible that a secondary origin-like sequence drives circular plasmid replication from RP7. The Rep-iteron organisation would be similar to that found for pSLA2 (Chang *et al.*, 1996), SCP1



(Redenbach *et al.*, 1999), pSCL1 (Huang *et al.*, 2003), SLP2 (Wu and Roy, 1993), pCLP (Picardeau *et al.*, 2000) and pRHL3 (De Mot *et al.*, 1997).

Based on the findings related to pBSSB2 replication, the following hypothesis can be put forward:

1. Fragment RP7 encodes an initiator of replication and fragment RP2 contains the origin of replication as well as a replication repressor.
2. In the linear plasmid the RP27 region, i.e. combined RP7 and RP2, constitutes the basic replicon.
3. It is possible that the circularity of the pRP27 derivative represses replication. Hence the inviability of pRP27.
4. When the repressor and origin containing region (RP2) is omitted, replication of a circular plasmid is activated through an origin-like sequence in RP7.
5. Owing to the absence of the true origin in RP7 and because of its topology, replication driven by this fragment is inefficient, hence the low copy number of pRP7.

In the present study, the putative Rep protein is yet to be identified. Binding assays of any putative Rep protein with incompatibility determinants within RP2 could provide further supporting evidence of the role of iterons in replication control. Future studies also need to be carried out in order to identify elements associated with plasmid replication at the telomeres.

# Chapter 5

---

## Stability function of pBSSB2

---

## 5.1 Introduction

In order to ensure that plasmids are passed efficiently to daughter cells, stability function(s) must be present in the plasmid, particularly when the copy number of the plasmid is low and random distribution would lead to plasmid-free cells at high frequency.

Three mechanisms commonly used by low-copy number plasmids to ensure segregational stability are:

1. Partitioning systems that distribute plasmids actively into daughter cells e.g. the partitioning system of F plasmid (Ogura and Hiraga, 1983).
2. Toxin - antitoxin systems which code for a stable toxin and an unstable antitoxin. Plasmid-free cells are killed by the toxin. Plasmid - containing cells are viable as only these contain the antitoxin. An example of this is the *hok-sok* system in plasmid R1 (Gerdes *et al.*, 1990).
3. Plasmid multimer resolution systems which convert multimers to monomers by site-specific recombination. This increases number of available plasmid entities to be partitioned during cell division. An example is the *Cre-lox* system of P1 (Sternberg and Hamilton, 1981).

Because of their structure, multimer formation due to recombination between linear plasmids is not a threat. The presence of a partitioning system, a post-segregational killing system, or both, is more likely for a linear plasmid. One of the major objectives of the present investigation was to identify and characterize regions involved with stable inheritance of pBSSB2.

Limited studies have been conducted on stability functions of linear plasmids. Of these the stability functions of the linear plasmid pCLP of *Mycobacterium celatum* have been investigated in circular derivatives (Dantec *et al.*, 2001). Similarly circular plasmids have been used to identify the region(s) carrying pBSSB2 stability function(s). This makes the assumption that partition and replication functions are not interlinked and identification of stability determinants is possible in a circular vector not carrying the

linear plasmid replicon. The independence of replication and stability functions is supported by a number of studies in circular plasmids. The *parS* *cis*-acting site of P1 has been shown to stabilize heterologous replicons (Austin and Abeles, 1983; Austin *et al.*, 1986). P1 can stabilize mini-F plasmids that do not contain any partition region (Austin and Abeles, 1983). Plasmid F *par/sop* can stabilize *oriC* plasmids that lack stability functions (Ogura and Hiraga, 1983). The *hok/sok* system of R1 has also been shown to be independent of the replicon (Boe *et al.*, 1987; Gerdes, 1988; Gerdes *et al.*, 1985). This system stabilizes F and R1 plasmids about 50-100 fold and high copy number plasmids such as p15A and pBR322 1000-10,000 fold. Based on these observations, it was assumed that linear plasmid stability functions should also function independently of replication mechanisms. In other words, stability functions of pBSSB2 would be expected to stabilize an unstable heterologous circular replicon.

## 5.2 Identification of regions containing stability functions of pBSSB2

Mutations in *pcnB* (plasmid copy number control gene) decrease the copy number of ColE1 (Lopilato *et al.*, 1986). *pcnB* codes for a poly(A) polymerase (Cao and Sarkar, 1992), which polyadenylates RNAI. The RNAI antisense suppressor of replication initiation is initially degraded by ribonuclease E (Tomcsanyi and Apirion, 1985; Lin-Chao and Cohen, 1991) and the polyadenylated cleaved product undergoes rapid degradation (Xu *et al.*, 1993). In the absence of polyadenylation, RNAI undergoes slow decay and the copy number of ColE1-like plasmids is reduced. Copy number reduction in a *pcnB* mutant means that pUC18 is lost more rapidly. To maintain itself in the host population an additional stability system becomes necessary. The linear plasmid library pRP2-pRP10 was used to identify regions that carried stability functions. This screen envisages that the presence of a pBSSB2 stability function in any of these pUC18 derivatives will improve their stability in a *pcnB* host.

Each construct was transformed into the *pcnB* mutant and stability was tested using the assay described by Martin *et al.* (1987). In this method, 8 transformant colonies were picked randomly and streaked onto L-agar containing carbenicillin. The plate was incubated at 37°C overnight. In the next step 8 colonies were randomly picked and plated onto L-agar without carbenicillin. This procedure was repeated once more. A total of 64 colonies were picked from the final L-agar plate after incubation overnight

at 37°C and patched onto L-agar with and without selection. The number of colonies that grew in the presence and absence of selection was used to calculate the percentage of cells that retained the plasmid. In this assay, pUC18 was used as a negative control.

In the case of pUC18, 59% of cells retained the plasmid (Figure 5.1). All constructs carrying regions of pBSSB2, with the exception of pRP2 and pRP4, had a smaller percentage of cells retaining the plasmid when compared to pUC18. It is notable that the presence of the inserts often reduces stability further. This may be explained by the added metabolic load imposed by the larger plasmid. Construct pRP2 showed variable stability (74%) with a high standard deviation. pRP4 (93.75%) was consistently highly stable. It was clear that fragment RP4 contributed to the stabilization of pUC18 in the *pcnB* host and may therefore play a role in the stability of pBSSB2. It was decided to work primarily with fragment RP4. pRP2 was also included in further stability assays for comparison.

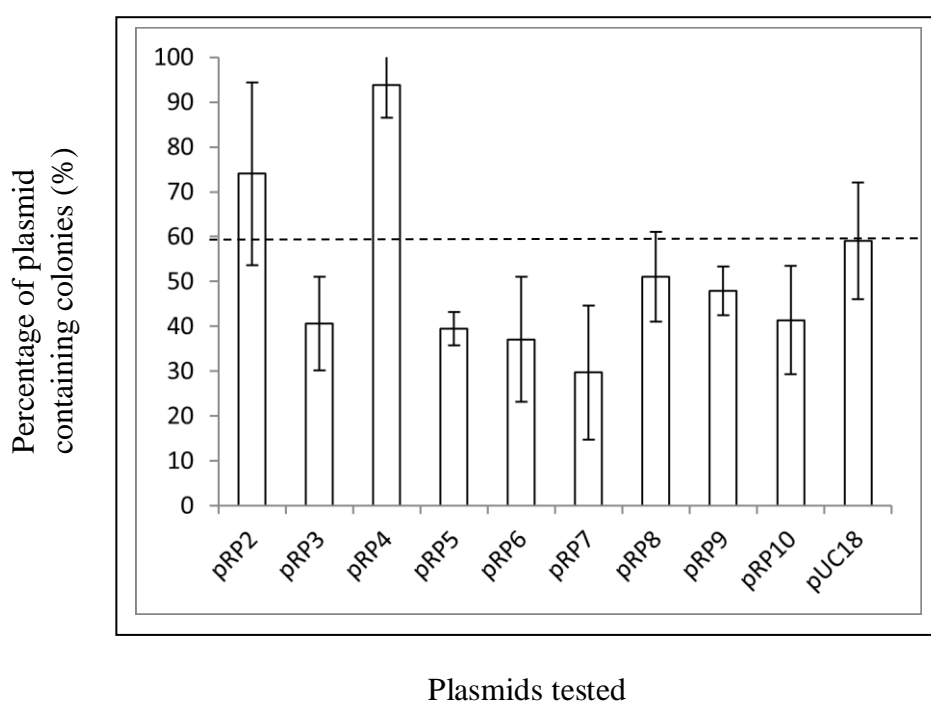


Figure 5.1 Stability of constructs pRP2 to pRP10 in a *pcnB* mutant. Vertical bars show average percentage of cells retaining a construct  $\pm$  standard deviation (S.D.) of at least three independent experiments. The dotted line shows the cut-off value and corresponds to the percentage of cells retaining pUC18.

### 5.3 Further analysis of stability functions in regions RP2 and RP4

In a further analysis of pBSSB2 stability functions an unstable bi-replicon plasmid, pFH450 (Hayes, 1998), was used. This plasmid carries the medium copy number ColE1 replicon, and a low copy number P1 replicon, but no stability function. It was used as a cloning vector to clone the fragments RP2 and RP4 that had stabilized pUC18 in the *pcnB* mutant. In wild-type *E. coli*, this vector replicates from the ColE1 replicon and random distribution is sufficient to ensure stable maintenance. However, in a DNA polymerase I mutant (BR825), the ColE1 replicon is non-functional and replication proceeds from the low copy number P1 replicon which leads to segregational instability. However, presence of a pBSSB2 stability function would be expected to stabilize the vector. This also helped to overcome any concern that RP2 or RP4 raised the copy number and hence stability of pUC18 in the *pcnB* mutant rather than independently assisting stable inheritance.

RP2 and RP4 were individually cloned in pFH450. The same PCR primers that were used to generate RP2 and RP4 amplicons for cloning into pUC18 were used for amplification and insertion of these fragments into pFH450, except that instead of EcoRI tailed primers, SacI tailed primers were used. BamHI and SacI restriction sites were used for cloning. Ligated DNA was transformed into commercially available chemically competent XL1 - Blue cells according to the manufacturer's instructions (Agilent technologies). Transformed cells were spread on L-agar containing chloramphenicol ( $10\mu\text{g ml}^{-1}$ ) and incubated at  $37^{\circ}\text{C}$  overnight. About five transformants were randomly selected and grown in L-broth containing chloramphenicol ( $10\mu\text{g ml}^{-1}$ ) at  $37^{\circ}\text{C}$  overnight. Extracted plasmid was analysed by gel electrophoresis following digestion with BamHI and SacI. Confirmation of the insert was further obtained by DNA sequencing.

Having obtained pFH450 clones containing RP2 (designated pFHRP2) or RP4 (designated pFHRP4), their stability was assayed in a DNA polymerase I mutant (BR825). The optimized protocol for transformation into BR825 (section 2.18.2) was employed to introduce pFHRP2 or pFHRP4. pREG531 (that contained the partition cassette of plasmid TP228 inserted in pFH450) was used as a positive control. The stability assay was performed as described in section 5.1, with the exception that

chloramphenicol was used instead of carbenicillin and pFH450 was used instead of pUC18 as a negative control.

Figure 5.2 shows the percentage of cells that retained the plasmid after growth for about 25 generations without antibiotic selection. In case of the positive control, pREG531, 66% of the colonies retained the plasmid, as opposed to 8% in case of the negative control, pFH450. Fragment RP4 in plasmid pFHRP4 increased the stability of pFH450 such that 83% of the colonies retained the plasmid. In contrast, no significant stability was conferred by pRP2 which was retained by only 3% of the colonies. In fact, the presence of fragment RP2 reduced the stability of the negative control further, possibly through increased metabolic load on the host. Since fragment RP2 exhibited variable stability in the *pcnB* mutant system and did not stabilize pFH450, it was decided to focus primarily on fragment RP4 which increased stability in both assays.

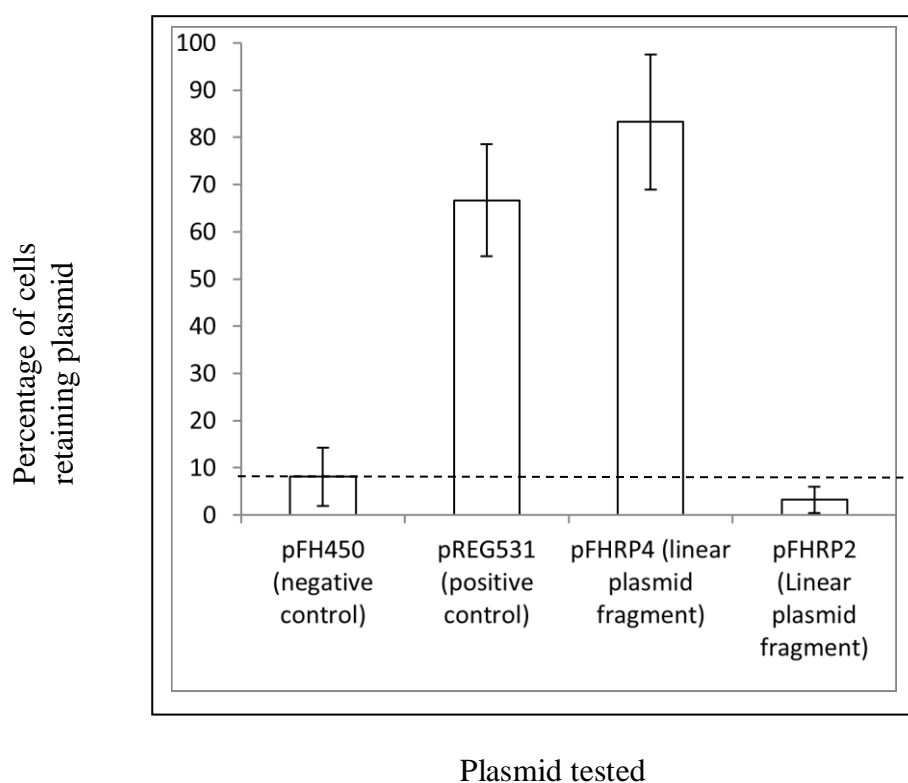


Figure 5.2 Stability of plasmid pFH450 and derivatives containing fragments RP4 and RP2. The result is the mean  $\pm$  S.D. of at least three independent assays. The dotted line is the cut-off value for the experiment and corresponds to the percentage of cells retaining the negative control, pFH450.

#### 5.4 Deletion analysis of fragment RP4 to identify minimum region necessary for stability

Fragment RP4 is approximately 4.5 kb. An initial deletion analysis to identify the minimum region essential for stability was made by restriction digestion. The restriction map for RP4 shows the presence of a HindIII site within the sequence. Another HindIII site is present within the multiple cloning site of pUC18. A HindIII digestion of pRP4 would therefore remove about 2.3kb of RP4 sequence as shown in Figure 5.3. This would leave approximately 2.2kb of RP4 sequence attached to the vector backbone.

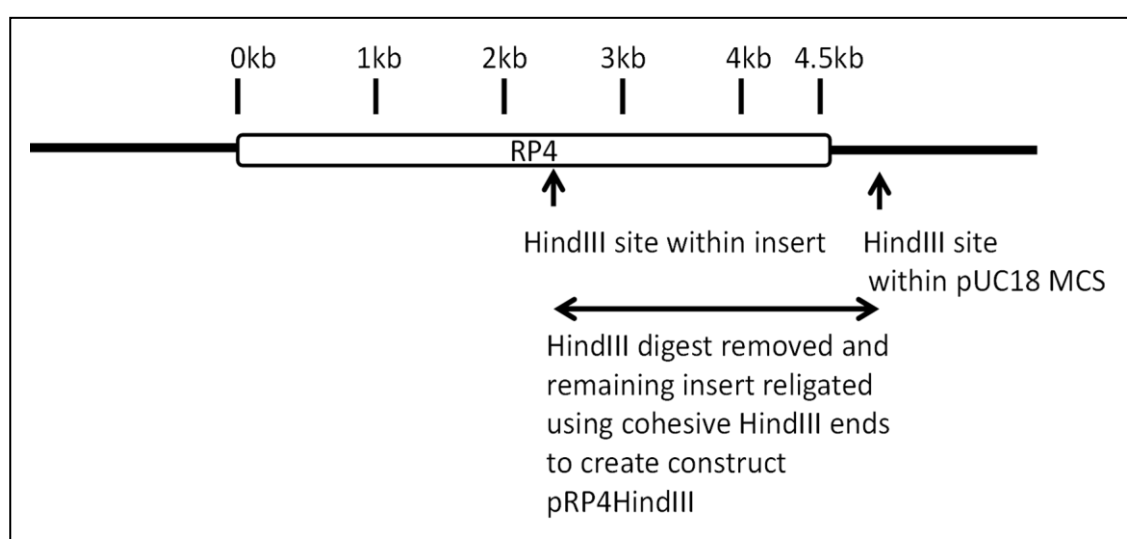


Figure 5.3 Schematic diagram of pRP4. HindIII sites within RP4 (indicated by box) and within the MCS of pUC18 (black lines) are indicated. For simplicity pUC18 sequence has been shown by straight lines.

pRP4 was digested with HindIII and the digestion mix was run on an agarose (1%) gel. The DNA band corresponding to pUC18 plus the remaining 2.2kb RP4 sequence was purified from the gel using the Qiagen gel purification kit (Qiagen). The purified DNA was self-ligated, drop dialyzed and transformed into chemically competent XL1 Blue cells following the manufacturer's recommendations (Agilent technologies). The resultant pRP4 derivative was named pRP4HindIII. Plasmid DNA was prepared from white XL1 - Blue colonies using the Qiagen mini prep plasmid purification kit (Qiagen). Purified plasmid was analysed on an agarose (1%) gel for size confirmation.



Plasmid stability assays in a *pcnB* host were carried out as detailed in section 5.2 comparing the stability of pRP4HindIII with pRP4 as a positive control and pUC18 as negative control. It was found that fragment RP4HindIII still stabilized a circular plasmid (Figure 5.4). The position of fragment RP4HindIII (2.2kb) in the linear plasmid map is shown in figure 5.5.

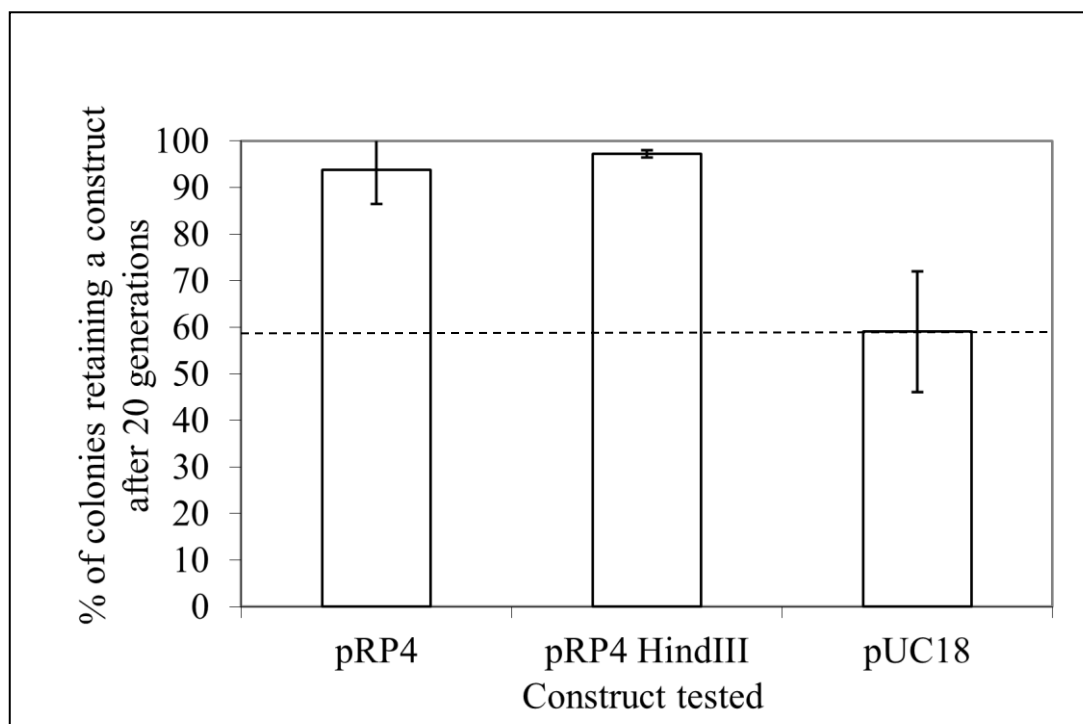


Figure 5.4 Stability of pRP4HindIII as compared to pRP4 and pUC18. The results are the means  $\pm$  S.D. of at least three independent assays. The dotted line represents the percentage of cells retaining the negative control, pUC18.

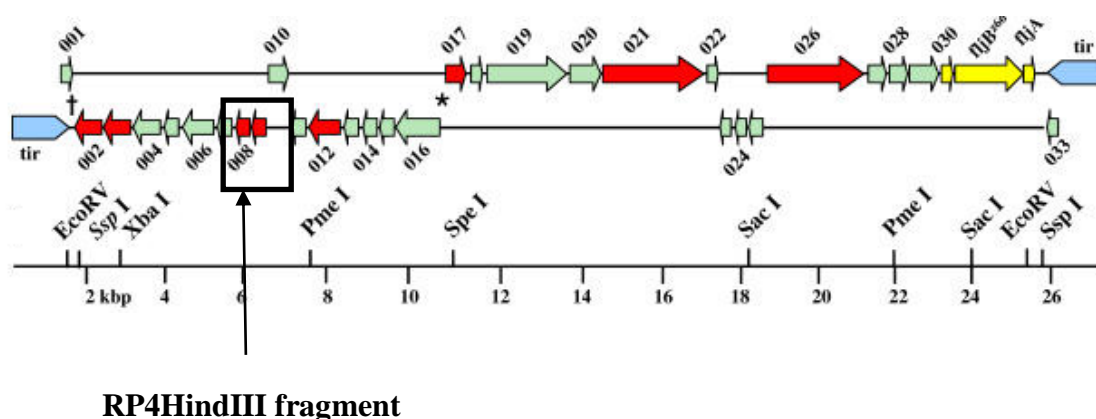


Figure 5.5. The location of the RP4HindIII region within pBSSB1 DNA sequence. The black box represents RP4HindIII. The pBSSB2 genetic map is taken from Baker *et. al.*, 2007b. The arrows indicate pBSSB2 ORFs and point in the direction in which they are transcribed. The star shows the position of the GC skew. A size scale is also shown in the diagram including selected restriction sites.

Since RP4HindIII was sufficient to confer stability to pUC18 in *pcnB* mutant, it was used as the basis for further deletion studies in order to identify the minimum fragment containing the stability function.

A series of deletions of RP4HindIII (Figure 5.6) was generated by PCR, using primers listed in Table 2.4. The designation and nucleotides covered by each PCR product is indicated in figure 5.6. For PCR reactions the Expand Long Template PCR (Roche) system detailed in section 2.10.4 was used. Each PCR product was cloned in pUC18 using EcoRI and BamHI restriction sites. Ligated DNA was subjected to drop dialysis before transformation into chemically competent *E. coli* XLI-Blue cells. Transformant colonies were grown in L-agar containing carbenicillin ( $50 \mu\text{g ml}^{-1}$ ) and also containing X-gal ( $300 \mu\text{g ml}^{-1}$ ) and IPTG ( $0.3 \text{mM}$ ) for blue-white screening. Five white colonies were screened for the presence of the desired plasmids. Constructs were validated by both restriction digestion and sequencing.

All constructs were tested in triplicate in stability assays in the *pcnB* mutant (Figure 5.7). The deletion analysis showed that all deletions from the right end of RP4HindIII (constructs Del1, 2 and 3, figure 5.6) reduced stabilization of the circular plasmid. Deletions of up to 1173bp from the left end (construct DelB, figure 5.6) still enabled RP4HindIII to stabilize a circular derivative. This indicated that the right end of the

fragment was necessary for stability functions in the circular mode. A 1023 bp fragment (DelB, figure 5.6) was the smallest that contained pBSSB2 stability function. A further 73bp deletion from the 5' end of this minimum fragment removed the stabilization effect. On the other hand, a further 40bp deletion from the 3' end also reduced stability function.

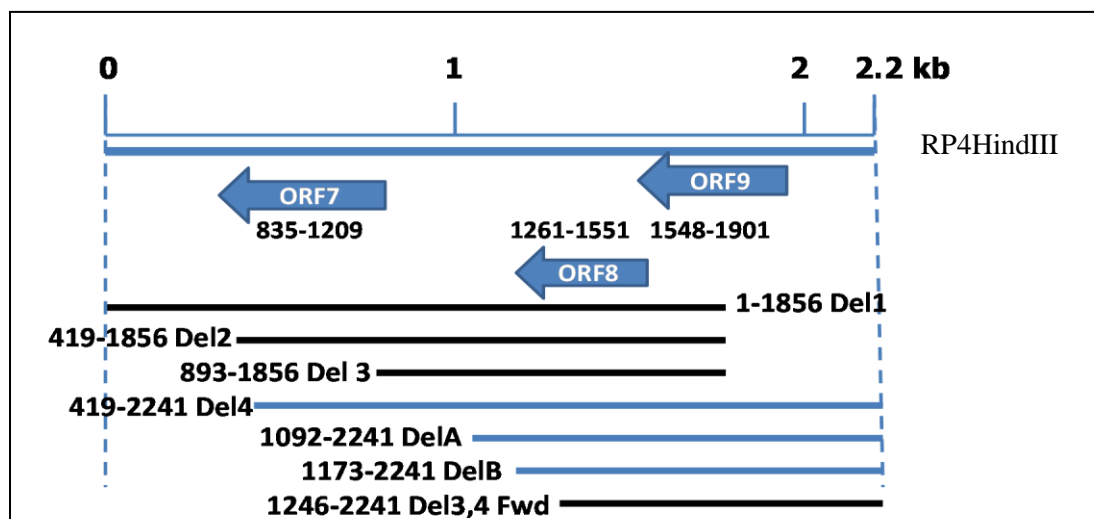


Figure 5.6 Deletion derivatives of the RP4HindIII 2.2kb fragment. Numbers indicate nucleotide co-ordinates, taking the start nucleotide of the left end of RP4HindIII to have a co-ordinate of 1. Blue lines indicate fragments that were stable and black lines indicate those that were unstable in a *pcnB* mutant.

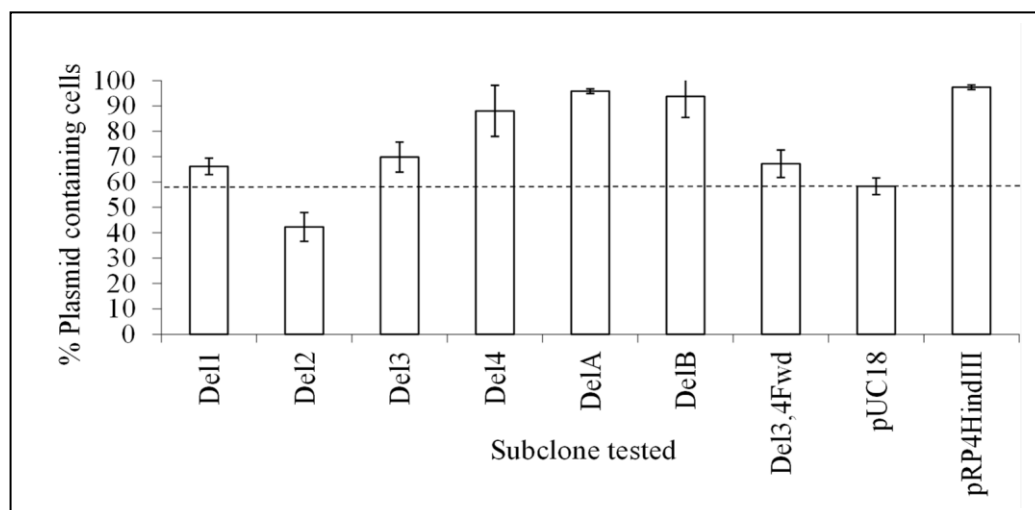


Figure 5.7. Stability of RP4HindIII deletion series. Results are the average  $\pm$  S.D. of at least three independent assays. Del1-4, DelA, DelB and Del3,4Fwd are deletion derivatives. pUC18 and pRP4HindIII were negative and positive controls, respectively. The dotted line indicates the cut-off value for instability and corresponds to the percentage for pUC18.

## 5.5 Mutational analysis of the pBSSB2 stability region

Sequence analysis of RP4HindIII indicated the presence of ORF07, ORF08 and ORF09 (Figure 5.6). In order to investigate the potential role of these ORFs in stability, stop codons were introduced into each by site directed mutagenesis. Oligonucleotides used to introduce the desired base changes are listed in table 2.4. Site-directed mutagenesis was carried out according to the overlap extension method (Ho *et al.* 1989) outlined in section 2.20.1. For PCR, the Expand Long Template PCR system described in section 2.10.4 was used. Amplicons with the desired base change were inserted into pUC18 using EcoRI and BamHI restriction sites. Ligated DNA was transformed into chemically competent *E. coli* XL1-blue cells. Blue-white screening was used to identify transformants containing inserts. Plasmid DNA was extracted from selected white colonies using the QIAprep miniprep plasmid kit (Qiagen). Extracted plasmids were validated for the correct insert by sequencing.

### 5.5.1 Engineering a stop codon in ORFs 07

The original sequence and mutation introduced in ORF07 are shown in figure 5.8.



Nucleic acid sequence: 1 ATG ATA AAG AGA TAC AGT GAA AGA CTT GCA GAA CTG GAA CTA CTA  
Amino acid sequence: 1 M I K R Y S E R L A E L E L L  
Nucleic acid sequence: 46 CAA AAC AGG CTG TTG CTG TCG TTC CGA ACG GAT GAC GAC GAT TAT  
Amino acid sequence: 16 Q N R L L L S F R T D D D D Y  
Nucleic acid sequence: 91 ATT TTA TCT GTA TGC CGT CAA ATC AGC GAT ACA GGG CTA AAT CTC  
Amino acid sequence: 31 I L S V C R Q I S D T G L N L  
Nucleic acid sequence: 136 AAA TAC AGT AAT ACA GAT TAT ATT TTC CAT TAT ATA AAC TGT TGC  
Amino acid sequence: 46 K Y S N T D Y I F H Y I N C C  
Nucleic acid sequence: 181 TCT TAC CAC AGA GAA CCC AGT TTT ATT GTG ATT GGT TTA CTG TTG  
Amino acid sequence: 61 S Y H R E P S F I V I G L L L  
  
Nucleic acid sequence: 226 TCA TTA CAG GCT AAA AAA ACG GTC ATG GCA TAC AGA TTA TTT AAG  
Amino acid sequence: 76 S L Q A K K T V M A Y\* R L F K  
  
Nucleic acid sequence: 271 AAG CTG TAT ATA GAT AAA AAA GAT AGC CAC AGC TTA ACT GAC AAC  
Amino acid sequence: 91 K L Y I D K K D S H S L T D N  
Nucleic acid sequence: 116 ATT CAA AGA ACG GCT GGT AGT TTA CTG ACT GTT ATG AAC AGA AAG  
Amino acid sequence: 106 I Q R T A G S L L T V M N R K  
Nucleic acid sequence: 161 GAT AGT GCA GCT TGA  
Amino acid sequence: 121 D S A A Stop

Figure 5.8 DNA and corresponding amino acid sequence of ORF07. The altered DNA sequence and amino acid are indicated by arrows. Y\* corresponds to the amino acid tyrosine.

## 5.5.2 Introduction of nonsense mutation in ORF08

The DNA and amino acid sequences corresponding to ORF08 and the mutation introduced are shown in figure 5.9.

DNA sequence: 1 ATG AGA AAA GAA CTC TTT GAC GAC CTG ATC GAG AGT GTT AAC GAA  
Amino acid sequence: 1 M R K E L F D D L I E S V N E  
DNA sequence: 46 GCC GTA GCC ATC AAA CGG GGT CTT GTA GCC CCA GCT CGT GTT ACG  
Amino acid sequence: 16 A V A I K R G L V A P A R V T  
DNA sequence: 91 ACT TTT GAG ATG CCC GAT GTG AAA GCT ATC CGC ACG AAA GCG AAG  
Amino acid sequence: 31 T F E M P D V K A I R T K A K  

Changed to **TAG**  
  
DNA sequence: 136 CTC AAA CAA AAT GAG TTT GCC CAG GTA GTG GGT GTC TCC CCT TCT  
Amino acid sequence: 46 L K Q N E F A Q\* V V G V S P S  
  

Changed to **Stop**

DNA sequence: 181 CTG GTG CAA GCC TGG GAA CAG CAA AAG AGA ACA CCC ACC GGC AGC  
Amino acid sequence: 72 L V Q A W E Q Q K R T P T G S  
DNA sequence: 246 AGT CTT AAA CTG TTG CGG GTC ATT GAA AAG CAT CCT GAC ATG CTT  
Amino acid sequence: 97 S L K L L R V I E K H P D M L  
DNA sequence: 291 AAC CTG TTC AGA ACA GCA TAA  
Amino acid sequence: 112 N L F R T A **Stop**

Figure 5.9. DNA and amino acid sequences of ORF08. The position in which mutation was introduced is indicated by asterisk. Q\* represents the amino acid glutamine, which was changed to a stop code by the introduced mutation.

### 5.5.3 Introduction of stop code in ORF09

The wild-type sequence of ORF09 and position mutated is shown in figure 5.10.

DNA sequence: 1 ATG AAT TTA CTA CAC TTC ATT GAA ACC AAA GTC TTT CAG AGA GTA ATC  
Amino acid sequence: 1 M N L L H F I E T K V F Q R V I  
DNA sequence: 49 GAT GGA TTG CTT TCA CCC GAC GAA CTC CGG GAG TTT CAG GAA GTG TTA  
Amino acid sequence: 17 D G L L S P D E L R E F Q E V L  
DNA sequence: 97 AGG CAA GAT CCT ACC GCT GGC GAT ACC ATA TCA GGT ACG GGC GGT TGC  
Amino acid sequence: 39 R Q D P T A G D T I S G T G G C  
DNA sequence: 145 CGT AAG ATT CGC TGG GCT TTG CCT GGT ATG GGT AAA AGT GGT GGT ATC  
Amino acid sequence: 49 R K I R W A L P G M G K S G G I  
DNA sequence: 193 AGG GTC ATA TAC TAC TAT CTG ACA GCA GAT AAC GAA GTA TTC CTT CTT  
Amino acid sequence: 77 R V I Y Y Y L T A D N E V F L L  
DNA sequence: 241 ATA GCT TAT CCC AAG AAC GAG AAA GAC AAT CTC ACT AAT GCT GAA AAG  
Amino acid sequence: 81 I A Y P K N E K D N L T N A E K  

**Changed to TAG**  
↑  
DNA sequence: 289 AAC CAG TTG AAG AAA GCC GTA GAA GGT ATA GAG CAA GCC TCC CAG GAG  
Amino acid sequence: 97 N Q L **K\*** K A V E G I E Q A S Q E  
↓  
**Changed to Stop**

DNA sequence: 337 GCA AGG GGG AAG AAA TGA  
Amino acid sequence: 113 A R G K K **Stop**

Figure 5.10. The nucleotide and amino acid sequences of ORF09 showing the introduction of mutation. The codon AAG that normally codes for lysine was changed to a stop code by introduction of the mutation.

### 5.6 Role of ORFs 07-09 in pBSSB2 stability

The stability of pRP4HindIII derivatives with ORFs 07-09 individually mutated was assayed in a *pcnB* mutant as described previously. Construct pRP4HindIII was used as positive control and pUC18 was used as negative control. The stability assay indicated that 97% of the colonies retained pRP4HindIII. The corresponding percentages of

colonies containing the plasmids were 89 and 100 for plasmids containing mutant ORF07 and ORF08, respectively. These values were significantly higher than the 58% plasmid-containing colonies in the case of the negative control, pUC18. A plasmid containing a mutation in ORF09 had 64% plasmid-containing colonies which is comparable with the 58% for pUC18 (Figure 5.17). It was therefore concluded that ORF09 is necessary for stabilization of the circular derivative and by implication the linear plasmid, pBSSB2.

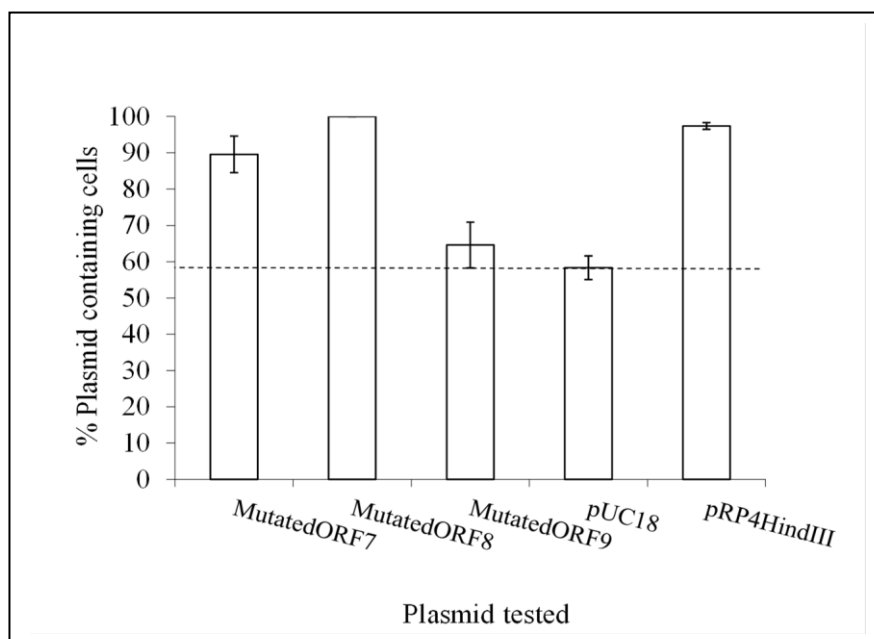


Figure 5.11. Stability assay of constructs with ORFs 07-09 mutated. Stability assays were performed in triplicate for each plasmid. The dotted line indicates the percentage for the negative control, pUC18.

In summary, the minimum region required to stabilize a circular derivative was a 1023 bp fragment of pBSSB2. This region was obtained by deletions from the left end of RP4HindIII. Any deletion from the right end reduced stability of the circular plasmid. This indicated that information contained in the right end was necessary for stabilization functions. On the other hand, mutagenesis studies showed that mutation of ORF09 affected stability of pRP4HindIII. ORF09 is present at the right end of the fragment RP4HindIII (Figure 5.6). The result of the mutagenesis experiment is, therefore, consistent with that of the deletion analysis.



## 5.7. Characterization of the RP4 stability function

Of two possible plasmid stability systems, viz. partitioning system and toxin-antitoxin, the latter seemed more likely for RP4 since only one ORF is required. In partitioning systems, two ORFs are usually involved, coding for ATPase and DNA-binding proteins. On the other hand, in a toxin-antitoxin system, the antitoxin can either be protein or RNA. In the latter case, therefore, it is possible to have either one ORF and one transcript, or two protein - coding ORFs. It was therefore hypothesized that the stability system contained in RP4HindIII is possibly a toxin-antitoxin system and that ORF09 was a potential toxin.

In order to test the hypothesis, ORF09 was cloned in an expression vector, pBAD18 (Guzman *et al.*, 1995). Plasmid pBAD18 carries the promoter of the arabinose operon,  $P_{BAD}$ , a multiple cloning site, a pBR322 origin of replication and chloramphenicol resistance marker.

Oligonucleotide primers tailed with EcoRI and PstI restriction sites (section 2.10.4) were used to amplify ORF09 from pBSSB2. The restriction sites were used to clone ORF09 in the pBAD18 vector. The presence of the insert was validated by amplification of the insert using the original pair of primers and by sequencing. The construct was designated pBAD18-ORF09.

For the toxicity assay, *E. coli* XLI-Blue containing pBAD18-ORF09 was grown overnight in L-broth at 37°C. An aliquot of overnight culture (250-300µl) was transferred to 50ml fresh L-broth medium and allowed to grow to  $OD_{600} \approx 0.3$ . The culture was then split into two. Two conditions were set up: one in which cells containing pBAD18-ORF09 were grown in L-broth plus glucose (0.2%) to repress  $P_{BAD}$  promoter, and one in which arabinose (0.2%) was added to the L-broth so that the arabinose promoter was induced and ORF09 was expressed. The time the culture was split was  $t=0$  (Figure 5.12). Cultures were also set up in which pBAD18 was contained in the host to control for the possibility of any toxic effect of the vector itself. As a measure of growth  $OD_{600}$  was read 30min, 60min, 90min and 150min. Samples were taken simultaneously for estimation of viable counts.

Figure 5.12 shows the growth of cultures containing the vector pBAD18 and pBAD18-ORF09 grown in the presence of glucose or arabinose. The cultures containing pBAD18 showed a similar pattern of growth in the presence of glucose and arabinose (Figure 5.12). This showed that no toxic effect was exerted by the vector alone. Compared to cells containing the vector only, those containing pBAD18-ORF09 exhibited a reduced rate of increase in OD when arabinose was present. This effect was not seen when the culture was supplemented with glucose rather than arabinose. This indicated that some growth-inhibitory effect was exerted by the product of ORF09.

In order to find out the effect of ORF09 expression on cell viability, duplicate samples collected during the growth experiment were diluted and spread on L-agar without additional glucose or arabinose. Plates were incubated at 37°C overnight and viable counts were determined. Figure 5.20 shows the viable counts obtained for samples of cultures containing pBAD18 or pBAD18-ORF09 grown in L-broth supplemented with glucose (0.2%) or arabinose (0.2%).

In the presence of glucose, similar viable counts were observed for cells containing either pBAD18 or pBAD18-ORF09. However, in the presence of arabinose, the pattern changed remarkably for pBAD18-ORF09. Viable counts for cells containing pBAD18-ORF09 were  $10^4$ -fold less after 30 minutes in the culture supplemented with arabinose. Thus, the expression of ORF09 resulted in the death but not lysis of most of the cells within 30 minutes of induction.

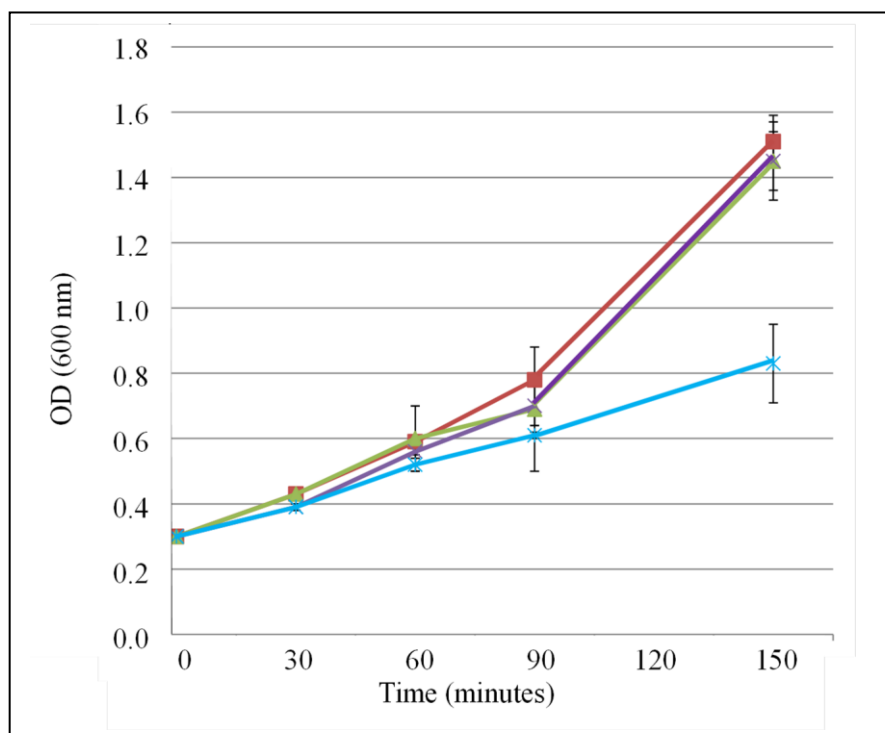


Figure 5.12. Growth of cultures containing pBAD18 or pBAD18-ORF09 in the presence of glucose (0.2%) or arabinose (0.2%). The red and the purple lines correspond to cells containing pBAD18 and pBAD18-ORF09, respectively, grown in L-broth containing glucose. Green and blue lines represent cells containing pBAD18 and pBAD18-ORF09, respectively, grown in L-broth with arabinose. All assays were performed in triplicate and results are expressed as average  $\pm$ S.D.

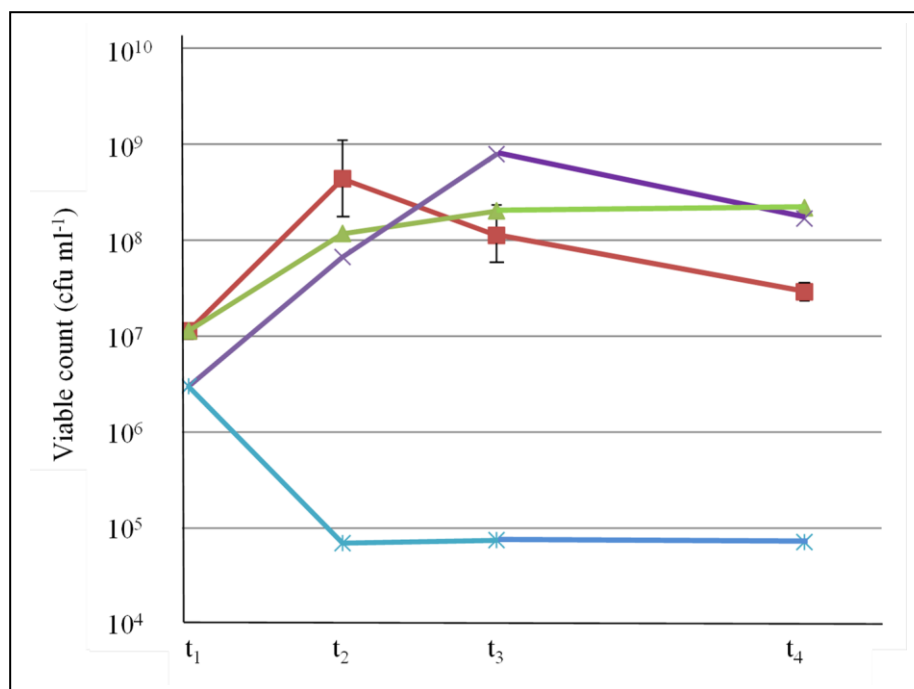


Figure 5.13. Growth of cultures containing pBAD18 and pBAD18-ORF09 in the presence of glucose or arabinose. Viable counts (cfu ml<sup>-1</sup>) of samples from cultures of XL1-Blue containing pBAD18 grown in L-broth with glucose (green) or arabinose (red) and pBAD18-ORF09 grown in L-broth containing glucose (purple) or arabinose (Blue).

### 5.8. The effect of ORF09 expression on cell morphology

In order to observe any changes in cell morphology due to the expression of ORF09, culture samples (5µl approx.) of *E. coli* XLI-Blue containing pBAD18 or pBAD18-ORF09 grown in L-broth with glucose (0.2% final concentration) or arabinose (0.2% final concentration) were embedded in molten, slightly cooled agarose (1% w/v) in phosphate buffer saline (PBS) and observed under the phase contrast microscope (Figure 5.14).

There was no change in cell morphology when *E. coli* XL1-Blue transformed with pBAD18 was grown for 30 minutes in L-broth containing glucose (Figure 5.14) (a) or arabinose (Figure 5.14) (b). In case of cells containing pBAD18-ORF09, no morphological change was evident following growth for 30 minutes in the presence of

glucose (Figure 5.14) (c). However, after induction with arabinose for 30 minutes, cells were larger and many appeared to be ‘empty’ or ‘ghost cells’ (Figure 5.14) (d). Such cells were visible within half an hour of the induction of ORF09 expression. A magnified image of cells with altered morphology is also shown in Figure 5.14 (e) and (f).

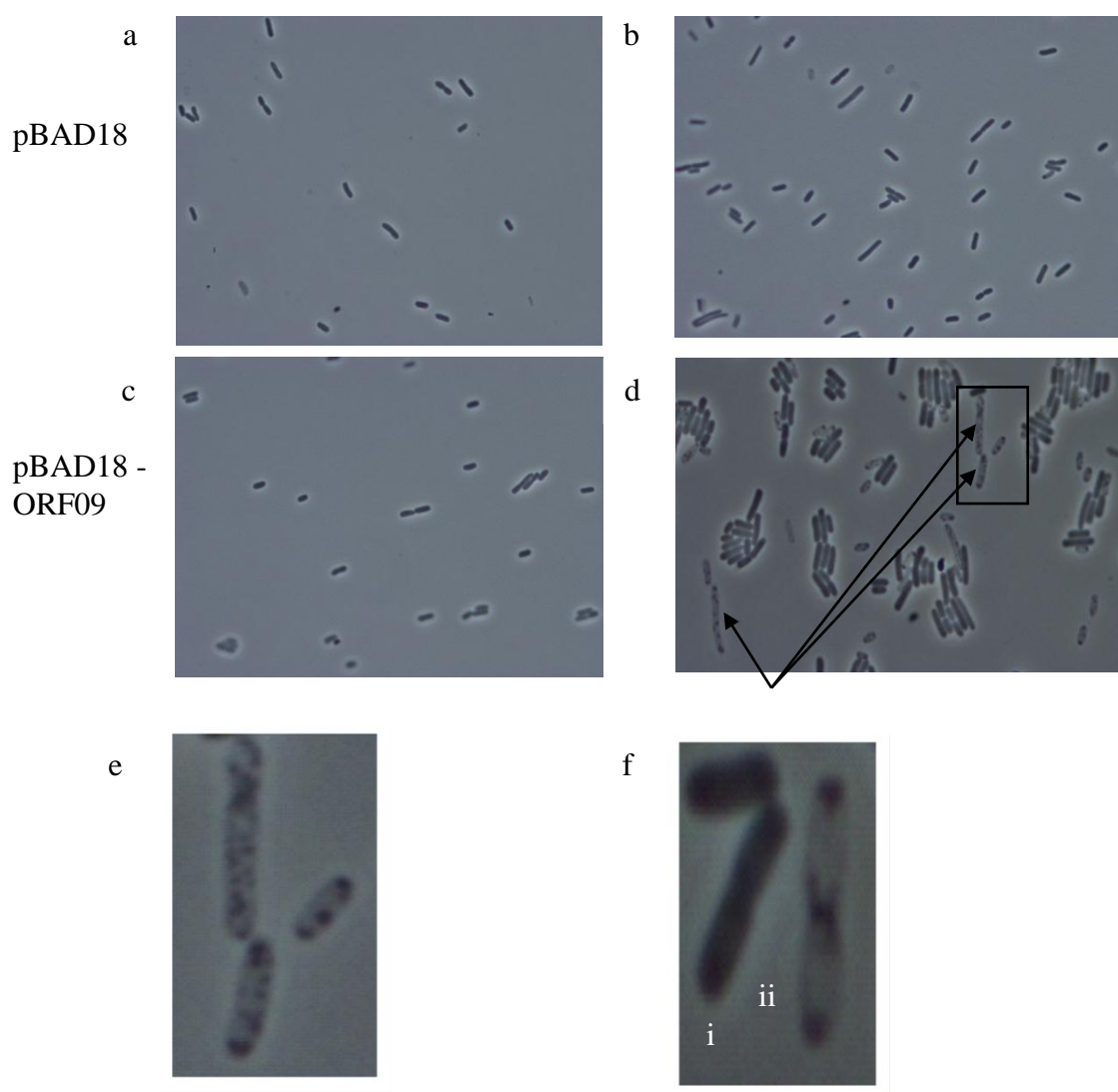


Figure 5.14. Phase contrast micrographs of *E. coli* XL1-Blue cells containing pBAD18 grown in L-broth with glucose (a) or arabinose (b), and pBAD18-ORF09 grown in L-broth containing glucose (c) or arabinose (d). Images were taken after 30min of growth in the presence of glucose or arabinose. Enlarged images of cells with altered morphology are shown in (e) and (f).

## 5.9. Possible candidates for an antitoxin

Mutation of ORF09 reduces the stability of pRP4HindIII, indicating that the ORF plays a role in stabilization of the circular plasmid by the 1023 bp region of pBSSB2. Furthermore, ectopic expression of ORF09 from an expression vector (pBAD18) dramatically reduces viability of the host. These observations support the hypothesis that the product of ORF09 is the toxin component of a toxin-antitoxin system. In most toxin antitoxin systems, the antitoxin gene precedes the toxin gene (Hayes, 2003). ORFs 07 and 08 are down-stream of ORF09 in fragment RP4HindIII, and it has been shown experimentally that they are not required for the stabilization of a circular plasmid. Further sequence analysis of the 1023 bp fragment was performed to identify a candidate for the antitoxin.

### 5.9.1 Promoter search in the 1023 bp stability region

An antitoxin can either be a protein or a countertranscript. For the expression of either, a promoter is required. Promoter search software <http://linux1.softberry.com/> was used to find promoters in the 1023 bp RP4 sequence. Details of the potential promoters are presented in Table 5.1 and a schematic diagram showing the positions of the promoters within the 1023 bp fragment are shown in figure 5.15.

**Table 5.1. Promoter prediction for the 1023 bp RP4 region**

Serial	-35 sequence	Position	-10 sequence	Position	Spacing	Transcription start nucleotide
<b>Promoters in the complementary strand to that containing ORF09</b>						
1	TTATCA	31-36	TAATAC	52-57	15	T
2	TTGTCT	461-466	TAAGCT	482-487	15	G
3	TTGATA	870-875	GATAAT	890-895	14	G
<b>Promoters in the same strand as that containing ORF09</b>						
4	TTGAAA	892-897	TAACCT	914-919	16	A
5	TTGCCG	435-440	TTGCCT	457-462	16	T
6	TTCATT	126-131	TAATAT	146-151	14	A

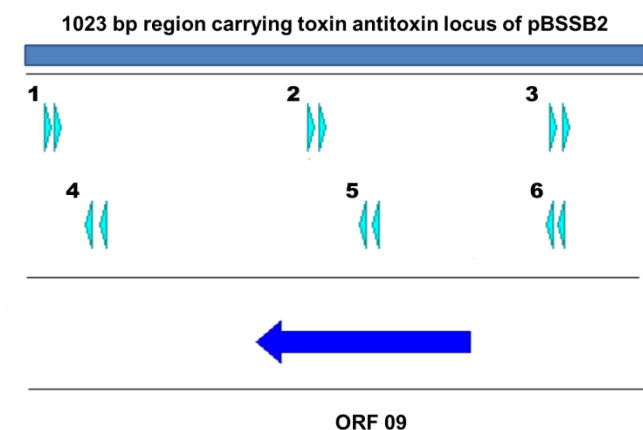


Figure 5.15. Potential promoters within the 1023 bp fragment carrying the toxin-antitoxin locus of pBSSB2

Promoter 6 is 143 bases upstream of the translation start codon of ORF09. It is possible that this promoter transcribes the toxin mRNA. The -10 box, TAATAT, conforms closely to the consensus TATAAT. However, the -35 sequence of TTCATT differs significantly from the consensus TTGACA. The spacer length of 14 bases is short. In spite of a weak -35 sequence and a short spacer length, promoter 6 may be the promoter for ORF09 mRNA since this is the only predicted promoter close to this ORF. Intrinsic terminators of transcription were searched by looking for GC-rich sequences that were followed by T-rich segments (Farnham and Platt, 1981; reviewed by Platt, 1986; Wilson and Hippel, 1995). In the absence of these parameters, rho-dependent terminators were searched by looking for C-rich regions (reviewed by Richardson and Greenblatt, 1996). However, no candidate transcription termination sequences could be found for the proposed toxin transcript (Figure 5.16).

-35	-10
<b><u>TTCATT</u></b> ATCAGGAAATAAAT <b><u>TAATAT</u></b> CAATATGTACCAGTTATGTTTCTGGATATACGGCTTACCCA CTTAAAAGAACGGTTAACTTATCATCAAAAAATGGCTTGCAACTCTTACACAGTACCACTAAGATAGA CTCAT ACAGTGTGCGAGTGAGAT <b>AGAGTTAAGGGG</b> ATGAATTTACTACACTTCATTGAAACCAA <b>GTCTTTCAGAGAGTAATCGATGGATTGCTTTCACCCGACGAACTCCGGGAGTTTCAGGAAGTGTTAA</b> <b>GGCAAGATCCTACCGCTGGCGATACCATATCAGGTACGGGCGGTTGCCGTAAGATTTCGCTGGGCTTT</b> <b>GCCTGGTATGGGTAAAAGTGGTGGTATCAGGGTCATATACTACTATCTGACAGCAGATAACGAAGTA</b> <b>TTCTTCTTATAGCTTATCCCAAGAACGAGAAAGACAATCTCACTAATGCTGAAAAGAACCAGTTGA</b> <b>AGAAAGCCGTAGAAGGTATAGAGCAAGCCTCCCAGGAGGCAAGGGGGAAGAAATGA</b> GAAAAGAAGCTC TTTGACGACCTGATCGAGAGTGTTAACGAAGCCGTAGCCATCAAACGGGGTCTTGTAGCCCCAGCTC	

Figure 5.16. Nucleotide sequence containing putative promoter for pBSSB2 toxin. The putative -35 and -10 sequences are shown in bold and underlined, the proposed Shine-Dalgarno sequence is shown in blue and ORF09 is highlighted in yellow.

The RNA-seq transcription analysis carried out at the Wellcome Trust Sanger Institute by Chinyere Okoro was used for additional information. This analysis indicated transcription from one strand only (Figure 5.17). The positions of the promoters were aligned with the transcription profile in order to evaluate the likelihood that a promoter was functional.



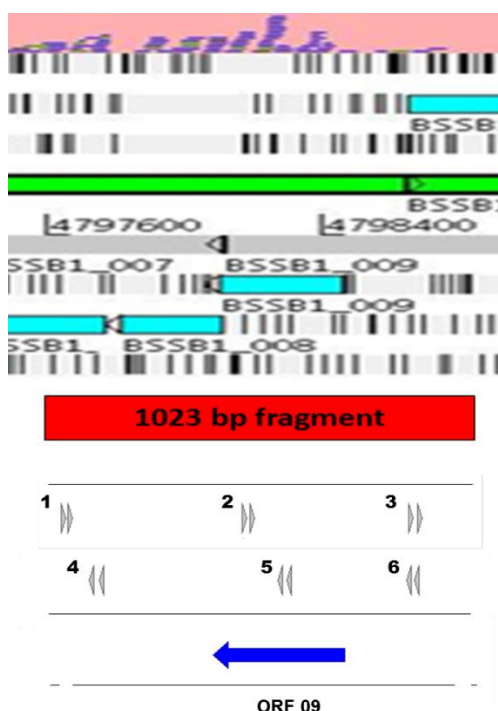


Figure 5.17. Putative promoters in relation to the RNA-seq transcription pattern obtained for the 1023 bp RP4 derivative carrying the toxin antitoxin locus of pBSSB2. The blue graph on pink background represents the transcription pattern from the same strand that contains ORF09; each pair of grey arrowhead indicates a putative promoter. The position of ORF09 has been shown with a dark blue arrow in order to compare the relative positions of the promoters with respect to this ORF.

The antitoxin can be a countertranscript that can bind to the toxin mRNA and prevent its synthesis. For this to occur, the countertranscript must be complementary to the toxin mRNA at least over some of its sequence. Looking at the promoters in the anti-sense strand (complementary to the strand containing ORF09) in Table 5.1 and Fig. 5.17, promoters 1 and 2 were obvious candidates for further analysis because they were present either upstream (promoter 1) or within the sequence (promoter 2) of the region complementary to ORF09. Promoter 3 is downstream of the region complementary to ORF09 and is therefore unlikely to be a potential candidate for synthesis of a candidate countertranscript.

For promoter 1 the -35 sequence, TTATCA, and the -10 sequence, TAATAC, were significantly different from the consensus (TTGACA for the -35 and TATAAT for the -

10). In addition the spacer length between the -35 and the -10 sequences of 15 bp in size was shorter than the optimum of 17 bp. Because of the poor match to the consensus sequences, spacer length and its position, which is significantly upstream of the region concerned, it was not considered to be a strong candidate promoter for a countertranscript.

Promoter 2 lies within the region complementary to ORF09 (Figure 5.18). When compared to the consensus (TATAAT) the -10 box, TAAGCT, has the invariants (Harley and Reynolds, 1987) at 1<sup>st</sup>, 2<sup>nd</sup> and 6<sup>th</sup> positions conserved. The -35 sequence of TTGTCT has the invariants (Harley and Reynolds, 1987) at 1<sup>st</sup> to 3<sup>rd</sup> positions conserved compared to the consensus TTGACA. The spacer length of 15 bp is, however, short for a functional promoter. Because of its strong -10 sequence, promoter 2 was considered a stronger candidate for further analysis than promoter 1.

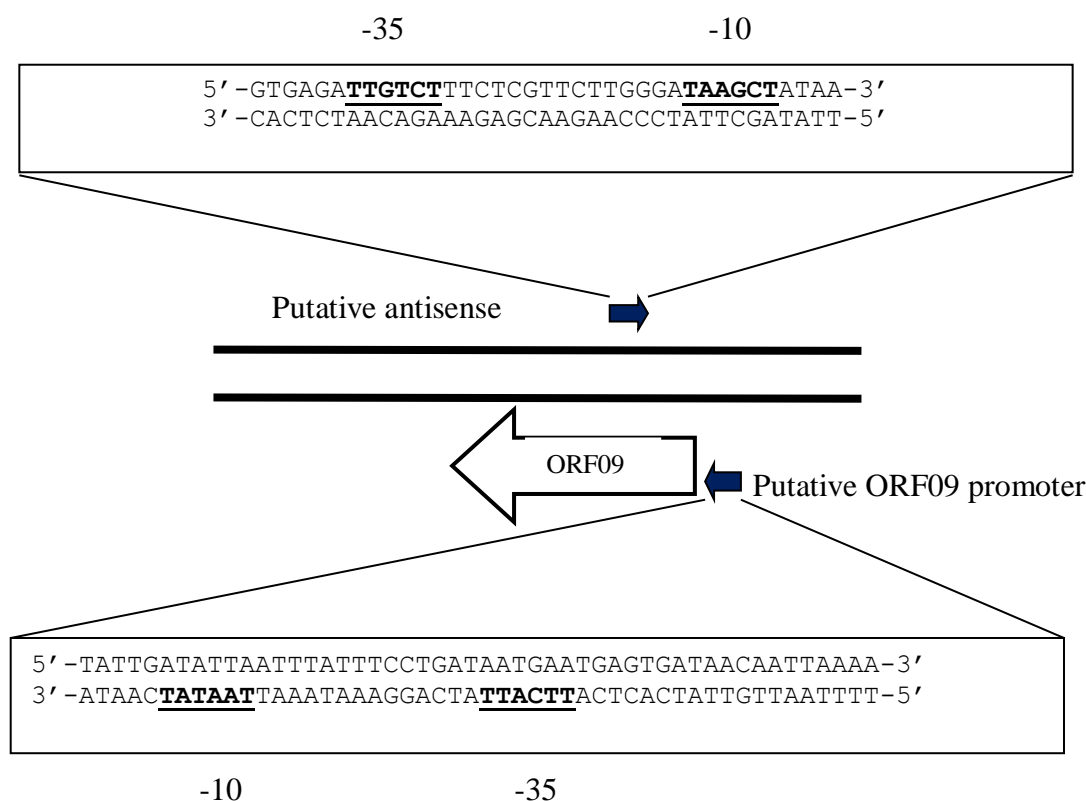


Figure 5.18. Promoter 2, a candidate promoter for antitoxin countertranscript formation.

A GC- rich region with the sequence CCCGGAGTTCGTCGGG (651-667 bp) followed by a relatively AT rich region might act as a transcription terminator for this promoter

(Figure 5.19). Secondary structure prediction indicated that the GC rich region can form a stem-and-loop region consisting of four GC bonds in the stem (Figure 5.20).

TTGTCTTTCTCGTTCTTGGGATAAGCTATAAGAAGGAATACTTCGTTATCTGCTGTCAGATAGTAGTA  
TATGACCCTGATACCACCACTTTTACCCATACCAGGCAAAGCCCAGCGAATCTTACGGCAACCGCCCG  
TACCTGATATGGTATCGCCAGCGGTAGGATCTTGCCTTAACACTTCCTGAAACTCCCGAGTTCGTCG  
GGTGAAAGCAAT

Figure 5.19. Possible transcription termination sequence (bold and underlined) for transcription initiated from promoter 2 (highlighted in blue).

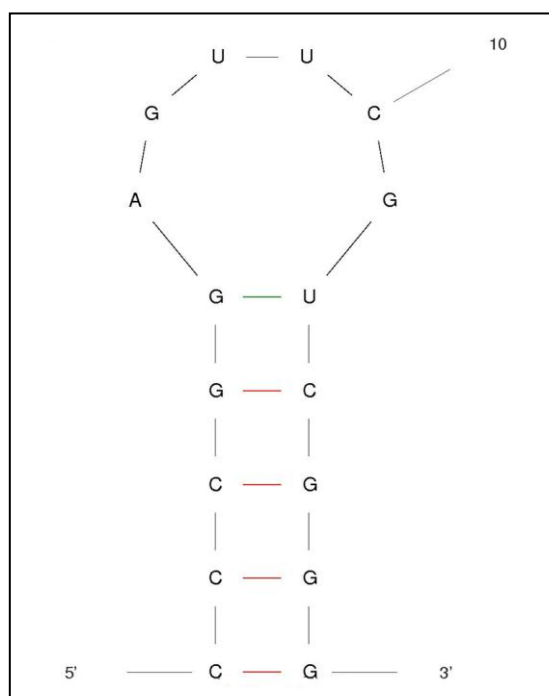


Figure 5.20. Secondary structure of possible terminator of transcription from promoter 2. GC bonds are indicated in red.

A countertranscript from promoter 2 might bind to the toxin mRNA and inhibit translation. For this to happen, the region of the toxin that is complementary to the anti-toxin RNA has to be accessible for the antitoxin to bind. It was, therefore, necessary to study the secondary structure of the toxin mRNA to estimate how likely it would be for a countertranscript to bind to it. The secondary structure of the toxin was predicted using the software available at <http://mfold.rna.albany.edu> (Figure 5.21).

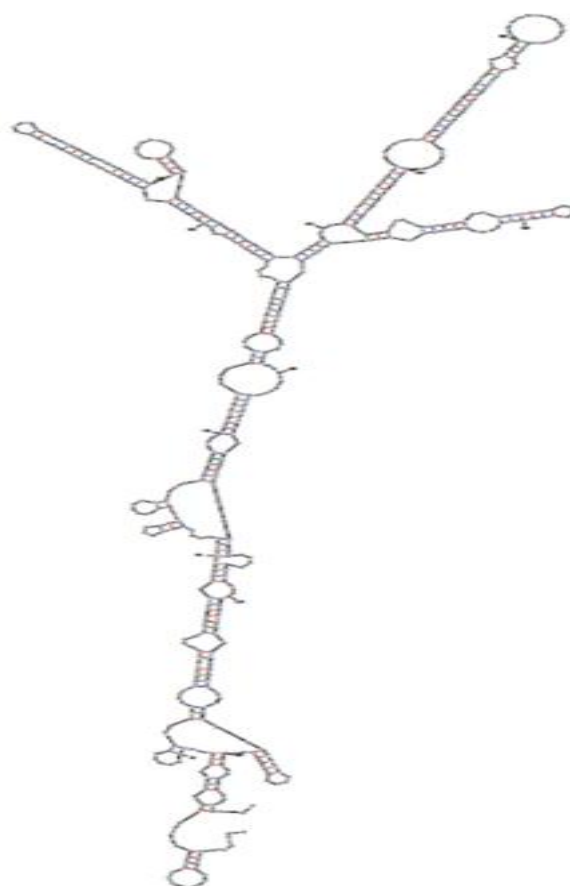


Figure 5.21. Predicted secondary structure for the ORF09 transcript.

The putative countertranscript from promoter 2 is 175 bases. The program mfold ([www.mfold.rna.albany.edu](http://www.mfold.rna.albany.edu)) was used to predict the secondary structure of this putative anti-sense RNA (Figure 5.22). It was found that the antisense RNA also forms stem-loop structures as has also been observed with the *sok* RNA of R1 (Section 1.4.2.2, figure 1.12).

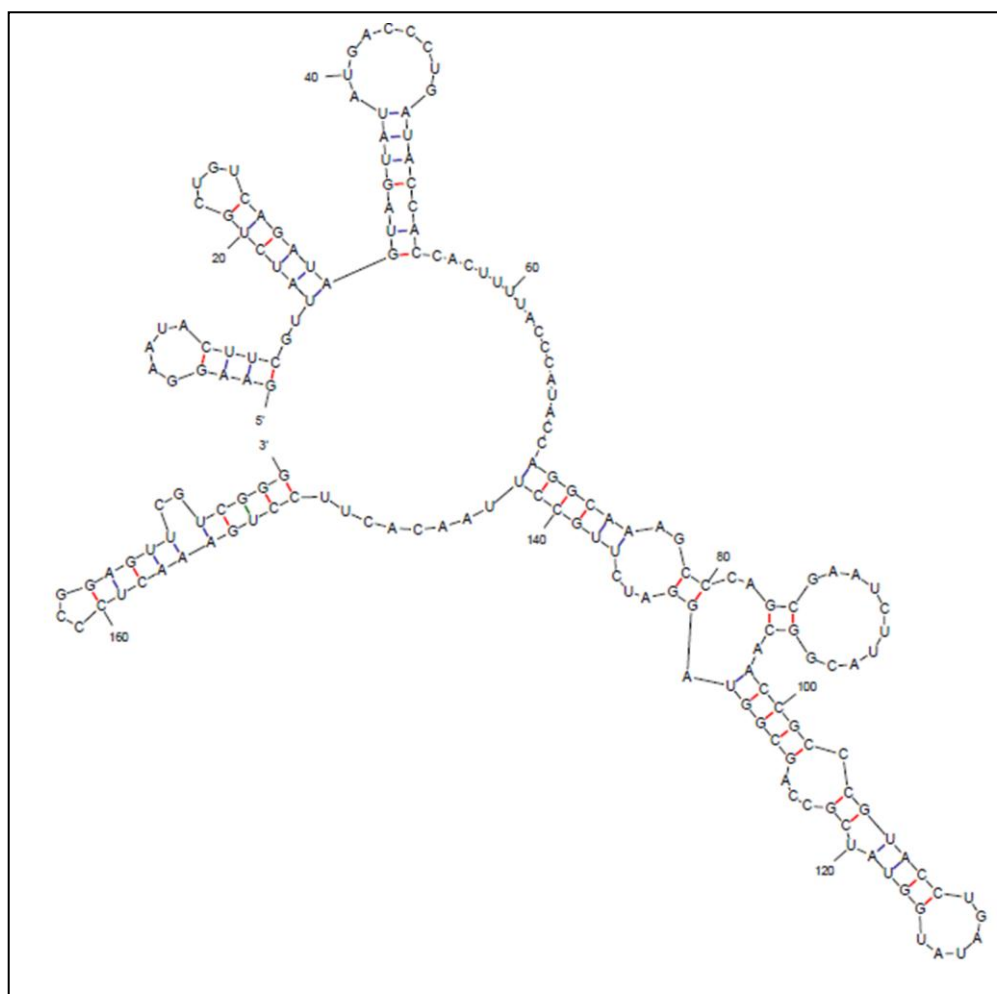


Figure 5.22. Predicted secondary structure for candidate antitoxin transcript. GC bonds are shown in red.

Figure 5.23 shows the region of the predicted toxin mRNA which is complementary to the putative antitoxin transcript from promoter 2. According to this figure, the part of the toxin mRNA which is complementary to the candidate countertranscript is involved in the formation of extensive stem and loop structures. It is possible that initial interaction between the toxin and the candidate antitoxin occurs through loops.

The possibility that the suggested countertranscript binds to the toxin mRNA needs to be tested in future studies. qPCR could be done to determine whether the countertranscript is actually synthesized. Introduction of mutation at promoter 2 sequences could also determine whether this promoter is functional. If the promoter were functional, mutation would prevent transcription. In the absence of the antitoxin, a

reduction in viability of the host cells and a phenotype as was observed in section 5.8, figure 5.14 would be observed.

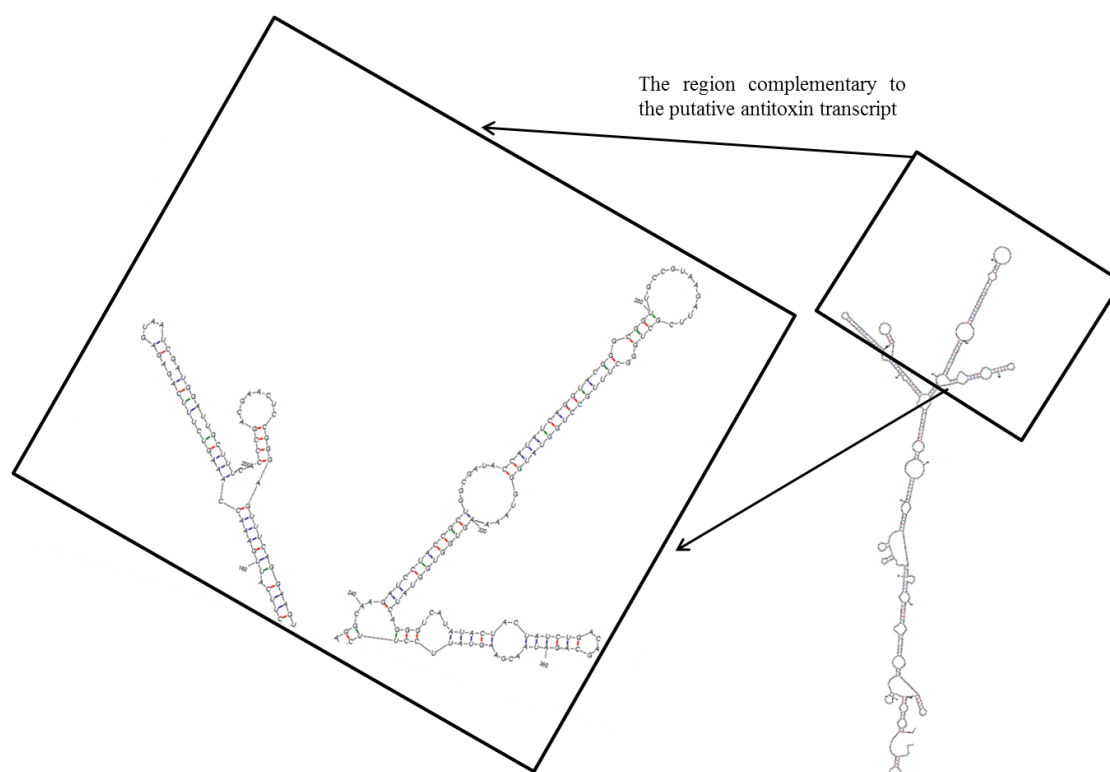


Figure 5.23. Predicted secondary structure of pBSSB2 toxin. The region complementary to hypothetical countertranscript is enlarged.

### **5.9.2 Prediction of ORFs in the 1023 bp region**

Another option for an antitoxin is that it could be a protein. The ORFs located in both the strands were analysed with a view to finding a protein anti-toxin using the ORF finder tool available at [www.ncbi.nlm.nih.gov/gorf/gorf.html](http://www.ncbi.nlm.nih.gov/gorf/gorf.html). Bacterial code was selected under the 'Genetic Code' option. Other parameters were kept at default. ORFs found using this tool are shown in Table 5.2. Two ORFs, v and vi, were present in this strand. The nucleotide sequences of these ORFs are shown in figures 5.23 and 5.24. The relative positions of these ORFs and putative promoters identified in this strand are indicated in figure 5.25.

**Table 5.2. ORFs identified within the 1023 bp fragment as identified by the program at <http://www.ncbi.nlm.nih.gov/gorf/gorf.html>.**

ORF designation	Position	Size (bp)	Reading frame	Shine-Dalgarno sequence
ORFs in the strand complementary to that with ORF09				
i	114-455	342	+3	None
ii	122-484	363	+2	None
iii	132-455	132	+3	None
iv	530-646	117	+2	AGATAG
ORFs in the strand that contains the toxin				
v	5-112	108	+2	None
ORF09	295-648	354	+1	AGAGTT
vi	466-648	183	+1	None
ORF08	645-935	291	+3	GGGGAA

TTAAATGGGGTTCGTGGTTTTACTTCATAAGTGCAATCAAAGTATTCAGGAATAAATTTACTCTTTC  
TTTGTGTCTCGGATTCTTAATACCATATTTATCTCCTTTTTAA

Figure 5.23. Nucleotide sequence of ORF v. The nucleotide sequence is highlighted in blue.

GATCCTACCGCTGGCGATACCATATCAGGTACGGGCGGTTGCCGTAAGATTCGCTGGGCTTTGCCTGG  
TATGGGTAAAAGTGGTGGTATCAGGGTCATATACTACTATCTGACAGCAGATAACGAAGTATTCCTTC  
TTATAGCTTATCCCAAGAACGAGAAAGACAATCTCACTAATGCTGAAAAGAACCAGTTGAAGAAAGCC  
GTAGAAGGTATAGAGCAAGCCTCCCAGGAGGCAAGGGGAAGAAATGA GAAAAGAACTCT TTGA CG

Figure 5.24. Nucleotide sequence of ORF vi. The nucleotide sequence is highlighted in blue.

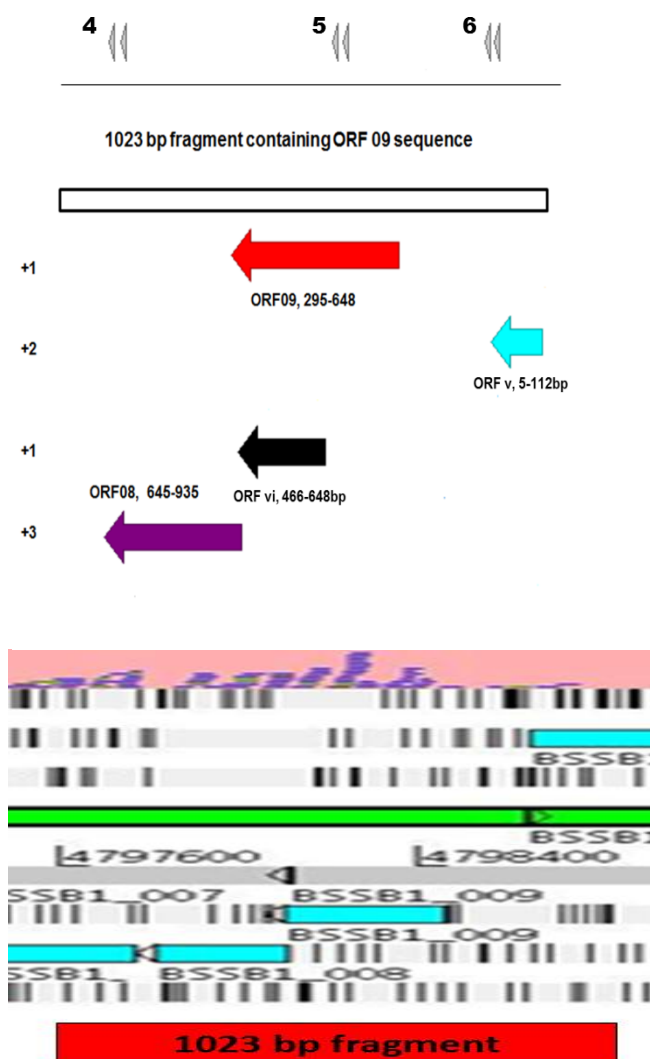


Figure 5.25. ORFs in the same strand as ORF09. The hollow square and the red box represent the 1023 bp fragment. The coloured arrows indicate ORFs. The numbers on the left that follow the ‘+’ sign are the reading frames of the ORFs. The grey arrowhead pairs indicate the predicted promoters.

ORF v was only four bases away from the end ligated to pUC18 MCS. This was too short to look for SD sequence. ORF vi, which is in the same reading frame as ORF09 (Figure 5.25), can also be transcribed from promoter 6. Although, no putative Shine-Dalgarno sequence could be detected <10 bp of the start codon of ORF vi (Figure 5.24) the possibility that ORF vi is translated cannot be ruled out. Translation of ORF vi would support the transcription pattern obtained for this region at the Sanger Institute



which indicates transcription from one strand only. However, whether ORF vi is translated needs to be determined in future studies.

The nucleotide sequences of the ORFs in the complementary strand are shown in figures 5.26-5.29.

```
AATAAAGCCCCAGTAAAGGGGCTTATGCTGTTCTGAACAGGTTAAGCATGTCAGGATGCTTTTCAA
TGACCCGCAACAGTTTAAGACTGCTGCCGGTGGGTGTTCTCTTTTGCTGTTCCAGGCTTGCACCA
GAGAAGGGGAGACACCCACTACCTGGGCAAATCATTGTTTGTGAGCTTCGCTTTCGTGCGGATAG
CTTTCACATCGGGCATCTCAAAAGTCGTAACACGAGCTGGGGCTACAAGACCCCGTTTGATGGCTA
CGGCTTCGTTAACACTCTCGATCAGGTCGTCAAAGAGTTCTTTTCTCATTCTTCCCCCTTGCCTC
CTGGGAGGCTTGCTCTATACCTTCTACGGCTTCTTCAACTGGTTCTTTTCAGCATTAGTGAGATT
```

Figure 5.26. Nucleotide sequence of ORF I highlighted in blue.

```
GTTAAGCATGTCAGGATGCTTTTCAATGACCCGCAACAGTTTAAGACTGCTGCCGGTGGGTGTTCTC
TTTTGCTGTTCCAGGCTTGCACCAGAGAAGGGGAGACACCCACTACCTGGGCAAATCATTGTT
TGAGCTTCGCTTTCGTGCGGATAGCTTTCACATCGGGCATCTCAAAAGTCGTAACACGAGCTGGGGC
TACAAGACCCCGTTTGATGGCTACGGCTTCGTTAACACTCTCGATCAGGTCGTCAAAGAGTTCTTTT
CTCATTCTTCCCCCTTGCCTCCTGGGAGGCTTGCTCTATACCTTCTACGGCTTCTTCAACTGGTT
CTTTTCAGCATTAGTGAGATTGCTTTCTCGTTCTTGGGATAAGCTATAAGAAGGAATACTTCGTTA
```

Figure 5.27. Nucleotide sequence of ORF ii highlighted in blue.

```
TCAGGATGCTTTTCAATGACCCGCAACAGTTTAAGACTGCTGCCGGTGGGTGTTCTCTTTTGCTGTTC
CCAGGCTTGCACCAGAGAAGGGGAGACACCCACTACCTGGGCAAATCATTGTTTGTGAGCTTCGCTT
TCGTGCGGATAGCTTTCACATCGGGCATCTCAAAAGTCGTAACACGAGCTGGGGCTACAAGACCCCGT
TTGATGGCTACGGCTTCGTTAACACTCTCGATCAGGTCGTCAAAGAGTTCTTTTCTCATTCTTCCCC
CTTGCCCTCCTGGGAGGCTTGCTCTATACCTTCTACGGCTTCTTCAACTGGTTCTTTTCAGCATTAGT
GAGATTGTCTTTCT
```

Figure 5.28. Nucleotide sequence of ORF iii highlighted in blue.

```

-35          -10
TTGTCTTTCTCGTTCTTGGGATTAAGCTATAAGAAGGAATACTTCGTTATCTGCTGTCAGATAGTAGT
ATATGACCCTGATACCACCACTTTTACCCATACCAGGCAAAGCCCAGCGAATCTTACGGCAACCGCC
CGTACCTGATATGGTATCGCCAGCGGTAGGATCTTGCTTAACTTCTCTGAAACTCCCGGAGTTTCG
TCGGGTGAAAGCAATCCATCGAT
```

Figure 5.29. Nucleotide sequence of ORF iv highlighted in blue. The putative SD sequence is highlighted in yellow. The -35 and -10 sequences are in bold and underlined.

The relative positions of the ORFs and promoters in the complementary strand are shown in figure 5.30.

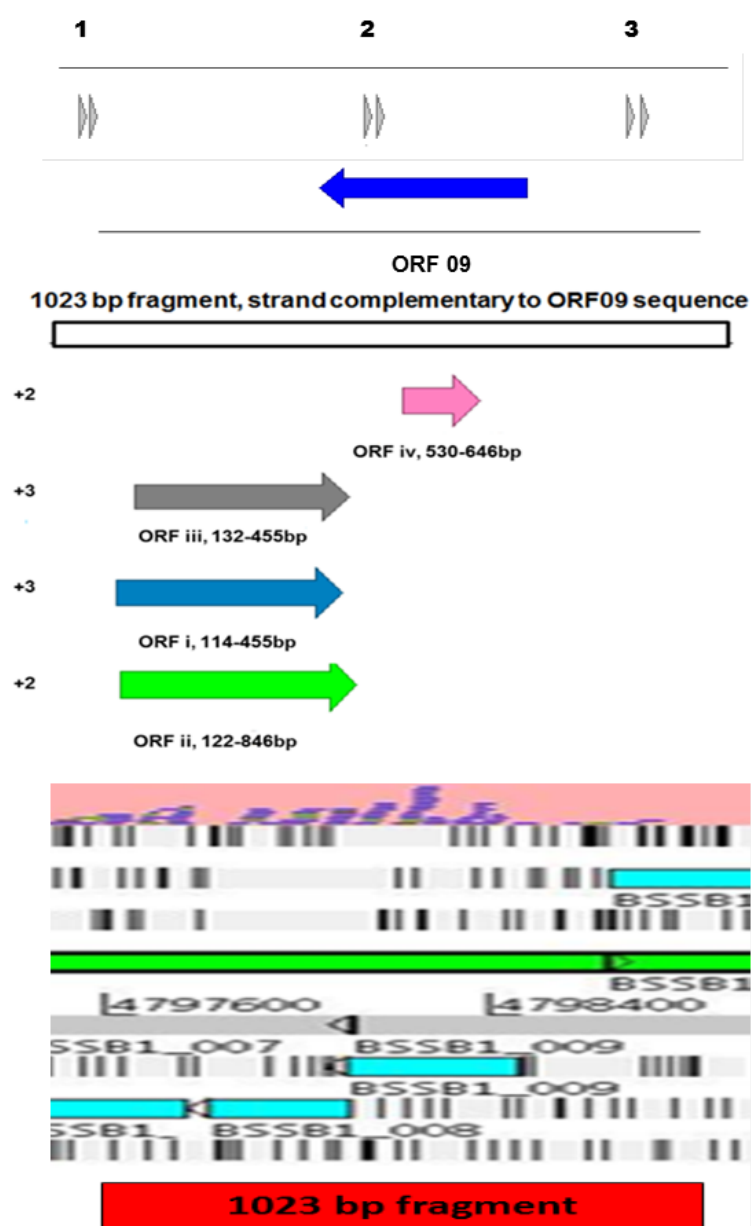


Figure 5.30. ORFs identified within the 1023 bp fragment carrying the toxin-antitoxin locus of pBSSB2. The hollow square and the red box represent the 1023 bp fragment. The coloured arrows indicate ORFs. The numbers on the left that follow the ‘+’ sign are the reading frames of the ORFs. The grey arrowhead pairs indicate predicted promoters.

With respect to ORFs identified within the complementary strand (Table 5.2), only ORF iv had a putative RBS (the sequence AGATAG six bases upstream of the start codon; Figure 5.29). This could be a potential candidate for encoding a protein antitoxin. Promoter 2, which is immediately upstream with its -10 Pribnow box 37 bases upstream of the start codon of ORF iv could potentially act as a promoter for transcription of ORF iv. The relative positions of ORF iv and 09 are indicated in figure 5.31.

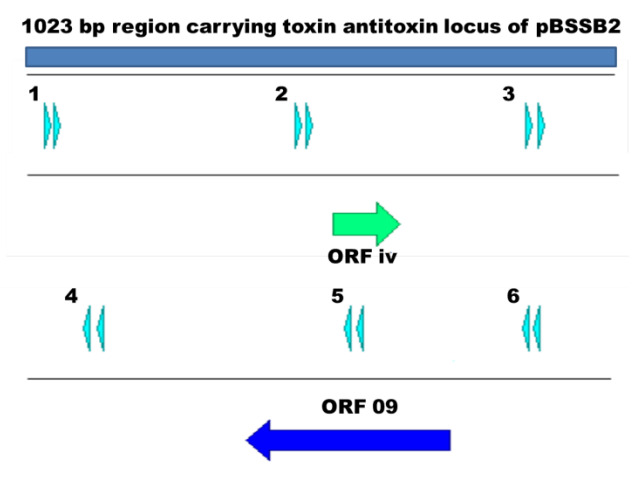


Figure 5.31. Position of ORF iv relative to the toxin ORF09. The solid blue bar represents the 1023 bp fragment carrying the toxin-antitoxin locus. The light blue arrowhead pairs are the -35 and -10 consensus of promoters. The green and blue arrows indicate ORFs iv and 09, respectively.

Based on the sequence analysis described above, it appears that either the transcript from promoter 2 or the translation of the same transcript (ORF iv) could be potential candidates for the antitoxin. Future studies will need to be carried out to determine the validity of these hypotheses.

### 5.9.3 BLAST analysis of nucleotide and protein sequences

A search for sequence similarity and simple structure prediction was done using the software at <http://www.sbg.bio.ic.ac.uk/~phyre/> to predict the structure of both the

toxin and the candidate antitoxin. This software uses amino acid sequence as input. A psi blast carried out by the same software indicated similarity of the toxin sequence to RelE like proteins. The *relEB* is a chromosomal toxin-antitoxin locus found in *E. coli*. The RelE protein toxin inhibits translation through mRNA cleavage and its activity is inhibited by binding of the RelB protein antitoxin (reviewed by Gerdes, 2005). A screenshot of the hits with similarities is shown in (Figure 5.32).

Toxin_structure		bits	E-value	N	100.0%
1 UniRef50_B1Y113	RelE-like cytotoxic transla...	123	2e-27	1	41.5%
2 UniRef50_D2MJ10	RelE-like cytotoxic transla...	122	2e-27	1	39.1%
3 UniRef50_C6N0S4	RelE-like cytotoxic transla...	119	2e-26	1	39.6%
4 UniRef50_A8AR77	Putative uncharacterized pr...	119	2e-26	1	50.9%
5 UniRef50_B8INT0	Putative uncharacterized pr...	118	4e-26	1	36.9%
6 UniRef50_Q5P4X4	RelE-like Cytotoxic transla...	118	4e-26	1	37.4%
7 UniRef50_C3X7Y8	Helix-turn-helix domain-con...	117	8e-26	1	42.7%
8 UniRef50_C5BEN8	Putative uncharacterized pr...	117	8e-26	1	51.4%
9 UniRef50_B1Z9S3	RelE-like cytotoxic transla...	117	8e-26	1	41.9%
10 UniRef50_Q39YW4	Putative uncharacterized pr...	115	3e-25	1	36.9%
11 UniRef50_C0VLK1	RelE-like cytotoxic transla...	114	5e-25	1	48.1%
12 UniRef50_A1WQC4	RelE-like cytotoxic transla...	113	1e-24	1	33.3%
13 UniRef50_A5G7I9	Putative uncharacterized pr...	113	1e-24	1	40.4%
14 UniRef50_B6ITK4	Putative uncharacterized pr...	113	1e-24	1	41.2%
15 UniRef50_A0S0K1	Putative uncharacterized pr...	111	5e-24	1	35.4%
16 UniRef50_B4S0M0	RelE-like Cytotoxic transla...	110	7e-24	1	37.6%
17 UniRef50_C7JIU2	Translation repressor RelE/...	110	8e-24	1	33.6%
18 UniRef50_A7BRZ1	RelE-like cytotoxic transla...	109	2e-23	1	41.6%
19 UniRef50_B7L4T2	Putative uncharacterized pr...	109	2e-23	1	50.5%
20 UniRef50_A9BUS1	Putative uncharacterized pr...	109	3e-23	1	27.9%
21 UniRef50_A9IKD5	RelE protein n=1 Tax=Barton...	108	5e-23	1	36.9%
22 UniRef50_A8GWH8	Putative transcriptional re...	107	7e-23	1	36.4%
23 UniRef50_C3JY94	Putative uncharacterized pr...	107	1e-22	1	40.0%
24 UniRef50_C7I4F4	Putative uncharacterized pr...	106	1e-22	1	33.3%
25 UniRef50_Q7N0B2	Similar to unknown protein ...	105	3e-22	1	49.1%
26 UniRef50_A1WP31	Putative uncharacterized pr...	105	3e-22	1	39.6%
27 UniRef50_Q9KMA6	Toxin hlgB-2 n=31 Tax=Prote...	105	3e-22	1	43.6%
28 UniRef50_A0LNU9	Putative uncharacterized pr...	105	3e-22	1	28.8%
29 UniRef50_UPI0001C348DB	hypothetical protein PretD1...	105	4e-22	1	39.6%

Figure 5.32. Selected results of psi blast of the ORF09 sequence at [http:// www. sbg. bio. ic. ac. uk /~phyre/](http://www.sbg.bio.ic.ac.uk/~phyre/).

The same software also searches for similarity between the predicted folding pattern of the query sequence with folding patterns deposited for other proteins in the database. The putative secondary structure indicated here is predicted with 60% probability for the toxin and was drawn by analysing the possible folding patterns of the toxin and comparison with the folding patterns deposited for other proteins in the database. The folding indicating the greatest match with existing protein information in the database is shown here (Figure 5.33).



Figure 5.33. A proposed secondary structures of the toxin encoded by the 1023 bp RP4 derivative of pBSSB2. This structure was derived with 60% probability based on the similarity of its folding pattern with RelE-like bacterial chromosomal toxin.

The amino acid sequence of the candidate antitoxin did not show any significant match with any other protein in the database. The only folding similarity was with a human signalling protein and the structure output was provided with a 10% probability only. Because there was no similarity to bacterial proteins, this structure was not taken into consideration.

Future studies need to be conducted to investigate whether the candidate ORF suggested for the protein antitoxin is functional. If it is, then introduction of a

premature nonsense mutation into the ORF would be expected to generate the same phenotype as ectopic overexpression of the ORF09 toxin.

## 5.10 Discussion

This chapter constitutes the first report of the identification and preliminary characterization of a stability function of plasmid pBSSB2. Using plasmid vectors that were either unstable naturally (pFH450) or were rendered unstable because of the genotype of their host (pUC18 in *pcnB* mutant), fragment RP4, covering nucleotide 4594-9106 of pBSSB2, was found to contain a stability function. Upon further deletion studies, the stability function was found to be contained in a 1023 bp fragment spanning nucleotides 7236 to 8258 in the pBSSB2 genetic map. This region contained three ORFs but a mutational analysis revealed that only one (ORF09) was required to confer stability on a circular plasmid.

Since a partitioning system would most likely require two ORFs, one coding for an ATPase and the other for a DNA-binding protein, it was hypothesized that the stability function was a toxin - antitoxin system. This hypothesis was tested by ectopic expression of ORF09. It was found that cultures that over-expressed ORF09 contained mostly non-viable cells. These cells appeared to be devoid of cytoplasmic contents when viewed using a phase contrast microscope. Cells showing a central clearing with condensed poles as a result of the effect of a toxin were coined 'Ghost cells' by Gerdes, (1986). A psi blast carried out at <http://www.sbg.bio.ic.ac.uk/~phyre/> indicated similarity to RelE-like cytotoxic translation inhibitors (E-value =  $2e-27$ ). The *relEB* is a chromosomal toxin antitoxin locus found in *E. coli*. The protein toxin, RelE, degrades mRNA and inhibits translation. As with pBSSB2, a toxin - antitoxin system has also been found as the stability determinant in the linear plasmid, pCLP, of *M. celatum* (Dantec *et al.*, 2001). The plasmid pCLP system is similar to PemI - PemK of *E. coli* plasmid R100 (Tsuchimoto *et al.*, 1988). The *pem* operon of *M. celatum* encodes 98 and 84 amino acid peptides that act as PemK (toxin) and PemI (antitoxin), respectively. When *M. smegmatis* was transformed with the shuttle vector pMIP12 containing the 373bp *pemK* gene, it was unable to grow on selective medium. Growth was restored

when a point mutation was introduced into the *pemK* gene. Like RelE, the toxin component of pCLP is also an endoribonuclease. No similarity was, however, found between the nucleotide sequences of ORF09 of pBSSB2 and the PemK of pCLP (Accession number: AF312688.1).

The absence of any toxicity resulting from mutations of ORFs 07 and 08 seemed to discount them as antitoxins. A candidate ORF (ORF vi) in the same strand as that containing ORF09 and a potential countertranscript or a putative protein antitoxin in the strand complementary to that containing ORF 09 has been identified and are proposed as candidates for the antitoxin in future investigations. Transcription of the candidate antisense RNA could potentially prevent transcription of ORF09 mRNA and hence stop translation of the toxin. Alternatively, translation of a candidate ORF can lead to the production of a candidate antitoxin. ORF09 shows similarity to RelE-like toxins. The antitoxin of RelE is the protein RelB. Thus a protein antitoxin for ORF09 toxin might be expected. However, at present there is no evidence that either the promoter or the putative protein antitoxins from either strand (translated from ORF vi or ORF iv) is functional and these hypotheses need to be tested in future studies.

In the present study, an unsuccessful attempt was made to knock out ORF09 in the parental linear plasmid using the lambda red recombinase method (Datsenko and Wanner, 2000). A chloramphenicol marker flanked by sequences specific for ORF09 was amplified by PCR and transformed into *E. coli* cells containing pBSSB2, with the expectation that the marker will insert into ORF09 sequence due to homologous recombination between the flanking sequences and the corresponding sequences in the linear plasmid. Although chloramphenicol resistant colonies were obtained, they did not contain the cassette in the right position. This was determined by amplifying a region containing the cassette. Colonies containing the cassette in the right position were expected to give an amplicon that was about a kb longer than the negative control that did not contain the cassette. In the absence of the right sized amplicon, it was concluded that the cassette had inserted in a different location and that no conclusion could be drawn from this experiment.

In addition to a toxin-antitoxin system a partitioning system is found in most plasmids. The partitioning system ensures that the daughter plasmids are distributed evenly during cell division. The toxin - antitoxin system is more likely to be a secondary

system that leads to death of cells not containing the plasmid on rare occasions when the partitioning system has failed. A similar arrangement is expected for pBSSB2. Sequencing and annotation of the pBSSB2 genome was done at the Wellcome Trust Sanger Institute and the sequence is deposited at the National Centre for Biotechnology Information (accession number AM419040). The annotation suggests that ORF017 possibly codes for a nucleotide binding protein that has similarity to ATPase-like ParA protein. The sequence has homology to the sequence of other bacterial plasmid partitioning systems such as those from *Vibrio caribbenthicus* (63% homology) and *Corynebacterium glutamicum* (51% homology) (Baker *et al.*, 2007b). The function of the ParA homolog in pBSSB2 and neighbouring sequence has not been investigated in detail in the present study. However, fragment RP2, which contained ORF017, exhibited some improvement in stability when inserted into pUC18 into a *pcnB* mutant. As with the *M. celatum* pCLP linear plasmid ParA homolog (Dantec and Picardeau, 2001), fragment RP2 is also located immediately adjacent to the RP7 replicon region. However, studies with fragment RP2 cloned in pFH450 did not provide support for the hypothesis that it contains stability function. Consequently work on this fragment was discontinued. Further studies with this fragment may yet reveal a partitioning function in pBSSB2. It is possible that a partitioning function of a linear plasmid might not effectively stabilize a circular plasmid.



# Chapter 6

---

## Discussion and Future work

---

## **6.1 Introduction**

This study focussed on the replication and stability functions of the linear plasmid, pBSSB1. The medical significance of studying pBSSB1 lies in the fact that it carries the gene for a flagellar antigen, H:z66. H:z66 is a phase II antigen in some *S. Typhi* Indonesian isolates that also express the H:d or H:j phase I antigen. Flagella are important for salmonella pathogenesis and understanding the biology of pBSSB1 will aid in the understanding of the epidemiology of typhoid in the Indonesian archipelago.

## **6.2 A method for the purification of protein-capped linear plasmids**

An initial objective of the study was the development of a method for the purification of the linear plasmid. The method used by Baker *et al.* (2007b) suffered from extreme lack of reproducibility. It also gave a very poor yield of the linear plasmid, such that detection of the plasmid required Southern blotting of gels (Stephen Baker, personal Communication). It was therefore important to develop a more efficient method to purify pBSSB2. According to the existing literature, PFGE and a variety of extraction protocols have been employed for extraction of linear plasmids (section 3.1). There is no standard protocol for linear plasmid purification. Conventional plasmid purification protocols do not always include a proteinase K treatment. For linear plasmids with covalently attached protein this would lead to loss of linear plasmids during any phenol-chloroform step used for the removal of cellular proteins. Column based methods also do not normally include a protease treatment step and linear plasmid DNA may be lost with cellular debris during the clearing spin. Any plasmid left in the supernatant that binds to the column may be removed during the subsequent medium-salt wash for the removal of protein, RNA and other impurities.

The method developed in this study, which owes its success to the inclusion of a proteinase K treatment step, gives good yield of pBSSB2. The optimized method has been used routinely for the purification of the linear plasmid during the course of the study. This method should be suitable for purification of protein-capped linear plasmids in general. This method is similar to that used by Baszczynski and Kemble (1987) for

the extraction of linear mitochondrial plasmids that had similar structural features to pBSSB2 and were purified from sorghum and maize.

### **6.3 Replication of pBSSB2**

In this study, although a 2831bp region has been found to be sufficient for replication of pBSSB2 in a circular mode (albeit inefficiently), a candidate Rep protein or initiator RNA has not been identified. Candidate ORFs have been proposed for further investigation upon sequence analysis. An alternative hypothesis based on initiation of replication by RNA primer has also been proposed. However, the hypothesis for both the candidate ORFs and the potential RNA primer need to be tested in future studies.

An adjacent 176bp region has also been found to exert incompatibility against pBSSB2 when present in a circular multicopy plasmid. Based on bioinformatic approaches and finding of a GC skew, a putative origin of replication was previously suggested to be located upstream of ORF017 (Baker *et al.*, 2007b). The ORF017 initiation codon lies about 300bp outside of the 2831bp fragment (Figure 6.1). It was found that although the putative origin of replication identified by Baker *et al.*, (2007b) lies outside of the minimum replication region it is actually contained within the incompatibility region, which lies upstream of ORF017 (Figure 6.1).

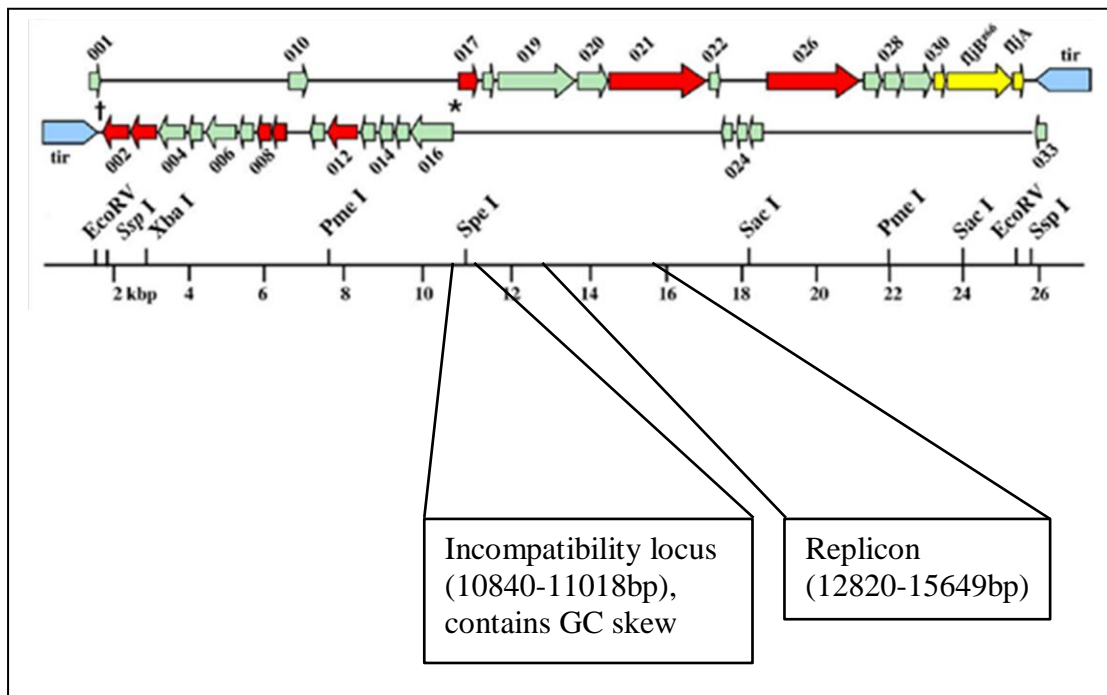


Figure 6.1 Schematic diagram showing the minimum region of amplicon RP7 that can drive replication of a circular pBSSB2 derivative and the region of amplicon RP2 conferring incompatibility functions in a circular mode.

The fact that the GC skew lies within the region conferring incompatibility supports the possibility that the origin of replication of pBSSB2, in addition to regulatory iterons, lies within this region. On the other hand, the ability of pRP7 (a circular plasmid containing the RP7 region) to replicate points to the possibility that this fragment carries the information for the initiator (e.g. a Rep protein). Hence, RP27 (combination of RP2 and RP7) must contain the actual basic replicon for the linear plasmid. However, this region is not sufficient to support replication in the circular mode. In the absence of RP2, repression of replication in the circular mode is removed and pRP7 probably replication from an origin-like sequence within RP7. RP7-supported replication is inefficient and this would be consistent with the need to use a secondary origin. In plasmid pSLA1 the presence of a locus, *rlrA*, which is required for replication in a linear mode has been found to be inhibitory to this function in the circular mode (Qin *et al.*, 2003). At least in this respect the function of RP2 is analogous to *rlrA*.

Incompatibility regions carry either information for a repressor (often RNA) or the origin or sites that titrate Rep protein. Four putative 6bp imperfect repeats were found in the 176 bp region conferring incompatibility against the parental plasmid. These short repeats were significantly different from the 21-mer iterons present in the well-characterized linear plasmid pSLA2 (Huang *et al.*, 2003). Future studies need to be carried out to find out whether the 6bp imperfect repeats found in this study contribute to titration of a hypothetical Rep protein.

As the components of the minimum replicon of the pBSSB2 circular derivative still remain enigmatic, it is pertinent to consider what the basic replicons of linear plasmids with structures similar to pBSSB2 contain. It was found that in *Streptomyces* linear plasmids with protein-capped 5' termini, the basic replicon usually contains one or more genes required for replication (Huang *et al.*, 2003). For example, in plasmid pSLA1, two genes (*rep1* and *rep2*) and two 21-mer iterons are required for replication in a circular mode (Huang *et al.*, 2003). Looking at the sequence of the 2831bp fragment of pBSSB2, four candidate ORFs (120, 201, 213 and 1224bp) still remain to be tested as encoding potential Rep proteins. An initial attempt made to insert a stop codon in the 1224bp ORF was unsuccessful but could not be repeated owing to time constraints. Direct repeats and two ORFs have also been found in the basic replicon of the linear plasmid SCP1 (Redenbach *et al.*, 1993). The Rep-iteron structure also exists in the minimum replicon of the linear plasmid, pSHK1 (Zhang *et al.*, 2009). Based on these findings it is hypothesized that one of the four hitherto untested ORFs might encode a Rep protein for pBSSB2 replication.

An interesting previous report identified two origins of replication in the plasmid pFRL1 from the *Streptomyces sp. FRI* (Zhang *et al.*, 2010). One of these origins can propagate in the circular mode. The other can propagate in the linear mode with the aid of additional loci containing the genes *rlrA/rorA*. In the case of pBSSB2 only one region has been found to support replication of a circular vector that is otherwise unable to replicate in the selected host. The possibility of another replicon which supports linear, but not circular, plasmid replication cannot be excluded. It is possible that the experimental approach used in the present investigation was not able to identify all

linear plasmid origins. Putting together all the information available at this stage, one possible model is that the 2831bp region encodes a Rep protein and the adjacent 176bp region contains the origin of replication and regulatory elements required for control of replication by titrating Rep protein (Figure 6.2).

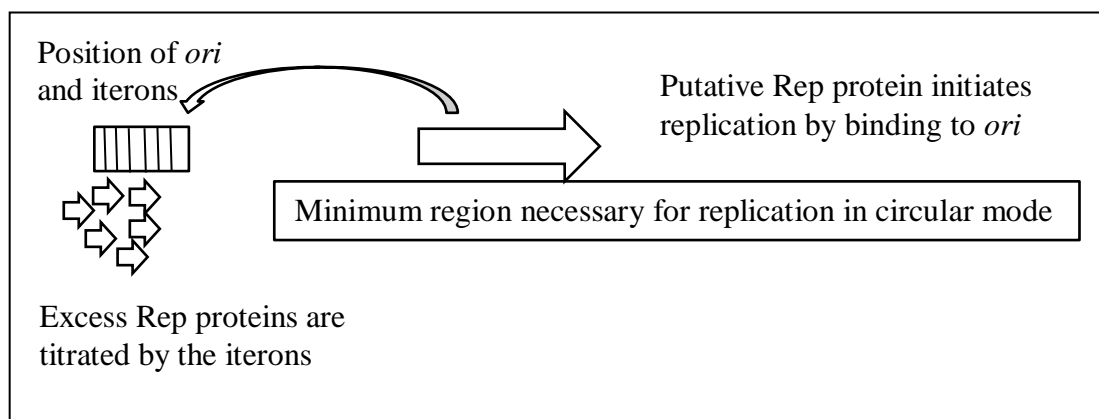


Figure 6.2. A tentative model for replication and its control in pBSSB2

#### 6.4 Future work necessary to characterize the pBSSB2 basic replicon

Future work on replication functions of pBSSB2 will need to focus primarily on determining whether a Rep protein or initiator RNA is involved in initiation of replication. Binding studies of a Rep protein to the direct repeats identified in this study could test whether they act as iterons in the regulation of pBSSB2 replication. In the event that an initiator RNA is involved, the role of RP2 in exerting incompatibility will also need to be determined. An alternative possibility is that the replicon for pBSSB2 may be similar to the ‘Rep-noncoding sequence’ configuration of some linear plasmids including plasmid pSLA1-L from *Str. rochei* (Hiratsu *et al.*, 2000), pSCL2 from *S. clavuligerus* (Wu *et al.*, 2006), pRL2, pRL4 and pFRL2 (Zhang *et al.*, 2009), pCLP from *Mycobacterium celatum* linear plasmid and pRHL3 from *Rhodococcus erythropolis* (De Mot *et al.*, 1997).

## 6.5 Stability function of pBSSB2

The present study has identified a potential toxin-antitoxin system from pBSSB2 that is able to stabilize an otherwise unstable circular plasmid. It is proposed that postsegregational killing contributes to the stabilization of pBSSB2 in linear mode. Very little progress has been made so far in studying linear plasmid stability systems. The only report that exists describes a toxin-antitoxin system in the linear plasmid pCLP of *Mycobacterium celatum* (Dantec and Picardeau, 2001). This is similar to PemI - PemK system of *E. coli* plasmid R100. The authors have shown that transformation of *M. smegmatis* with the *pemK* gene in a shuttle vector pMIP12 kills host cells. Confirmation that death of cells was occurring due to the presence of *pemK* was obtained when a point mutation in this gene restored growth of host cells. PemK is an endoribonuclease and degrades mRNA, inhibiting protein synthesis and eventually leading to death of host cells (Zhang *et al.*, 2004). In the present study, although the toxicity of the product of ORF09 has been experimentally demonstrated, the mode of action of the toxin has not been investigated.

Plasmid toxin-antitoxin systems are usually found in addition to a partitioning system. Whilst the former ensures that plasmid-free cells are killed, the latter minimises the probability of their formation. Thus, it is expected that a partitioning system exists in pBSSB2 but remains to be identified. The sequence similarity of ORF017 (obtained through the blast program at [www.ncbi.nlm.nih.gov/blast](http://www.ncbi.nlm.nih.gov/blast)) to bacterial ParA genes indicates the possibility of a partitioning system in pBSSB2. However, no sequence similarity to ParB has been identified. Fragment RP2, containing ORF017 and ORF018, was unable to stabilize an unstable circular plasmid. This may indicate this is not a *par* locus or that the information contained in RP2 is not sufficient for partitioning function. Further assays combining regions adjacent to RP2 in order to look for partitioning functions was not pursued due to lack of time. Finally it is possible that RP2 (or some other region of pBSSB2) does encode a partitioning system but that this is non-functional in a circular plasmid.

The testing of the toxin-antitoxin system, as well as any potential role of ORF017 and other associated ORFs in plasmid partitioning, remains to be investigated for a linear plasmid. If both postsegregational killing and partitioning systems exist in pBSSB2, knocking-out of ORF09, that codes for the toxin, may not significantly reduce the stability of pBSSB2, since a partitioning system alone might be sufficient to ensure proper distribution of the plasmid at most cell divisions. If ORF017 plays a role in partitioning, its knock-out should reduce stability of pBSSB2 although plasmid-free cells might still be killed by the toxin-antitoxin system. Simultaneous knockout of both ORF09 and of ORF017 and comparison of the segregational stability with the parental plasmid, as well as the effect of individual knock-outs, should be attempted to understand the role of ORF017 in pBSSB2 stability. In addition systematic deletions of each annotated ORF of pBSSB2, and testing of each derivative for stability, may help to identify any other regions involved with stability of pBSSB2.

The mechanism of action of ORF09 and identification of the antitoxin also remain the subjects of future studies. Future studies are necessary to identify the putative antitoxin. The antitoxin has not yet been identified but a candidate countertranscript and two potential ORFs have been proposed for future studies. A possible way of testing the candidate antitoxin is through construction of a *lacZ* fusion to the toxin gene and exploring the effect of putative regulator transcripts on *lacZ* expression. Further characterization of a hypothetical antitoxin may be achieved through RT-PCR, qPCR methods and mutagenesis analysis.

## **6.6 Conclusion**

In conclusion, this study has been successful in developing a method for the purification of protein-capped linear plasmids with good yield and reproducibility. The study has also identified the minimum replicon and an additional incompatibility region which function in a circular plasmid. It is hypothesized that the initiator of replication lies in the minimum replicon whereas the incompatibility region is associated with control of replication as well as containing the origin of replication. In addition, a toxin-



antitoxin system capable of stabilizing a circular unstable vector has also been found. A figure summarizing the findings from this study is shown (Figure 6.3).

An important point to bear in mind is that all work in this study has involved circular derivatives of pBSSB2. It is possible that there are differences in the requirements for replication in circular and linear modes. Whether the basic replicon identified in this study suffices to support replication in a linear mode, and hence in pBSSB2, is the major question to be answered in future phases of this study. This can be achieved through reconstruction of a linear plasmid using the minimum region identified for replication, incompatibility and stability regions identified in this study. Controlled replication and stable maintenance of this linear construct will provide evidence as to whether these components are sufficient for replication in the linear mode or whether additional loci are necessary.

The next step of this study would be to test whether the regions identified in this study also function in a linear mode. For this it is necessary to reconstruct a linear plasmid using pBSSB2 telomeres, the basic replicon identified in this study and the toxin-antitoxin module that would potentially stabilize the construct.

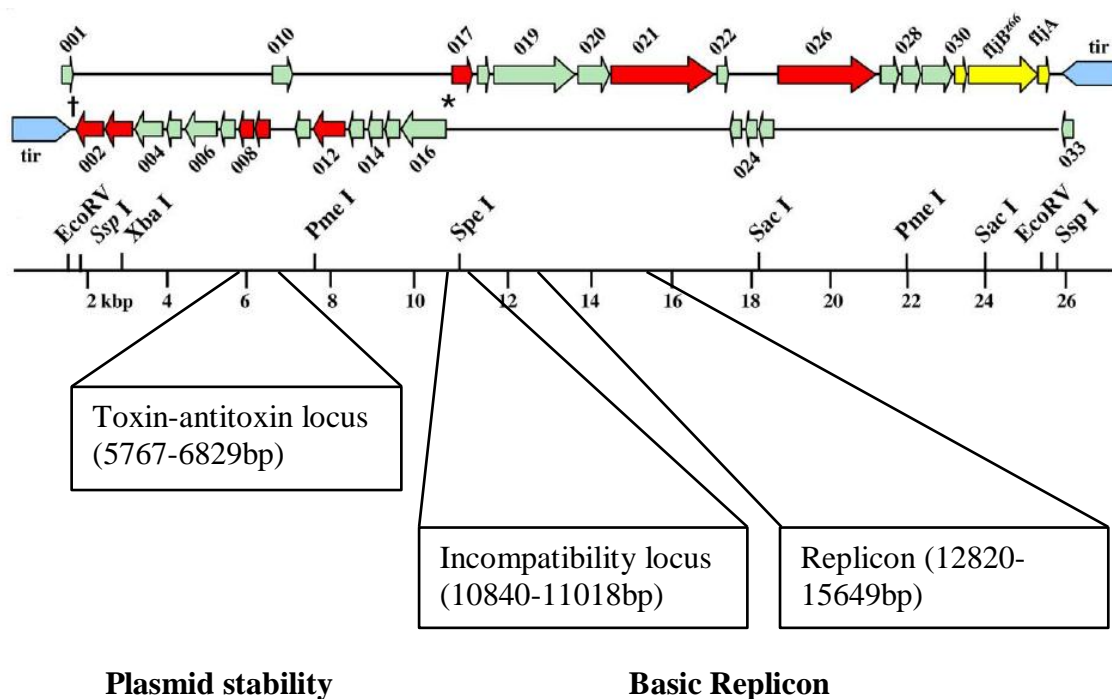


Figure 6.3. Overview of the regions associated with replication and stability of pBSSB2 identified in this study.

---

## Reference list

---

**Abeles, A.L. Reaves, L.D. Youngren-Grimes, B. Austin, S.J. (1995).** Control of P1 plasmid replication by iterons. *Mol. Microbiol.* **18(5)**: 903-912

**Abeles, A.L. and Austin, S.J. (1991).** Antiparallel plasmid-plasmid pairing may control P1 plasmid replication. *Proc. Natl. Acad. Sci. U.S.A.* **88(20)**:9011-5.

**Abeles, A. L., Reaves, L.D. and Austin, S.J. (1990).** A single DnaA box is sufficient for initiation from the P1 plasmid origin. *J. Bacteriol.* **172(8)**: 4386-91.

**Abeles, A. L., Friedman, S. A. and Austin, S.J. (1985).** Partition of unit-copy miniplasmids to daughter cells. III. The DNA sequence and functional organization of the P1 partition region. *J. Mol. Biol.* **185(2)**: 261-72.

**Abeles, A. L., Snyder, K.M. and Chatteraj, D.K. (1984).** P1 plasmid replication: replicon structure. *J. Mol.Biol.* **173(3)**: 307-24.

**Abramoff, M.D., Magelhaes, P. J., and Ram, S. J. (2004).** Image Processing with ImageJ. *Biophotonics International.* **11**: 36-42.

**Actis, L.A., Tolmasky, M.E. and Crosa, J.H. (1999).** Bacterial Plasmids: Replication of Extrachromosomal Genetic Elements Encoding Resistance to Antimicrobial Compounds. *Front. Biosci.* **4**: 43-62.

**Allen, J. M., Simcha, D. M., Ericson, N. G., Alexander, D. L., Marquette, J. T., Van Biber, B. P., Troll, C. J., Karchin, R., Bielas, J. H., Loeb, L. A., and Camps, M. (2011).** Roles of DNA polymerase I in leading and lagging-strand replication defined by a high-resolution mutation footprint of ColE1 plasmid replication. *Nuc. Acids Res.* **39(16)**: 7020-33.

**Alverson, J., Bundle, S.F., Sohaskey, C.D., Lybecker, M.C. and Samuels, D.S. (2003).** Transcriptional regulation of the *ospAB* and *ospC* promoters from *Borrelia burgdorferi*. *Mol Microbiol* **48(6)**:1665-77.

**Austin, S., Friedman, S. and Ludtke, D. (1986).** Partition functions of unit-copy plasmids can stabilize the maintenance of plasmid pBR322 at low copy number. *J Bacteriol.* **168(2)**:1010-3.

**Austin, S. and Abeles, A. (1983).** Partition of unit-copy miniplasmids to daughter cells. II. The partition region of miniplasmid P1 encodes an essential protein and a centromere-like site at which it acts. *J Mol Biol.* **169**: 373-87.

**Austin, S., Hart, F., Abeles, A. and Sternberg, N. (1982).** Genetic and physical map of a P1 miniplasmid. *J. Bacteriol.* **152**(1): 63-71.

**Baker, S., Holt, K., van de Vosse, E., Roumagnac, P., Whitehead, S., King, E., Ewels, P., Keniry, A., Weill, F.X., Lightfoot, D., van Dissel, J.T., Sanderson, K.E., Farrar, J., Achtman, M., Deloukas, P. and Dougan, G. (2008).** High-throughput genotyping of *Salmonella enterica* serovar Typhi allowing geographical assignment of haplotypes and pathotypes within an urban District of Jakarta, Indonesia. *J.Clin. Microbiol.* **46**(5): 1741-1746.

**Baker, S, Hardy, J., Sanderson, K.E., Quail, M., Goodhead, I., Kingsley, R.A., Parkhill, J., Stocker, B. and Dougan, G. (2007a).** A novel linear plasmid mediates flagellar variation in *Salmonella* Typhi. *PLoS Pathog.* **3**(5):605-610.

**Baker, S., Holt, K., Whitehead, S., Goodhead, I., Perkins, T., Stocker, B., Hardy, J. and Dougan, G. (2007b).** A linear plasmid truncation induces unidirectional flagellar phase change in H: z66 positive *Salmonella* Typhi. *Mol. Microbiol.* **66**(5):1207-18.

**Bannerman, T.L., Hancock, G.A., Tenover, F.C., Miller, J.M. (1995).** Pulsed-field gel electrophoresis as a replacement for bacteriophage typing of *Staphylococcus aureus*. *J Clin Microbiol.* **33**(3): 551-5.

**Bao, K. and Cohen, S.N. (2004).** Reverse transcriptase activity innate to DNA polymerase I and DNA topoisomerase I proteins of *Streptomyces* telomere complex. *Proc. Natl. Acad. Sci. USA* **101**(40): 14361-6.

**Bao, K. and Cohen, S.N. (2003).** Recruitment of terminal protein to the ends of *Streptomyces* linear plasmids and chromosomes by a novel telomere-binding protein essential for linear DNA replication. *Genes Dev.* **17**(6): 774-85.

**Bao, K. and Cohen, S.N. (2001).** Terminal proteins essential for the replication of linear plasmids and chromosomes in *Streptomyces*. *Genes Dev.* **15(12)**: 1518-27.

**Barbour, A.G., Carter, C.J., Bundoc, V. and Hinnebusch, J. (1996).** The nucleotide sequence of a linear plasmid of *Borrelia burgdorferi* reveals similarities to those of circular plasmids of other prokaryotes. *J. Bacteriol.* **178(22)**: 6635-9.

**Barbour, A.G. (1988).** Antigenic variation of surface proteins of *Borrelia* species. *Rev Infect Dis.* **10 Suppl 2**:S399-402.

**Barbour, A.G. and Garon, C.F. (1987).** Linear plasmids of the bacterium *Borrelia burgdorferi* have covalently closed ends. *Science.* **237(4813)**: 409-11.

**Baril, C., Richard, C., Baranton, G., Girons, S.I.S. (1989).** Linear chromosome of *Borrelia burgdorferi*. *Res. Microbiol.* **140(8)**:507-16.

**Boe, L., Gerdes, K. and Molin, S. (1987).** Effects of genes exerting growth inhibition and plasmid stability on plasmid maintenance. *J. Bacteriol.* **169(10)**: 4646-50.

**Baszczynski, C. L. and Kemble, R. J. (1987).** Terminal protein association and sequence homology in linear mitochondrial plasmid-like DNAs of sorghum and maize *Plant cell reports* **6(2)**: 127-130

**Beaurepaire, C. and Chaconas, G. (2005).** Mapping of essential replication functions of the linear plasmid lp17 of *B. burgdorferi* by targeted deletion walking. *Mol. Microbiol.* **57**:132–142

**Birnbaum, S. and Bailey, J.E. (1991).** Plasmid presence changes the relative levels of many host cell proteins and ribosome components in recombinant *Escherichia coli*. *Biotechnol. Bioeng.* **37**: 736-45.

**Birnboim, H.C. and Doly, J. (1979).** A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nuc. Acids Res.* **7(6)**:1513-23.

**Blakely, G., May, G., Mcculloch, R., Arciszewska, L.K., Burke, M., Lovett, S.T., and Sherratt, D.J. (1993).** Two related recombinases are required for site-specific recombination at *dif* and *cer* in *E. coli* K12. *Cell* **75**: 351-61.

**Bonifield, H.R. and Hughes, K.T. (2003).** Flagellar Phase Variation in *Salmonella enterica* is Mediated by a Posttranscriptional Control Mechanism. *J. Bacteriol.* **185(12)**: 3567-3574.

**Bouet, J. Y., Surtees, J.A., and Funnell, B.E. (2000).** Stoichiometry of P1 plasmid partition complexes. *J Biol Chem.* **275**: 8213-9.

**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. *EMBO J.* **18**: 1415-24.

**Bramhill, D. and Kornberg, A. (1988).** A Model for Initiation at Origins of DNA Replication. *Cell* 54: 915-918

**Bravo, A., De Torrontegui, G., and Diaz, R. (1987).** Identification of components of a new stability system of plasmid R1, ParD, that is close to the origin of replication of this plasmid. *Mol Gen Genet.* **210**: 101-10.

**Brendler, T., Abeles, A.L. and Austin, S. (1995).** A protein that binds to the P1 origin core and the *oriC* 13mer region in a methylation-specific fashion is the product of the host *seqA* gene. *EMBO J.* **14(16)**: 4083-9.

**Brendler, T., Abeles, A.L. and Austin, S. (1991).** Critical sequences in the core of the P1 plasmid replication origin. *J. Bacteriol.* **173(13)**: 3935-42.

**Cao G.J. and Sarkar, N. (1992).** Identification of the gene for an *Escherichia coli* poly(A) polymerase. *Proc. Natl. Acad. Sci. U. S. A.* **89(21)**:10380-4.

**Carle, G.F. and Olson, M.V. (1984).** Separation of chromosomal DNA molecules from yeast by orthogonal-field-alternation gel electrophoresis. *Nuc. Acids Res.* **12**:5647–5664

**Carleton, S., Projan, S.J., Highlander, S.K., Moghazeh, S.M. and Novick, R.P. (1984).** Control of pT181 replication II. Mutational analysis. *EMBO J.* **3(10)**: 2407-14.

**Casjens, S., Palmer, N., van Vugt, R., Huang, W.M., Stevenson, B., Rosa, P., Lathigra, R., Sutton, G., Peterson, J., Dodson, R.J., Haft, D., Hickey,**

**E., Gwinn, M., White, O. and Fraser, C.M. (2000).** A bacterial genome in flux: the twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete *Borrelia burgdorferi*. *Mol. Microbiol.* **35(3)**:490-516.

**Cesareni, G. (1982).** Control of ColE1 DNA Replication: The *rop* Gene Product Negatively Affects Transcription from the Replication Primer Promoter. *Proc. Natl. Acad. Sci. USA* **79(20)**: 6313-6317.

**Chaconas, G. and Kobryn, K. (2010).** Structure, function, and evolution of linear replicons in *Borrelia*. *Ann. Rev. Microbiol.* **64**: 185-202.

**Chaconas, G. (2005).** Hairpin telomeres and genome plasticity in *Borrelia*: all mixed up in the end. *Mol. Microbiol.* **58(3)**: 625-35.

**Chaconas, G., Stewart, P.E., Tilly, K., Bono, J.L. and Rosa, P. (2001).** Telomere resolution in the Lyme disease spirochete. *EMBO J.* **20(12)**: 3229-37.

**Chang, T., Kramer, M. G., Ansari, R. A. and Khan, S. A. (2000).** Role of Individual Monomers of a Dimeric Initiator Protein in the Initiation and Termination of Plasmid Rolling Circle Replication. *J. Biol. Chem.* **275(18)**: 13529 -13534.

**Chang, P.C., Kim, E.S. and Cohen, S.N. (1996).** *Streptomyces* linear plasmids that contain a phage-like, centrally located replication origin. *Mol. Microbiol.* **22(5)**:789-800.

**Chang, P.C. and Cohen, S.N. (1994).** Bidirectional replication from an internal origin in a linear *streptomyces* plasmid. *Science* **265 (5174)**: 952-4.

**Chardon-Lauriaux, I., Charpentier, M. and Percheron, F. (1986).** Isolation and characterization of a linear plasmid from *Streptomyces rimosus*. *FEMS Microbiol. Lett.* **35**:153–155

**Chater, K. F. and Kinashi, H. (2007).** *Streptomyces* Linear Plasmids: Their Discovery, Functions, Interactions with Other Replicons, and Evolutionary Significance. *Microbiol. Monogr.* **(7)**: 1-31

**Chattoraj, D.K., Mason, R.J. and Wickner, S.H. (1988).** Mini-P1 plasmid

replication: the autoregulation-sequestration paradox. *Cell* **52(4)**: 551-7.

**Chattoraj, D.K., Snyder, K.M. and Abeles, A. L. (1985).** P1 plasmid replication: multiple functions of RepA protein at the origin. *Proc. Natl. Acad. Sci. USA* **82(9)**: 2588-92.

**Chattoraj, D., Cordes, K. and Abeles, A.N.N. (1984).** Plasmid P1 replication: Negative control by repeated DNA sequences Integrative Suppression: An *in Vivo* Assay for Initiation of k-P1: 5' A. *Cell* **81(2)**: 6456-6460.

**Cheah, U.E. (1987).** Effects of Recombinant Plasmid Size on Cellular Processes in *Escherichia coli*. *Plasmid* **18(2)**:127-34

**Chen, C.W. (1996).** Complications and implications of linear bacterial chromosomes. *Trends Genet.* **12(5)**: 192-6.

**Clewell, D. B. and Helinski, D. R. (1969).** Supercoiled circular DNA-protein complex in *Escherichia coli*: purification and induced conversion to an open circular DNA form. *Proc Natl Acad Sci U S A.* **62(4)**: 1159–1166.

**Colloms, S. D., Sykora, P., Szatmari, G., and Sherratt, D.J. (1990).** Recombination at ColE1 *cer* requires the *Escherichia coli xerC* gene product, a member of the lambda integrase family of site-specific recombinases. *J Bact.* **172**: 6973-80.

**Cooper, T. F., Paixao, T. and Heinemann, J. A. (2010).** Within-host competition selects for plasmid-encoded toxin-antitoxin systems. *Proc. Biol. Sci.* **277**: 3149-3155.

**Cooper, T.F. and Heinemann, J.A. (2005).** Selection for plasmid post-segregational killing depends on multiple infection: evidence for the selection of more virulent parasites through parasite-level competition. *Proc Biol Sci.* **272**: 403-10.

**Cooper, T. F. (2000).** Postsegregational killing does not increase plasmid stability but acts to mediate the exclusion of competing plasmids. *Proc. Natl. Acad. Sci. USA* **97**: 12643-12648.



**Crespi, M., Messens, E., Caplan, A.B., van Montagu, M. and Desomer, J. (1992).** Fasciation induction by the phytopathogen *Rhodococcus fascians* depends upon a linear plasmid encoding a cytokinin synthase gene. *EMBO J.* **11(3)**:795-804.

**Crestfield, A.M., Smith, K.C. and Allen, F.W. (1955).** The preparation and characterization of ribonucleic acids from yeast. *J. Biol. Chem.* **216(1)**: 185-93.

**Dabrock, B., Kebeler, M., Averhoff, B. and Gottschalk, G. (1994).** Identification and characterization of a transmissible linear plasmid from *Rhodococcus erythropolis* BD2 that encodes isopropylbenzene and trichloroethene catabolism. *Appl. Environ. Microbiol.* **60**: 853-860.

**Dam, M. N. and Gerdes, K. (1997).** Sok antisense RNA from plasmid R1 is functionally inactivated by RNase E and polyadenylated by poly(A) polymerase I. *Mol Microbiol.* **26**: 311-20.

**Dantec, L. C., Winter, N., Gicquel, B., Vincent, V. and Picardeau, M. (2001).**

Genomic sequence and transcriptional analysis of a 23-kilobase mycobacterial linear plasmid: evidence for horizontal transfer and identification of plasmid maintenance systems. *J. Bacteriol.* **183(7)**:2157-64.

**Das, N. (2005).** Multiple homeostatic mechanisms in the control of P1 plasmid replication. *Proc. Natl. Acad. Sci. USA* **102(8)**: 2856-61.

**Dasgupta, S., Masukata, H. and Tomizawa, J.I. (1987).** Multiple mechanisms for initiation of ColE1 DNA replication: DNA synthesis in the presence and absence of ribonuclease H. *Cell* **51(6)**: 1113-1122.

**Datsenko, K.A. and Wanner, B.L. (2000).** One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl. Acad. Sci. U. S. A.* **97(12)**: 6640-5.

**Davey, M. J., and Funnell, B.E. (1994).** The P1 plasmid partition protein ParA. A role for ATP in site-specific DNA binding. *J. Biol. Chem.* **269**: 29908-13.

**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. *Mol Microbiol.* **21**: 1029-36.

**Davis, M. A., Martin, K.A., and Austin, S.J. (1992).** Biochemical activities of the parA partition protein of the P1 plasmid. *Mol Microbiol.* **6**: 1141-7.

**Davis, M. A., Martin, K.A., and Austin, S.J. (1990).** Specificity switching of the P1 plasmid centromere-like site. *EMBO J.* **9**: 991-8.

**De La Cueva-Mendez, G. and Pimentel, B. (2007).** Gene and cell survival: lessons from prokaryotic plasmid R1. *EMBO reports.* **8**: 458-464.

**De Mot, R., Nagy, I., DeSchrijver, A., Pattanapitpaisal, P., Schoofs, G. and Vanderleyden, I. (1997).** Structural analysis of the 6 kb cryptic plasmid pFA 52600 from *Rhodococcus erythropolis* NI86/21 and construction of *Escherichia coli*-*Rhodococcus* shuttle vectors. *Microbiol.* **143**: 3137-3147.

**del Solar, G., Giraldo, R., Ruiz-Echevarri'a, M.J., Espinosa, M. and Di'az-Orejas, R. (1998).** Replication and control of circular bacterial plasmids. *Microbiol. Mol. Biol. Rev.* **62**(2): 434-64

**Dempsey, L. A., Birch, P. and Khan, S.A. (1992).** Six Amino Acids Determine the Sequence-specific DNA Binding and Replication Specificity of the Initiator Proteins of the pTl8I Family. *J. Biol. Chem.* **267**: 24538-24543.

**Deneke, J., Burgin, A.B., Wilson, S.L. and Chaconas, G. (2004).** Catalytic residues of the telomere resolvase ResT: a pattern similar to, but distinct from tyrosine recombinases and type IB topoisomerases. *J. Biol. Chem.* **279**:53699–706

**Deneke, J., Ziegelin, G., Lurz, R. and Lanka, E. (2002).** Phage N15 telomere resolution: target requirements for recognition and processing by the protelomerase. *J. Biol. Chem.* **277**:10410–19

**Deneke, J., Ziegelin, G., Lurz, R., and Lanka, E. (2000).** The protelomerase of temperate *Escherichia coli* phage N15 has cleaving-joining activity. *Proc. Natl. Acad. Sci. USA* **97**:7721–26

**DeRoeck, D., Jodar, L. and Clemens, J. (2007).** Putting Typhoid Vaccination on the Global Health Agenda. *N. Engl. J. Med.* **357(11)**: 1069-1071.

**Dibbens, J. A, Muraiso, K. and Chatteraj, D.K. (1997).** Chaperone-mediated reduction of RepA dimerization is associated with RepA conformational change. *Mol. Microbiol.* **26(1)**: 185-95.

**Eliasson, A.A., Bernander, R., Dasgupta, S., and Nordstrom, K. (1992).** Direct visualization of plasmid DNA in bacterial cells. *Mol. Microbiol.* **6**: 165–170

**Engelberg-Kulka, H., and Glaser, G. (1999).** Addiction modules and programmed cell death and antideath in bacterial cultures. *Ann. Rev. Microbiol.* **53**: 43-70.

**Espinosa, M., del Solar, G., Rojo, F. and Alonso, J.C. (1995).** Plasmid rolling circle replication and its control. *FEMS Microbiol. Lett.* **130(2-3)**: 111-20.

**Evans, M., Kaczmarek, F.S., Stutzman-Engwall, K. and Dyson, P. (1994).** Characterization of a *Streptomyces-lividans*-type site-specific DNA modification system in the avermectin-producer *Streptomyces avermitilis* permits investigation of two novel giant linear plasmids, pSA1 and pSA2. *Microbiol.* **140 (6)**: 1367-71.

**Farnham, P. J. and Platt, T. (1981).** Rho-independent termination: dyad symmetry in DNA causes RNA polymerase to pause during transcription in vitro. *Nucl. Acids Res.* **9**: 563-577

**Flan, F. and Macnab, R.M. (1996).** Enzymatic characterization of FliI. An ATPase involved in flagellar assembly in *Salmonella typhimurium*. *J Biol. Chem.* **271(50)**: 31981-8.

**Franch, T., Gulyaev, A. P. and Gerdes, K. (1997).** Programmed Cell Death by hok/sok of Plasmid R1: Processing at the hok mRNA 3'-end Triggers Structural Rearrangements that Allow Translation and Antisense RNA Binding. *J. Mol. Biol.* **273**: 38-51.

**Franch, T. and Gerdes, K. (1996).** Programmed cell death in bacteria: translational repression by mRNA end-pairing. *Mol. Microbiol.* **21**: 1049-60.

**Frankel, G., Newton, S.M., Schoolnik, G.K. and Stocker, B.A. (1989).** Characterization of the *H1-j* gene of *Salmonella typhi*. *EMBO J* **8(1)**: 3149 - 3152.

**Fung, E., Bouet, J.Y., and Funnell, B.E. (2001).** Probing the ATP-binding site of P1 ParA: partition and repression have different requirements for ATP binding and hydrolysis. *EMBO J.* **20**: 4901-11.

**Funnell, B. E., and Gagnier, L. (1993).** The P1 plasmid partition complex at *parS*. II. Analysis of ParB protein binding activity and specificity. *J. Biol. Chem.* **268**: 3616-24.

**Funnell, B. E. (1991).** The P1 plasmid partition complex at *parS*. The influence of *Escherichia coli* integration host factor and of substrate topology. *J. Biol. Chem.* **266**: 14328-37.

**Funnell, B. E. (1988).** Participation of *Escherichia coli* integration host factor in the P1 plasmid partition system. *Proc. Natl. Acad. Sci. USA* **85**: 6657-61.

**Gerdes, K., Christensen, S.K. and Løbner-Olesen, A. (2005).** Prokaryotic toxin-antitoxin stress response loci. *Nat Rev Microbiol.* **3(5)**:371-82

**Gerdes, K. (2000).** Toxin-antitoxin modules may regulate synthesis of macromolecules during nutritional stress. *J. Bacteriol.* **182**: 561-72.

**Gerdes, K., Gulyaev, A.P., Franch, T., Pedersen, K. and Mikkelsen, N. D. (1997).** Antisense RNA-regulated programmed cell death. *Annu. Rev. Genet.* **31**: 1–31.

**Gerdes, K. (1988).** The *parB* (*hok/sok*) Locus of Plasmid R1: A General Purpose Plasmid Stabilization System. *Nat. Biotechnol.* **6**: 1402 - 1405

**Gerdes, K., Rasmussen, P. B. and Molin, S. (1986).** Unique type of plasmid maintenance function: postsegregational killing of plasmid free cells. *Proc. Natl. Acad. Sci. USA* **83**: 3116–3120.

**Gerdes, K., Larsen, J.E. and Molin, S. (1985).** Stable inheritance of plasmid R1 requires two different loci. *J. Bacteriol.* **161(1)**: 292-8.

**Gordon, S., Rech, J., Lane, D., and Wright, A. (2004).** Kinetics of plasmid segregation in *Escherichia coli*. *Mol Microbiol.* **51**: 461-9.

**Gravius, B., Glocker, D., Pigac, J., Pandza, K., Hranueli, D. and Cullum, J. (1994).** The 387 kb linear plasmid pPZG101 of *Streptomyces rimosus* and its interactions with the chromosome. *Microbiol.* **140 (9)**:2271-7.

**Greene, P. J., Guptam, M., Boyer, H.W., Brown, W.E., Rosenberg, J.M. (1981).** Sequence analysis of the DNA encoding the Eco RI endonuclease and methylase. *J.Biol. Chem.* **256**: 2143-53.

**Guinee, P.A. Jansen, W.H. Maas, H.M. Le Minor, L. and Beaud, R. (1981).** An unusual H antigen (z66) in strains of *Salmonella* Typhi. *Ann Microbiol.* **132(3)**: 331-334

**Guzman, L.M., Belin, D., Carson, M.J. and Beckwith, J. (1995).** Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177(14)**: **4121-4130**

**Hansen, E.B. and Yarmolinsky, M.B. (1986).** Host participation in plasmid maintenance: Dependence upon *dnaA* of replicons derived from P1 and F. *Proc. Natl. Acad. Sci. USA* **83**: 4423-4427.

**Hao, J. and Yarmolinsky, M. (2002).** Effects of the P1 Plasmid Centromere on Expression of P1 Partition Genes. *J. Bact.* **184(17)**: 4857-4867

**Harley, C. B. and Reynolds, R. P. (1987).** Analysis of *E. coli* promoter sequences. *Nuc. Acids. Res.* **15(5)**: 2343-2361

**Harrington, T. R., Glick, B.R. and Lem, N.W. (1986).** Molecular cloning of the phosphoenolpyruvate carboxylase gene of *Anabaena variabilis*. *Gene* **45**: 113-6.

**Hashimoto-Gotoh, T. and Timmis, K.N. (1981).** Properties of Col E1 and pMB1 Derivative Plasmids : Random Replication of Multicopy Replicons. *Cell.* **23**: 229-238.

**Hatta, M., Sultan, A. R., Pastoor, R. and Smits, H. L. (2011).** New flagellin gene for *Salmonella enterica* serovar Typhi from the East Indonesian archipelago. *Am. J. Trop. Med. Hyg.* **84(3)**: 429-34.

**Hattman, S., Brooks, J.E. and Masurekar, M. (1978).** Sequence specificity of the P1 modification methylase (M.Eco P1) and the DNA methylase (M.Eco dam) controlled by the *Escherichia coli* dam gene. *J. Mol. Biol.* **126(3)**: 367-80.

**Hayes, F. and Barilla, D. (2006).** The bacterial segrosome: a dynamic nucleoprotein machine for DNA trafficking and segregation. *Nat. Rev. Microbiol.* **4**: 133-143.

**Hayes, F. (2003).** Toxins-Antitoxins: Plasmid Maintenance, Programmed Cell Death, and Cell Cycle Arrest. *Science* **301**: 1496-1499.

**Hayes, F. (1998).** A family of stability determinants in pathogenic bacteria. *J. Bacteriol.* **180**: 6415–6418

**Henderson, S.C., Bounous, D.I. and Lee, M.D. (1999).** Early events in the pathogenesis of avian salmonellosis. *Infect Immun.* **67(7)**: 3580-6.

**Hiratsu, K., Mochizuki, S. and Kinashi, H. (2000).** Cloning and analysis of the replication origin and the telomeres of the large linear plasmid pSLA2-L in *Streptomyces rochei*. *Mol. Gen. Genet.* **263(6)**: 1015-21.

**Ho, S.N., Hunt, H.D., Horton, R. M., Pullen, J. K. and Pease, L.R. (1989).** Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene.* **77**: 51-59

**Holmes, D.S. and Quigley, M. (1981).** A rapid boiling method for the preparation of bacterial plasmids. *Anal. Biochem.* **114(1)**: 193-7.

**Huang, T.W. and Chen, C.W. (2008).** DNA polymerase I is not required for replication of linear chromosomes in *streptomyces*. *J. Bacteriol.* **190(2)**: 755-8.

**Huang, F. C. Werne, A. Li, Q. Galyov, E. E. Walker, W. A. and Cherayil, B. J. (2004).** Cooperative Interactions between Flagellin and SopE2 in the Epithelial

Interleukin-8 Response to *Salmonella enterica* Serovar Typhimurium Infection. *Infect Immun.* **72(9)**: 5052-5062.

**Huang, C.H., Chen, C.Y., Tsai, H.H., Chen, C., Lin, Y.S. and Chen CW. (2003).** Linear plasmid SLP2 of *Streptomyces lividans* is a composite replicon. *Mol. Microbiol.* **47(6)**: 1563- 76.

**Huang, C.H., Lin, Y.S., Yang, Y.L., Huang, S.W. and Chen, C.W. (1998).** The telomeres of *Streptomyces* chromosomes contain conserved palindromic sequences with potential to form complex secondary structures. *Mol. Microbiol.* **28(5)**: 905-16.

**Iordanescu, S. and Basheer, R. (1991).** The *Staphylococcus aureus* mutation *pcrA3* leads to the accumulation of pT181 replication initiation complexes. *J. Mol. Biol.* **221(4)**: 1183-9.

**Itoh, T. and Tomizawa, J.I. (1980).** Formation of an RNA Primer for Initiation of Replication of ColE1 DNA by Ribonuclease H. *Proc. Natl. Acad. Sci. USA* **77(5)**: 2450-2454.

**Itoh, T. and Tomizawa, J. (1978).** Initiation of replication of plasmid Col EI DNA by RNA polymerase, ribonuclease H, and DNA polymerase I, Cold Spring Harbor Symp. *Quant. Biol.* **43**: 409-417.

**Ivanoff, B. and Levine, M. (1997).** Typhoid fever: continuing challenges from a resilient bacterial foe. *Bull. Inst. Pasteur* **95(3)**: 129-142.

**Jaffe, A., Ogura, T. and Hiraga, S. (1985).** Effects of the *ccd* function of the F plasmid on bacterial growth. *J. Bacteriol.*, **163**: 841-849.

**Jakimowicz, D., Majka, J., Messer, W., Speck, C., Fernandez, M., Martin, M.C., Sanchez, J., Schauwecker, F., Keller, U., Schrempf, H., and Zakrzewska-Czerwinska, J. (1998).** Structural elements of the *Streptomyces oriC* region and their interactions with the DnaA protein. *Microbiol.* **144**: 1281-90.

**Johnson, R.C. and Simon, M.I. (1985).** Hin-mediated site-specific recombination requires two 26 bp recombination sites and a 60 bp recombinational enhancer. *Cell* **41(3)**:781-91.

**Kalkus, J., Reh, M. and Schlegel, H.G. (1990).** Hydrogen autotrophy of *Nocardia opaca* strains is encoded by linear megaplasmids. *J. Gen. Microbiol.* **136(6)**:1145-51.

**Kado, C.I. and Liu, S.T. (1981).** Rapid procedure for detection and isolation of large and small plasmids. *J. Bacteriol.* **145(3)**: 1365-73.

**Kay, E. R. M., Simmons, N. S., Dounce, A. L. (1952).** An Improved Preparation of Sodium Desoxyribonucleate. *J. Am. Chem. Soc.* **74 (7)**: 1724–1726

**Keen, C. L., Mendelovitz, S., Cohen, G., Aharonowitz, Y. and Roy, K. L. (1988).** Isolation and characterization of a linear DNA plasmid from *Streptomyces clavuligerus*. *Mol. Gen. Genet.* **212**:172–176

**Kelly, A. (2004).** A global role for Fis in the transcriptional control of metabolism and type III secretion in *Salmonella enterica* serovar Typhimurium. *Microbiol.* **150(7)**: 2037-53.

**Kendall, K.J. and Cohen, S.N. (1987).** Plasmid transfer in *Streptomyces lividans*: identification of a *kil-kor* system associated with the transfer region of pIJ101. *J. Bacteriol.* **169(9)**: 4177-83.

**Kennedy, C. K. (1971).** Induction of Colicin Production by High Temperature or Inhibition of Protein Synthesis. *J. Bacteriol.* **108**: 10-19

**Khan, S.A. (2005).** Plasmid rolling-circle replication: highlights of two decades of research. *Plasmid.* **53(2)**:126-36.

**Khan, S. A. (1997).** Rolling-circle replication of bacterial plasmids. *Microbiol. Mol. Biol. Rev.* **61(4)**: 442-55.

**Khan, S. A., Adler, G.K. and Novick, R.P. (1982).** Functional origin of replication of pT181 plasmid DNA is contained within a 168-base-pair segment. *Proc. Natl. Acad. Sci. USA* **79(15)**: 4580-4.



**Khan, S.A., Carleton, S.M. and Novick, R.P. (1981).** Replication of Plasmid pT181 DNA *In vitro*: Requirement for a Plasmid-Encoded Product. *Proc. Natl. Acad. Sci. USA* **78(8)**: 4902-4906.

**Khosravi, M. et al., (1990).** Variation of Oxygen Requirement with Plasmid Size in Recombinant *Escherichia coli*. *Plasmid*. **143**: 138-143.

**Kobayashi, I. (2001).** Behavior of restriction-modification systems as selfish mobile elements and their impact on genome evolution. *Nuc. Acids Res.* **29**: 3742-56.

**Kinashi, H., Doi, M. and Nimi, O. (1995).** Isolation of large linear plasmids from beta-lactam producing actinomycete strains. *Biotech. Lett.* **17(3)**: 243-246.

**Kinashi, H., Mori, E., Hatani, A. and Nimi, O. (1994).** Isolation and characterization of linear plasmids from lankacidin-producing *Streptomyces* species. *J. Antibiot.* (Tokyo) **47**:1447–1455.

**Kinashi, H. and Shimaji, M. (1987).** Detection of giant linear plasmids in antibiotic producing strains of *Streptomyces* by the OFAGE technique. *J. Antibiot.* **40**:913–916

**Kinashi, H., Shimaji, M. and Sakai, A. (1987).** Giant linear plasmids in *Streptomyces* which code for antibiotic biosynthesis genes. *Nature*. **328(6129)**:454-6.

**Kingsbury, D.T. and Helinski, D.R. (1970).** DNA polymerase as a requirement for the maintenance of the bacterial plasmid colicinogenic factor E1. *Biochem. Biophys. Res. Commun.* **41(6)**:1538-44.

**Kobryn, K. and Chaconas, G. (2005).** Fusion of hairpin telomeres by the *B. burgdorferi* telomere resolvase ResT implications for shaping a genome in flux. *Mol. Cell* **17(6)**: 783-91.

**Kobryn, K. and Chaconas, G. (2002).** ResT, a telomere resolvase encoded by the Lyme disease spirochete. *Mol. Cell* **9(1)**: 195-201.

**Koepsel, R.R., Murray, R.W. and Khan, S. A. (1986).** Sequence-specific interaction between the replication initiator protein of plasmid pT181 and its origin of replication. *Proc. Natl. Acad. Sci. USA* **83(15)**: 5484-8.

**Koepsel, R.R. and Khan, S.A. (1986).** Static and initiator protein-enhanced bending of DNA at a replication origin. *Science*. **233(4770)**: 1316-1318

**Koepsel, R.R., Murray, R.W., Rosenblum, W.D. and Khan, S.A. (1985).** The replication initiator protein of plasmid pT181 has sequence-specific endonuclease and topoisomerase-like activities. *Proc. Natl. Acad. Sci. USA* **82**: 6845-6849.

**Kogoma, T. (1984).** Absence of RNase H allows replication of pBR322 in *Escherichia coli* mutants lacking DNA polymerase I. *Proc. Natl. Acad. Sci. USA* **81(24)**: 7845-9.

**Kollek, R. Oertel, W. and Goebel, W. (1978).** Isolation and characterization of the minimal fragment required for autonomous replication ("basic replicon") of a copy mutant (pKN102) of the antibiotic resistance factor R1. *Mol. Gen. Genet.* **162(1)**:51-7

**Kraus, E. and Femfert, U. (1976)** Proteinase K from the mould *Tritirachium album* Limber. Specificity and mode of action. *Hoppe. Seylers. Z. Physiol. Chem.* **357(7)**: 937-47.

**Kues, U. and Stahl, U. (1989).** Replication of plasmids in gram-negative bacteria. *Microbiol. Rev.* **53(4)**: 491-516.

**Kumar, C.C. and Novick, R.P. (1985).** Plasmid pT181 Replication is Regulated by Two Countertranscripts. *Proc. Natl. Acad. Sci. USA* **82(3)**: 638-642.

**Kutsukake, K. Nakashima, H. Tominaga, A. and Abo, T. (2006).** Two DNA Invertases Contribute to Flagellar Phase Variation in *Salmonella enterica* Serovar Typhimurium Strain LT2. *J. Bacteriol.* **188(3)**: 950-957.

**Kyslik, P., Dobisova, M., Maresova, H. and Sobotkova, L. (1993).** Plasmid burden in chemostat culture of *Escherichia coli*: Its effect on the selection for overproducers of host enzymes. *Biotechnol. Bioeng.* **41**: 325-329

**Lee, K. and Cohen, S.N. (2003).** A *Streptomyces coelicolor* functional orthologue of *Escherichia coli* RNase E shows shuffling of catalytic and PNPase-binding domains. *Mol. Microbiol.* **48(2)**: 349-60.

**Lin, Y.S., Kieser, H.M., Hopwood, D.A. and Chen, C.W. (1993).** The chromosomal DNA of *Streptomyces lividans* 66 is linear. *Mol. Microbiol.* **10(5)**: 923-33.

**Lin-Chao, S. and Cohen, S.N. (1991).** The rate of processing and degradation of antisense RNAI regulates the replication of ColE1-type plasmids *in vivo*. *Cell.* **65(7)**:1233-42.

**Lobocka, M., and Yarmolinsky, M. (1996).** P1 plasmid partition: a mutational analysis of ParB. *J. Mol. Biol.* **259**: 366-82.

**Lopilato J, Bortner S, Beckwith J. (1986).** Mutations in a new chromosomal gene of *Escherichia coli* K-12, *pcnB*, reduce plasmid copy number of pBR322 and its derivatives. *Mol. Gen. Genet.* **205(2)**: 285-90.

**Ludtke, D. N., Eichorn, B. G. and Austin, S. J. (1989).** Plasmid-partition functions of the P7 prophage. *J. Mol. Biol.* **209(3)**: 393-406

**Marko, A.M. and Butler, G. C. (1951).** The isolation of sodium desoxyribonucleate with sodium dodecyl sulfate. *J. Biol. Chem.* **190(1)**: 165-76.

**Martin, K.A., Friedman, S.A. and Austin, S.J. (1987).** The partition site of the P 1 plasmid. *Proc. Natl. Acad. Sci. U.S.A.* **84**: 8544-8547

**Masukata, H. and Tomizawa, J. (1990).** A mechanism of formation of a persistent hybrid between elongating RNA and template DNA. *Cell* **62(2)**: 331-8.

**Masukata, H. and Tomizawa, J.I. (1984).** Effects of point mutations on formation and structure of the RNA primer for ColE1 DNA replication. *Cell* **36(2)**:513-522.

**McLeod, M.P., Warren, R.L., Hsiao, W.W., Araki, N., Myhre, M., Fernandes, C., Miyazawa, D., Wong, W., Lillquist, A.L., Wang, D., Dosanjh, M., Hara, H., Petrescu, A., Morin, R.D., Yang, G., Stott, J.M., Schein, J.E., Shin, H., Smailus, D., Siddiqui, A.S., Marra, M.A., Jones, S.J., Holt, R., Brinkman, F.S., Miyauchi, K., Fukuda, M., Davies, J.E., Mohn, W.W. and Eltis, L.D. (2006).** The complete genome of *Rhodococcus* sp. RHA1 provides insights into a catabolic powerhouse. *Proc. Natl. Acad. Sci. USA* **103**: 15582-15587.

**Morita, M. and Oka, A. (1979).** The structure of a transcriptional unit on colicin E1 plasmid. *Eur. J. Biochem.* **97(2)**: 435-43.

**Mukhopadhyay, S. and Chattoraj, D.K. (2000).** Replication-induced transcription of an autorepressed gene: the replication initiator gene of plasmid P1. *Proc. Natl. Acad. Sci. USA* **97(13)**: 7142-7.

**Mukhopadhyay, G. and Chattoraj, D.K. (1993).** Conformation of the Origin of P1 Plasmid Replication. *J. Mol. Biol.* **231(1)**: 19-28.

**Naito, S., Kitani, T., Ogawa, T., Okazaki, T., and Uchida, H. (1984).** *Escherichia coli* mutants suppressing replication-defective mutations of the ColE1 plasmid. *Proc. Natl. Acad. Sci. USA* **81(2)**: 550-4.

**Noirot, P., Bargonetti, J. and Novick, R.P. (1990).** Initiation of rolling-circle replication in pT181 plasmid: initiator protein enhances cruciform extrusion at the origin. *Proc. Natl. Acad. Sci. USA* **87(21)**: 8560-4.

**Nomura, M. (1964).** Mechanism of action of colicines. *Proc Natl Acad Sci U S A.* **52**: 1514-21.

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

**Novick, R.P., Iordanescu, S., Projan, S.J., Kornblum, J. and Edelman, I. (1989).** pT181 plasmid replication is regulated by a countertranscript-driven transcriptional attenuator. *Cell* **59(2)**: 395-404.

**Novick, R.P. (1987).** Plasmid incompatibility. *Microbiol. Rev.* **51(4)**: 381-95.

**Novick, R. P., Adler, G. K., Majumder, S., Khan, S. A., Carleton, S., Rosenblum, W. D. and Iordanescu, S. (1982).** Coding sequence for the pT181 *repC* product: a plasmid-coded protein uniquely required for replication. *Proc. Natl. Acad. Sci. USA* **79(13)**: 4108-12.

- Ogura, T., Niki, H., Kano, Y., Imamoto, F. and Hiraga, S. (1990).** Maintenance of plasmids in HU and IHF mutants of *Escherichia coli*. *Mol. Gen. Genet.* **220(2)**:197-203.
- Pal, S.K. and Chattoraj, D.K. (1988).** P1 plasmid replication: initiator sequestration is inadequate to explain control by initiator-binding sites. *J. Bacteriol.* **170(8)**: 3554-60.
- Pang, T., Bhutta, Z.A., Finlay, B.B., Altwegg, M. (1995).** Typhoid fever and other Salmonellosis: a continuing challenge. *Trends Microbiol.* **3(7)**: 253-255.
- Park, K., Han, E., Paulsson, J., and Chattoraj, D.K. (2001).** Origin pairing ('handcuffing') as a mode of negative control of P1 plasmid copy number. *EMBO J.* **20(24)**: 7323-32.
- Parry, C.M., Hien, T.T., Dougan, G., White, N.J. and Farrar, J.J. (2002).** Typhoid fever. *N. Engl. J. Med.* **347(22)**: 1770-1782.
- Patient, M. E., and Summers, D.K. (1993).** ColE1 multimer formation triggers inhibition of *Escherichia coli* cell division. *Mol. Microbiol.* **9**: 1089-95.
- Phillips, G. and Funnell, B. E. (2004).** Plasmid Biology. American Society of Microbiology.
- Picardeau, M., Le Dantec, C. and Vincent, V. (2000).** Analysis of the internal replication region of a mycobacterial linear plasmid. *Microbiol.* **146 (2)**: 305-13.
- Picardeau, M., Lobry, J.R. and Hinnebusch, B.J. (1999).** Physical mapping of an origin of bidirectional replication at the centre of the *Borrelia burgdorferi* linear chromosome. *Mol. Microbiol.* **32(2)**: 437-45.
- Picardeau, M. and Vincent, V. (1998).** Mycobacterial linear plasmids have an invertron-like structure related to other linear replicons in actinomycetes. *Microbiol.* **144**: 1981-1988
- Platt, T. (1986).** Transcription termination and the regulation of gene expression. *Annu. Rev. Biochem.* **55**: 339-372

**Podkovyrov, S.M. and Larson, T.J. (1995).** A new vector-host system for construction of lacZ transcriptional fusions where only low-level gene expression is desirable. *Gene*. **156(1)**:151-2.

**Pogliano, J., Ho, T.Q., Zhong, Z. and Helinski, D.R. (2001).** Multicopy plasmids are clustered and localized in *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S.A.* **98(8)**:4486-91

**Qin, Z., Shen, M. and Cohen, S.N. (2003).** Identification and Characterization of a pSLA2 Plasmid Locus Required for Linear DNA Replication and Circular Plasmid Stable Inheritance in *Streptomyces lividans*. *J. Bacteriol.* **185(22)**: 6575-6582.

**Qin, Z. and Cohen, S.N. (1998).** Replication at the telomeres of the *Streptomyces* linear plasmid pSLA2. *Mol. Microbiol.* **28(5)**: 893-903.

**Rasmussen, P. B., Gerdes, K. and Mlolin, S. (1987).** Genetic analysis of the *parB*+ locus of plasmid R1. *Mol. Gen. Genet.* **209**: 122-128.

**Rasooly, A., Projan, S.J. and Novick, R.P. (1994a).** Plasmids of the pT181 family show replication-specific initiator protein modification. *J. Bacteriol.* **176(8)**: 2450-3.

**Rasooly, A., Wang, P.Z. and Novick, R.P. (1994b).** Replication-specific conversion of the *Staphylococcus aureus* pT181 initiator protein from an active homodimer to an inactive heterodimer. *EMBO J.* **13(21)**: 5245-5251.

**Rasooly, A. and R. P. Novick. (1993).** Replication-specific inactivation of the pT181 plasmid initiator protein. *Science* **262**:1048–1050

**Redenbach, M., Bibb, M., Gust, B., Seitz, B. and Szychaj, A. (1999).** The linear plasmid SCP1 of *Streptomyces coelicolor* A3(2) possesses a centrally located replication origin and shows significant homology to the transposon Tn4811. *Plasmid* **42(3)**: 174-85.

**Richardson, L. V. and Richardson, J. P. (1996).** Rho-dependent termination of transcription is governed primarily by the upstream Rho utilization (rut) sequences of a terminator. *J. Biol. Chem.* **271**: 21597–21603.

**Rohe, M., Schrage, K. and Meinhardt, F. (1991).** The linear plasmid pMC3-2 from *Morchella conica* is structurally related to adenoviruses. *Curr. Genet.* **20**: 527-533.

**Rosado, A. and Seldin, L. (1993).** Isolation and partial characterization of a new linear DNA plasmid isolated from *Bacillus polymyxa* SCE2. *J. Gen. Microbiol.* **139**: 1277-1282.

**Rush, M.G., Gordon, C.N., Novick, R.P. and Warner, R.C. (1969).** Penicillinase plasmid DNA from *Staphylococcus aureus*. *Proc. Natl. Acad. Sci. U.S.A.* **63(4)**:1304-10.

**Ryan, W., Parulekar, S.J. and Stark, B.C. (1989).** Expression of beta-lactamase by recombinant *Escherichia coli* strains containing plasmids of different sizes-effects of pH, phosphate, and dissolved oxygen. *Biotechnol. Bioeng.* **34(3)**: 309-19.

**Rybchin, V.N. and Svarchevsky, A. N. (1999).** The plasmid prophage N15: a linear DNA with covalently closed ends. *Mol. Microbiol.* **33(5)**: 895-903.

**Saeki, H., Akira, M., Furuhashi, K., Averhoff, B. and Gottschalk, G. (1999).** Degradation of trichloroethene by a linear-plasmid-encoded alkene monooxygenase in *Rhodococcus corallinus* (*Nocardia corallina*) B-276. *Microbiol.* **145(7)**: 1721-30.

**Sakakibara, Y. and Tomizawa, J.I. (1974).** Termination point of replication of colicin E1 plasmid DNA in cell extracts. *Proc. Natl. Acad. Sci. USA* **71(12)**: 4935-9.

**Smith, C. (2007).** Cloning and mutagenesis: tinkering with the order of things. *Nat. Methods* **4**: 455-461

**Sambrook, J. and Russell, D. (2001).** *Molecular Cloning: A Laboratory Manual* (3rd ed.). Cold Spring Harbor Laboratory Press.

**Schrempf, H., Bujard, H., Hopwood, D.A., Goebel, W. (1975).** Isolation of covalently closed circular deoxyribonucleic acid from *Streptomyces coelicolor* A3(2). *J. Bacteriol.* **121(2)**:416-21.

**Schwan, T.G., Burgdorfer, W. and Garon, C.F. (1988).** Changes in infectivity and plasmid profile of the Lyme disease spirochete, *Borrelia burgdorferi*, as a result of *in vitro* cultivation. *Infect Immun.* **56(8)**:1831-6.

**Schwartz, D.C. and Cantor, C.R. (1984).** Separation of yeast chromosome-sized DNAs by pulsed field gradient gel electrophoresis. *Cell.* **37(1)**:67-75.

**Selzer, G., Som, T., Itoh, T., and Tomizawa, J. (1983).** The origin of replication of plasmid p15A and comparative studies on the nucleotide sequences around the origin of related plasmids. *Cell* **32(1)**: 119-29.

**Selzer, G. and Tomizawa, J.I. (1982).** Specific cleavage of the p15A primer precursor by ribonuclease H at the origin of DNA replication. *Proc. Natl. Acad. Sci. USA* **79(23)**: 7082-6.

**Sengupta, M., Nielsen, H.J., Youngren, B., and Austin, S. (2010).** P1 plasmid segregation: accurate redistribution by dynamic plasmid pairing and separation. *J Bacteriol.* **192**: 1175-83.

**Seo, J.H. and Bailey, J.E. (1985).** Effects of recombinant plasmid content on growth properties and cloned gene product formation in *Escherichia coli*. *Biotechnol. Bioeng.* **27(12)**:1668-74.

**Shiffman, D. and Cohen, S.N. (1992).** Reconstruction of a *Streptomyces* linear replicon from separately cloned DNA fragments: existence of a cryptic origin of circular replication within the linear plasmid. *Proc. Natl. Acad. Sci. U. S. A.* **89(13)**: 6129-33.

**Simon, M. Zieg, J. Silverman, M. Mandel, G. and Doolittle. (1980).** Phase variation: evolution of a controlling element. *Science* **209(4463)**: 1370-1374

**Stein, D. and Cohen, S.N. (1989).** A cloned regulatory gene of *Streptomyces lividans* can suppress the pigment deficiency phenotype of different developmental mutants. *J. Bacteriol.* **171(4)**: 2258-61.



**Sternberg, N. and Austin, S. (1981).** The maintenance of the P1 plasmid prophage. *Plasmid* **5(1)**: 20-31.

**Sternberg, N. and Hamilton, D. (1981).** Bacteriophage P1 site-specific recombination. I. Recombination between *loxP* sites. *J. Mol. Biol.* **150(4)**:467-86.

**Stewart, P.E., Chaconas, G. and Rosa, P. (2003).** Conservation of Plasmid Maintenance Functions between Linear and Circular Plasmids in *Borrelia burgdorferi*. *J. Bacteriol.* **185(10)**: 3202-3209.

**Stirling, C. J., Colloms, S.D., Collins, J.F., Szatmari, G., and Sherratt, D.J. (1989).** *xerB*, an *Escherichia coli* gene required for plasmid ColE1 site-specific recombination, is identical to *pepA*, encoding aminopeptidase A, a protein with substantial similarity to bovine lens leucine aminopeptidase. *EMBO J.* **8**: 1623-7.

**Stirling, C. J., Szatmari, G., Stewart, G., Smith, M.C., and Sherratt, D.J. (1988).** The arginine repressor is essential for plasmid-stabilizing site-specific recombination at the ColE1 *cer* locus. *EMBO J.* **7**: 4389-95.

**Summers, D. K. (1996).** The biology of plasmids, Blackwell Science.

**Summers, D. K., Beton, C.W., and Withers, H.L. (1993).** Multicopy plasmid instability: the dimer catastrophe hypothesis. *Mol. Microbiol.* **8**: 1031-8.

**Summers, D. K. (1991).** The kinetics of plasmid loss. *Trends Biotechnol.* **9**: 273-8.

**Summers, D. K. and Sherratt, D.J. (1988).** Resolution of ColE1 dimers requires a DNA sequence implicated in the three-dimensional organization of the *cer* site. *EMBO J.* **7**: 851-8.

**Summers, D. K., and Sherratt, D.J. (1984).** Multimerization of high copy number plasmids causes instability: ColE1 encodes a determinant essential for plasmid monomerization and stability. *Cell* **36**: 1097-103.

**Surtees, J. A., and Funnell, B.E. (2003).** Plasmid and chromosome traffic control: how ParA and ParB drive partition. *Curr. Top. Dev. Biol.* **56**: 145-80.

**Surtees, J. A., and Funnell, B.E. (2001).** The DNA binding domains of P1 ParB and the architecture of the P1 plasmid partition complex. *J. Biol. Chem.* **276**: 12385-94.

**Taghavi, S., van der Lelie, D., Mergeay, M. (1994).** Electroporation of *Alcaligenes eutrophus* with (mega) plasmids and genomic DNA fragments. *Appl. Environ. Microbiol.* **60(10)**: 3585-91.

**Tamura, K., Sakazaki, R., Kuramochi, S. and Nakamura, A. (1988).** Occurrence of H-antigen z66 of R phase in cultures of *Salmonella* serovar typhi originated from Indonesia. *Epidem. Inf.* **101**: 311 -314.

**te Riele, H., Michel, B. and Ehrlich, S. D. (1986).** Are single-stranded circles intermediate in plasmid DNA replication? *EMBO J.* **5**:631–637

**Thisted, T., Sørensen, N.S. and Gerdes, K. (1995).** Mechanism of post-segregational killing: secondary structure analysis of the entire Hok mRNA from plasmid R1 suggests a fold-back structure that prevents translation and antisense RNA binding. *J. Mol.Biol.* **247(5)**: 859-73.

**Thisted, T., Sorensen, N. S., Gerhart, E., Wagner, H. and Gerdes, K. (1994).** Mechanism of post-segregational killing: Sok antisense RNA interacts with Hok mRNA via its 5'-end single-stranded leader and competes with the 3'-end of Hok mRNA for binding to the mok translational initiation region. *EMBO J.* **13**: 1960- 1968.

**Thisted, T. and Gerdes, K. (1992).** Mechanism of Post-segregational Killing by the *hok/sok* System of Plasmid R1. *J. Mol. Biol.* **223**: 41-54.

**Thomas, C. D., Balsonj, D.F. and Shawn, W. V. (1990).** *In Vitro* Studies of the Initiation of *Staphylococcal* Plasmid Replication. *J. Biol. Chem.* **265**: 5519-5530.

**Thomas, C. D., Balson, D. F. and Shaw, W. V. (1988).** Identification of the tyrosine residue involved in bond formation between replication origin and the initiator protein of plasmid pC221. *Biochem. Soc. Trans.* **16**:758–759

**Tomcsányi, T. and Apirion, D. (1985).** Processing enzyme ribonuclease E specifically cleaves RNA I. An inhibitor of primer formation in plasmid DNA synthesis. *J. Mol. Biol.* **185(4)**:713-20.

**Tomizawa, J. (1990).** Control of ColE1 plasmid replication. Interaction of Rom protein with an unstable complex formed by RNA I and RNA II. *J Mol. Biol.* **212(4)**:695-708.

**Tomizawa, J.I. and Masukata, H. (1987).** Factor-Independent Termination in a Stretch of Deoxyadenosine in the Template DNA of Transcription Residues. *Cell* **51**: 623-630.

**Tomizawa, J.I. (1986).** Control of ColE1 plasmid replication: Binding of RNA I to RNA II and inhibition of primer formation. *Cell* **47(1)**: 89-97.

**Tomizawa, J.I. (1984).** Control of ColE1 Plasmid Replication: The Process of Binding of RNA I to the Primer Transcript. *Cell* **38**: 861-870.

**Tomizawa, J. I. and Som, T. (1984).** Control of ColE1 plasmid replication: Enhancement of binding of RNA I to the primer transcript by the rom protein. *Cell* **38(3)**: 871-878.

**Tomizawa, J.-ichi and Itoh, T. (1982).** The importance of RNA secondary structure in ColE1 primer formation. *Cell* **31(3)**: 575-583.

**Tomizawa, J., Itoh, T., Selzer, G., and Som, T. (1981).** Inhibition of ColE1 RNA primer formation by a plasmid-specified small RNA. *Proc. Natl. Acad. Sci. USA* **78(3)**: 1421-5.

**Tomizawa, J.I., Ohmori, H. and Bird, R.E. (1977).** Origin of replication of colicin E1 plasmid DNA. *Proc. Natl. Acad. Sci. USA* **74(5)**: 1865-9.

**Tsai, H. H., Huang, C. H., Lin, A. M. and Chen, C. W. (2008).** Terminal proteins of *Streptomyces* chromosome can target DNA into eukaryotic nuclei. *Nuc. Acids Res.* **36**: e62.

**Tsuchimoto, S., Ohtsubo, H. and Ohtsubo, E. (1988).** Two genes, *pemK* and *pemI*, responsible for stable maintenance of resistance plasmid R100. *J. Bacteriol.* **170(4)**:1461-6.

**Tsutsui, H., Fujiyama, A., Murotsu, T. and Matsubara, K. (1983).** Role of nine repeating sequences of the mini-F genome for expression of F-specific incompatibility phenotype and copy number control. *J. Bacteriol.* **155(1)**: 337-44.

**Walz, D. and Caplan, S.R. (2000).** An electrostatic mechanism closely reproducing observed behaviour in the bacterial flagellar motor. *Biophys. J.* **78(2)**: 626-51.

**Wang, P.Z., Projan, S.J., Henriquez, V. and Novick, R.P. (1992).** Specificity of origin recognition by replication initiator protein in plasmids of the pT181 family is determined by a six amino acid residue element. *J. Mol. Biol.* **223(1)**: 145-58.

**Warren, R., Hsiao, W.W., Kudo, H., Myhre, M., Dosanjh, M., Petrescu, A., Kobayashi, H., Shimizu, S., Miyauchi, K., Masai, E., Yang, G., Stott, J.M., Schein, J.E., Shin, H., Khattra, J., Smailus, D., Butterfield, Y.S., Siddiqui, A., Holt, R., Marra, M.A., Jones, S.J., Mohn, W.W., Brinkman, F.S., Fukuda, M., Davies, J., Eltis, L.D. (2004).** Functional Characterization of a Catabolic Plasmid from Polychlorinated- biphenyl-degrading *Rhodococcus* sp. strain RHA1. *J. Bacteriol.* **186(22)**: 7783-7795.

**Weitao, T., Dasgupta, S., and Nordstrom, K. (2000).** Plasmid R1 is present as clusters in the cells of *Escherichia coli*. *Plasmid* **43**: 200-4.

**Wickner, S., Hoskins, J., Chatteraj, D.K., and McKenney, K. (1990).** Deletion analysis of the mini-P1 plasmid origin of replication and the role of *Escherichia coli* DnaA protein. *J. Biol. Chem.*: **265(20)**: 11622-7.

**Wickner, S.H. and Chatteraj, D.K. (1987).** Replication of mini-P1 plasmid DNA *in vitro* requires two initiation proteins, encoded by the *repA* gene of phage P1 and the *dnaA* gene of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **84(11)**: 3668-72.

**Wilson, K. S. & von Hippel, P. H. (1995).** Transcription termination at intrinsic terminators: the role of the RNA hairpin. *Proc. Natl Acad. Sci. USA* **92**: 8793-8797

**Wong, E.M. and Polisky, B. (1985).** Alternative conformations of the ColE1 replication primer modulate its interaction with RNA I. *Cell* **42(3)**: 959-966.

**Wu, W., Leblanc, S.K., Piktel, J., Jensen, S.E. and Roy, K.L. (2006).** Prediction and functional analysis of the replication origin of the linear plasmid pSCL2 in *Streptomyces clavuligerus* *Can. J. Microbiol.* **52**:293-300

**Wu, X. and Roy, K.L. (1993).** Complete nucleotide sequence of a linear plasmid from *Streptomyces clavuligerus* and characterization of its RNA transcripts. *J.Bact.* **175(1)**: 37-52.

**Xu, M., Zhu, Y., Zhang, R., Shen, M., Jiang, W., Zhao, G. and Qin, Z. (2006).** Characterization of the genetic components of *Streptomyces lividans* linear plasmid SLP2 for replication in circular and linear modes. *J. Bacteriol.* **188(19)**: 6851-7.

**Yamamoto, S. and Kutsukake, K. (2006).** FlhT acts as an anti-FlhD2C2 factor in the transcriptional control of the flagellar regulon in *Salmonella enterica* serovar typhimurium. *J. Bacteriol.* **188(18)**: 6703-8.

**Yanisch-Perron, C., Vieira, J. and Messing, J. (1985).** Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**: 103-119.

**Zakrzewska-Czerwińska, J. and Schrempf, H. (1992).** Characterization of an autonomously replicating region from the *Streptomyces lividans* chromosome. *J Bacteriol.* **174(8)**: 2688-93.

**Zavitz, K.H. and Marians, K.J. (1991).** Dissecting the functional role of PriA protein-catalysed primosome assembly in *Escherichia coli* DNA replication. *Mol. Microbiol.* **5(12)**: 2869-2873.

**Zhang, R., Peng, S. and Qin, Z. (2010).** Two internal origins of replication in *Streptomyces* linear plasmid pFRL1. *Appl. Environ. Microbiol.* **76(17)**: 5676-83.

**Zhang, R., Xia, H., Guo, P. and Qin, Z. (2009).** Variation in the replication loci of *Streptomyces* linear plasmids. *FEMS Microbiol. Lett.* **290(2)**: 209-16.

**Zhao, A. C., Ansari, R. A., Schmidt, M. C. and Khan, S. A. (1998).** An Oligonucleotide Inhibits Oligomerization of a Rolling Circle Initiator Protein at the pT181 Origin of Replication. *J. Biol. Chem.* **273**: 16082–16089.

**Zieg, J. and Simon, M. (1980).** Analysis of the nucleotide sequence of an invertible controlling element. *Proc. Natl. Acad. Sci. USA* **77(7)**: 4196-200.

**Zieg, J. Silverman, M. Hilmen, M. and Simon, M. (1977).** Recombinational switch for gene expression. *Science*. **196(4286)**: 170-172