http://researchspace.auckland.ac.nz

ResearchSpace@Auckland

Copyright Statement

The digital copy of this thesis is protected by the Copyright Act 1994 (New Zealand).

This thesis may be consulted by you, provided you comply with the provisions of the Act and the following conditions of use:

- Any use you make of these documents or images must be for research or private study purposes only, and you may not make them available to any other person.
- Authors control the copyright of their thesis. You will recognise the author's right to be identified as the author of this thesis, and due acknowledgement will be made to the author where appropriate.
- You will obtain the author's permission before publishing any material from their thesis.

To request permissions please use the Feedback form on our webpage.
http://researchspace.auckland.ac.nz/feedback

General copyright and disclaimer

In addition to the above conditions, authors give their consent for the digital copy of their work to be used subject to the conditions specified on the Library Thesis Consent Form.
Such laboured nothings, in so strange a style,  
Amaze th' unlearn'd, and make the learned smile.

Alexander Pope, 1688-1744
Clepsydra

Water from a stream (A) is fed into an upper reservoir (B) and maintained at a constant pressure head by a small over-fill pipe (C). The water flows into the main reservoir (D) which slowly raises a float (E) carrying a pointer (F). The pointer marks the time on a vertical scale (G) of twelve hours. Once a cycle has been completed, the main reservoir is drained through a siphon tube (H).
Molecular and genetic analysis of RepA from the P307 RepFIB replicon

Andrew J. Spiers

Cellular and Molecular Biology, School of Biological Sciences

This thesis is submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

December 1992
ACKNOWLEDGMENTS

I would like to express my appreciation to my supervisor, Professor Peter Bergquist, for his advice and guidance throughout the course of this degree. I would like to extend my appreciation to Drs. David Christie and David Palmer for their interest, help and encouragement, especially in the biochemical analysis of RepA. My thanks also to Penelope Lawson and Nila Bhana for the technical assistance they provided from time to time. Finally, I would like to express my deep gratitude to my parents for their love and their support of my perpetual studentship.
# SHORT CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td></td>
<td>i</td>
</tr>
<tr>
<td>Contents</td>
<td></td>
<td>ii</td>
</tr>
<tr>
<td><strong>CHAPTER ONE</strong></td>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td><strong>CHAPTER TWO</strong></td>
<td>Active promoter elements in the E11 fragment</td>
<td>21</td>
</tr>
<tr>
<td><strong>CHAPTER THREE</strong></td>
<td>The identification and control of the repA promoter</td>
<td>29</td>
</tr>
<tr>
<td><strong>CHAPTER FOUR</strong></td>
<td>In vivo repression of three promoters by RepA</td>
<td>43</td>
</tr>
<tr>
<td><strong>CHAPTER FIVE</strong></td>
<td>The translational start codon of repA</td>
<td>59</td>
</tr>
<tr>
<td><strong>CHAPTER SIX</strong></td>
<td>RepA DNA-binding in vitro</td>
<td>67</td>
</tr>
<tr>
<td><strong>CHAPTER SEVEN</strong></td>
<td>The functional relevance of the DNA repeat groups</td>
<td>79</td>
</tr>
<tr>
<td><strong>CHAPTER EIGHT</strong></td>
<td>Review and Discussion</td>
<td>91</td>
</tr>
<tr>
<td><strong>APPENDIX ONE</strong></td>
<td>Protocols</td>
<td>105</td>
</tr>
<tr>
<td><strong>APPENDIX TWO</strong></td>
<td>Plasmids</td>
<td>121</td>
</tr>
<tr>
<td><strong>APPENDIX THREE</strong></td>
<td>Sequence analysis of the E11 fragment from P307</td>
<td>133</td>
</tr>
<tr>
<td><strong>APPENDIX FOUR</strong></td>
<td>Attempts to inhibit RepA-DNA binding in vitro</td>
<td>151</td>
</tr>
<tr>
<td><strong>APPENDIX FIVE</strong></td>
<td>Evidence for a ccd-like function</td>
<td>155</td>
</tr>
<tr>
<td><strong>APPENDIX SIX</strong></td>
<td>Additional reading: replicon genetics</td>
<td>163</td>
</tr>
<tr>
<td><strong>APPENDIX SEVEN</strong></td>
<td>Papers published or submitted for publication</td>
<td>199</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>203</td>
</tr>
</tbody>
</table>
The work in this Thesis concerns the replication control system of the P307 plasmid RepFIB replicon. The basic replicon occupies = 1.6kb of DNA and contains a single large open reading frame (repA) flanked on either side by a series DNA repeat elements. The organisational structure of the replicon has placed RepFIB into the Step function class of replicons. The placement of RepFIB within this group, as well as a strong homology between RepFIB and mini-P1, has resulted in a series of predictions concerning the control elements of RepFIB replication. The aim of this work was to test some of these predictions and to characterise the fundamental control elements utilised by the replicon.

This Thesis describes three different active promoter elements found embedded within the repeat elements flanking repA. Although the functional significance of two of the promoters is unknown (oriP and EFp), the third is responsible for the expression of RepA and has been designated 'repAp'. All three promoters are sensitive to RepA in trans, demonstrating that repA is autoregulated and that RepA is a DNA-binding protein capable of recognising copies of the repeat elements. RepA DNA-binding has also been demonstrated in vitro using a modification of the Western analysis technique (referred to a ‘Western-DNA’) in order to complement the in vivo experimental results. Although the coding region of repA had been determined in earlier work, the identification of the translational start codon was uncertain. This uncertainty has been resolved by limited N-terminal sequence analysis of a RepA:β-galactosidase fusion protein which has demonstrated that translation begins from a CTG codon located upstream of the predicted start sites. Finally, a series of genetic experiments have been used to determine the functional significance of RepA binding to the repeat elements. The repeat group upstream of repA are involved in autoregulation and also form part of the origin of replication, whilst the downstream repeats appear to be involved in the sensing and setting of plasmid copy number.

Although the work presented in this Thesis does not directly test the applicability for RepFIB of various control models proposed to explain the behaviour of Step function replicons, the nature and type of control elements identified in RepFIB support the placement of RepFIB within the Step function class. As a result of this work, it is clear that RepFIB is confronted by the same kind of control paradox faced by replicons such as mini-P1 and mini-F. All three replicons use autoregulation and titration to control the supply of initiator protein required for replication. However, concurrent autoregulation and titration appear to be incompatible in current control models, and the identification of both mechanisms in these replicons has lead to a control paradox. Some of the results presented here suggest potentially valuable avenues of future research which may help resolve the paradox faced by RepFIB and other Step function replicons.
# CONTENTS

Abstract

Contents
List of figures
List of tables
List of abbreviations and symbols
List of company abbreviations
Notes

<table>
<thead>
<tr>
<th>CHAPTER ONE</th>
<th>Introduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Plasmid genetics</td>
<td>1</td>
</tr>
<tr>
<td>1.1.1 The plasmid replicon</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Analysis of plasmid replicons</td>
<td>5</td>
</tr>
<tr>
<td>1.2.1 Division of replicon types into the hyperbolic and step function classes</td>
<td>5</td>
</tr>
<tr>
<td>1.2.2 Initiation of DNA synthesis</td>
<td>7</td>
</tr>
<tr>
<td>1.2.3 Control of the initiator</td>
<td>7</td>
</tr>
<tr>
<td>1.2.4 The control paradox</td>
<td>9</td>
</tr>
<tr>
<td>1.2.5 Possible solutions to the control paradox</td>
<td>9</td>
</tr>
<tr>
<td>1.2.6 Invoking host factors to solve the control paradox</td>
<td>11</td>
</tr>
<tr>
<td>1.2.7 New models which seek to overcome the control paradox</td>
<td>13</td>
</tr>
<tr>
<td>1.2.8 The steric hindrance model</td>
<td>13</td>
</tr>
<tr>
<td>1.2.9 The anti-parallel pairing model</td>
<td>13</td>
</tr>
<tr>
<td>1.2.10 Comparison of the two new models</td>
<td>13</td>
</tr>
<tr>
<td>1.3 The RepFIB replicon</td>
<td>15</td>
</tr>
<tr>
<td>1.3.1 The initial investigation of RepFIB</td>
<td>15</td>
</tr>
<tr>
<td>1.3.2 Sequence analysis of the EcoRI E11 fragment from P307</td>
<td>17</td>
</tr>
<tr>
<td>1.3.3 A simplistic RepFIB model</td>
<td>18</td>
</tr>
<tr>
<td>1.4 Aims of this thesis</td>
<td>19</td>
</tr>
</tbody>
</table>
CHAPTER TWO

Active promoter elements in the E11 fragment

2.1 Introduction 21
  2.1.1 Promoter elements of interest 21
  2.1.2 Sequence analysis of the E11 fragment 21

2.2 Promoter-probe plasmids used to identify active promoters 22

2.3 Identification of active promoters in the E11 fragment 23
  2.3.1 Expression of a RepA:ß-galactosidase fusion protein 23
  2.3.2 Expression of ORF-3:ß-galactosidase fusion proteins 24
  2.3.3 Expression of an ORF-4:ß-galactosidase fusion protein 25
  2.3.4 Expression of an ORF-5:ß-galactosidase fusion protein 25
  2.3.5 Expression of an ORF-7:ß-galactosidase fusion protein 25
  2.3.6 ORF-2 and ORF-6 26
  2.3.7 Promoters not associated with ORFs 26

2.4 Summary of active and inactive promoter elements identified in the E11 fragment 26

CHAPTER THREE

The identification and control of the repA promoter

3.1 Location of the repA promoter: sequence analysis and the protocol used to identify the promoter in vivo 29

3.2 Construction of the plasmids required to identify repA_p 31
  3.2.1 Initial promoter-probe plasmids 31
  3.2.2 Promoter-probe plasmids generated using PCR amplification 32

3.3 Analysis of the promoter-probe plasmids expressing the RepA:ß-galactosidase fusion protein 32

3.4 A test of repA autoregulation 34
  3.4.1 Design of the repA autoregulation assay 34
CHAPTER FOUR

In vivo repression of three promoters by RepA

4.1 Introduction
   4.1.1 Regulation of three promoters found within the minimal replicon
   4.1.2 Promoter-probe plasmids used to test the regulation of oriP, repA_p and EF_p

4.2 Demonstration of regulation: use of agar plate assays
   4.2.1 A description of the agar plate assays
   4.2.2 Repression of cm1 gene fusions by a RepFIB mini-plasmid
   4.2.3 Repression of β-galactosidase protein fusions by a RepFIB mini-plasmid
   4.2.4 A summary of the repression test using agar plate assays

4.3 Demonstration of regulation: use of liquid culture assays
   4.3.1 A description of the liquid culture assay
   4.3.2 Development of the liquid culture assay
   4.3.3 Repression of ori_p, repA_p and EF_p by RepA supplied by pAS60
   4.3.5 Summary of the repression tests using the liquid culture assay

4.4 RepA DNA-binding affinities for the different groups of repeat elements
   4.4.1 Introduction
   4.4.2 The three-plasmid assay
CHAPTER FIVE

The translational start codon of repA

5.1 Introduction
  5.1.1 Considerations concerning the N-terminal sequence determination of RepA
  5.1.2 Practical problems in the isolation of a particular protein
  5.1.3 The utility of β-galactosidase fusion proteins

5.2 Limited N-terminal sequence analysis of the RepA:β-galactosidase fusion protein
  5.2.1 Isolation of the RepA:β-galactosidase fusion protein
  5.2.2 Sequence analysis

5.3 The efficiency of initiation of translation from the CTG start codon of repA
  5.3.1 Introduction
  5.3.2 Alteration of the repA start codon by PCR mutagenesis
  5.3.3 A comparison of the level of expression of the RepA:β-galactosidase fusion protein from genes containing CTG and ATG start codons

5.4 Comment on the translational start codon used in the expression of repA

CHAPTER SIX

RepA DNA-binding in vitro

6.1 Introduction

6.2 The over-expression of RepA
  6.2.1 RepA expression plasmid
  6.2.2 Solubilisation of RepA
  6.2.3 RepA rabbit antiserum
CHAPTER SEVEN

The functional relevance of the DNA repeat groups

7.1 Functional significance of the repeat groups
   7.1.1 Predictions based on homology with step function replicons
   7.1.2 A summary of observations of incompatibility reactions

7.2 The RepFIB origin of replication
   7.2.1 Location of the origin of replication
   7.2.2 Host factors required for RepFIB replication

7.3 Control function of the EFGHIJ repeats
   7.3.1 Introduction
   7.3.2 Determination of the copy number of RepFIB mini-plasmids
   7.3.3 Comment on copy number determinations

7.4 Summary of the functional relevance of the DNA repeat groups

CHAPTER EIGHT

Review and Discussion

8.1 Review of experimental results
   8.1.1 Promoter sequences in the E11 fragment
   8.1.2 RepA repression of three distinct promoters
   8.1.3 The initiation codon used in the translation of repA
   8.1.4 RepA DNA-binding in vitro
8.1.5 Three-plasmid assay 93
8.1.6 The functional significance of RepA binding to the repeat elements 94

8.2 Summary of the research work 95

8.3 Discussion 96
8.3.1 The repA promoter 96
8.3.2 Initiation of translation from a CTG codon 97
8.3.3 RepA DNA-binding 98
8.3.5 Comparison of RepFIB and mini-P1 100

8.4 Future RepFIB research 102
8.4.1 Examination of the promoters 102
8.4.2 RepA DNA-binding 103
8.4.3 Similarities with mini-P1 103

8.5 Epilogue 104

APPENDIX ONE

A1.1 Bacterial strains 105
A1.2 Growth media 105
A1.2.1 Media 105
A1.2.2 Antibiotics 106
A1.2.3 β-galactosidase indicators 106
A1.2.4 Methylmethane sulphonate (MMS) 106

A1.3 Transformation 106
A1.3.1 Heat shock transformation 106
A1.3.2 DMSO transformation 107

A1.4 Isolation of plasmid DNA from bacterial cultures 107
A1.4.1 Mini-preparations of plasmid DNA 107
A1.4.2 Large-scale preparation of plasmid DNA 107
A1.4.3 CsCl purification of plasmid DNA 108

Protocols
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1.5</td>
<td>Isolation of DNA after electrophoresis &lt;br&gt; A1.5.1 DNA isolation by diffusion and phenol-extraction &lt;br&gt; A1.5.2 DNA isolation by melting and phenol-extraction</td>
<td>108</td>
</tr>
<tr>
<td>A1.6</td>
<td>Electrophoresis of DNA</td>
<td>108</td>
</tr>
<tr>
<td>A1.7</td>
<td>Manipulation of DNA &lt;br&gt; A1.7.1 Enzymes &lt;br&gt; A1.7.2 Polishing reactions &lt;br&gt; A1.7.3 Labelling DNA by nick-translation &lt;br&gt; A1.7.4 Labelling DNA by random-priming &lt;br&gt; A1.7.5 PCR synthesis of DNA</td>
<td>109</td>
</tr>
<tr>
<td>A1.8</td>
<td>Colony hybridisation</td>
<td>111</td>
</tr>
<tr>
<td>A1.9</td>
<td>Copy number determination</td>
<td>111</td>
</tr>
<tr>
<td>A1.10</td>
<td>Preparation of cell extracts &lt;br&gt; A1.10.1 8M urea extracts containing RepA &lt;br&gt; A1.10.2 Extracts containing the RepA:ß-galactosidase fusion protein &lt;br&gt; A1.10.3 Preparation of tubing for the dialysis of protein samples</td>
<td>112</td>
</tr>
<tr>
<td>A1.11</td>
<td>Cleavage of protein samples</td>
<td>113</td>
</tr>
<tr>
<td>A1.12</td>
<td>Protein concentration &lt;br&gt; A1.12.1 Precipitation with ammonium sulphate &lt;br&gt; A1.12.2 Precipitation with acetone &lt;br&gt; A1.12.3 Freeze drying</td>
<td>113</td>
</tr>
</tbody>
</table>
A1.14.6  Immunodetection using anti-β-galactosidase antibodies and an anti-mouse-HRP conjugate 115
A1.14.7  Immunodetection using RepA antiserum, anti-rabbit-Ig-biotin and a streptavidin-HRP complex 116
A1.14.8  Immunodetection using RepA antiserum and an ECL kit 116
A1.14.9  Limited N-terminal sequence analysis from PVDF membranes 116

A1.15  Western-DNA analysis of RepA DNA-binding 116

A1.16  Protein assays 117
A1.16.1  Estimation of the amount of total protein in samples 117
A1.16.2  Assay for β-galactosidase activity 117
A1.16.3  In vivo 14C-labelling of proteins 117

A1.17  RepA rabbit antiserum 118
A1.17.1  Preparation of RepA antigen 118
A1.17.2  Immunisation with RepA 118
A1.17.3  Preparation of the antiserum 118

A1.18  Immunoadfinity chromatography 118

A1.19  DNA sequence analysis 119

A1.20  Imaging techniques 119

**APPENDIX TWO**

**Plasmids**

A2.1  Plasmid used for cloning 121
A2.1.1  General plasmids used in this thesis 121
A2.1.2  Promoter-probe plasmids used in this thesis 121
A2.1.3  Expression plasmids used in this thesis 121

A2.2  Plasmids made for this thesis 123
      pAS1, pAS2, pAS4, pAS7, pAS9, pAS10 and pAS15 125
APPENDIX THREE

Sequence analysis of the E11 fragment from P307

A3.1 General features of the E11 EcoRI fragment
A3.2 Open reading frames
   A3.2.1 repA (ORF-1)
   A3.2.3 Other open reading frames
A3.3 DNA repeat elements
   A3.3.1 The A-K repeat elements
   A3.3.2 Other DNA repeat elements
A3.4 Sequences with homology to the E. coli consensus promoter sequence
A3.5 The RepF1B minimal replicon
A3.6 The RepF1B origin of replication
   A3.6.1 Functional single-strand initiation sequence
   A3.6.2 Sequences with homology to the dnaA box
   A3.6.3 Dam methylation sites
   A3.6.4 Sequences with homology to the oriC 13mer sequence
   A3.6.5 Sequences with homology to the primosome assembly site
   A3.6.6 Origin repeat elements
   A3.6.7 Origin AT and GC-rich sequences
A3.7 RepF1B and E11 fragment homology with other DNA
   A3.7.1 Homology between RepF1B and mini-P1 RepA proteins
   A3.7.2 ORF-3 homology to the resolvase gene of mini-F
A3.9  Phylogenetic analysis of the relationship between different RepFlB replicons  148

APPENDIX FOUR

Attempts to inhibit RepA-DNA binding \textit{in vitro}

A4.1  Introduction  151
A4.1  Antiserum inhibition of RepA DNA-binding  151
A4.1  Identification of the RepA DNA-binding domain  152

APPENDIX FIVE

Evidence for a ccd-like function

A5.1  A brief description of the ccd system  155
A5.2  Experimental evidence for the presence of a ccd-like system in the E11 fragment  156
A5.3  Probable location of the ccd-like system  156
A5.4  Attempts to complement the ccd-like function  161
A5.5  Summary of the analysis of the ccd-like function in the E11 fragment  161

APPENDIX SIX

Additional reading: replicon genetics

A6.1  Replication control of ColE1
   A6.1.1  Overview  163
   A6.1.2  Primer-RNA  163
   A6.1.3  RNA-I  163
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A6.1.4</td>
<td>RNA-1/ primer-RNA interaction and the involvement of Rom (Rop)</td>
<td>164</td>
</tr>
<tr>
<td>A6.1.5</td>
<td>DNA synthesis</td>
<td>166</td>
</tr>
<tr>
<td>A6.1.6</td>
<td>Summary of ColE1 replication control</td>
<td>167</td>
</tr>
<tr>
<td>A6.2</td>
<td>Replication control of mini-R1</td>
<td>168</td>
</tr>
<tr>
<td>A6.2.1</td>
<td>Overview</td>
<td>168</td>
</tr>
<tr>
<td>A6.2.2</td>
<td>Identification of the repressors</td>
<td>169</td>
</tr>
<tr>
<td>A6.2.3</td>
<td>Level one: CopA/RepA mRNA interaction</td>
<td>170</td>
</tr>
<tr>
<td>A6.2.4</td>
<td>Level two: CopB rescue by de-repression of repAp</td>
<td>171</td>
</tr>
<tr>
<td>A6.2.5</td>
<td>RepA</td>
<td>171</td>
</tr>
<tr>
<td>A6.2.6</td>
<td>Origin of replication</td>
<td>171</td>
</tr>
<tr>
<td>A6.2.7</td>
<td>Summary of mini-R1 replication control</td>
<td>172</td>
</tr>
<tr>
<td>A6.3</td>
<td>Replication control of pT181</td>
<td>172</td>
</tr>
<tr>
<td>A6.3.1</td>
<td>Overview</td>
<td>172</td>
</tr>
<tr>
<td>A6.3.2</td>
<td>CopA</td>
<td>173</td>
</tr>
<tr>
<td>A6.3.3</td>
<td>RepC</td>
<td>174</td>
</tr>
<tr>
<td>A6.3.4</td>
<td>CopA interaction with RepC</td>
<td>174</td>
</tr>
<tr>
<td>A6.3.5</td>
<td>DNA synthesis</td>
<td>175</td>
</tr>
<tr>
<td>A6.3.6</td>
<td>Summary of pT181 replication control</td>
<td>175</td>
</tr>
<tr>
<td>A6.4</td>
<td>Replication control of mini-F (RepFIA)</td>
<td>176</td>
</tr>
<tr>
<td>A6.4.1</td>
<td>Overview</td>
<td>176</td>
</tr>
<tr>
<td>A6.4.2</td>
<td>Mini-F</td>
<td>176</td>
</tr>
<tr>
<td>A6.4.3</td>
<td>The E gene</td>
<td>177</td>
</tr>
<tr>
<td>A6.4.4</td>
<td>Autoregulation of the E gene</td>
<td>178</td>
</tr>
<tr>
<td>A6.4.5</td>
<td>RNA polymerase-o^{32} transcription of pE</td>
<td>178</td>
</tr>
<tr>
<td>A6.4.6</td>
<td>E protein activities</td>
<td>179</td>
</tr>
<tr>
<td>A6.4.7</td>
<td>A brief summary of the mini-F replication control models</td>
<td>179</td>
</tr>
<tr>
<td>A6.4.8</td>
<td>Hierarchical binding</td>
<td>179</td>
</tr>
<tr>
<td>A6.4.9</td>
<td>Post-translational modification (the two-stage model)</td>
<td>180</td>
</tr>
<tr>
<td>A6.4.10</td>
<td>9kDa repressor action</td>
<td>181</td>
</tr>
<tr>
<td>A6.4.11</td>
<td>Steric hindrance</td>
<td>181</td>
</tr>
<tr>
<td>A6.4.12</td>
<td>Summary of mini-F replication control</td>
<td>183</td>
</tr>
<tr>
<td>A6.5</td>
<td>Replication control of mini-P1</td>
<td>184</td>
</tr>
</tbody>
</table>
A6.5.1 Overview 184
A6.5.2 Isolation of mini-P1 184
A6.5.3 Sequence analysis of mini-P1 184
A6.5.4 RepA 186
A6.5.5 The incA repeat elements 186
A6.5.6 Involvement of host factors in mini-P1 replication 187
A6.5.7 Origin of replication 188
A6.5.8 The mini-P1 control paradox 189
A6.5.9 The mini-Rts1 replicon 189
A6.5.10 Summary of mini-P1 and mini-Rts1 replication control 190

A6.6 Other plasmid replicons 191

A6.7 The DnaA protein and oriC replication 191
A6.7.1 DnaA involvement in the initiation of oriC replication 191
A6.7.2 DnaA in the cell cycle 191
A6.7.3 Physical role of DnaA at oriC 192
A6.7.4 A second function of DnaA DNA-binding 195
A6.7.5 DnaA-independent initiation of chromosomal replication 195
A6.7.6 Summary of DnaA involvement in chromosomal replication 196

A6.8 DnaA involvement in plasmid replication 196
A6.8.1 DnaA dependent and independent replicons 196
A6.8.2 Summary of DnaA involvement in plasmid replication 198

APPENDIX SEVEN Papers published or submitted for publication

Nucleotide sequence and replication characteristics of RepFIB, a basic replicon of IncF plasmids 199

Expression and regulation of the RepA protein of the RepFIB replicon from plasmid P307 199

RepFIB: a basic replicon of large plasmids 200
Regulatory interactions between RepA, an essential replication protein, and the DNA repeats of RepFIB from the plasmid P307
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FRONTISPIECE</th>
<th>Clepsydra</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHAPTER ONE</td>
<td></td>
</tr>
<tr>
<td>1:1</td>
<td>The membrane attachment hypothesis</td>
</tr>
<tr>
<td>1:2</td>
<td>Kinetics of negative control of plasmid replication</td>
</tr>
<tr>
<td>1:3</td>
<td>Three models of the initiation of plasmid replication</td>
</tr>
<tr>
<td>1:4</td>
<td>Inter and intra-molecular coupling</td>
</tr>
<tr>
<td>1:5</td>
<td>Control via anti-parallel pairing</td>
</tr>
<tr>
<td>1:6</td>
<td>The minimal RepFIB replicon</td>
</tr>
<tr>
<td>CHAPTER TWO</td>
<td></td>
</tr>
<tr>
<td>2:1</td>
<td>Different types of promoter-probe plasmids</td>
</tr>
<tr>
<td>2:2</td>
<td>Promoter sequences: active promoters and <em>TargSearch</em> scores</td>
</tr>
<tr>
<td>CHAPTER THREE</td>
<td></td>
</tr>
<tr>
<td>3:1</td>
<td>Location of potential promoter sequences upstream of <em>repA</em></td>
</tr>
<tr>
<td>3:2</td>
<td>Identification of the <em>repA</em> promoter</td>
</tr>
<tr>
<td>3:3</td>
<td>Experimental assay for <em>repA</em> autoregulation</td>
</tr>
<tr>
<td>3:4</td>
<td>Autoregulation of <em>repA</em>&lt;sub&gt;p&lt;/sub&gt; by RepA provided <em>in trans</em></td>
</tr>
<tr>
<td>3:5</td>
<td>Possible repression mechanism of <em>repA</em>&lt;sub&gt;p&lt;/sub&gt;</td>
</tr>
<tr>
<td>3:6</td>
<td>Repression of <em>repA</em>&lt;sub&gt;p&lt;/sub&gt; does not require the BCD repeat elements</td>
</tr>
<tr>
<td>3:7</td>
<td>The origin repeat elements BCDD'D&quot; and <em>repA</em>&lt;sub&gt;p&lt;/sub&gt;</td>
</tr>
<tr>
<td>CHAPTER FOUR</td>
<td></td>
</tr>
<tr>
<td>4:1</td>
<td>Promoter-probe plasmids used in this section</td>
</tr>
<tr>
<td>4:2</td>
<td>Relative promoter strengths of <em>repA</em>&lt;sub&gt;p&lt;/sub&gt; and <em>ori</em>&lt;sub&gt;p&lt;/sub&gt;</td>
</tr>
<tr>
<td>4:3</td>
<td>Effect of RepA expression on cell growth</td>
</tr>
<tr>
<td>4:4</td>
<td>Effect of RepA expression on the behaviour of three minimal replicon promoters</td>
</tr>
<tr>
<td>4:5</td>
<td>RepA affinity for different repeat groups</td>
</tr>
<tr>
<td>4:6</td>
<td>Repeat elements present on the 'third' plasmids</td>
</tr>
<tr>
<td>CHAPTER FIVE</td>
<td></td>
</tr>
<tr>
<td>5:1</td>
<td>Expression of a RepA:<em>β</em>-galactosidase fusion protein from <em>pAS7</em></td>
</tr>
<tr>
<td>5:2</td>
<td>RepA:<em>β</em>-galactosidase fusion protein isolation and N-terminal sequencing strategy</td>
</tr>
<tr>
<td>5:3</td>
<td>PCR-directed mutagenesis of the RepA start codon</td>
</tr>
</tbody>
</table>

continued on the following page
<table>
<thead>
<tr>
<th>LIST OF FIGURES (continued)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHAPTER SIX</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>CHAPTER SEVEN</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>CHAPTER EIGHT</td>
</tr>
<tr>
<td>APPENDIX TWO</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>APPENDIX THREE</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>APPENDIX FOUR</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>APPENDIX FIVE</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>APPENDIX SIX</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

continued on the following page
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A6:7</td>
<td>Organisation of the mini-P1 replicon</td>
<td>185</td>
</tr>
<tr>
<td>A6:8</td>
<td>DnaA-dependent replication from oriC</td>
<td>195</td>
</tr>
</tbody>
</table>

**TAILPIECE**

Modified clepsydra
<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>TABLE</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHAPTER ONE</td>
<td>1:1</td>
<td>The control mechanisms in some plasmid replicons</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>1:2</td>
<td>Some representative RepFIB replicons</td>
<td>16</td>
</tr>
<tr>
<td>CHAPTER TWO</td>
<td>2:1</td>
<td>Promoter-probe analysis of the expression of open reading frames</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>2:2</td>
<td>Active and inactive promoter sequences</td>
<td>28</td>
</tr>
<tr>
<td>CHAPTER THREE</td>
<td>3:1</td>
<td>Promoter-probe plasmids used to identify repAp</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>3:2</td>
<td>Vector-initiated transcription</td>
<td>32</td>
</tr>
<tr>
<td>CHAPTER FOUR</td>
<td>4:1</td>
<td>Repression of pAS21 and pAS22</td>
<td>46</td>
</tr>
<tr>
<td></td>
<td>4:2</td>
<td>Repression of pAS7 and pAS9</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>4:3</td>
<td>RepA binding to repeat elements</td>
<td>56</td>
</tr>
<tr>
<td>CHAPTER FIVE</td>
<td>5:1</td>
<td>Alignment of the N-terminal fusion protein sequence to the predicted sequence of RepA</td>
<td>63</td>
</tr>
<tr>
<td></td>
<td>5:2</td>
<td>Relative expression of the fusion proteins from pAS2 and pAS56</td>
<td>66</td>
</tr>
<tr>
<td>CHAPTER SIX</td>
<td>6:1</td>
<td>Extraction of RepA</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>6:2</td>
<td>The western-DNA protocol</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>6:3</td>
<td>Summary of western-DNA information</td>
<td>76</td>
</tr>
<tr>
<td>CHAPTER SEVEN</td>
<td>7:1</td>
<td>Summary of incompatibility reactions</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td>7:2</td>
<td>Origin 'rescue' assay</td>
<td>85</td>
</tr>
<tr>
<td></td>
<td>7:3</td>
<td>Requirement for Dam methylase</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td>7:4</td>
<td>Requirement for DnaA</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td>7:5</td>
<td>Copy numbers of RepFIB mini-plasmids</td>
<td>88</td>
</tr>
<tr>
<td></td>
<td>7:6</td>
<td>Affect on RepFIB mini-plasmid copy number by pAS18 in trans</td>
<td>89</td>
</tr>
<tr>
<td>CHAPTER EIGHT</td>
<td>8:1</td>
<td>Comparison of RepFIB and mini-P1</td>
<td>101</td>
</tr>
<tr>
<td></td>
<td>8:2</td>
<td>Areas of research in this thesis</td>
<td>103</td>
</tr>
<tr>
<td>APPENDIX ONE</td>
<td>A1:1</td>
<td><em>Escherichia coli</em> strains</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td>A1:2</td>
<td>PCR oligonucleotide primers</td>
<td>110</td>
</tr>
</tbody>
</table>

continued on the following page
<table>
<thead>
<tr>
<th>APPENDIX TWO</th>
<th>A2:1 Cloning plasmids used in this thesis</th>
<th>122</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>A2:2 RepFlB plasmids used in this thesis</td>
<td>123</td>
</tr>
<tr>
<td></td>
<td>A2:3 Plasmid replicons, RepFlB inc determinants and antibiotic resistances</td>
<td>124</td>
</tr>
<tr>
<td>APPENDIX THREE</td>
<td>A3:1 Statistical analysis of the open reading frames in the E11 fragment</td>
<td>135</td>
</tr>
<tr>
<td></td>
<td>A3:2 Potential ORF-1 polypeptides</td>
<td>136</td>
</tr>
<tr>
<td></td>
<td>A3:3 Amino acid sequence variations in RepA</td>
<td>136</td>
</tr>
<tr>
<td></td>
<td>A3:4 The A-K repeat elements</td>
<td>138</td>
</tr>
<tr>
<td></td>
<td>A3:5 Potential promoter sequences</td>
<td>142</td>
</tr>
<tr>
<td></td>
<td>A3:6 dnaA box-like sequences</td>
<td>143</td>
</tr>
<tr>
<td></td>
<td>A3:7 GATC-repeat elements and 13mer-like sequences</td>
<td>144</td>
</tr>
<tr>
<td></td>
<td>A3:8 AT-richness of the origin region</td>
<td>145</td>
</tr>
<tr>
<td></td>
<td>A3:9 Comparison of the P1 37bp GC-rich spacer and the analogous region in RepFlB</td>
<td>146</td>
</tr>
<tr>
<td></td>
<td>A3:10 Amino acid sequence homology between RepA of mini-P1, mini-Rts1 and RepFlB</td>
<td>147</td>
</tr>
<tr>
<td>APPENDIX FIVE</td>
<td>A5:1 Repression of ORF-7p</td>
<td>160</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
<td></td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td>BCDD’D'</td>
<td>DNA repeat group located upstream of repA</td>
<td></td>
</tr>
<tr>
<td>8-gal</td>
<td>8-galactosidase</td>
<td></td>
</tr>
<tr>
<td>8-me</td>
<td>8-mercaptoethanol</td>
<td></td>
</tr>
<tr>
<td>bp</td>
<td>base pairs</td>
<td></td>
</tr>
<tr>
<td>CAPS</td>
<td>3-(cyclohexylamino)-1-propane-sulphonic acid</td>
<td></td>
</tr>
<tr>
<td>CAT</td>
<td>chloramphenicol acetyl transferase</td>
<td></td>
</tr>
<tr>
<td>cbp</td>
<td>co-ordinate base pairs</td>
<td></td>
</tr>
<tr>
<td>cmI</td>
<td>CAT gene</td>
<td></td>
</tr>
<tr>
<td>CnBr</td>
<td>cyanogen bromide</td>
<td></td>
</tr>
<tr>
<td>cRNA</td>
<td>counter transcript RNA</td>
<td></td>
</tr>
<tr>
<td>CTD</td>
<td>calf thymus DNA</td>
<td></td>
</tr>
<tr>
<td>°C</td>
<td>degrees celsius</td>
<td></td>
</tr>
<tr>
<td>DAB</td>
<td>3,3'-diaminobenzidine</td>
<td></td>
</tr>
<tr>
<td>Dam</td>
<td>dam methylase</td>
<td></td>
</tr>
<tr>
<td>dCTP</td>
<td>deoxycytidine 5'-triphosphate</td>
<td></td>
</tr>
<tr>
<td>DMSO</td>
<td>dimethyl sulphoxide</td>
<td></td>
</tr>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
<td></td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxyribonucleotide triphosphates</td>
<td></td>
</tr>
<tr>
<td>ECL</td>
<td>enhanced chemiluminescence</td>
<td></td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetra-acetate</td>
<td></td>
</tr>
<tr>
<td>EFG</td>
<td>DNA repeat group located downstream of repA</td>
<td></td>
</tr>
<tr>
<td>EFp</td>
<td>promoter element located in the EF repeats</td>
<td></td>
</tr>
<tr>
<td>EtBr</td>
<td>ethidium bromide</td>
<td></td>
</tr>
<tr>
<td>HIJ</td>
<td>DNA repeat group located downstream of repA and EFG</td>
<td></td>
</tr>
<tr>
<td>HPLC</td>
<td>high pressure liquid chromatography</td>
<td></td>
</tr>
<tr>
<td>hr</td>
<td>hours</td>
<td></td>
</tr>
<tr>
<td>HRP</td>
<td>horse raddish peroxidase</td>
<td></td>
</tr>
<tr>
<td>inc</td>
<td>incompatibility</td>
<td></td>
</tr>
<tr>
<td>IPTG</td>
<td>isopropylthio-β-D-galactoside</td>
<td></td>
</tr>
<tr>
<td>Kleow</td>
<td>large fragment of DNA Pol 1</td>
<td></td>
</tr>
<tr>
<td>LMP</td>
<td>low melting point</td>
<td></td>
</tr>
<tr>
<td>kb</td>
<td>kilobase pairs</td>
<td></td>
</tr>
<tr>
<td>kDa</td>
<td>kilodaltons</td>
<td></td>
</tr>
<tr>
<td>lacZ</td>
<td>β-galactosidase gene</td>
<td></td>
</tr>
<tr>
<td>λPl,R</td>
<td>hybrid λ right and left promoter</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>molar</td>
<td></td>
</tr>
<tr>
<td>min.</td>
<td>minutes</td>
<td></td>
</tr>
<tr>
<td>MMS</td>
<td>methylethane sulphonate</td>
<td></td>
</tr>
<tr>
<td>n</td>
<td>nucleotide</td>
<td></td>
</tr>
<tr>
<td>NFMP</td>
<td>non-fat milk powder</td>
<td></td>
</tr>
<tr>
<td>Ω</td>
<td>omega resistance fragment</td>
<td></td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
<td></td>
</tr>
<tr>
<td>ori</td>
<td>origin of replication</td>
<td></td>
</tr>
<tr>
<td>oriP</td>
<td>promoter element producing transcripts which cross the origin of replication</td>
<td></td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
<td></td>
</tr>
<tr>
<td>PEG</td>
<td>polyethylene glycol</td>
<td></td>
</tr>
<tr>
<td>PMSF</td>
<td>phenylmethyl-sulphonyl fluoride</td>
<td></td>
</tr>
<tr>
<td>Pol I</td>
<td>DNA polymerase I</td>
<td></td>
</tr>
<tr>
<td>PTC</td>
<td>trp/lac hybrid promoter</td>
<td></td>
</tr>
<tr>
<td>PVDF</td>
<td>polyvinylidene difluoride</td>
<td></td>
</tr>
<tr>
<td>repA</td>
<td>replication/initialisation gene of RepFIB</td>
<td></td>
</tr>
<tr>
<td>repAp</td>
<td>protein product of repA</td>
<td></td>
</tr>
<tr>
<td>repA promoter</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
<td></td>
</tr>
<tr>
<td>rpm</td>
<td>revolutions per minute</td>
<td></td>
</tr>
<tr>
<td>SD</td>
<td>Shine and Dalgarno sequence</td>
<td></td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodeyl sulphate</td>
<td></td>
</tr>
<tr>
<td>SDS-PAC</td>
<td>SDS-polyacrylamide</td>
<td></td>
</tr>
<tr>
<td>Spec</td>
<td>spectinomycin</td>
<td></td>
</tr>
<tr>
<td>STET</td>
<td>sucrose, Triton X-100, EDTA, Tris buffer</td>
<td></td>
</tr>
<tr>
<td>SSC</td>
<td>sodium chloride, sodium citrate</td>
<td></td>
</tr>
<tr>
<td>TA</td>
<td>Trit-buffered acetate</td>
<td></td>
</tr>
<tr>
<td>TBS</td>
<td>Tris-buffered saline</td>
<td></td>
</tr>
<tr>
<td>TE</td>
<td>Tris-EDTA buffer</td>
<td></td>
</tr>
<tr>
<td>TEN</td>
<td>Tris, NP-40 buffer</td>
<td></td>
</tr>
<tr>
<td>TEP</td>
<td>Tris, EDTA, PMSF buffer</td>
<td></td>
</tr>
<tr>
<td>Tet</td>
<td>tetracycline</td>
<td></td>
</tr>
<tr>
<td>Tris</td>
<td>tris (hydroxymethyl) aminomethane</td>
<td></td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet light</td>
<td></td>
</tr>
<tr>
<td>xg</td>
<td>relative gravities</td>
<td></td>
</tr>
<tr>
<td>Xgal</td>
<td>5-bromo-4-chloro-3-indolyl-β-D-galactoside</td>
<td></td>
</tr>
</tbody>
</table>

**LIST OF COMPANY ABBREVIATIONS**

<table>
<thead>
<tr>
<th>Company</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amersham</td>
<td>Amersham International</td>
</tr>
<tr>
<td>Bio 101</td>
<td>Bio 101 Inc.</td>
</tr>
<tr>
<td>BM</td>
<td>Boehringer Mannheim GmbH</td>
</tr>
<tr>
<td>BRL</td>
<td>Bethesda Research Laboratories</td>
</tr>
<tr>
<td>Difco</td>
<td>Difco Laboratories</td>
</tr>
<tr>
<td>Millipore</td>
<td>Millipore Corporation</td>
</tr>
<tr>
<td>NEN</td>
<td>New England Biologs, Inc.</td>
</tr>
<tr>
<td>NEB</td>
<td>New England Nuclear Research Products</td>
</tr>
<tr>
<td>Promega</td>
<td>Promega Biotec</td>
</tr>
<tr>
<td>Sigma</td>
<td>Sigma Chemical Company</td>
</tr>
<tr>
<td>Stratagene</td>
<td>Stratagene Cloning Systems</td>
</tr>
<tr>
<td>Whatman</td>
<td>Whatman International Ltd.</td>
</tr>
</tbody>
</table>

**Buckinghamshire, UK**
- La Jolla, CA, USA
- Mannheim, Germany
- Life technologies, Inc., Gaithersburg, MD, USA
- Detroit, MI, USA
- Bedford, MA, USA
- Beverly, MA, USA
- Boston, MA, USA
- Madison, WI, USA
- St. Louis, MO, USA
- San Diego, CA, USA
- Maidstone, UK
NOTES

Standard SI units have been used throughout this Thesis, with the exception of temperature (°C) and time (hr, min.).

Where appropriate, data has been presented as a mean value ± standard error. In some figures, the error bar is smaller than the symbol used to mark the mean value.

Restriction endonuclease recognition sites are referred to using standard format, although in some figures the location of sites might be marked with a single letter or symbol. Unless otherwise stated, nucleotide positions (cbp) correspond to the sequence of the P307 EcoRI E11 fragment. For a key to any unlabelled figure of RepFIB, see Chapter One, Figure 1:6, or Appendix Three, Figure A1:1.

All plasmids constructed for this thesis are designated by the prefix 'p' and are listed or described in Appendix Two.
Introduction, Research Chapters and Review and Discussion
Prokaryotic plasmids are covalently closed circular DNA elements. They are stably inherited, extrachromosomal elements which have been isolated from many bacterial genera. Plasmids are characteristically different from chromosomal and viral DNAs by virtue of being dispensable to the cell, lacking cellular pathogenicity and having no extracellular stage in their life cycles. The selective advantage to the host of some plasmids remain cryptic but many contain genes encoding obvious advantages. These advantages include resistance to antibiotics, colicins and heavy metals, the production of toxins, virulence factors and metabolic enzymes as well as allowing conjugation between bacterial cells of the same or similar species.\footnote{Plasmid biology has been described or reviewed by Novick (1969); Clowes (1972), Datta (1979 and 1985); Hardy (1986); Levy \textit{et al.}, (1981); Couturier \textit{et al.}, (1988); and by Grinsted and Bennett (1988).}

The first plasmid to be identified was the \textit{Escherichia coli} fertility factor ("F") which was discovered in the late 1950s due to its ability to mediate the transfer of chromosomal markers from one strain of \textit{E. coli} to another (well described by Stent and Calendar, 1978). Although the identification of F as a plasmid was of fundamental importance for the development of bacterial genetics, the discovery generated little interest outside of academic circles. However, this attitude was changed by the discovery of infectious resistance to antibiotics in epidemic strains of \textit{Shigella} in the late 1950s, and by the fact that the resistances were encoded by plasmids (see Grinsted and Bennett, 1988).

Since the 1960s plasmid research has progressed in three areas. The first has been an interest in the genetic and biochemical mechanisms employed by plasmids to ensure stable inheritance within the host cell. These mechanisms are collectively referred to as 'maintenance functions' and include replication and partition elements which interact with host factors to ensure that cells of each generation of the host contain copies of the plasmid. The second area has been an investigation of various phenotypes associated with plasmids, such as the expression of antibiotic resistances, toxins, virulence factors, or enzymes involved in unusual metabolic pathways (for a more extensive list see Stanisich, 1988). The third area has been the manipulation of several well-characterised plasmids for use in other branches of molecular genetics. This thesis is focused in the first area, and is concerned with the genetic control of replication of a particular class of plasmid replicon referred to as RepFIB.
1.1.1 The plasmid replicon

The term 'replicon' is often used to refer to that part of a plasmid (which may include more than a single gene) which is responsible for the replication of the plasmid within the host cell. The concept of a replicon was first defined by Jacob and Brenner (cited in Jacob et al., 1963) in which the replicon was defined as a "DNA molecule which can replicate, and the capacity to act as a replicon depended on the presence and activity of certain specific determinants which set up a system of signals allowing or preventing the replication of the replicon". Therefore, a simple plasmid might consist of a replicon responsible for the controlled replication of the plasmid, and a gene conferring a selective advantage to the host, such as a resistance to a particular antibiotic.

![Diagram of the membrane attachment hypothesis](image)

**FIGURE 1:1 The membrane attachment hypothesis**

The first model proposed for the control of plasmid replication was the Membrane Attachment hypothesis (Jacob et al., 1963). During most of the cell cycle the plasmid is not attached to the membrane (A). Plasmid replication is initiated by: i) the binding of the plasmid origin of replication (thick arc) to the attachment factor via the initiator; and ii) by the appearance of a new, unoccupied attachment factor expressed by the host (B). The plasmid replicates from the origin (a theta-replication mode with two replication forks is shown here) (C). A new attachment factor binds to the second origin as soon as it is synthesised. Partition of the cell membrane would separate the daughter plasmids, resulting in one plasmid in each of the two new cells (D).

The replicon in its simplest form consists of two elements: a structural gene which expresses a replicon-specific initiator of replication, and a cis-acting operator of replication (origin of replication) on which the initiator would act allowing replication to occur. The simple replicon model described by Jacob and Brenner was further developed by Jacob et al. (1963) to give the Membrane Attachment hypothesis (Figure 1:1). In this model, the initiator binds to both the origin of replication and to a
membrane-bound host 'attachment' factor. The frequency of replication and the timing of replication within the cell cycle would then be determined by the appearance and numbers of attachment factors. This model was unable to explain certain incompatibility phenomena observed between F+ and Hfr strains, and with the isolation of a number of bacteriophage and F factor replication and copy number mutants, consideration was given to the different ways in which plasmid copy numbers could be regulated within a cell. Three main suggestions were made: i) passive regulation by the host; ii) positive control by the host or the plasmid; or iii) negative control by the host or plasmid. Of these three suggestions, the first two were quickly discounted for biochemical and semantic reasons (discussed by Nordström, 1985), and research focused on the concept of negative regulation as a means of copy control.

Negative regulation requires the involvement of a repressor or inhibitor of replication which would either control the expression of a plasmid encoded initiator of replication or prevent the initiation of plasmid replication by a host replication factor. The most widely accepted early negative control model was the 'Inhibitor-dilution' model of Pritchard et al. (1969).

The Inhibitor-dilution model used a simple specific-volume titration mechanism involving a repressor (inhibitor) of replication as an alternative to the 'passive control' Membrane Attachment hypothesis. In this model, the synthesis of the repressor was coupled with the initiation of replication such that each initiation event lead to the production of a burst of repressor molecules. Once chromosomal replication was completed, the cell would divide and the repressor concentration in the daughter cells would be diluted over time by the increase in cell volume. Once the repressor concentration fell below a certain level there would be insufficient regulator present to repress the initiator gene. When this point was reached, the initiator would be expressed and a new round of replication would begin.

The main emphasis of this model was that cell growth resulted in the dilution of the repressor, and that once the cell had reached a certain size, plasmid replication would take place. Although the model required a restricted period of repressor synthesis, little consideration was given to how this might be accomplished. Sompayrac and Maaløe (1973) provided a solution by proposing a model of regulation in which the synthesis of the repressor and initiator were linked (the two genes were in the same operon). The repressor regulated the activity of the operon promoter resulting in a steady level of repressor and initiator in the cell; as the cell volume increased, the concentration of the repressor falls causing de-repression of the promoter, which in turn allows the expression of more repressor and initiator molecules. Over the period of cell growth, the amount of initiator within the cell increases. The initiator would bind to the origin, and once a sufficient initiation mass had accumulated, a round of replication would occur.

These two models should not be considered separately, as it is apparent that each model requires elements of the other in order to work: the Inhibitor-dilution model requires some form of autoregulation which will restrict the expression of the repressor to the 'replication' phase of the cell, whilst the
autoregulation model utilises the 'dilution' mechanism to vary the expression of the operon promoter during cell growth.

However, by this stage the debate about the regulation of plasmid replication was no longer simply an intellectual argument, and various groups had begun to investigate the control mechanisms of a number of different plasmid replicons. Over the following years, the control systems of a variety of replicons have been elucidated, and it has become clear that although all replicons are negatively regulated, plasmid replicons tend to use either the constitutive expression of a repressor, or an autoregulated combination of repressor and initiator to control replication.

**Figure 1:2** Kinetics of negative control of plasmid replication

Plasmid copy number is ideally maintained at a relative copy number of one, with one round of replication per plasmid copy in each cell generation (coordinates of [1,1] in this graph). Negative control is active within the shaded areas of the graph and is defined by the kinetics of no control (A) and the strongest control (C). The curve of the strongest control is often referred to as the 'Step Function', whilst control at an intermediate level is referred to as the 'Hyperbolic Function' (B). Novick (1987) defines a step function as a "Regulatory function in which a small change in inhibitor concentration brings about an all-or-none response in target activity"; and hyperbolic as a "Regulatory function in which there is an inverse proportionality between inhibitor concentration and target activity". The graph is from Nordström (1990).

Despite the fact that the replicon model proposed by Jacob and Brenner was quickly superseded, the investigation of plasmid control mechanisms has been restricted to those elements clearly encoded by the replicon. As a result, many host factors which are involved in chromosomal replication and in the
replication of plasmid DNA were ignored or undervalued. However, increasingly, efforts now are being made to understand how interactions between plasmid control mechanisms and host factors combine to initiate and limit plasmid replication.

1.2 Analysis of plasmid replicons

1.2.1 Division of replicon types into the hyperbolic and step function classes

It has been possible to obtain copy mutants (cop or rep) of a wide variety of plasmid replicons where the mutation affects the ability of the replicon to maintain a normal copy number within the cell. Some of these cop mutations are temperature sensitive, allowing the maintenance of the replicon at a wild-type copy number (by growth at the permissive temperature), or at higher mutant copy numbers (by growth at the non-permissive temperature). The copy control mechanisms of a number of replicons has been investigated by looking at how quickly a cop mutant can adjust the mutant copy number to the wild-type copy number following a shift in the growth temperature of the cells.

Despite the fact that there are a variety of replicon types, low copy number plasmid replicons can be broadly divided into two classes according to the copy number dynamics of the replicon recovering from an artificially high copy number (Nordström et al., 1984; Nordström, 1985 and 1990). In this manner, well-studied plasmid replicons such as ColE1 and mini-R1 fall into the 'hyperbolic' class of replicons, whereas plasmid replicons such as mini-F and mini-P1 fall into the 'Step function' class (Figure 1.2).

This simple division of replicons reflects a more fundamental difference in the copy control mechanisms used by the replicons to sense and regulate their copy numbers (Table 1.1). The hyperbolic class of replicons are typified by the involvement of counter-transcript RNA (ctRNA)², molecules which act as repressors at the primary level of control of the initiation of replication (reviewed by Cesareni and Banner, 1985). In the case of ColE1, the ctRNA molecule (RNA-I) will hybridise with the complementary i-RNA thus altering the downstream conformation of the i-RNA/DNA hybrid molecule which is required for the initiation of replication. Both the ctRNA and i-RNA are expressed at the same rate. However, the ctRNA half-life is much lower than that of the i-RNA. If for some reason the replicon copy number falls, the amount of ctRNA within the cell will be

² The Hyperbolic function replicons can be divided into three subgroups, exemplified by ColE1, mini-R1 and pT181. In all three subgroups, control is exercised through the interaction of a short repressor RNA molecule with a larger RNA molecule required for initiation. The small RNA molecule is referred to as a counter-transcript repressor (ctRNA). The coding region of the ctRNA is in the 5′ coding region of the larger RNA; transcription of the ctRNA is in the opposite direction to the transcription of the larger RNA with which it interacts. The larger initiator RNA molecule (i-RNA) may be present in various size classes although each retains the 5′ complementary sequence necessary to interact with the ctRNA. In the case of ColE1, translation of the larger RNA is not required for the initiation of replication. However, in mini-R1 and pT181 translation is required to express a protein essential for replication and is sometimes referred to as an initiator mRNA (i-mRNA).
reduced more quickly than the amount of i-RNA. The resulting imbalance of ctRNA and i-RNA will allow the initiation of DNA replication and a recovery of copy number.

**TABLE 1.1**  
The control mechanisms in some plasmid replicons

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Repressor</th>
<th>Initiator</th>
<th>Nature of the Control Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hyperbolic Function</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ColE1</td>
<td>RNA-I (ctRNA)</td>
<td>RNA-II (RNA)</td>
<td>ctRNA/RNA hybridisation leading to a change in the origin conformation.</td>
</tr>
<tr>
<td>pT181</td>
<td>CopA (ctRNA)</td>
<td>RepC protein</td>
<td>RNA/mRNA hybridisation leading to (i) a change in the secondary structure of the RepC mRNA in the region of the ribosome binding site and (ii) mRNA attenuation.</td>
</tr>
<tr>
<td>Step Function</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>mini-F</td>
<td>E protein</td>
<td>E protein</td>
<td>DNA binding resulting in autoregulation, titration and initiation.</td>
</tr>
<tr>
<td>mini-P1</td>
<td>RepA protein</td>
<td>RepA protein</td>
<td>DNA binding resulting in autoregulation, titration and initiation.</td>
</tr>
</tbody>
</table>

The replication control systems of these plasmids are described in Appendix Six (ColE1: A6.1; mini-R1: A6.2; pT181: A6.3; mini-F: A6.4 and mini-P1: A6.5).

In the case of the pBR322 replicon which is similar to ColE1, Chiang and Bremer (1991) have demonstrated that without the interaction between ctRNA and i-RNA, plasmid replication and cell growth are uncoupled. The interaction of the two RNA molecules is the only means of maintaining a stable copy number within the cell. A second example can be found in mini-R1, where the ctRNA hybridises with the leader sequence of the i-mRNA of a protein required for replication, so that normally accessible ribosomal binding sites are no longer available and the reading frame is no longer translated. This type of RNA-RNA interaction is the primary regulatory control for the IncFII and RCR (rolling circle replication) plasmids (such as pT181).

The copy control mechanism for the Hyperbolic class of replicons is well understood. However, the same cannot be said of the Step function replicons (reviewed by Thomas, 1988). In the Step function class, copy control is believed to be achieved through the interaction of an initiator protein and a set of DNA repeat sequences arrayed on either side of the gene (so called 'tandem repeats'). Two well-investigated examples of this type of replicon are mini-F and mini-P1. Although mini-F and mini-P1 share no sequence similarity, the organisation of mini-P1 is similar to that of mini-F (Abeles et al., 1984; and Chattoraj et al., 1985a). The Step function class relies on autoregulation of the initiator gene and titration of the gene product to control replication.
Although Nordström's functional dichotomy is useful in a simplistic analysis of replicons, a more detailed classification of replicons can be obtained by dividing replicons on the basis of how DNA synthesis is initiated, and on how the initiation of synthesis is regulated (reviewed by Nordström, 1990).

1.2.2 Initiation of DNA synthesis
The initiation of DNA synthesis for each round of plasmid replication appears to occur by one of three different mechanisms (Figure 1:3). In some replicons such as ColE1, DNA synthesis is initiated by DNA polymerase I (DNA Pol I) through the addition of nucleotides to the 3'-OH terminus of a cleaved initiator RNA molecule. In the group of plasmids known as the rolling circle replication (RCR) plasmids, DNA Pol I adds to a DNA 3'-OH terminus which has been produced by the double-stranded DNA nicking activity of a protein encoded by the replicon. The third mechanism is as yet unclear, but involves the action of a single initiator protein encoded by the replicon. The initiator protein recognises the replicon origin of replication and identifies the region to host factors involved in DNA replication. DNA replication from origins utilising this last mechanism most closely resembles the replication from oriC of the host chromosome requiring the involvement of the host DnaA protein.

1.2.3 Control of the initiator
Plasmid replicons appear to use one of two different primary systems to regulate plasmid copy number (Nordström, 1990).4 The majority of plasmids utilise the interaction of a small repressor ctRNA molecule with the initiator RNA (such as ColE1) or initiator mRNA (mini-R1 and pT181) to prevent the initiation of replication. In the case of ColE1, hybridisation of the two RNA molecules distorts the downstream conformation of the hybrid formed between the initiator RNA and its complementary DNA strand. If the RNA/DNA hybrid is distorted, RNase HI will not cleave the RNA molecule in the origin region to provide a 3'-OH necessary for the initiation of DNA synthesis.

In the case of mini-R1 and pT181, the interaction of the ctRNA with the initiator mRNA is more complex. Ultimately, such interaction prevents the expression of the mRNA message and the initiator protein is not produced, thus preventing the initiation of DNA synthesis. In the case of mini-R1, ctRNA hybridisation alters the secondary structure of the 5' end of the initiator mRNA in such a manner that the ribosome binding site (rbs) is moved from a single stranded loop region to a double stranded stem structure. It is believed that this transition prevents ribosomal recognition of the rbs thus preventing translation and RepA expression. The translation of the RepC initiator protein of pT181 from the initiator mRNA is prevented in a similar manner. In addition, the hybridisation of ctRNA and RepC mRNA can alter the secondary structure of the mRNA DNA template strand in such a manner that a terminator loop can form before the RNA polymerase reaches the translation start codon of the initiator gene. If this occurs, the polymerase will abort transcription in a manner similar to classical attenuation.

3 DnaA involvement in oriC and plasmid replication is reviewed in Appendix Six, A6.7 and A6.8.
4 Nordström suggests that there are three systems; the two described here plus that of λdv. However, λdv is arguably an artificial construction and functionally the control mechanism responsible for the maintenance of λdv copy number is similar to the Step function replicons.
The initiation of plasmid replication appears to occur by one of three different mechanisms. Initiation from an i-RNA as in ColEl where a DNA/RNA hybrid is cleaved at a specific site (△) by RNase H (A). DNA Pol I then begins to synthesise DNA from the 3'-OH provided by the cleaved i-RNA resulting in unidirectional replication. Initiation from a nicked DNA as in pT181 where the initiator protein (RepC) nicks the top strand at a specific site (△) and remains bound to the 5' end of the DNA (B). DNA Pol I then begins to synthesise DNA from the exposed 3'-OH, displacing the 5' end of the top strand resulting in unidirectional replication. Initiation involving origin recognition as in mini-P1 where the origin of replication is recognised by the initiator protein (RepA) and DnaA (C). The binding of these proteins is thought to melt the DNA and allow bidirectional synthesis in a manner similar to oriC replication.

Plasmid replicons such as mini-F and mini-P1 appear to use a different system of control. These replicons rely on autoregulation of the initiator gene and titration of the gene product to control replication. In mini-F, the E gene product (E protein)\(^5\) is required for replication. The E gene is flanked on either side by a series of repeat sequences which bind the E protein, and a smaller inverted repeat of the same sequence is located close to the E gene promoter. The E protein binds to the inverted repeat to autoregulate the expression of the gene, whilst binding to the set of repeats upstream of the E gene (incB or origin repeats) is required to initiate replication. The E protein also binds to the repeats located downstream of the E gene (incC repeats). By binding at these sites, the effective

---

5 Despite the redundancy, in this thesis the protein is referred to as the 'E protein' and the gene referred to as the 'E gene' rather than 'E' and 'E' respectively.
concentration of E protein is reduced (titrated) and the origin repeats cannot bind E protein to initiate replication.

In mini-P1, the repA gene product (RepA) is required for replication. The repA gene is flanked on either side by a series of repeat sequences which bind RepA, and the upstream repeat element (incC) includes the repA promoter. RepA protein binds to incC to autoregulate the expression of the gene and to initiate replication. RepA also binds to the repeats located downstream of repA (incA) in a manner similar to that of the E protein binding to the mini-F incC repeats.

1.2.4 The control paradox
Although mini-F and mini-P1 share no sequence similarity, the organisation of the replicons is genetically very similar. Despite the fact that autoregulation and titration have been well defined individually in mini-F and mini-P1, attempts to combine these elements in a single control model have produced an unresolved paradox: it appears impossible to have the autoregulation of a gene, yet still be able to titrate the gene product at a different site.

1.2.5 Possible solutions to the control paradox
Solutions to the paradox posed by mini-F and mini-P1 were initially based on one of two suggestions. In the first suggestion the assumption that the repressor and initiator activities of the replication protein were carried out by the same molecule was questioned. In the case of mini-F, there were the possibilities that: i) the E protein had different affinities for the three binding sites (which would allow leaky autoregulation, then titration and finally initiation); ii) the E protein was modified in some manner such that the initial form was the repressor and the subsequent form the initiator (which was also titrated); and finally, iii) a second gene in the same segment of DNA carrying the E gene was responsible for the expression of the repressor, and that the E protein had only initiator activity⁷. Although these possibilities have not been completely ruled out (for example, see Kline et al., 1992), they appear to be unlikely solutions to the paradox and are now in contest with a new generation of control models.

1.2.6 Invoking host factors to solve the control paradox
The second solution to the paradox was the suggestion that the primary control element was not encoded by the plasmid, but was in fact a host factor. Although this suggestion went against the direction set by Jacob and Brenner's definition of a replicon, the DnaA protein involved in the control and initiation of replication from the host origin (oriC) seemed to be a likely candidate. DnaA involvement in oriC and plasmid replication is reviewed by Messer (1987), Bramhill and Kornberg (1988b) and Georgopoulos (1989) (see also Bernander et al., 1991)⁸. The current consensus of opinion is that DnaA controls the initiation of replication at oriC and that through a variety of

---

⁶ Mini-F and mini-P1 replication is reviewed in Appendix Six, A6.4 and A6.5, respectively.  
⁷ See Appendix Six, A6.4.8, A6.4.9 and A6.4.10.  
⁸ See Appendix Six, A6.7 and A6.8.
channels, chromosomal replication is linked with segregation and cell division (see Newton and Ohta, 1992, for a recent review). The timing of oriC initiation is controlled by the level of DnaA which is determined by DnaA degradation, DnaA autoregulation and the influence of other factors such as Dam methylation and the presence of repressors like IciA (Inhibition of chromosomal initiation).

The role of DnaA in the initiation of replication in many plasmids where it is essential for replication is not well understood\(^9\). It was assumed that DnaA bound to dnaA boxes at the origin melted the DNA to form a initial complex which then lead to the formation of a replisome and the initiation of DNA synthesis. However, some DnaA-dependent replicons are capable of replication even though the origin dnaA boxes have been altered or deleted (Murakami et al., 1987; Itoh and Terawaki, 1989; and Ortega-Jiménez et al., 1992). Although replication in these cases is less efficient compared to replicons containing the normal complement of dnaA boxes, these results imply that the involvement of DnaA in the replication of some replicons is different mechanistically to the action of DnaA in oriC replication. In the case of mini-R1, DNA-melting of the origin (oriR) and the formation of the replisome complex is bought about by RepA activity alone. However, if DnaA is available, it will interact with a RepA-oriR complex, rather than specifically with the oriR dnaA box, and result in a more efficient replication of mini-R1 (Ortega-Jiménez et al., 1992).

It is not yet clear whether DnaA has a role in the regulation or timing of initiation of replication, or whether DnaA is simply required for the efficient replication of DnaA-dependent plasmid replicons.

A number of other host factors are clearly involved in the initiation and control of DNA synthesis (see Masters, 1991). Of these, Dam methylase and IciA may be involved in the regulation of replication of some replicons. Dam methylates adenosine in the sequence 5'-GATC-3' with a variety of effects (reviewed by Sternberg, 1985; including a strong enhancement of DNA bending in oriC, Kimura et al., 1989). Such sequences are not randomly distributed in the E. coli genome (Barras and Marinus, 1988) or in plasmid DNA, and often appear in clusters within origins of replication (mini-P1: Abeles and Austin, 1988). A number of studies have shown that hemimethylated DNA is a less efficient substrate than either unmethylated or fully methylated DNA for replication in vivo and in vitro (Huges et al., 1984; Messer et al., 1985; Landoulsi et al., 1989; Huges et al., 1989; and Boye, 1991). It appears that the hemimethylation of origin sequences resulting from a round of replication (between the passage of a replication fork and methylation of the nascent strand by Dam) provides a refractory period during which no further replication can occur (Smith et al., 1985; Messer et al., 1985; Campbell and Kleckner, 1990; and Boye and Løbner-Olesen, 1990).

Recent investigation of oriC replication has also identified a protein of 33kDa (IciA) which binds to oriC DNA and prevents DnaA-mediated melting of the origin and subsequent replication (Hwang and Kornberg, 1990; and Thøng et al., 1991). It is possible that such a factor might also be involved in

\(^9\) See Appendix Six, A6.8.
the regulation of replication from origins which also contain the AT-rich '13mer' sequence required to bind the protein.

The final set of host factors which may effect the control of replication are the RNA polymerase sigma factors which are responsible for the recognition of promoter sequences for the holoenzyme. In the case of mini-F, transcription of the E gene is from a promoter which appears to be recognised by a minor species of RNA polymerase which contains σ^{30} rather than the normal σ^{72} subunit\(^{10}\). However, this possibility has only been reported in mini-F and does not appear to effect the expression of the replication gene of mini-P1 (an unusual σ factor has also been implicated in the expression of cTRNA repressors in pT181).

1.2.7 New models which seek to overcome the control paradox

A new set of models were needed to explain mini-P1, mini-F, and R6K replication control as the apparent paradox presented by concurrent autoregulation and titration of the replicon initiator proteins could not be solved by existing models. For example, Pal and Chattoraj (1988) demonstrated that by increasing the concentration of RepA in trans, the copy number of a mini-P1 ori plasmid increased. However, if the ori plasmid also contained a copy of the incA titration repeats, the copy number of the plasmid remained the same despite increasing RepA concentrations. This observation was clearly inconsistent with the suggestion that incA inhibits mini-P1 replication by sequestering RepA which would otherwise bind to the origin repeats to initiate replication (if the titration model was correct, then the ori plasmid copy number should have increased once the binding capacity of the incA DNA had been reached by increasing concentrations of RepA)\(^{11}\). Pal and Chattoraj therefore suggested that, in addition to a titration function, incA can also restrain replication by causing steric hindrance to the origin function. Electron microscopy has revealed that incA can bind to a RepA-origin complex in vitro (Chattoraj et al., 1988; and Pal and Chattoraj, 1988).

A similar system of steric hindrance or intermolecular coupling may restrict the origin function of two other replicons (RK2 and R6K) which have a similar arrangement of origin and titration repeat elements (although they share no DNA sequence similarity with mini-P1). In the case of the RK2 replicon, Kittell and Helinski (1991) have demonstrated that a minimum of two titration repeats are capable of specifically inhibiting the in vitro replication of a functional RK2 ori plasmid, and that this inhibition can not be overcome by increasing the concentration of initiator protein (TrfA).

The observations of Pal and Chattoraj, Kittell and Helinski suggest that it is possible for two copies of a replicon to be physically linked during part of the cell cycle, and that during this time of physical

---

10 See Appendix Six, A6.4.5.
11 This argument is the basis of the Hierarchal Binding model suggested by Chattoraj et al. (1984) for the control of mini-F (Appendix Six, A6.4.8). This model incorporates the Clepsydra analogy (Chattoraj et al., 1985a) where time is kept by the gradual filling of a container by a steady stream of water. Once the container was full, it automatically emptied and the cycle began again (see the Frontispiece). The analogy in mini-F (or mini-P1) suggests that once the inc repeats are filled by bound initiator protein, new protein then binds to the ori repeats. Once these repeats are full, replication is initiated and the process of DNA synthesis clears both sets of repeats to allow the cycle to begin again.
linkage, further replication can not occur. The two replicon copies can theoretically be linked in either a parallel or anti-parallel configuration (Figure 1.4, A, B). An equally possible alternative to inter-molecular linkage is that of intra-molecular linkage or looping (Figure 1.4, C).

A third example of DNA linkage can be seen in the case of R6K which is unrelated to both mini-P1 and R2K. Like mini-P1 and R2K, the origin of replication (γ-ori) contains repeat elements which bind the R6K π initiator protein. Mukherjee et al. (1988b) have demonstrated that two γ origins separated by as much as 2kb (when cloned into the same plasmid) can be linked together to form a loop through the action of the π protein. An analysis of a π mutant unable to form DNA loops in vitro suggests that in R6K, DNA looping is required to activate the distant α and β origins which are located up to 3kb from the γ origin12 (Miron et al., 1992).

In mini-P1, the repA promoter is located within the incC repeat elements. In R6K, the analogous π gene promoter is located = 100bp away from the origin repeat elements and is also capable of binding π protein (Shafferman et al., 1982; Filutowicz et al., 1985a, 1985b; and Kelley and Bastia, 1985). The promoter binding site and the origin repeats are located on opposite sides of the DNA helix and do not form a DNA loop when π protein is present. However, if a half-helix turn, 6bp sequence is inserted between the two elements, looping can occur (Mukherjee et al., 1988b).

---

12 The γ origin is normally silent and can be activated by the deletion of the α and β origins (Kolter et al., 1978). The γ origin appears to have a dual role as both α and β origins require the presence in cis of γ-ori for activation (Crosa et al., 1978; and Mukherjee et al., 1988a).
1.2.8 The steric hindrance model
In both inter and intra-molecular linkages, initiator protein binding to various repeats allows the possibility of concurrent autoregulation and titration. In the Steric Hindrance model proposed by Pal and Chattoraj (1988), a single copy of mini-P1 exists in equilibrium between linear and looped forms. As the mass of RepA builds up during the cell cycle, more protein binds to both repeats and the linear form of the replicon is favoured. When sufficient RepA is present, mini-P1 replicates and the two copies are linked in a parallel configuration via RepA interactions. As the cell volume increases during growth, the equilibrium between inter and intra-molecular pairing favours the formation of individual looped replicons. Towards the end of the cell cycle, most replicons will form loops rather than inter-molecular pairings. The cell divides and the process is repeated before sufficient RepA is built up to enable the daughter replicons to replicate.

This model has three important features. Firstly, replication is inhibited when intra-molecular linkages or loops are formed. Secondly, replication is initiated when sufficient RepA is bound to both sets of repeats, and such binding tends to favour a linear configuration rather than the inhibitory ('steric interference' or 'hindrance') looped configuration of the replicon. Finally, daughter replicons form inter-molecular linkages in the parallel configuration. Autoregulation is not addressed per se in this model and occurs as a result of protein binding to the ori repeats. Although autoregulation may affect the rate of protein synthesis, it is titration and loop formation which regulate replication.

1.2.9 The anti-parallel pairing model
Abeles and Austin (1991) have suggested an alternative model for mini-P1 control ('Anti-parallel Pairing', Figure 1.5). In this model, mini-P1 replicates immediately after cell division and the mini-P1 copies are linked by RepA in the anti-parallel configuration. The daughter replicons remain in this configuration for the majority of the cell cycle, until the partition of the cell immediately prior to cell division physically separates the two copies of mini-P1. During the period of anti-parallel binding, both daughter replicons are prevented from initiating replication due to steric hindrance of the origins by their partners' titration repeats.

Like the Steric Hindrance model, autoregulation of the initiator protein is a minor feature of the control system; initiation of replication is regulated primarily through changes in DNA-protein binding between replicon copies.

1.2.10 Comparison of the two new models
The new models differ in two major areas. Firstly, in the Steric Hindrance model, repression is maintained by intra-molecular looping, whereas in the Anti-parallel Pairing model repression is maintained by inter-molecular pairing. The second difference is in the initiation of replication. In the Steric Hindrance model, initiation depends on an adequate RepA mass. During normal growth, the mini-P1 replicon should replicate once per cell generation. However, during periods of more rapid growth (or in vitro), mini-P1 replication should increase with increasing amounts of RepA.
In the Anti-parallel Pairing model, mini-P1 replication is restricted to the very early stages of the cell cycle and cannot be repeated until the replicon configurations are reset by cellular partition. This implies that either multiple rounds of initiation at fast growth rates is not possible without the production of plasmid minus segregants (which is not the case), or that the cell provides additional partition elements to deal with multiple copies of mini-P1 within the cell (plasmid partition is reviewed by Austin, 1988). This last prediction is reminiscent of the Membrane Attachment hypothesis suggested by Jacob et al. (1963), in that a major element of mini-P1 replication control now appears to be the result of a host function, rather than a function of the replicon itself.

Both the Steric Hindrance and Anti-Parallel Pairing models accommodate current mini-P1 experimental evidence. The results of incompatibility experiments using origin or titration repeats, autoregulation and binding experiments do not falsify either of the two new models, so long as each feature is interpreted as resulting from a partially disrupted system. In particular, titration or binding at the inc and ori repeats are both events required for inter- or intra-replicon binding; the degree of protein binding at these sites influences the type of DNA-DNA interaction and subsequently the initiation of replication.

**FIGURE 1:5** Control via anti-parallel pairing

Early after the formation of the new cell, DNA replication is initiated from the plasmid origin (A) to give two copies of the plasmid (B). The two plasmid copies bind to one another in an anti-parallel configuration (C), through interactions between the ori (open box) and titration repeats (black box) via the action of the initiator protein (dashed lines). Once the plasmids are in this configuration, further replication cannot occur. Near the end of the cell cycle during partition, the plasmid copies are physically separated (D) prior to cell division. The probability of the initiation of plasmid DNA synthesis during the cell cycle is shown in the bottom portion of the figure. From Abeles and Austin (1991).
Autoregulation is not seen to be major factor of regulation in these models (especially in the Anti-parallel Pairing model), and occurs as a result of protein binding at the ori repeats limiting RNA Pol recognition of the promoter sequence. However, especially in the case of the Steric Hindrance model, the continued expression of the 'repressed' gene is required to allow the gradual build-up of initiator protein to occur (which will allow the saturation of the binding repeats and subsequent changes in DNA-DNA interactions leading to the initiation of replication). With this interpretation, the paradox generated by the apparent concurrent and equally important autoregulation and titration is resolved, as titration is seen to be the more important and dominant aspect of the control system.

In both models, the initiator protein exists as a regulatory element allowing (or preventing) the formation of specific DNA configurations for the majority of the cell cycle. For a small period in each generation, the initiator protein acts in its second role which is simply to mark the origin of replication to allow recognition by the host replisome (the multi-enzyme complex responsible for DNA synthesis during chromosomal replication). The two models differ in terms of when this second activity is allowed to occur in the cell cycle. The Steric Hindrance model allows initiation at any time and suggests that the primary regulatory determinant is the amount of initiator present. However, the Anti-parallel Pairing model requires that initiation occurs at the beginning of the cell cycle (or at fixed times at higher growth rates). This restriction, as well as the necessity of partition to 'reset' the replicon configuration to allow further replication, suggests that the amount of initiator protein is not the sole regulatory determinant of replication.

Recent experiments have demonstrated that mini-F plasmid replication occurs in a cell-cycle specific manner when a constant mass per origin is achieved and at the same time chromosomal replication takes place (Keasling et al., 1991). These observations do not fit well with either of the two replication models. In contrast, mini-R6K replication is in a cell cycle independent manner and initiation occurs throughout the cell cycle in agreement with the Steric Hindrance model (Keasling et al., 1992). Although the mini-F, mini-P1 and R6K replicons are organisationally similar, it is apparent that R6K replicates in a cell cycle independent manner more similar to high copy number plasmids such as ColE1 (Zeuthen et al., 1972, and Leonard and Helmstetter, 1986).

1.3 The RepFIB replicon

1.3.1 The initial investigation of RepFIB
The first RepFIB replicon to be isolated resulted from an attempt to isolate the f5 EcoRI fragment which carried the primary F plasmid replicon (mini-F, Lane and Gardner, 1979). Lane and Gardner demonstrated that the new replicon was functionally distinct from the mini-F replicon, and that it was carried by the f7 fragment generated by cleavage of F plasmid DNA with EcoRI (32.8-40.3Fkb coordinates). Although the f7 fragment is 7.5kb and the RepFIB replicon included in the fragment is substantially smaller (Gardner et al., 1985), the f7 fragment has been used to identify by Southern
hybridisation other examples of the replicon from a wide range of diverse plasmids (See Table 1.2; Bergquist et al., 1982, and 1986).

The majority of plasmids which contained RepFIB are found within the IncF incompatibility group. The two exceptions to this are R621a and pH502, belonging to the IncE and IncP groups, respectively (Bukhari et al., 1977; and Jacob et al., 1977). The RepFIB replicon appears to be a fairly common replicon and does not appear to be associated with any particular plasmid phenotype (Bergquist et al., 1986). However, in the case of some of the ColV plasmids, the presence of the RepFIB replicon appears to be linked with aerobactin iron uptake genes. Waters and Crosa (1986) have suggested that the aerobactin operon is genetically mobile, and that the linkage between the operon and RepFIB has been instrumental in the preservation and spread of the aerobactin genes among ColV plasmids.

Initial RepFIB research was undertaken to establish the general features of the RepFIB class of replicons. The early analysis of several examples of RepFIB revealed that the replicon was responsible for the expression of incE incompatibility. This feature, along with sequence similarity with the RepFIB portion of the F/ fragment was used to define the RepFIB class.

Functionally, mini-RepFIB plasmids show a Step function response to variations in copy number (demonstrated experimentally in the case of RepFIB from P307, Maas et al., 1989), all mini-plasmids have a copy number of between one and two copies per host chromosome, and all mini-plasmid replication is independent of DNA polymerase I.  

**Table 1.2**  
Some representative RepFIB replicons

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>F</td>
<td>Escherichia coli sex factor. First example of RepFIB isolated as a mini-plasmid. RepFIB is located on the 7.5kb F7 fragment.</td>
</tr>
<tr>
<td>P307</td>
<td>Enterotoxin plasmid isolated from a case of porcine diarrhoea (Gyles et al., 1977). RepFIB is located on the 4.3kb E11 fragment, which is the smallest EcoRI fragment containing RepFIB.</td>
</tr>
<tr>
<td>pMA4322</td>
<td>A spontaneous P307 RepFIB copy mutant (R. Maas). A single amino acid alteration in RepA is responsible for the copy number mutation.</td>
</tr>
<tr>
<td>R386</td>
<td>Tetracycline resistance plasmid isolated from a human patient (Dennison, 1972). R386 is used as an incE-type plasmid in incompatibility testing. RepFIB is located on the 5.8kb F9 fragment.</td>
</tr>
</tbody>
</table>

a All plasmids have been isolated from E. coli strains. All plasmids (unless indicated) belong to the IncF Inc group.

The minimal RepFlB replicon contains three major elements: a gene encoding a replication protein (repA), and two sets of repeat elements on either side of repA (the origin repeats, BCD; and part of the incE or titration repeats, EFG). The origin repeats are adjacent to a number of sequences associated with replicon origins (boxed). Deletion analysis of the P307 E11 fragment has demonstrated that the HIJ incE repeat elements are not required for RepFlB replication. However, the EFG and HIJ repeats probably form a functional unit which suggests that the right hand border of the basic replicon is located to the right of the J repeat element (†).

Although the RepFlB replicon fell in the Step function class, various RepFlB replicons showed no sequence similarity with other Step function replicons. This observation lead to the suggestion that perhaps the RepFlB replication control mechanism was unique. Subsequent RepFlB analysis was directed towards testing this suggestion by obtaining sequence information of the replicon. The P307 RepFlB replicon was the first candidate for sequencing (in this laboratory) as the EcoRI fragment (E11) containing RepFlB was one of the smallest mini-RepFlB mini-plasmids available.

1.3.2 Sequence analysis of the EcoRI E11 fragment from P307

The F plasmid RepFlB replicon was located within the f7 fragment by a series of partial PstI and BamHI digestions (Gardner et al., 1985), and a similar analysis of the P307 E11 fragment using deletions produced for sequencing purposes defined the minimal RepFlB replicon (Spiers, MSc. thesis, 1989). The minimal replicon region for both examples of RepFlB was ≈ 1.6kb. An analysis of the E11 fragment sequence in this region identified a single large open reading frame (ORF-1) which was flanked on either side by a series of 21bp DNA repeat elements. Three repeats were located upstream of ORF-1 and formed a direct-repeat group (labelled as ‘BCD’). These repeats are associated with Dam methylation sites and a single dnaA box which suggests that the BCD repeat region of the replicon represents the origin of replication (Spiers, MSc thesis, 1989). Six repeat elements were located downstream of ORF-1 (labelled as ‘EFGHIJ’). The HIJ repeat elements formed a direct repeat cluster, whilst F and G elements (in the same orientation as HIJ) face repeat E to form an inverted repeat (Figure 1:6). A full analysis of the E11 sequence (including a schematic diagram of the entire E11 fragment with annotated sequence showing the locations of repeat elements, open reading frames and promoter sequences) can be found in Appendix Three.
Transposon mutagenesis has been used to demonstrate that an integral ORF-1 is required for mini-RepFlB replication. Maxicell analysis of translation products from the E11 fragment failed to identify a protein with an appropriate molecular weight corresponding to ORF-1 (Saul et al., 1989). However, a protein of = 40kDa was identified from a spontaneous copy mutant (pMA4322) generated from a mini-RepFlB plasmid. The sequence of the minimal replicon of the copy mutant has been obtained and demonstrates that the copy mutation is the result of a single residue change in the ORF-1 coding sequence. Saul et al. suggest that the = 40kDa protein corresponds to the ORF-1 product, and was detectable by maxicell analysis due to its higher expression resulting from a higher mini-plasmid copy number. Open reading frame one (ORF-1) was therefore tentatively named repA (RepA is the protein expressed by the gene).

Deletion analysis of both the F plasmid f7 and P307 plasmid E11 fragments demonstrated that the incE determinant characteristic of the RepFlB replicon corresponded to the E11 EFGHIJ repeat elements (Gardner et al., 1985; Spiers, MSc. thesis, 1989; and Saul et al., 1989). In an analysis of the RepFlB-carrying fragment of ColV3-K30, Perez-Casal and Crosa (1989) have suggested that incE incompatibility is determined by the EFG repeats, and that the HIJ repeats form a second and distinct inc locus (incF). However, in trans incompatibility experiments do not support this division.

An analysis of the sequence of E11 demonstrated that although RepFlB sequence was unique, the genetic organisation of repeat elements and the open reading frame within the minimal replicon was very similar to several Step function class replicons (such as mini-F, mini-P1 and R6K). In addition, the predicted amino acid sequence of the protein expressed by the ORF has a significant degree of homology with the repA initiator gene of mini-P1. Subsequent sequence analysis of other RepFlB replicons has shown that there is remarkably little variation among replicons (Kim, 1988; Saul et al., 1989; Perez-Casal and Crosa, 1989; Perez-Casal et al., 1991; Gibbs, MSc thesis, 1991; see Appendix Three, A1.8)\textsuperscript{14}.

1.3.3 A simplistic RepFlB model

The arrangement of ORF-1, DNA repeats and sequence features associated with other plasmid replicon origins allowed an early description of the RepFlB replicon, despite relatively little experimental data other than sequence information\textsuperscript{15}. The arrangement of these features also suggested the following basic model for the control of RepFlB replication in which ORF-1 expresses an autoregulatory initiator protein (RepA) capable of binding to both the incE and origin repeat elements. RepA would be titrated at the incE repeats in a manner which determines the correct plasmid copy number, and when bound to the origin repeats initiates plasmid replication. However, even in this simplistic model the autoregulation of ORF-1 and titration of RepA at incE would lead to the control paradox faced by replicons such as mini-P1, Rts1, R6K and mini-F.

\textsuperscript{14} See Appendix Three, A3.8.
\textsuperscript{15} For a full sequence analysis of E11, see Appendix Three.
Since this simplistic model of RepFIB replication control suffers from the same problems as do current models for mini-P1 and mini-F, the model in itself is not particularly good in describing the control of RepFIB replication. The value of the simplistic model is that it relates RepFIB to mini-P1 and mini-F replicons in terms of the analogous organisation of origin, initiator gene and flanking repeat elements in all three replicons. Regardless of whether this relationship is analogous (in the case of mini-F) or homologous (possibly in the case of mini-P1), the relationship itself can be used to predict elements or features of the control system of RepFIB. These predictions cover the activities of RepA; the location and control of the repA promoter, the function of the repeat elements and the location and requirements of the origin of replication.

1.4 Aims of this thesis

The past and present research of RepFIB has been aimed at gaining an understanding of the genetic elements involved in the control of RepFIB replication. Although no satisfactory control model for mini-P1 has yet been described, the organisational similarity between mini-P1 and RepFIB has allowed a number of predictions to be made about features or aspects of the RepFIB control system. In particular, the similarity has suggested the location of the origin of replication and the general function of the repeat elements arrayed on either side of repA as well as characteristics of the initiator protein RepA.

This thesis tests some of the predictions made about RepFIB based on the organisational similarity between RepFIB and mini-P1. In particular, this thesis examines the origin of replication and some of the requirements for origin activity; the general nature of the two groups of repeat elements flanking repA; the expression and control of repA; and the interaction between RepA and RepFIB DNA in vivo and in vitro.

The aims of this thesis have been to establish the fundamental features of the control system in RepFIB. Although the nature of some of these features may lead to unresolved control paradoxes, an understanding of each of the elements in the control system is necessary before a complete model can be formulated. If such a model becomes available, a comparison with other Step function replicon models may allow a general control model, independent of replicon sequence, to be formulated.

NOTE
The organisation of this Thesis differs from the format normally seen in Ph.D. theses. The 'Materials and Methods' chapter has been divided in two and is found in Appendices One (Protocols) and Two (Plasmids). Appendix Two contains a description of the construction of all plasmids made specifically for the work presented in this Thesis. A detailed analysis of the sequence features of the P307 EcoRI E11 fragment is given in Appendix Three. The statistical and sequence examination of RepFIB
contained within this appendix is referred to constantly in the research chapters of this Thesis.
Appendix Three also contains an annotated sequence of the E11 fragment which may be of help to
the reader. A more extensive description of several well-studied replicons (including oriC) is provided in
Appendix Six for those readers requiring information in addition to that provided in this chapter.
CHAPTER 2
Active promoter elements in the E11 fragment

2.1 Introduction

2.1.1 Promoter elements of interest
A knowledge of active promoter elements in the E11 fragment is a necessary step in the identification of transcriptional and translational units of the fragment. The identification of such units is required for the understanding of RepFLB replication and copy number control mechanisms. Since the E11 fragment contains a functional replicon, the fragment has the potential to include two different types of promoter elements: those which express genes resulting in the production of proteins, and promoters which produce untranslated RNA transcripts which have a functional role.

Maxicell analysis has shown that the E11 fragment expresses at least three proteins which correspond well with the predicted molecular weights of proteins from three of the seven ORFs and implies that at least three promoter sequences are active within the fragment. The presence of the RepFLB replicon within the E11 fragment also raises the possibility that untranslated RNA transcripts might also be expressed from some promoter sequences. Although untranslated RNA molecules (RNA counter-transcripts and initiator molecules) are usually associated with Hyperbolic function replicons and not the Step function replicons (which appear to include RepFLB), RNA transcripts may be involved in origin activation and control of RepFLB replication.

2.1.2 Sequence analysis of the E11 fragment
Regions of the E11 fragment were analysed for potential promoter sequences according to Mulligan et al. (1984) and likely sequences assigned a probability reflecting the degree of homology with the Escherichia coli consensus promoter sequence. Although the nature of most promoter mutations suggested that the consensus sequence corresponds to maximal E. coli promoter function, the identification of a promoter within a sequence cannot be determined solely by sequence comparison. Promoter sequence, Shine and Dalgarno (SD) sequence, start codon, spacer size and sequence all affect the expression of a gene, yet no general consensus sequence including all of these elements is available for use. For practical reasons, the entire E11 fragment was not analysed for promoter sequences. However, the upstream regions of all open reading frames were examined.

1 Potential promoter sequences are listed in Appendix Three, Table A3:5.
The analysis of these regions identified a number of potential promoter sequences with significant TargSearch scores. Open reading frames one (repA), two, five and seven were preceded by one or more potential promoter sequences. In the case of repA, the gene was preceded by twelve different sequences sharing various degrees of homology with the consensus sequence. In addition, the upstream regions of repA and ORF-5 which include the origin (BCD) and titration repeat elements (EFGHIJ) respectively, also contain promoters which are orientated in the opposite direction to that of repA and ORF-5. In order to determine which of the eighteen promoter sequences identified in the E11 fragment were functional promoter elements, an in vivo test of activity using promoter-probe plasmids was initiated.

2.2 Promoter-probe plasmids used to identify active promoters

Two different types of promoter-probe plasmids were used to test whether particular pieces of DNA contained active promoter sequences (Figure 2:1). Gene fusions were formed between a potential promoter sequence and a plasmid 'reporter' gene using pMU575 (Yang and Pittard, 1987). The reporter gene in pMU575 is itself a galK: lacZ protein-fusion gene. When galK: lacZ is linked to an active promoter, an easily assayed hybrid β-galactosidase enzyme is expressed. pMU575 provides a SD sequence and a start codon for the fusion. Hence, the expression of the hybrid β-galactosidase relies solely on the provision of an active promoter sequence inserted upstream of the reporter gene.

Gene fusions were also formed using pKK232-8 (Brosius, 1984). When an active promoter element is inserted into this plasmid, chloramphenicol acetyltransferase (CAT) is expressed resulting in chloramphenicol resistance of the host.

In contrast, protein fusions using plasmids such as pMLB1034 (Shultz et al., 1982; see also Berman, 1983) requires that the inserted piece of DNA includes a functional promoter, SD sequence, start codon and a small section of coding sequence which is in-frame with the plasmid reporter gene. In the case of pMLB1034, the reporter gene is a truncated lacZ gene with the first six codons deleted. If the coding sequence associated with the potential promoter sequence is in-frame with lacZ, then a fusion protein will be translated from the hybrid mRNA. In many cases the fusion protein retains the β-galactosidase enzymatic activity, and in some cases the fusion protein may also retain the activity of the other protein used in the fusion.

In general, gene fusions were used when it seemed unlikely that the potential promoter sequence was associated with an open reading frame, or when the ORF had a low statistical probability of expressing a protein. Protein fusions were used when the potential promoter was associated with an ORF with a high probability of expression which also included a suitable restriction site with which to form an in-frame fusion when ligated into the promoter-probe plasmid. In several cases, the pNM480...
A series of plasmids were used to produce protein fusions instead of pMLB1034. These plasmids are very similar to pMLB1034, but contain a more extensive set of restriction sites which can be used in ligation and allow the formation of fusions in any of the three reading frames (Minton, 1984).

**Figure 2:1** Different types of promoter-probe plasmids

The use of different promoter-probe plasmids allows the formation of gene fusions (A) and protein fusions (B). The only requirement for a gene fusion is that the insert DNA contain a functional promoter sequence. The mRNA from the gene fusion is translated to produce the reporter enzyme. Protein fusions require that the insert contains a functional promoter sequence and the start of an open reading frame (ORF). This ORF must have a Shine and Dalgarne (SD) sequence and a translation start codon (ATG). In addition, the ORF must be in the same frame as the reporter gene of the plasmid to form a fusion protein. The mRNA from the protein fusion is translated to produce a hybrid protein comprised of the N-terminal domain of the ORF fused to the reporter enzyme.

**2.3 Identification of active promoters in the E11 fragment**

**2.3.1 Expression of a repA:β-galactosidase fusion protein**

The BamHI 0.97kb fragment from pNZ945 was used to form an ORF-1 (repA) fusion protein in pMLB1034 (pAS7). This fragment contains the BCD repeat elements and approximately half of the repA coding region. The analysis of β-galactosidase activity expressed by cells containing pAS7 indicates that the 0.97kb fragment contains an active promoter element (Table 2:1). However, computer analysis of the sequence has shown that the fragment contains twelve potential promoter elements capable of directing the expression of repA. The repA promoter (repAp) was identified using a series of promoter-probe plasmids based on pMLB1034 and is described in Chapter Three.

---

2 Details of all plasmids constructed for this thesis are supplied in Appendix Two.

3 Potential promoter sequences are identified by numbers listed in Appendix Three, Table A3:5. Active promoters determined in this Thesis are named and are designated with a subscript 'p'.
The *BamHI*-*PstI* 200bp fragment containing *repA* has also been used to form gene fusions in pMU575 (pAS33) and pKK232-8 (pAS21).

### TABLE 2:1 Promoter-probe analysis of the expression of open reading frames

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Fragment</th>
<th>β-Gal. a</th>
<th>Protein fusion b</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Promoter-probe plasmids containing protein fusions</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>repA</td>
<td>pAS7</td>
<td><em>BamHI</em> 0.97kb</td>
<td>1271 ± 89</td>
</tr>
<tr>
<td>ORF-3</td>
<td>pAS52</td>
<td><em>EcoRI</em>-<em>BamHI</em> 0.28kb</td>
<td>4.9 ± 0.4^d</td>
</tr>
<tr>
<td>ORF-4</td>
<td>pAS53</td>
<td><em>EcoRI</em>-<em>BamHI</em> 0.28kb</td>
<td>2.4 ± 0.5^d</td>
</tr>
<tr>
<td>ORF-5</td>
<td>pAS54</td>
<td><em>BamHI</em> 0.33kb</td>
<td>3.0 ± 0.3^d</td>
</tr>
<tr>
<td>ORF-7</td>
<td>pAS69</td>
<td><em>PstI</em>-<em>PfMI</em> 1.13kb</td>
<td>-2.9 ± 1.0</td>
</tr>
<tr>
<td>ORF-7</td>
<td>pAS69</td>
<td><em>PstI</em>-<em>PfMI</em> 1.13kb</td>
<td>-2.9 ± 1.0</td>
</tr>
<tr>
<td>ORF-7</td>
<td>pAS50</td>
<td><em>SnaBI</em>-<em>BamHI</em> 0.27kb</td>
<td>1.5 ± 0.8^d</td>
</tr>
<tr>
<td>ORF-7</td>
<td>pAS44</td>
<td><em>BglII</em> 0.67kb</td>
<td>1483 ± 44</td>
</tr>
</tbody>
</table>

| Promoter-probe plasmids (with no inserts) c  | | | |
| pNM480 | -3.9 ± 1.0 | No insert DNA. |
| pNM481 | -2.5 ± 0.8 | No insert DNA. |
| pNM482 | -3.3 ± 0.5 | No insert DNA. |
| pMLB1034 | -1.2 ± 0.7 | No insert DNA. |

a Plasmids were used to transform DH5α. β-Galactosidase activities were determined from mid-log phase cells. Mean Miller units with standard errors are shown.

b All protein fusion plasmids except pAS7 have been constructed using pNM480, pNM481 or pNM482. pAS7 has been constructed using pMLB1034.

c The negative β-galactosidase activities are the result of light scattering from cell debris (determined at 550nm) and no detectible β-galactosidase activity (determined at 420nm). In comparison, where there is some β-galactosidase activity (such as with pAS53), the 550nm measurement is similar, but the 420nm measurement is great enough to result in a positive value according to the equation for Miller units of β-galactosidase activity.

d Despite the low level of expressed β-galactosidase activity, colonies with these plasmids on L-plates with Xgal will turn pale blue after an overnight incubation at 37°C and between 1-2 days incubation at room temperature.

#### 2.3.2 Expression of ORF-3:β-galactosidase fusion proteins

ORF-3 is believed to extend across the *EcoRI* site at 4298cbp and as a result contains neither the start codon nor the gene promoter^4. The *EcoRI*-*BamHI* 0.28kb and *BamHI* 0.33kb fragments from pNZ945 covering the N-terminal portion of ORF-3 have been tested for their ability to form protein fusions with pNM480, pNM481 and pNM482. Both of these internal ORF-3 fragments contain in-frame translation start codons which may allow the expression of a truncated ORF-3 protein. Colonies which weakly expressed β-galactosidase fusion proteins were isolated containing recombinant plasmids with both the 0.28kb and 0.33kb fragments (pAS52 and pAS54, respectively).

---

4 See Appendix Three, A3.2.2 and A3.7.2.
In addition, the 0.28kb fragment also directed the very weak expression of a fusion protein when in the opposite orientation to that of pAS52 (in pAS53)\(^5\) (Table 2:1).

An analysis of cells containing these plasmids demonstrated that the β-galactosidase activity of the cells was significantly greater than that of cells containing only the promoter-probe plasmids (Table 2:1). However, the magnitude of the β-galactosidase activities suggests that the fragments only contribute a translation start codon to the fusion and do not contain active promoter elements. In these cases the RNA transcript expressing the fusion protein originates from a vector promoter element, not from a RepFlIB promoter element introduced into the promoter-probe plasmid.

### 2.3.3 Expression of an ORF-4:β-galactosidase fusion protein

The upstream region of ORF-4 does not include any promoter elements within 200bp of the start codon. However, an active promoter element situated = 700bp upstream of ORF-4 has been located (EF\(_p\), see below). In order to determine whether an EF\(_p\) directed transcript was capable of expressing ORF-4, a fusion protein was formed using a derivative of pAS7 (pAS69). Analysis of the β-galactosidase activity expressed by cells containing pAS69 suggest that the fusion protein is not expressed at all (Table 2:1).

However, the expression of β-galactosidase from pAS31 (a gene fusion plasmid) indicates that the promoter EF\(_p\) is active. These two pieces of information suggest that either i) the EF\(_p\) transcript does not reach the beginning of ORF-4; or that ii) although the transcript extends through ORF-4, the reading frame is not translated.

### 2.3.4 Expression of an ORF-5:β-galactosidase fusion protein

An ORF-5 protein fusion was formed by cloning the Smal-BamHI 0.27kb fragment from pAS18 into pNM481 (to give pAS50). This fragment of DNA includes the HIJ repeat elements and almost half of the ORF-5 coding region. pAS50 expresses a very low level of β-galactosidase activity, suggesting that promoter 18 is inactive (Table 2:1) and that the Smal-BamHI fragment only contributes a translation start codon to the fusion protein.

### 2.3.5 Expression of an ORF-7:β-galactosidase fusion protein

ORF-7 was also demonstrated to have an functional promoter sequence by cloning the BglII 0.67kb fragment from pNZ945 into pNM480 to give pAS44. This fragment contains almost all of the ORF-7 coding sequence, and the expression of a fusion protein by the plasmids indicates that either promoter 1 or 2 is active (Table 2:1). Promoter 1 is located 30bp upstream of the start of ORF-7 whilst the sequence of promoter 2 overlaps both the SD sequence and the start of ORF-7. This suggests that promoter 1 is more likely to be responsible for the expression of ORF-7, rather than promoter element 2 (therefore, ORF-7\(_p\) refers to either promoter element; Table 2:1).

---

\(^5\) Initiation of translation is likely to occur at one of three in-frame 'GTG' codons located at 4201, 4210 and 4216bp.
2.3.6 ORF-2 and ORF-6
Despite repeated attempts, an ORF-2 fusion was not obtained using pNM481 and either a BgII-BamHI 0.91kb fragment or an EcoRI-BamHI 2kb fragment containing almost all of the ORF-2 coding region. ORF-6 lacks a suitable restriction enzyme site which would allow the formation of an in-frame fusion, and as a result, ORF-6 expression has not been tested.

2.3.7 Active promoters not associated with ORFs
Promoter elements 6 and 17 were shown to active promoter elements through the use of gene fusions in pMU575. Since neither promoter sequence was closely linked to an ORF, protein fusions could not be produced. Promoter 6 was cloned into pMU575 as a BamHI fragment from pAS4 to give pAS32, and promoter 17 was cloned as a HindIII-BamHI fragment from pNZ945 to give pAS31. Due to the positioning and orientation of these promoter elements, they have been named orип (promoter 6) and EFп (promoter 17) respectively (these promoters are described in Chapter Four).

The BamHI-PstI 200bp fragment containing repAп and orип has been inserted into pKK232-8 in both orientations, resulting in plasmids which express CAT (repAп in pAS21; and orип in pAS22).

2.4 Summary of active and inactive promoter elements identified in the E11 fragment

A number of active promoter elements have been identified in vivo using promoter-probe plasmids. Of the eighteen potential promoter sequences identified by computer analysis6, three have been shown to be active and 13 are inactive (one or both of the remaining promoters is also active). In general, active promoter elements had been assigned high scores (Figure 2:2). Of the seven major open reading frames in the E11 fragment, only repA and ORF-7 appear to be preceded by an active promoter sequence (repAп and ORF-7п; Table 2:2 and Figure 2:2). These two open reading frames were two of the three open reading frames predicted most likely to express a protein based on a statistical analysis of the ORF coding sequences7. The third ORF (ORF-5) has an associated potential promoter sequence, but the sequence does not appear to be transcriptionally active.

Previous maxicell analysis of the translational products of the E11 fragment suggested that at least three proteins (= 40kDa, 10.8kDa and 9.8kDa) were expressed (Saul et al., 1989). Of these three, the ~ 40kDa protein is expressed by repA and the 9.8kDa protein is most likely to be expressed by ORF-7. The statistical analysis of the open reading frames had suggested that ORF-5 was likely to express an 11.63kDa protein, but an ORF-5 protein fusion in pNM481 indicates that the open reading frame is not expressed. The next most likely coding region for the 10.8kDa protein is ORF-6 which

6 Listed in Appendix Three, Table A3:5.
7 Listed in Appendix Three, Table A3:1.
originally was rated as having a poor likelihood of expression. However, the ORF can code for a protein of 11.21kDa. ORF-6 expression has not been investigated directly through a use of gene or protein fusion plasmids.

\[\text{Figure 2.2} \quad \text{Promoter sequences: active promoters and TargSearch scores}\]

The E11 fragment contains a number of potential promoter sequences. The use of promoter-probe plasmids have identified three active promoters (\(\rightarrow\)) in the RepFIB replicon-portion of the E11 fragment (A). Each of these promoters are embedded within copies of the A-K repeat elements. A fourth active promoter has been identified which is not associated with the RepFIB replicon (not shown, ORF-7\(_p\)). Potential promoter elements were initially identified using TargSearch and were assigned percentage score where a value above 30% was considered significant (B). Of the 18 potential sequences in the E11 fragment, only four were found to be active (C). Inactive promoters are represented by '●' or '○' (where potential sequences have identical -35 sequences and different -10 sequences). ORF-7\(_p\) is either promoter 1 or 2. See Appendix Three, Table A3:5 for a list of potential promoter sequences.

Two additional active sequence elements not associated with open reading frames have also been identified. One promoter appears to produce transcripts which will cross the origin of replication and has therefore been designated 'ori\(_p\}'. The second promoter produces transcripts directed against repA\(_p\). Because the promoter is embedded within the EF repeat elements, it has been designated 'EF\(_p\}'. Transcripts originating from EF\(_p\) might extend as far as ORF-4. However, a ORF-4: lacZ promoter-probe plasmid does not express the fusion protein and suggests that the EF\(_p\) transcript is not translated.

Three of the active promoters identified through use of gene and protein fusions are located within the boundaries of the basic RepFIB replicon. All three promoter elements are associated with DNA repeat elements presumed to bind RepA protein. Two promoter elements are embedded within the sequences

---

8 Since the EF\(_p\) transcript is in the opposite sense to the repA\(_p\) transcript and covers the same stretch of DNA, the EF\(_p\) transcript could be described as a repA counter-transcript RNA. However, due to copy control system connotations (see Appendix Six, A6.1, A6.2 and A6.3), this term has not been used here.
of the repeat elements, and the third, repAp, is located immediately adjacent to the nearest repeat (element D). The placement of the promoters suggests the possibility that each promoter is affected by RepA, and that each promoter may play a role in RepFlB replication or copy number control.

Of the three promoter elements, repAp most clearly has a role in replication and, possibly, in the control of replication (Figure 2:2). The possibility of repAp autoregulation by RepA is addressed in Chapter Three. Due to the location and orientation of orip, the promoter may also be regulated by RepA and the promoter may play a role in replication (the regulation of orip by RepA is addressed in Chapter Four).

### Table 2:2 Active and inactive promoter sequences

<table>
<thead>
<tr>
<th>Promoter</th>
<th>Association</th>
<th>DNA fragment</th>
<th>Plasmid</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Active Promoter Elements</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1, 2</td>
<td>ORF-7p</td>
<td>ORF-7</td>
<td>BglII 0.67kb</td>
</tr>
<tr>
<td>6</td>
<td>orip</td>
<td>origin</td>
<td>BamHI-PstI 0.2kb</td>
</tr>
<tr>
<td>8</td>
<td>repAp</td>
<td>repA (ORF-1)</td>
<td>BamHI 0.97kb</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>BamHI-PstI 0.2kb</td>
</tr>
<tr>
<td>17</td>
<td>EFp</td>
<td>unknown</td>
<td>HindIII-BamHI 0.5kb</td>
</tr>
<tr>
<td><strong>Inactive Promoter Elements</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4, 5, 7, 9, 10, 11, 12, 13, 14, 15, 16</td>
<td>repA (ORF-1)</td>
<td>BamHI 0.97kb</td>
<td>pAS7d</td>
</tr>
<tr>
<td>18</td>
<td>ORF-5</td>
<td>SnaBl-BamHI 0.27kb</td>
<td>pAS50</td>
</tr>
</tbody>
</table>

- **a** Potential promoter sequences are listed in Table A3:5 (Appendix Three). Promoter 3 associated with ORF-3, has not been tested.
- **b** Fragment used to produce the promoter-probe plasmid.
- **c** Promoter probe plasmids were made using pMLB1034, pNM480, pNM481, pNM482, pMU575 and pKK232-8.
- **d** A series of plasmids based on the BamHI 0.97kb fragment was used to identify the repA promoter which is located within the 200bp BamHI-PstI fragment.
- **e** This plasmid does not express significant amounts of fusion protein.

**NOTE**
A more extensive analysis of the repA promoter is given in the next chapter (Chapter Three). Chapter Four describes the behaviour of three promoters (repAp, orip and EFp) when RepA is provided in trans.
3.1 Location of the repA promoter: sequence analysis and the protocol used to identify the promoter in vivo

A number of potential repA promoter sequences were identified by a search for sequences showing similarity with the Escherichia coli promoter consensus sequence (Hawley and McClure, 1983). Although the initial search was limited to 2231-2832 cbp upstream of repA, it identified a number of possible promoter sequences, each of which were assigned a probability score according to Mulligan et al. (1984). On the basis of these scores and the relative placement of each sequence with the start of repA, the most likely repA promoter sequence was located just upstream of the PfMI site at 2267 cbp (sequence 9; see Figure 3:1; Saul et al., 1989). However, some unease at attributing it as the repA promoter (repA\textsubscript{p} was felt, as although a Shine and Dalgarno (SD) sequence was close to promoter sequence 9, the spacer distance from the SD sequence to the expected repA start codon was significantly larger than normal (= 50bp cf 5-9bp). In addition, promoter sequence 9 also lacked direct or inverted repeat elements which would be required for the expected autoregulation of repA\textsuperscript{2}.

As a result of this unease, a second search for repA promoter sequences was extended to include the BCD repeats located several hundred base pairs upstream of the first start codon. This search identified four new potential promoter sequences (promoter sequences 4, 5, 7 and 8; Figure 3:1\textsuperscript{3} located between the BamH1 site at 2001 cbp and the PfMI site at 2267 cbp. Because the promoter sequences have similar levels of homology to the E. coli consensus sequence, it was not clear which of the four was more likely to be repA\textsubscript{p}. As a result, a decision was made to identify the repA promoter through the use of promoter-probe plasmids in an in vivo genetic analysis of the transcriptional activities of each of the potential promoter sequences\textsuperscript{4}.

---

1. This region begins 100bp upstream of the first repA start codon (GTG, 2321 cbp) and finishes at the last likely repA start codon (GTG, 2832 cbp).
2. RepA autoregulation was predicted based on the close organisational homology between RepFIB and Step function replicons. Autoregulation in these replicons involves a promoter sequence embedded within the repeat elements upstream of the initiator gene (as in the case of mini-P1) or the presence of a small inverted repeat near the -35 sequence of the promoter which share homology with the larger upstream repeat elements (as in the case of mini-F).
3. Listed in Appendix Three, Table A3:5.
4. An in vivo genetic analysis was chosen over an in vitro technique such as primer extension since repA transcripts have not been identified by Northern analysis (W. Maas, personal communication to P. Bergquist) and therefore must be present at very low numbers. The expression of repA transcripts could not be increased by using a hybrid high copy number RepFIB plasmid (such as pNZ3945) because of the prediction that repA would be autoregulated (see Footnote 2). If repA is autoregulated, then hybrid plasmids carrying RepFIB at greater than normal RepFIB copy numbers would be expected to produce even lower levels of RepA mRNA than a RepFIB mini-plasmid.
The transcriptional strength of any promoter sequence can be determined by linking the promoter to a suitable promoter-probe plasmid and measuring the expression of the plasmid 'reporter' gene (discussed in Chapter 2, 2.2). Although a number of promoter-probe plasmids with different reporter genes were available, a RepA:β-galactosidase fusion was considered to be the most useful as β-galactosidase activity is readily assayed and the fusion protein could be used in further work (see Chapter Five). As a result, repA fusions under the control of one or more potential repA promoter sequences were formed using pMLB1034 (Shulte et al., 1982; see also Berman, 1983) and portions of the BamHI 0.97kb fragment which included the N-terminal region of repA (Figure 3:1 and Table 3:1).

**Figure 3:1** Location of potential promoter sequences upstream of repA

A number of potential promoter elements are located upstream of the repA gene. A schematic representation of the BamHI 0.97kb fragment containing this region includes the BCD repeat elements (triangles) on the right and the N-terminal region of repA on the left (black line, the striped region covers the most likely initiation start codons for translation) (A). The location of potential repA promoter sequences (→) are shown although promoters 10-16 are not mapped. Various portions of the BamHI fragment have been cloned into pMLB1034 to produce a number of fusion plasmids (B). The fusion between repA and lacZ in these plasmids occurs at the BamHI site marked by '✓' (protein fusions are described in Chapter Two, Figure 2:1, and a larger scale diagram of the fusion in pAS7 is provided in Chapter Five, Figure 5:1). Three of the fragments were generated using restriction enzymes and two by PCR using oligonucleotide primers #8 (located in repeat element C) and #9 (located to the right of repeat element D). The second plasmid in each pair listed to the right of the fragments contains the Ω fragment which has been inserted upstream of the repA/lacZ fusion. See also Table 3:1.
### Table 3:1  Promoter-probe plasmids used to identify repA<sub>p</sub>

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Potential promoter sequences</th>
<th>Fragment&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Ω</th>
</tr>
</thead>
<tbody>
<tr>
<td>pAS7</td>
<td>4 5 7 8 9 10-16</td>
<td>0.97kb BamHI</td>
<td></td>
</tr>
<tr>
<td>pAS25</td>
<td></td>
<td>0.97kb BamHI</td>
<td>Ω</td>
</tr>
<tr>
<td>pAS39</td>
<td></td>
<td>PCR fragment (#8)</td>
<td>Ω</td>
</tr>
<tr>
<td>pAS40</td>
<td></td>
<td>PCR fragment (#8)</td>
<td>Ω</td>
</tr>
<tr>
<td>pAS37</td>
<td></td>
<td>PCR fragment (#9)</td>
<td>Ω</td>
</tr>
<tr>
<td>pAS38</td>
<td></td>
<td>PCR fragment (#9)</td>
<td>Ω</td>
</tr>
<tr>
<td>pAS2</td>
<td></td>
<td>0.77kb PstI-BamHI</td>
<td>Ω</td>
</tr>
<tr>
<td>pAS29</td>
<td></td>
<td>0.77kb PstI-BamHI</td>
<td>Ω</td>
</tr>
<tr>
<td>pAS1</td>
<td></td>
<td>0.70kb PflMI-BamHI</td>
<td>Ω</td>
</tr>
<tr>
<td>pAS24</td>
<td></td>
<td>0.70kb PflMI-BamHI</td>
<td>Ω</td>
</tr>
</tbody>
</table>

<sup>a</sup> The number refers to the PCR forward primer used to generate the fragment.

Ω Inserted in the EcoRI site upstream of the protein fusion in each plasmid.

Specific promoter sequences identified in Figure 3:1.

All promoter sequences downstream of the PflMI site capable of expressing the fusion protein (sequences 10-16; listed in Appendix Three, Table A3:5).

## 3.2 Construction of the plasmids required to identify repA<sub>p</sub>

### 3.2.1 Initial promoter-probe plasmids

Possible repA promoters were initially sub-divided by ligating portions of the BamHI 0.97kb fragment into pMLB1034 (Figure 3:1 and Table 3:1)<sup>5</sup>. Analysis of these plasmids indicated that the majority, if not all, of the promoter activity responsible for the expression of the fusion protein was located within the 200bp BamHI-PstI portion of the BamHI fragment. The residual activity expressed by plasmids not carrying this portion of RepFIB DNA raised the question of whether the low-level expression was the result of vector-initiated transcription or simply the activity of an additional, very weak downstream RepFIB promoter. The former was shown to be the cause of the low-level expression by introducing the Omega-resistance fragment (Ω, Prentki and Krisch, 1984)<sup>6</sup> into the EcoRI site of each plasmid, resulting in the elimination of all β-galactosidase activity from the PstI-BamHI and PflMI-BamHI constructions (Table 3:2).

The difference in the levels of β-galactosidase activity between cells carrying pAS7 and pAS25 might be due to a reduction in plasmid copy number resulting from an increase in the size of the plasmid, or due to the prevention of vector-initiated transcription which is augmenting the expression of the fusion protein.

---

<sup>5</sup> Details of all plasmids constructed for this Thesis are supplied in Appendix Two.

<sup>6</sup> Ω contains the R100.1 spectinomycin-resistance gene useful for both selection and maintenance, and is flanked at either end by short inverted repeats containing transcription and translation terminators to prevent both transcription and translation read-through across the fragment’s borders (Prentki and Krisch, 1984).
Table 3:2  
Vector-initiated transcription

<table>
<thead>
<tr>
<th>Plasmids without Ω</th>
<th>Ω-derivative Plasmids</th>
<th>Differencea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasmid</td>
<td>β-Gal. activity</td>
<td>Plasmid</td>
</tr>
<tr>
<td>pAS1</td>
<td>4.9 ± 0.3</td>
<td>pAS24</td>
</tr>
<tr>
<td>pAS2</td>
<td>28.9 ± 0.9</td>
<td>pAS29</td>
</tr>
<tr>
<td>pAS7</td>
<td>904.3 ± 16.3</td>
<td>pAS25</td>
</tr>
<tr>
<td>pMLB1034</td>
<td>0.4 ± 0.4</td>
<td></td>
</tr>
<tr>
<td>PB1905</td>
<td>1.7 ± 0.6</td>
<td></td>
</tr>
</tbody>
</table>

a Percentage difference between the mean β-galactosidase activity of a PB1905 strain carrying an unmodified plasmid and a strain carrying the Ω-derivative.

PB1905 cells were transformed with a variety of repA: lacZ fusion plasmids formed using pMLB1034. Cultures (inoculated with a 1/100 dilution of an overnight culture) were grown to mid-log phase before β-galactosidase activity was assayed. Mean Miller units and standard errors are given. Note that the β-galactosidase activities reported in this table should not be compared with values elsewhere due to the use of old Z buffer in the Miller assay.

Since both pAS1 and pAS2 express significantly more β-galactosidase activity than either of the two Ω-derivative plasmids and both controls (pMLB1034 and PB1905), it is clear that the fusion protein can be expressed from a weak plasmid (non-RepFlB) promoter.

3.2.2  
Promoter-probe plasmids generated using PCR amplification

The BamHI-PstI 200bp fragment contains four possible promoter sequences (4, 5, 7 and 8) which could not be divided by restriction enzyme cloning. The polymerase chain reaction was therefore used to amplify sections of the BamHI 0.97kb fragment containing one or more of these sequences. Two forward primers were designed with sequences identical to the -35 sequences of promoters 7 and 8. PCR amplification with a suitable reverse primer (#2) generated fragments which were cloned into pMLB1034 after digestion with BamHI. The resultant plasmids were then modified by the insertion of the Ω fragment upstream of the protein fusion to produce pAS38 and pAS40.

3.3  
Analysis of the promoter-probe plasmids expressing the RepA: β-galactosidase fusion protein

β-galactosidase assays of PB1905 transformed with each of the Ω-derivative promoter-probe plasmids were used to determine which promoter sequence was active and expressed the fusion protein. The analysis of the activities is shown in Figure 3:2. From the figure, it is apparent that only pAS25, pAS38 and pAS40 express significant quantities of the fusion protein. By considering the standard errors for the mean activities for cells containing pAS40 and pAS25, it can be concluded...
that there is no significant difference in the expression of the fusion protein between these two plasmids. These results indicate that promoter sequences 4 and 5 are not involved in the expression of the fusion protein, and therefore, of repA.

The difference between β-galactosidase activities of cells with pAS38 and pAS40 or pAS25 is small, but may be significant. Cells harbouring pAS35 express 91% of the β-galactosidase activity expressed by cells containing pAS25. This difference may imply that repA expression involves two promoter elements; the majority of transcripts originate from promoter sequence 3 whilst a smaller number originate from promoter sequence 7. However, if the difference between pAS38 and pAS40 is insignificant, then promoter sequence 8 is clearly the repA gene promoter (repAp).

![Diagram](image)

**FIGURE 3:2** Identification of the repA promoter

The location of potential repA promoter sequences are shown (A) (see Figure 4:1 for the key; the left hand end points of RepFlB DNA used to produce each plasmid is shown). Promoter activity from these sequences results in the expression of the repA::lacZ fusion protein which were determined by β-galactosidase assay of mid-log phase cultures (B).

**Method:** Overnight cultures were used to inoculate 5ml L-broth cultures (1/100 dilution) which were grown to mid-log phase and assayed for β-galactosidase activity. Mean β-galactosidase activities are shown by a thick line; standard errors are shown by shading. Note that the β-galactosidase activities reported in this figure should not be compared with values elsewhere due to the use of old Z buffer in the Miller assay.
3.4 A test of repA autoregulation

3.4.1 Design of the repA autoregulation assay

The initial assay to demonstrate autoregulation of repA involved the use of two compatible plasmids, pAS9 and pAS49. pAS9 is a promoter probe plasmid similar to pAS7 formed by cloning the BamHI 0.97kb fragment from pMA4322 into pNZ338. As a result, the RepA:β-galactosidase fusion protein expressed from repAp contains a single residue alteration from the wild type RepA sequence. However, since the change in expression of the fusion protein was to be measured, the presence of the altered residue in the fusion would have no effect on the assay. pAS9 was chosen as the reporter plasmid for this experiment as it would also be compatible with the expression plasmid pAS498, whereas pAS7 would not be due to similar plasmid repicons (pMB1) and identical antibiotic resistances (to ampicillin).

Initial attempts to demonstrate autoregulation using pAS9 and pAS49 failed to produce consistent results. The major problem appeared to be related to the heat induction of pAS49 as cultures grew poorly at 42°C which was probably due to a combination of i) the cellular heat shock response, ii) the massive transcriptional activity from the λPdR promoter of pAS49 and iii) to the rapid accumulation of RepA within the cell. This meant that cells could not be grown for a sufficient period after heat induction to allow the degradation of β-galactosidase activity resulting from fusion protein synthesis prior to the elevation of temperature (persistent β-galactosidase activity would hide a decrease in fusion protein expression if autoregulation occurs).

In order to avoid the problems associated with heat induction as a means of controlling repA expression, a new expression vector was developed (pAS60). Autoregulation of repA could then be tested using this new plasmid and pAS9 (Figure 3:3).

3.4.2 Construction of a controlled source of RepA

An alternative to heat induction as a means of controlling gene expression was the use of a lac derivative promoter (pIC) and of IPTG to regulate the promoter activity. Since the RepA gene is not bounded by restriction enzyme sites which would allow the easy cloning of the coding region into an expression vector, the polymerase chain reaction was used to produce a RepA-coding DNA fragment.

The upstream primer used in this PCR was designed to make two alterations to the repA sequence: i) the repA start codon was changed from CTG to ATG; and ii) two bases before the CTG codon were altered to allow the formation of a Ncol site in the PCR fragment. These changes would allow the cloning of the fragment into pKK233.2 (Amann and Brosius, 1985) and ensure that the repA start codon was positioned at the optimal distance from the Shine and Dalgarno sequence provided by

---

8 See Chapter Six, 6.2.1.
pKK233.2. Although the DNA sequence of repA was altered, the protein sequence of the expressed form of RepA would be the same as the wild type protein.

A No IPTG induction

B IPTG induction, RepA expression

**FIGURE 3:3  Experimental assay for repA autoregulation**

The *in trans* assay of repA autoregulation consists of two plasmids, a RepA expression vector induced with IPTG, and a reporter plasmid, pAS9 (A). pAS9 carries a repA: lacZ fusion which is expressed from repA. The level of expression can easily be determined by measuring the β-galactosidase activity of the cells. When cells containing pAS60 are grown in the absence of IPTG, RepA is not expressed and the fusion protein is produced (●). However, if IPTG is added to the cells (B), RepA expressed (○) from pAS60 may bind to pAS9 and alter the expression of the fusion protein. Any RepA-effect on fusion protein expression should be reflected by a change in the β-galactosidase activity of the cells.

Initial attempts to insert the PCR fragment into pKK233.2 failed. Although repA-containing plasmids were identified by colony hybridisation, examination of the plasmid DNA by restriction enzyme digestion revealed small deletions in the repA coding region of the recombinant plasmids. These deletions may have been selected as a result of 'leaky' repA expression which was harmful to the transformant cells. If this were the case, then the provision of greater quantities of the P_{trp} repressor (LacI) might fully repress P_{trp} and allow the successful insertion of repA into pKK233.2. Greater quantities of LacI could be supplied *in trans* from a chromosomal lacI gene or from a second compatible plasmid carrying lacI.

However, both of these solutions may be awkward in any future multi-plasmid assays (due to incompatibility) or in experiments which might require specific host genotypes which lacked lacI. In order to avoid these problems, the lacI gene-fragment from pMC9 (Calos et al., 1983) was cloned.
into the EcoRI site of pKK233.2 to give pAS48. As a result of this manipulation, pAS48 has both repressor gene and promoter present at the same copy number.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure3_4.png}
\caption{Autoregulation of \textit{repA\textsubscript{p}} by RepA provided \textit{in trans}}
\end{figure}

Autoregulation of \textit{repA\textsubscript{p}} is demonstrated by the effect of RepA expression from pAS60 on the expression of a \textit{repA: lacZ} fusion protein from pAS9 in DH5\textalpha. Expression of RepA from pAS60 is induced by IPTG; expression of the fusion protein from pAS9 is controlled by the \textit{repA} promoter and the level of fusion protein expression is determined by an assay of \textit{\beta}-galactosidase activity. Mean \textit{\beta}-galactosidase activities (with standard errors) of strains carrying pAS9 (\(\square\)) or pAS9 and pAS60 (\(\bullet\) and \(\circ\)) are shown in the top graph. The bottom graph shows the corresponding culture density at the time of assay (two cultures of pAS60 + pAS9 were used; \(\bullet\) was assayed at a higher density than \(\circ\)). It is important to note that although the culture density of pAS9 + pAS60 (\(\bullet\)) at high IPTG concentrations is reduced compared to those for low IPTG concentrations, the Miller units describing \textit{\beta}-galactosidase activity already account for the difference in cell numbers. \textbf{Method:} An overnight culture was used to inoculate a 50mL L-broth culture (1/100 dilution). The culture was grown for 90 minutes before aliquots were taken and added to tubes containing various concentrations of IPTG. Culture growth and \textit{\beta}-galactosidase activity were measured 2hr after IPTG was added to the cells. The plasmids were maintained with the 40\(\mu\)g/ml kanamycin and 100\(\mu\)g/ml ampicillin.
Using pAS48, the PCR generated repA fragment was inserted downstream of the pTcr promoter to give pAS60. Analysis by SDS-PAGE of pAS60 cultures showed that the plasmid expresses RepA protein, and that RepA was visible in coomassie blue stained polyacrylamide gels thirty minutes after induction with 10μM IPTG.

3.4.3 The test of repA autoregulation
The autoregulation of repA was demonstrated using pAS9 and pAS60 in a recA strain (DH5α) to avoid potential problems resulting from recombination between the two plasmids (pAS9 and pAS60 have = 700bp of RepFlB DNA in common). The assay was carried out using mid-log phase cells in which RepA expression was induced by 0.1-150μM IPTG for two hours (Figure 3.4). Cells containing only pAS9 showed that the addition of IPTG has no effect on the expression of the fusion protein from repAp. At low IPTG concentrations (upto 1μM IPTG), the expression of the fusion protein in cells containing both pAS9 and pAS60 did not change. However, with higher concentrations of IPTG the expression of the fusion protein fell sharply to less than 1% of the unpressed level.

This observation represents strong evidence that repA is autoregulated. The design of the experiment (especially the source of RepA) is such that it can be suggested that autoregulation is via RepA protein, rather than the involvement of a second protein (in the same operon) or of a counter-transcript RNA molecule. Autoregulation is also demonstrated using different combinations of plasmids in the next section and in Chapter Four.

3.5 The minimum amount of DNA required for RepA repression of repAp

3.5.1 Introduction
Repression of a promoter element by a repressor protein generally occurs by one of two mechanisms: i) the repressor binds directly to the promoter element (or to a sequence very close to it) and prevents RNA polymerase from recognising and binding to the promoter; or ii) the repressor binds to sequences close to the promoter element and 'overhangs' the -35 and -10 promoter sequences preventing RNA polymerase recognition and binding (Figure 3.5). This last mechanism requires repressor-repressor interactions as well as repressor-DNA interactions. In the case of repAp repression by RepA, the mechanism could be identified by determining whether a promoter-probe plasmid with only repAp (pAS37) can be repressed by RepA, or whether autoregulation requires some of the repeat elements (pAS39 and pAS7) to be present for repAp repression.

3.5.2 Necessary modification of a RepA expression plasmid
An in trans test of repression using pAS37, pAS39 and pAS7 would require a new source of RepA as pAS60 is incompatible with these plasmids (due to similar replicons and antibiotic resistances). The translation cassette of pAS60 comprising of lacI9, pTcr, repA and the rrb termination sequences

CHAPTER THREE
was transferred to a new vector (pACYC184, Chang and Cohen, 1978) which is compatible with the promoter-probe plasmids. The cloning of the cassette into pACYC184 was made more simple by first introducing the \( \Omega \) fragment into pAS60 just upstream of the \( \lambda aq^R \) gene-fragment to form pAS62. This plasmid was then digested with \( SaI \) and \( Scal \) to release the translation cassette (which now includes \( \Omega \)) and the fragment inserted into the tetracycline resistance gene of pACYC184. Recombinants containing the cassette were selected using spectinomycin. The resultant expression plasmid pAS64 expresses RepA when induced with IPTG and like pAS60, carries the \( \lambda aq^R \) gene to insure adequate repression of \( p_{cr} \) when IPTG is absent.

3.5.3 RepA DNA-binding sites required for \( repA_p \) repression

The number of DNA repeat elements necessary for RepA repression of \( repA_p \) was determined using pAS7, pAS37 and pAS39 with pAS64 in DH5\( \alpha \) (Figure 3:6). RepA expression was induced with IPTG and the \( \beta \)-galactosidase activities of the cultures measured to determine the whether \( repA_p \) was repressed in any of the three promoter-probe plasmids. In all three cases, RepA was able to partially repress \( repA_p \), even though pAS37 does not contain the BCD repeat elements presumed to bind RepA. However, unlike the repression of pAS33 or pAS9 by pAS60, the repression of pAS7, pAS37 and pAS39 was only partial and the expression of the \( repA: \lambda aqZ \) fusion protein was lowered to \( \approx 50\% \) of the unrepessed expression level.

**FIGURE 3:5** Possible repression mechanism of \( repA_p \)

The \( repA \) promoter (\( repA_p \)) is autoregulated by RepA which binds to the BCD repeat elements. Two speculative mechanisms are shown in this figure. RepA may repress \( repA_p \) by binding to the BCD repeats and in doing so, 'overhang' the \( -35 \) and \( -10 \) sequences of \( repA_p \) (A). RepA overhang may therefore prevent RNA polymerase recognition and access to \( repA_p \). An alternative model of \( repA_p \) repression may rely on RepA binding to two additional DNA repeat elements located adjacent to repeat D (grey triangles). RepA bound to these elements would therefore prevent RNA polymerase binding to \( repA_p \) (B). The left-hand end-points of pAS7, pAS37 and pAS39 are shown under the schematic for A.
RepA repression of repA_p does not require the presence of the BCD repeat elements located immediately upstream of the repA promoter. The expression of a repA:lacZ fusion protein under the control of repA_p in three different promoter-probe plasmids can be repressed by RepA supplied in trans from pAS64. Mean β-galactosidase activities with standard errors of DH5α strains transformed with pAS7 (which carries the BCD repeat elements) and pAS64 (A), pAS37 (half of repeat C, plus repeat D) and pAS64 (B) and pAS37 (BCD repeats absent) and pAS64 (C) are shown. Method: Overnight cultures were used to inoculate 25ml L-broth cultures (1/100 dilution). These were grown for 90 minutes before aliquots were taken and added to tubes containing various concentrations of IPTG. β-galactosidase activity was measured 2hr after IPTG was added to the cells. The plasmids were maintained with 100μg/ml ampicillin and 25μg/ml chloramphenicol.
The inability of pAS64 to completely repress repAp (at least in pAS7) is probably due to the relative copy numbers of the promoter-probe plasmids and the expression plasmid. In the repression of pAS33 and pAS9 by pAS60, both promoter-probe plasmids were low copy number whilst pAS60 had a high copy number. In the case of the repression of pAS7, pAS37 and pAS39 by pAS64, the copy number of the promoter-probe plasmids and the expression plasmid are presumably more similar. In this instance, pAS64 could not produce sufficient RepA to fully repress all of the copies of repAp present within the cell.

The IPTG levels needed to observe repression of pAS7, pAS37 and pAS39 by pAS64 are considerably higher than that needed to observe repression using pAS60. This difference is probably an artefact of the experiment: i) if pTrc is fully derepressed in a high copy number plasmid (pAS60) by 1mM IPTG, then pTrc would be fully derepressed in a lower copy number plasmid (pAS64) where even less LacI is expressed; and ii) if the incubation period between adding IPTG and assaying for β-galactosidase activity were extended, repression by pAS64 should be detectable at lower IPTG levels.

The repA promoter (repAp) is embedded in the D* repeat element which is part of a set of five direct repeats in the origin region of RepFlB. The BCD repeats are the same as those reported in Saul et al. (1989); the D'D* repeats are the result of further analysis reported in this Thesis.

**Figure 3.7** The origin repeat elements BCDD'D* and repAp

3.6 Summary of the identification and control of repAp

Of twelve potential promoter sequences, one has been identified by *in vivo* genetic means as the primary repA promoter (repAp). Although a second promoter sequence may be responsible for the expression of some RepA, it is probably that RepA transcripts originate from a single promoter located downstream of the D' repeat element (-35: 2149-2164cbp, -10: 2178-2191cbp; see the annotated sequence provided in Appendix Three, Figure A3:1).

---

9 This presumption is based on the copy numbers of the replicons used to produce pACYC184 and pMLB1034. The actual copy numbers of the plasmids used here have not been determined.
The repA promoter is repressed by RepA supplied in trans which demonstrates that repA is autoregulated. RepA is also capable of partially repressing repAp in the absence of the BCD repeat elements. This observation suggests that either RepA recognises and binds to repAp in an unexpected manner (which is different to the recognition and binding to the repeat elements)\(^1\) or that the region between the D repeat element and repAp contains additional copies of the A-K repeat element consensus sequence. This later explanation is favoured and an analysis of the DNA sequence has identified two additional copies of the repeat element (D' and D'\(^\prime\)). D' and D'\(^\prime\) are adjacent to D and are in the same orientation as the BCD repeat group (Figure 3:7). The D' and D'\(^\prime\) repeat sequences share 15 and 14 identical bases with the repeat consensus sequence respectively\(^1\). This degree of overall similarity means that only repeats A and C share less homology than D' and D'\(^\prime\).

The -35 sequence of repAp overlies the D' element and presumably, RepA binding to D' prevents RNA polymerase access to the promoter and reduces the expression of repA.

\(^1\) The results discussed in this chapter do not prove that RepA repression of repAp is via recognition and binding to the BCDD'D'\(^\prime\) repeat elements. A second explanation might be that RepA recognises a different DNA sequence element and binds to that in order to repress repAp. However, from the results presented in Chapter Four, this appears unlikely and strongly suggests that the DNA binding of RepA requires some of the BCDD'D'\(^\prime\) repeat elements.

\(^1\) The D and D' repeats are listed in Appendix Three, Table A3:4. An additional repeat element is located to the right of D' (referred to as D''\(^\prime\)). However, D''\(^\prime\) shares less homology with the A-K repeat sequence consensus than repeat A, and as a result, has not been included in the diagrams of this Thesis.
The identification and control of the repA promoter
CHAPTER 4

In vivo repression of three promoters by RepA

4.1 Introduction

4.1.1 Regulation of three promoters found within the minimal replicon
The minimal RepFlB replicon contains three active promoters elements identified through use of gene fusion or protein fusion promoter-probe plasmids (Figure 4:1). One of these promoter elements has been identified as the repA promoter (repAp) and is embedded within the D" repeat element of the BCDD'D" origin repeat group. A second promoter element is embedded within the BC repeat elements and is orientated in the opposite sense to repAp. This promoter is referred to as the 'ori' promoter (oriP), as transcripts initiated from the promoter will extend across the RepFlB origin of replication. The third promoter is embedded within the EFGHIJ titration repeat group located at the carboxyl terminal end of the RepA gene. Since the promoter sequence is located within the E and F repeat elements, it is referred to as the EF promoter (EFp).

The repA promoter (repAp) has been shown to be autoregulated in Chapter Three. In this chapter, RepA regulation of the three promoter elements oriP, repAp and EFp are compared using agar plate assays and liquid culture assays.

4.1.2 Promoter-probe plasmids used to test the regulation of oriP, repAp and EFp
The promoter-probe plasmids used in this chapter have been described in Chapter Two and are shown in Figure 4:1. The promoter-probe plasmids pMU575 and pKK232-8 were used to construct gene fusions in which oriP, repAp or EFp control the expression of either a hybrid β-galactosidase gene (in pMU575; pAS31, pAS32 and pAS33) or the chloramphenicol acetyltransferase (cat) gene3 (in pKK232-8; pAS21 and pAS22). Protein fusions between RepA and β-galactosidase have also been formed using pMLB1034 (pAS7) and pNZ338 (pAS9). In pAS7 and pAS9 the expression of the fusion protein is controlled by the repA promoter (repAp).

The transcriptional activity of repAp and oriP are similar and can be demonstrated using a liquid-culture minimum inhibitory concentration assay (MIC) (Figure 4:2), an agar-plate MIC assay

1 Gene and protein fusions are described in Chapter Two.
2 Details of all plasmids constructed for this Thesis are supplied in Appendix Two.
3 The abbreviation used here for chloramphenicol acetyltransferase is 'CAT'.

43
(determined from data presented in Table 4:1) or through β-galactosidase assays (determined from data presented in Figure 4:4). The transcriptional activity of EF_<sub>p</sub> is approximately 10% of the activity of repA<sub>p</sub> (determined from data presented in Figure 4:4).

\[
\begin{align*}
\text{A} & \quad \text{ori}_{p}\text{repA}_p \quad \rightarrow \quad \text{repA} \quad \rightarrow \quad \text{EF}_p \\
\text{BCDDE'} \\
\text{EFG HJ} \\
\text{B} & \quad \text{ori}_{p} \rightarrow \text{repA}_p \quad \rightarrow \quad \text{pAS31} \\
\text{pAS32} \\
\text{EF}_p \\
\text{C} & \quad \text{ori}_{p} \rightarrow \text{repA}_p \quad \rightarrow \quad \text{pAS21} \\
\text{pAS22} \\
\text{EF}_p \\
\text{D} & \quad \text{repA}_p \quad \rightarrow \quad \text{pAS7, pAS9} \\

\text{FIGURE 4:1} \quad \text{Promoter-probe plasmids used in this section}
\]

The three active promoters in the minimal RepFIB replicon are shown (A). Three gene fusions have been made using pMU575 where ori<sub>p</sub>, repA<sub>p</sub> and EF<sub>p</sub> control the expression of a hybrid β-galactosidase protein (B). Two gene fusions have been made using pKK232-8 where ori<sub>p</sub> and repA<sub>p</sub> control the expression of chloramphenicol acetyltransferase (C). Two protein fusions have been made where repA<sub>p</sub> controls the expression of a hybrid RepA:B-galactosidase gene (D). pAS7 and pAS9 are high and low copy number plasmids, respectively.

4.2 Demonstration of regulation: use of agar plate assays

4.2.1 A description of the agar plate assays

A simple test of promoter regulation by RepA provided in trans can be made using plate assays where cultures are spread onto L-agar plates under different conditions. Two plate assays are described in this section and show: i) the regulation of ori<sub>p</sub> and repA<sub>p</sub> by pMA4322, as determined by the degree of host sensitivity to chloramphenicol; and ii) the regulation of repA<sub>p</sub> by pMA4322, as determined by the development of colony colour following transformation of the host. In both assays attempts were made to use pSS3928 (a RepFIB mini-plasmid) as a source of RepA. However, recA
host cells (DH5α) could not be transformed with both pSS3928 and the required promoter-probe plasmid⁴. As an alternative, pMA4322 was used to provide RepA in both assays. pMA4322 is a spontaneous copy mutant derived from pSS3928. The mutant phenotype is the result of a single residue change in the RepA sequence resulting in a higher copy number of the mini-plasmid and a more relaxed incompatibility reaction to in trans incompatibility (inc) determinants.

![Graph](image)

**Figure 4:2** Relative promoter strengths of repAp and oriP

A minimum inhibitory concentration (MIC) assay can be used to determine the relative strengths of the repAp and oriP promoters. Growth of cells containing the repAp gene fusion (pAS21) are inhibited by chloramphenicol above 100µg/ml (A). Growth of cells containing an oriP gene fusion (pAS22) are inhibited by a similar level of chloramphenicol (B). DH5α or DH5α containing pKK232-8 are inhibited by chloramphenicol concentrations above 25µg/ml (▲). **Method:** A 1/200 dilution of mid-log phase cultures were used to inoculate 100ml L-broth, 100µg/ml ampicillin. 2ml aliquots of this dilution were added to tubes containing varying amounts of chloramphenicol (0-250µg/ml). The cultures were grown overnight before culture densities were determined. Cultures were grown with 1%, 2% and 2.5% ethanol (●) to test whether cell growth was effected by the ethanol concentrations corresponding to 100, 200 and 250µg/ml chloramphenicol.

---

⁴ Co-transformation and sequential transformations of DH5α with resident pSS3928 were unsuccessful.
4.2.2 Repression of cml gene fusions by a RepFIB mini-plasmid

The BCDD'D" repeat elements contain two promoters, ori_p and repA_p. In order to determine whether ori_p was affected by RepA in trans in the same manner as repA_p, an attempt was made to repress the expression of chloramphenicol acetyltransferase (CAT) in pAS21 or pAS22 using RepA supplied in trans by pSS3928 or pMA4322. However, pAS21 or pAS22 CsCl DNA could not be used to produce transformants in DH5α containing resident pSS3928. In contrast, co-transformants containing pAS21 and pMA4322 or pAS22 and pMA4322 were easily selected on 100µg/ml ampicillin, 50µg/ml spectinomycin plates following transformation of DH5α with plasmid DNA.

Mid log-phase cultures containing pMA4322, pMA4322 with either pAS21 or pAS22, pAS21, and pAS22 were spread onto agar plates with the appropriate antibiotics to maintain the resident plasmid or plasmids. The cultures were plated at a suitable density to ensure that after a overnight incubation, a confluent lawn of bacteria would be produced. On some plates, thin strips of filter paper soaked with 20mg/ml chloramphenicol were placed in the centre of the plate⁵. Over the period of incubation, the antibiotic from the filter diffused into the surrounding agar and prevented the growth of chloramphenicol-sensitive bacteria.

Table 4:1 Repression of pAS21 and pAS22

<table>
<thead>
<tr>
<th>Plasmids</th>
<th>Promoter</th>
<th>Distancea</th>
<th>% Repression</th>
</tr>
</thead>
<tbody>
<tr>
<td>pAS21</td>
<td>repA_p</td>
<td>2.5 ± 0.1</td>
<td></td>
</tr>
<tr>
<td>pMA4322 + pAS21</td>
<td>repA_p</td>
<td>5.8 ± 0.3</td>
<td>43% repression</td>
</tr>
<tr>
<td>pAS22</td>
<td>ori_p</td>
<td>2.7 ± 0.2</td>
<td></td>
</tr>
<tr>
<td>pMA4322 + pAS22</td>
<td>ori_p</td>
<td>6.9 ± 0.2</td>
<td>39% repression</td>
</tr>
<tr>
<td>pMA4322</td>
<td></td>
<td>15.2 ± 0.2</td>
<td></td>
</tr>
</tbody>
</table>

a Distance from the edge of the filter strip to the edge of the bacterial lawn. Distances are expressed as means with standard errors.

Mid-log phase cultures were spread onto five agar plates at a sufficient density to form a confluent lawn. Whatman 540 filter paper was soaked in 20mg/ml chloramphenicol, drained and then laid across the middle of the plate. The agar plates were incubated overnight at 37°C. The following morning the zone of inhibition was measured (to the nearest 0.5mm). The plasmids were transformed into DH5α and were maintained with spectinomycin (pMA4322) and/or ampicillin (pAS21 and pAS22).

The presence of pAS21 or pAS22 allowed the development of the bacterial lawn to within a few millimetres of the filter strip (Table 4:1). However, where these plasmids were absent, the growth of the lawn was inhibited approximately 15mm away from the strip. The presence of pMA4322 appears to repress the expression of CAT in pAS21 and pAS22, as cells containing both pMA4322 and either pAS21 or pAS22 are sensitive to chloramphenicol further away from the strip than cells.

⁵ The best chloramphenicol concentration used to wet the 5mm-wide filter strips was determined by experiment.
containing only pAS21 or pAS22. The presence of pMA4322 appears to repress rep\textsubscript{Ap} and ori\textsubscript{p} by 43% and 39%, respectively\textsuperscript{6}.

4.2.3 Repression of β-galactosidase protein fusions by a RepFIB mini-plasmid

The repression of rep\textsubscript{Ap} by pMA4322 has also been demonstrated using a plate assay and two promoter-probe plasmids which express a RepA:β-galactosidase fusion protein under the control of rep\textsubscript{Ap} (pAS7 and pAS9). The two plasmids differ in copy number since pAS7 contains a pMB1 replicon (high copy number) whilst pAS9 contains a pSC101 replicon which results in a lower copy number. The difference in copy number might influence pMA4322 repression of rep\textsubscript{Ap} since the RepFIB mini-plasmid may only express sufficient RepA to repress rep\textsubscript{Ap} at low copy numbers, but not enough to repress rep\textsubscript{Ap} at higher levels.

As in Section 4.2.2 (above), pSS3928 could not be used to provide the necessary co-transformants for this assay. However, pAS7 or pAS9 co-transformants with pMA4322 in DH5α were readily obtained.

<table>
<thead>
<tr>
<th>Plasmids</th>
<th>Colony Colour</th>
</tr>
</thead>
<tbody>
<tr>
<td>pMA4322</td>
<td>all colonies were white</td>
</tr>
<tr>
<td>pMA4322 + pAS7</td>
<td>all colonies were blue</td>
</tr>
<tr>
<td>pMA4322 + pAS9</td>
<td>all colonies were white</td>
</tr>
<tr>
<td>pAS7</td>
<td>all colonies were blue</td>
</tr>
<tr>
<td>pAS9</td>
<td>all colonies were blue</td>
</tr>
</tbody>
</table>

Competent DH5α cells were transformed with pMA4322, pMA4322 + pAS7, pMA4322 + pAS9, pAS7 or pAS9 CsCl DNA and plated onto L-plates containing appropriate antibiotics and Xgal. The agar plates were incubated overnight at 37°C and the following morning colony colour determined.

After transformation of DH5α with plasmid DNA\textsuperscript{7}, the transformation culture was plated onto agar plates containing the appropriate antibiotics to select for both plasmids and Xgal to determine the expression of the RepA:β-galactosidase fusion protein. After an overnight incubation, the colour of the transformant colonies was scored (Table 4:2). All pAS9 + pMA4322 co-transformant colonies were white, suggesting that rep\textsubscript{Ap} controlled expression of the fusion protein had been effectively repressed by pMA4322. After several days storage at 4°C, the colonies started to turn a very pale blue, indicating that the repression was not absolute. In contrast, co-transformant colonies containing pAS7 + pMA4322 were all blue suggesting that rep\textsubscript{Ap} has not been repressed by pMA4322.

\textsuperscript{6} However, since the concentration of chloramphenicol as a function of the distance from the filter strip is not linear, the two percentages should not be directly compared with other experimentally determined measurements of repression.

\textsuperscript{7} Transformation was by the modified method of Lederberg and Cohen (1974). After the three minute heat shock treatment, the transformation culture was incubated at 37°C for one hour before plating.
In vivo repression of three promoters by RepA
4.2.4 A summary of the repression tests using agar plate assays

pSS3928 could not be used in an attempt to demonstrate RepA autoregulation of repA_p or repression of ori_p in trans using plate assays. However, the copy mutant pMA4322 was capable of co-transforming cells with a second plasmid carrying the BCDD'D repeats allowing a test of repression. pMA4322 is capable of repressing the expression of CAT under the control of either repA_p or ori_p present on a second plasmid (pAS21 and pAS22, respectively). pMA4322 is also capable of repressing the expression of a RepA:B-galactosidase fusion protein under the control of repA_p in pAS9. However, it appears that the relative copy number of the promoter-probe plasmids affects the ability of repression; in a low copy number vector, repression is complete (pAS9), whilst in a higher copy number vector (pAS7) repA_p is not repressed by RepA supplied in trans from pMA4322.

4.3 Demonstration of regulation: use of liquid culture assays

4.3.1 A description of the liquid culture assay

Although the regulation of ori_p and repA_p has been demonstrated using plate assays, the use of liquid culture assays also has been used to demonstrate RepA mediated repression of ori_p, repA_p and EF_p. The use of liquid culture assays allows a greater control of growth conditions; in particular, assays can be carried out during mid log phase growth when cell growth is maximal. In contrast, the growth conditions of bacteria plated onto agar may vary and the end result (for example, the expression of β-galactosidase and resultant colour development) may be the accumulative affect of several generations of growth.

The three promoters (ori_p, repA_p and EF_p) have been inserted into pMU575 where each promoter controls the expression of a hybrid β-galactosidase gene. These plasmids (pAS31, pAS32 and pAS33) are particularly useful as they are compatible in terms of antibiotic resistances and replicons with the RepA over-expression plasmid pAS60 and have low to medium copy numbers. Rather than rely on the supply of RepA from pMA4322, wild type RepA protein can be supplied on induction from the pAS60 p_tac promoter (induction is with IPTG).

FIGURE 4:3  Effect of RepA expression on cell growth

Opposite page  The induction of RepA from the expression plasmid pAS60 has a significant affect on culture growth rates between 2-3 hours after high concentrations of IPTG have been added to the culture medium. Method: An overnight culture of PB1905 was used to inoculate a 50ml L-broth culture (1/100 dilution). The culture was grown for 90 minutes before aliquots were taken and IPTG added to induce pAS60 expression. Cell growth was measured by determining the optical density at 600nm (OD600) at 1hr, 2hr and 3hr after IPTG was added to the cells (A). A marked repression of cell growth after 3hr with high IPTG concentrations (16-128μM) was seen using a PB1905 strain containing pAS60 (B). Cultures were taken from (B) after 3hr, washed to remove IPTG and used to inoculate new cultures (C). Cells derived from the original cultures with 64μM or 128μM IPTG exhibited a very poor recovery.

CHAPTER FOUR
4.3.2 Development of the liquid culture assay

Induction of proteins from pLcr in over expression plasmids usually involves the induction of a culture in mid-log phase growth with 1-10mM IPTG followed by up to five hours of incubation. Early experiments testing the expression of RepA from pAS60 had suggested that RepA was produced in significant amounts in the cell after thirty minutes after induction (as determined by SDS-PAGE), and that the level of RepA in the cell did not markedly increase after two hours of induction. In addition, after long periods of induction culture growth was inhibited and cell lysis occurred.

In order to carry out repression tests using pAS60 as a controlled source of RepA, the affect of RepA induction on PB1905 cell growth was investigated (Figure 4:5). The results of this test indicated that in order to avoid cell lysis and inhibition of growth, the optimal time to induce RepA expression was in early log phase and that cultures should be assayed after two hours of induction.

In vivo repression of three promoters by RepA
The two hour period after induction allows sufficient cell growth and division to remove any β-galactosidase present in the cells which had been synthesised before induction, yet minimises the difference in culture densities of low and high IPTG induction cultures and maintains rapid-growth cultures (low IPTG induction) in log phase whilst allowing slow-growth cultures (high IPTG induction) to reach adequate densities for the subsequent β-galactosidase assays.

PB1905 cultures containing pAS60 which had been induced with high levels of IPTG (greater than 4μM) displayed a reduction or repression in growth rate\(^8\). When cells from these cultures were washed to remove IPTG and used to inoculate new cultures, growth was very slow and may only have been the result of the development of mutants (Figure 4.3, A). Optical microscopy of cells with pAS60 induced for several hours showed the formation of inclusion bodies. The development of these bodies may be the cause of growth repression, and it is only if these bodies are small enough to be destroyed that cells can recover from RepA induction. In cases where the bodies are very large, mutants which loose the inclusion bodies during cell division may survive.

4.3.3 Repression of ori\(_p\), repA\(_p\) and EF\(_p\) by RepA supplied by pAS60

The liquid culture assay described above was used to determine whether ori\(_p\) and EF\(_p\) were affected by the expression of RepA in trans in a manner similar to the autoregulation of repA\(_p\). Cells containing one of the pMU575 derivative promoter-probe plasmids and pAS60 were grown to very early-log phase in L-broth with selection for both plasmids. The cultures were then subdivided, added to tubes containing various amounts of IPTG and allowed to grow for two hours. After this period, the cultures with very little IPTG had reached late-log phase, whilst cultures with large amounts of IPTG were at early-mid-log phase. Miller assays were used to determine the amount of β-galactosidase expressed by the cells and the data is presented in Figure 4.3.

All three promoters are regulated by RepA in trans and is shown by the decreasing expression of β-galactosidase activity in the cultures with increasing IPTG concentrations. The expression of β-galactosidase from pAS31 (containing EF\(_p\)) with no IPTG induction of pAS60 is = 45 units (Figure 4.3, A). Between 0.015-1μM IPTG the expression is not affected. However, with >1μM IPTG, EF\(_p\) controlled expression of the β-galactosidase gene is reduced to = 10% of the unreppressed level.

The expression of β-galactosidase from pAS32 and pAS33 (containing ori\(_p\) and repA\(_p\), respectively) follow similar patterns (Figure 4.3, B and C). If RepA expression from pAS60 is not induced, pAS32 expresses = 450 units of β-galactosidase activity and pAS33 expresses = 300 units. Between 0.015-1μM IPTG the expression of β-galactosidase from both promoter probe plasmids is not affected. However, with >1μM IPTG, ori\(_p\) and repA\(_p\) controlled expression of the β-galactosidase gene is reduced to = 20% (pAS32) and = 30% (pAS33) of the unrepressed levels.

---

\(^8\) This effect is not dependent on the genotype of the host strain as other strains (DH5\(\alpha\) and JM83) show the same effects when RepA is expressed from an over-expression plasmid such as pAS49, pAS60 or pAS64.
4.3.4 A summary of the repression tests using the liquid culture assay

Four important observations can be made from the data presented in Figure 4.3. The first is that the liquid culture assays demonstrate that RepA supplied in trans from an over-expression plasmid is able to repress the activity of the ori_p, repAp, and EFp promoters. The second observation is that the repression of EFp and ori_p begins when \( = 1 \mu M \) IPTG is added to the culture medium, whereas the repression of repAp begins with \( = 0.25 \mu M \) IPTG. This observation suggests that repAp is more sensitive to RepA than either of the other promoters, despite the fact that both repAp and ori_p share the same RepA-binding elements (the BCDD'D" repeats). The third observation is that although EFp and ori_p respond to increasing RepA expression quickly (indicated by the slope of the curve between 1-128 \( \mu M \) IPTG), the repAp response is a lot slower. This observation may imply that RepA binding to the repeat elements is co-operative. However, this explanation does not account for the fact that the repression slope of ori_p and repAp are not similar.

A second explanation may be that RepA binds with high affinity to the BC repeat elements and with low affinity to the D'D" repeat elements. If this were the case, then EFp would be more sensitive to low concentrations of RepA since the promoter is embedded in the EF repeats, whereas higher concentrations of RepA would be required to repress repAp because it is embedded within the D" repeat element.

The final observation is that although RepA represses ori_p, repAp, and EFp, the three promoters are not completely repressed by RepA supplied in trans. The partial repression of repAp determined in Section 4.3.3 is in contrast to the almost complete repression determined in Chapter Three, Section 3.4.3. In both tests RepA was supplied by pAS60. The repression of the promoter probe plasmid pAS9 (Section 3.4.3) appeared to begin with IPTG concentrations of greater than \( 1 \mu M \), whilst the repression of pAS33 (Section 4.3.3) began with concentrations greater than \( 0.125 \mu M \).

If the copy number of pAS9 was greater than that of pAS33, the amount of RepA needed to repress repAp in pAS9 should be greater than the amount required for repression of pAS33. The comparison of these two tests suggests that this may be true. However, the argument might be extended to suggest that the amount of RepA to fully repress repAp in pAS33 should be lower than that needed for pAS9; which in turn suggests that complete repression of pAS33 should be seen at a lower IPTG concentration than for pAS9. The data does not support this argument.

An alternative explanation relies on the difference of culture age (as determined by measurement of the OD600 values) between the two tests. In the repression of pAS9, the cultures were between 1.1 and 0.4 OD600 units at the time of assay, whereas in the repression of pAS33 the cultures were all approximately 0.3 units. The partial repression of pAS33 may therefore be the result of an inadequate growth period from the time of RepA induction to the time of assay (the actual time in both cases was two hours, but the number of generations will be different as the pAS33 cultures were in early-log phase when assayed). This period may have been too short to allow the degradation
of fusion protein which had been expressed in the cells prior to RepA induction before the cultures were assayed for β-galactosidase activity.

This argument reinforces the considerations regarding to the design of liquid culture assays of promoter regulation by a repressor supplied in trans (Section 4.3.2): the growth phase of the cultures and the timing of induction and assay all have an affect on the outcome of the test. However, despite these problems the liquid culture assays demonstrate that RepA is clearly able to repress the activity of three promoter elements, ori_p, repA_p and EF_p.

**FIGURE 4:5** RepA affinity for different repeat groups

RepA affinity for each of the different repeat groups (BCDD'D", EFG, HIJ and EFGHIJ) can be determined using a three-plasmid in vivo assay. RepA is expressed from pAS64 when induced with IPTG. pAS9 contains a repA: lacZ protein fusion under the control of the repA promoter (repA_p). RepA repression of repA_p occurs by RepA binding to the BCDD'D" repeat elements present in pAS9 and is simply determined by β-galactosidase assay. A third compatible plasmid carries additional copies of the repeat elements (either BCDD'D", EFG, HIJ or EFGHIJ). RepA expressed from pAS64 can either bind to the pAS9 BCDD'D" repeats and repress repA_p, or to the repeats present in the third plasmid, in which case repA_p in pAS9 will not be repressed.

4.4 RepA DNA-binding affinities for the different groups of repeat elements

4.4.1 Introduction

RepA binds both to the BCDD'D" and the EFG repeat elements and in doing so, represses the activity of three promoters embedded within the repeat element sequences. Although the consensus sequence of the BCDD'D" repeats does not markedly differ from either the EFG repeat consensus or the ABCDD'D"EFGHIJ repeat consensus, the orientation of the repeats within each group may lead to differences in RepA recognition and binding affinity to the repeat DNA. The BCDD'D" repeats are all aligned in the same orientation with similar spacings between each element, whilst the EFG repeats contain an inverted repeat (EF) and a direct repeat (FG) with a greater spacing between the EF
repeats than that found between repeats of the BCDD'D' group. Although the HIJ repeats are all in the same orientation, they do not form a perfect 'direct' repeat group. The H and I repeats overlap each other and there is a substantial spacer between the I and J repeat elements.

![Diagram](image)

**Figure 4:6** Repeat elements present on the 'third' plasmids

A number of different plasmids carrying copies of the A-K repeat elements have been produced using pBS+. The 2kb Omega fragment (Ω) was inserted into the SnaBI site located between the EFG and HIJ repeat groups in pAS15 to produce pAS65. pAS9 carries the same repeat elements as pAS4.

A three plasmid assay was developed to determine whether the BCDD'D", EFG, HIJ or EFGHIJ repeats bound RepA in a similar manner. In the assay, RepA is provided *in trans* from an expression plasmid (pAS64) when induced with IPTG. RepA is then able to bind to the BCDD'D" repeats present in pAS9 (the reporter plasmid), repress repA and thus reduce the expression of a repA: lacZ fusion protein. However, if a third plasmid is present carrying a copy of the BCDD'D" repeats, repA in pAS9 may only be partially repressed as RepA can bind to the repeats present in both plasmids. The relative RepA-binding affinities of the EFG, HIJ and EFGHIJ repeat groups could be determined by replacing the third plasmid carrying the BCDD'D" repeats with plasmids carrying instead the EFG, HIJ or EFGHIJ repeat elements. In this manner, RepA affinity for the different sets of repeat elements is determined by the binding competition between the repeats of the third plasmid with the BCDD'D" repeats present in pAS9.

### 4.4.2 The three-plasmid assay

The BCDD'D", EFG, HIJ and EFGHIJ repeat elements have been inserted into pBS+ to give pAS4, pAS16, pAS17 and pAS15 respectively. Although a three plasmid assay is complex in terms of plasmid compatibility, the pBS+-derived plasmids with pMB1-type replicons and ampicillin resistance genes are compatible with both the reporter plasmid, pAS9 with a pSC101 replicon and a kanamycin resistance gene, and with the RepA expression plasmid, pAS64 which has a p15A-type replicon and chloramphenicol and spectinomycin resistance genes. β-galactosidase assays were made of mid-log phase DH5α culture pairs, one of which had been induced with 64μM IPTG (Table 4:3).
The outcome of these assays was either a fully repressed or a fully de-repressed repA promoter, rather than a range of values falling between these two extremes. The Boolean state of repA<sub>p</sub> (either active or repressed) is probably the result of the difference in the relative copy number of pAS9 (containing repA<sub>p</sub>) and the 'third' plasmids. The copy number of the 'third' plasmid repeat elements would be significantly higher than the copy number of the BCDD"D" repeats present on pAS9. If RepA acted simply as a repressor capable of binding to only one set of repeats, the predicted experimental outcome using pAS4 as the 'third' plasmid would be the complete de-repression of repA<sub>p</sub> in pAS9, since the majority of RepA would bind to the pAS4 BCDD"D" repeats and relatively few RepA molecules would be free to bind the pAS9 BCDD"D" repeats and repress repA<sub>p</sub>.

### TABLE 4.3

<table>
<thead>
<tr>
<th>Plasmids&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Repeats</th>
<th>β-Galactosidase Activity</th>
<th>No IPTG</th>
<th>64μM IPTG</th>
<th>%Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>pAS9</td>
<td>pAS64</td>
<td>560.8 ± 11.7</td>
<td>23.3 ± 3.7</td>
<td>4.2</td>
<td></td>
</tr>
<tr>
<td>pAS9</td>
<td>pAS64 pAS4</td>
<td>556.7 ± 14.4</td>
<td>35.5 ± 2.5</td>
<td>6.3</td>
<td></td>
</tr>
<tr>
<td>pAS9</td>
<td>pAS64 pAS59</td>
<td>438.7 ± 26.0</td>
<td>11.9 ± 9.4</td>
<td>2.7</td>
<td></td>
</tr>
<tr>
<td>pAS9</td>
<td>pAS64 pAS58</td>
<td>425.8 ± 41.7</td>
<td>0.0 ± 8.5</td>
<td>&lt;0.2</td>
<td></td>
</tr>
<tr>
<td>pAS9</td>
<td>pAS64 pAS16</td>
<td>510.7 ± 7.7</td>
<td>29.1 ± 5.0</td>
<td>5.9</td>
<td></td>
</tr>
<tr>
<td>pAS9</td>
<td>pAS64 pAS17&lt;sup&gt;b&lt;/sup&gt;</td>
<td>729.6 ± 6.3</td>
<td>734.7 ± 10.2</td>
<td>100.7</td>
<td></td>
</tr>
<tr>
<td>pAS9</td>
<td>pAS64 pAS15</td>
<td>728.7 ± 5.6</td>
<td>691.5 ± 11.2</td>
<td>94.9</td>
<td></td>
</tr>
<tr>
<td>pAS9</td>
<td>pAS64 pAS65&lt;sup&gt;c&lt;/sup&gt;</td>
<td>417.7 ± 44.4</td>
<td>14.7 ± 8.8</td>
<td>2.1</td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> pAS9 carries a repA: lacZ fusion under the control of repA<sub>p</sub> as well as the BCDD"D" repeats. pAS64 is a RepA expression plasmid and is induced with IPTG. pAS4, pAS15, pAS16, pAS17, pAS58, pAS59 and pAS65 are pBS<sup>+</sup> plasmids carrying various sets of repeat elements listed in the adjacent column (See Figure 4.6).

<sup>b</sup> DH5α + pAS17 colonies on a plate with OPTG and Xgal are white.

<sup>c</sup> Ω has been inserted into the SnaBI site located between the EFG and HIJ repeat groups.

Overnight cultures were used to inoculate 50ml L-broth DH5α cultures (1/100 dilution). These were grown for two hours before aliquots were taken and added to tubes containing 0 or 64μM IPTG. β-galactosidase activities were measured 2hr after IPTG was added to the cells. Mean Miller units with standard errors are shown. Plasmids were maintained with the appropriate antibiotics (pAS9: 40μg/ml kanamycin, pAS64: 25μg/ml chloramphenicol; pAS4, 15, 16, 17, 58, 59 and 65: 100μg/ml ampicillin). This experiment has been repeated twice with the same pattern of % activity of repA<sub>p</sub>.

However, the presence of pAS4 with additional copies of the BCDD"D" repeats did not affect the repression of repA<sub>p</sub> in pAS9. The control experiment using only pAS9 and pAS64 indicates that pAS64 is able to express functional RepA in sufficient quantities to fully repress pAS9. These two experimental results are not incompatible if RepA bound to one copy of the BCDD"D" repeats is able to repress repA<sub>p</sub> present in cis and is able to repress in trans a second copy of repA<sub>p</sub> associated with the BCDD"D" repeat elements. This explanation is supported by the observation that both the CDD"D" repeats (pAS59) and the D" repeat alone (pAS58) show the same behaviour as pAS4 containing the BCDD"D" repeat elements.

<sup>In vivo</sup> repression of three promoters by RepA
A similar outcome is observed when the EFG repeats are present on the third plasmid (pAS16). In contrast, the presence of either the HIJ repeats alone, or the EFGHIJ repeat group in the third plasmid completely de-repress repAp. The de-repression of repAp by the HIJ repeats (pAS17) is in agreement with the predicted outcome of this assay: competitive binding of RepA to the repeat elements present in the highest copy number resulting in a de-repression of repAp due to the lower copy number of pAS9. This explanation is clearly contradictory to the explanation given above to cover the behaviour of pAS4, pAS58 and pAS59 in this assay.

When the third plasmid contains copies of the EFGHIJ repeat elements as in pAS15 (with the normal spacing and sequence between the two repeat groups), the third plasmid acts in a similar manner to pAS17 which only contains the HIJ repeats. This result implies in the context of this combination of repeat elements (BCDD'D" verses EFGHIJ) that the HIJ repeat group is functionally 'dominant' over the EFG repeat group. However, if the spacing between these two repeat groups is altered by the insertion of the Ω fragment (as in pAS65), the 'dominant' affect of the HIJ repeats is lost, and once again, RepA appears to be able to bind to one set of repeat elements and repress repAp associated with a second copy of repeats in trans.

Potentially, these conflicting results might be resolved using an assay where two different sets of repeat elements could be tested at once. The presence of two different 'third' plasmids would allow different combinations of repeat/repeat interactions which might then influence repAp de-repression. However, such an assay would be experimentally complex in order to avoid plasmid incompatibility reactions and in order to limit problems associated with differing plasmid copy numbers. The incompatibility problems may be limited by using a lysogen or similar system to provide an inducible source of RepA, but the reduced supply of RepA compared with that from pAS64 may then prevent the adequate repression of repAp when three sets of repeat elements are present in trans.

4.5 Summary of the in vivo repression of three promoters by RepA

A number of in vivo tests have been described in this chapter which demonstrate that RepA supplied in trans is able to repress the transcriptional activity of three different promoters. Two of the three promoter sequences are embedded within the BCDD'D" repeat group, whilst the third promoter is embedded within the EFG repeat group. The placement of the three promoters within the repeat groups which flank repA in the RepFIB minimal replicon lead to the prediction that if RepA recognised and bound to the repeat groups, then each of the three promoters should be regulated, regardless of their function in RepFIB replication.
One of the three promoters is the repA promoter (repA_p). Since RepA represses repA_p when supplied in trans from either a RepFIB mini-plasmid or a RepA expression plasmid, it is clear that repA is autoregulated (this has been demonstrated in Chapter Three as well).

There is no clear function for the other two promoters which are repressed by RepA (ori_p and EF_p). However, the presence of two active promoter elements in the BCDD'D'' and EFG repeat groups has provided in vivo proof that RepA binds to the BCDD'D'' repeats as well as to the EFG repeats. The placement of ori_p within the BC repeat elements and repA_p within the D'D'' repeat elements of the BCDD'D'' repeat group has also provided proof that there is a difference between RepA DNA-binding to the left hand elements and to the right hand elements of the repeat group. Presumably, differing RepA-repeat element affinities are responsible for the differential binding and repression of ori_p and repA_p. However, examination of the sequence of each of the repeats within the BCDD'D'' repeat group has failed to identify any significant sequence differences between the BC and D'D'' repeat elements9.

---

9 See Appendix Three, A3.3 for an analysis of the repeat elements.
5.1. Introduction

5.1.1. Considerations concerning the N-terminal sequence determination of RepA

The coding region of any gene which expresses a protein is simplistically defined by a start codon followed by an in-frame stop codon. However, in some cases a number of potential start codons may be present, in which case the actual start codon used in the initiation of translation must be determined experimentally. In the case of repA, the coding region is bounded by two stop codons which flank a number of potential start codons\(^1\). Preliminary investigation of repA by maxicell analysis suggested that RepA encoded by the gene had a relative molecular weight of between 35-40kDa (Saul et al., 1989). Although the size of RepA suggests that translation must begin within = 150bp of the first upstream stop codon, the relative size could not be used to determine which of four potential start codons within this region defined the beginning of the gene. An examination of the sequences surrounding the four codons (single ATG codon and three GTG start codons)\(^2\) failed to provide further information which might indicate which was used in the initiation of translation (see Dreyfus, 1988).

Therefore, a decision was made to determine the RepA start codon by limited N-terminal sequence analysis of the protein itself. Careful consideration was given to how sufficient RepA could be isolated for sequencing purposes since the method of expression and purification might have resulted in the isolation of an unusual or artificial form of the protein. In particular, over-expression of RepA using an expression plasmid might have caused considerable problems, since the placement of an inducible promoter may favour one start codon over another depending on the distance from the promoter sequence.

It was also possible that repA expressed more than one form of RepA. Since the expression of multiple RepA forms (isoforms) would be of considerable significance in terms of any RepFIB replication control model, it was important that the isolation strategy did not discriminate between potentially different RepA isoforms\(^3\).

---

\(^1\) See Appendix Three, A3.2.1.
\(^2\) DNA sequences will be used for all translation start codons rather than RNA sequences (the two are synonymous).
\(^3\) The possibility of different RepA isoforms was suggested by the similarity between RepFIB and other Step function replicons, and in particular, a control model proposed for mini-F (described in Appendix Six, A6.4.9). Although repA is
5.1.2 Practical problems in the isolation of a particular protein

The isolation or purification of a particular protein from a solution of proteins, such as those found within a cell, relies upon a means to detect and follow the protein through the various stages of purification. In general, the 'tracking' can be accomplished by monitoring an intrinsic activity of the protein using a specific enzymatic assay, or through the use of an antibody or another factor which will specifically bind to the protein. Unfortunately in the case of RepA, neither activity assay nor antibody was available to monitor the success of various purification protocols designed to isolate the protein for N-terminal sequencing.

These problems can be overcome by linking the protein to a well-characterised enzyme to form a fusion protein and monitoring the isolation procedure using an appropriate enzymatic assay (see Uhlén and Moks, 1990). Although a variety enzymes have been modified to allow the formation of enzymatically-active fusion proteins, fusions with *Escherichia coli* β-galactosidase are particularly successful.

![Diagram](image.png)

**Figure 5.1** Expression of a RepA:β-galactosidase fusion protein from pAS7

The promoter-probe plasmid pAS7 was formed by ligating the *BamH*I 0.97kb fragment into pMLB1034 to form a repA:lacZ fusion protein under the control of the repA promoter (repA) (A). The transcript produced by repA is a hybrid mRNA containing half of repA and all of the pMLB1034 lacZ coding sequence (C). Translation of the hybrid mRNA will produce a fusion protein of which the N-terminal portion is RepA and the C-terminal portion is β-galactosidase (D). Potential RepA translational initiation sites are indicated by 'V'.

---

most likely to express a single protein from a defined start codon, different isoforms may result i) from post-translational modification; ii) or when two start sites are used for expression where the choice of start sites is controlled by a regulatory switch; or iii) when a nested set of proteins is expressed from different start sites within repA.
5.1.3 The utility of β-galactosidase fusion proteins

It is possible to produce two forms of β-galactosidase fusion protein for use in genetic analysis. In the first form, the two proteins are joined by linking the C-terminal region of β-galactosidase to the N-terminal region of the protein of interest to give 'C-terminal fusions'. 'N-terminal fusions' are formed when the orientation of β-galactosidase and the protein are reversed (reviewed by Silhavy and Beckwith, 1985). In general, many features characteristic of β-galactosidase are retained by both C-terminal and N-terminal fusion proteins, including readily assayed β-galactosidase activity. The fusion protein may bind lactose and substrate or inhibitor analogues, and the fusion can often be recognised by β-galactosidase antibodies. Such binding by the fusion protein can be used in affinity chromatography to isolate the fusion whilst β-galactosidase assays and antibody recognition can be used to monitor the success of the purification procedures.

Since the N-terminal amino acid sequence of RepA was required, the RepA:β-galactosidase fusion must necessarily be a β-galactosidase N-terminal fusion. Unlike C-terminal fusions formed using commercially available fusion plasmids, the expression of the RepA:β-galactosidase fusion would not be under the control of an inducible, active promoter such as pLac, but under the relatively low-level expression of the repA promoter (repAp). A number of RepA:β-galactosidase fusions have been made using the promoter-probe plasmid pMLB1034 (see Chapter Three), one of which was used as the source of fusion protein (Figure 5:1).

5.2 Limited N-terminal sequence analysis of the RepA:β-galactosidase fusion protein

5.2.1 Isolation of the RepA:β-galactosidase fusion protein

The RepA:β-galactosidase fusion protein expressed by pAS7 in PB1905 was extracted as a crude cell lysate which was then enriched by precipitation with ammonium sulphate (Figure 5:2). The enriched extract was then used in a number of different strategies in an attempt to obtain sufficient fusion protein to allow successful N-terminal sequencing. These strategies involved APTG-affinity chromatography (p-aminophenyl β-D-thiogalactopyranoside), gel filtration and ultra-filtration using filters with suitable molecular weight exclusion limits to purify the fusion protein (detailed in Figure 5:3). Although the fusion protein was further enriched using these strategies, none resulted in successful N-terminal sequences.

After a change in approach, a quantity of RepA:β-galactosidase fusion protein was obtained using an anti-β-galactosidase immunoaffinity column to partially purify the ammonium sulphate-enriched extract.

---

4 Details of all plasmids constructed for this Thesis are supplied in Appendix Two.
5 As a result, the protocols and results of these strategies are not described in this Thesis.
PB1905 + pAS7 cell lysates containing the RepA:β-galactosidase fusion protein were enriched by precipitation with ammonium sulphate (A). The enrichment of the lysate is shown (B) with a maximum enrichment of 3.8x achieved with 40% of saturation of ammonium sulphate (▼). The precipitate from this cut retains 94% of the total β-galactosidase activity but only 23.5% of the total protein. The precipitate was then used as a basis for the further purification of the fusion protein (C). The initial isolation strategy followed the left arm of the diagram. Alternative strategies involved size-fractionation of the ammonium sulphate precipitate before affinity chromatography (centre), or the use of an immunoaffinity column (right). Sequence information was finally obtained using this procedure (●). Unsuccessful attempts to obtain N-terminal sequence (○) and proposed attempts which proved unnecessary (□) are indicated.

5.2.2 Sequence analysis

A portion of the immunoaffinity column-purified fusion protein was separated by SDS-PAGE, transferred to Immobilon membrane and the appropriate band was excised and sequenced for ten cycles. The sequence information obtained indicated that the section of the membrane contained two proteins unresolved by SDS-PAGE. As a result, two amino acid residues were detected for each degradation step of the sequencing cycle. However, because the proteins were present at different concentrations, the two sequences were easily differentiated from one another on the basis of signal intensity (Table 5:1). The primary N-terminal sequence maps exactly with the predicted amino acid sequence of a protein expressed from repA. The first residue was a methionine, and the next nine residues correspond exactly with the predicted amino acid sequence of repA (ORF-1) with the exception of residue 7 which was unresolved by the sequencer. However, the sequence...
correspondence of the primary fusion protein with repA indicates that the initiation of translation must have occurred from a CTG codon located upstream of the expected repA start codons at 2279cbp.

Although it is possible that the primary sequence is the result of post-translational modification of the fusion protein (such as proteolytic degradation), this is not the best explanation for several reasons. If the fusion protein was modified, the modification would have to involve proteolytic cleavage followed by the addition of a methionine to the end of the cleaved peptide. In addition to this unlikely modification, the protein still requires a start codon for translation which is in-frame with the rest of repA. There are no start codons located between the in-frame stop codon upstream of repA (TGA at 22286cbp) and the mapped position of the N-terminus of the fusion protein (CTG at 2279cbp).

No indication was found during the isolation of the RepA:β-galactosidase fusion protein to suggest that the repA gene expressed more than one protein, or that within the bacterium more than one form of the fusion exists. The presence of the second RepA sequence is almost certainly explained by in vivo proteolytic activity or degradation during the extraction procedure. Although the first residue of the second sequence was unresolved, residues 2-5 indicate that the N-terminus of the second protein maps to codon 56 of RepA (2443cbp). When mapped against the repA coding region, the first unresolved residue does not correspond to a possible translational start codon. This protein appears to be a proteolytic degradation product of RepA, after cleavage between two lysine residues.

**Table 5:1**

<table>
<thead>
<tr>
<th>Alignment of the primary sequence obtained from the blot</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA Sequence (5'→3')³</td>
</tr>
<tr>
<td>ORF-1 Translation</td>
</tr>
<tr>
<td>Fusion protein sequence #1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Alignment of the secondary sequence obtained from the blot</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA Sequence (5'→3')³</td>
</tr>
<tr>
<td>ORF-1 Translation</td>
</tr>
<tr>
<td>Fusion protein sequence #2</td>
</tr>
</tbody>
</table>

a) DNA sequence starting at 2276cbp in the reading frame of ORF-1, extended upstream from the expected GTG initiation codon.

b) DNA sequence starting at 2441cbp in the reading frame of ORF-1.
5.3 The efficiency of initiation of translation from the CTG start codon of *repA*

5.3.1 Introduction

The initiation of translation from non-ATG (or GTG) start codons is possible\(^6\) but is both rare and at lower efficiencies than from ATG or GTG codons. A clear prediction from this is that RepA expression should be increased if the *repA* start codon is changed from CTG to ATG. This prediction has been tested using RepA:β-galactosidase fusion protein promoter probe plasmids, where the start codon of the fusion protein has been changed to ATG by mutagenic PCR.

5.3.2 Alteration of the *repA* start codon by PCR mutagenesis

The CTG start codon of *repA* was altered to ATG by a two-stage PCR protocol (Figure 5:3). The 1.5kb PCR product was digested to release a *Bam*HI 0.97kb which contained the altered sequence. An attempt was made to ligate the 0.97kb fragment into pMLB1034. However, all colonies which appeared to express β-galactosidase activity did not have the expected recombinant plasmid. Restriction enzyme analysis of DNA isolated from these false positive colonies showed that either the RepFIB DNA was partially or completely deleted, or that the plasmid was in fact a pMLB1034 dimer or recombinant. The *Bam*HI fragment was finally cloned into pCGN565 to give pAS63. This construction avoided potential problems associated with high levels of fusion protein expression (discussed latter) and provided a more useful plasmid (than a pMLB1034-derived plasmid) for future sequence analysis of the mutated-PCR fragment\(^7\).

An attempt to clone the *Bam*HI fragment from pAS63 into pMLB1034 was made. However, DNA analysis of transformant colonies expressing β-galactosidase demonstrated that the plasmids either contained RepFIB DNA deletions or were pMLB1034-dimers or pMLB1034-deletion recombinants. Although the PCR synthesis of the *Bam*HI fragment may have introduced other changes in the DNA which prevented the expression of the fusion protein in pMLB1034\(^8\), it was also possible that the CTG→ATG alteration resulted in a sufficiently high level of expression which affected cell growth. This second possibility was avoided by cloning a portion of the *Bam*HI fragment which did not include the *repA* promoter into pMLB1034. By this means, deleterious over-expression was avoided, yet detection of a successful plasmid recombinant still might be possible if sufficient fusion protein was translated from mRNA transcripts initiated within pMLB1034\(^9\). Such a plasmid was produced by ligating the *PstI-Bam*HI fragment of pAS63 into pMLB1034 (to give pAS56). pAS56 is equivalent to...

---

6 Discussed in Chapter Eight, 8.3.2.
7 Sequence analysis of the fragment may have proved necessary if a fusion protein could not be expressed from a promoter probe plasmid. In order to determine whether the PCR synthesis had introduced other sequence alterations into the cloned fragment, the sequence of the fragment would need to be determined. Sequencing of the borders of the *Bam*HI fragment would be possible using general sequencing primers (M13 and M13 reverse) in pAS63, whereas sequencing from a pMLB1034 plasmid would require new sequencing primers made specifically for that plasmid.
8 Such as *repA* down mutations, or nonsense mutations in the *repA* coding sequence.
9 A discussion of vector-initiated transcription and its affect on the expression of a protein fusion in a promoter-probe plasmid is given in Chapter Three, 3.2.1.
pAS2 and the two plasmids differ only in the initiation codon of the fusion protein (CTG in pAS2, ATG in pAS56).

5.3.3 A comparison of the level of expression of the RepA:β-galactosidase fusion protein from genes containing CTG and ATG start codons

The RepA:β-galactosidase fusion protein in pAS2 (wild type or CTG start codon) and pAS56 (ATG start codon) is translated from a mRNA transcript originating from a promoter located upstream of the repA fragment in pMLB1034. Although the activity of this un-mapped promoter is weak compared to the repA promoter (repAp), the resultant expression of the fusion protein can be measured. The level of expression of the fusion protein in DH5α cells containing pAS2 and pAS56 was determined by β-galactosidase assay (Table 5:2). From this analysis it is apparent that pAS56 expresses approximately four times the amount of β-galactosidase activity as does pAS2. This increase suggests that the initiation of translation from an ATG start codon is more efficient than from the CTG start codon.

![Figure 5:3](image)

**Figure 5:3** PCR-directed mutagenesis of the RepA start codon

The polymerase chain reaction was used to alter the RepA translational initiation (start) codon from CTG to ATG. The alteration was achieved in two stages. In the first stage (A), two separate PCR reactions were performed generating DNA containing the CTG→ATG alteration and sharing a 10bp overlap region (bold line). In the first 10 cycles of the second stage PCR (B), annealing temperatures were kept low to allow annealing and ‘primer extension’ (dashed arrow). In the last 10 cycles, the annealing temperature was increased and the recombinant molecule amplified (C). Finally, the recombinant PCR DNA was digested with BamHI to produce a 0.97kb internal fragment which was then cloned into pCGN565 to give pAS63 (D). Oligonucleotide primers are shown by '→'.
The higher expression of the ATG-fusion protein compared to the CTG-fusion protein may indicate why the cloning of the *Bam*HI fragment of pAS63 into pMLB1034 was unsuccessful. The expression of the (CTG) fusion protein from *repAp* in pAS7 produces almost 1000 Miller units of β-galactosidase activity. It is possible that the continuous expression of four times this level of fusion protein would prove fatal to the cell.

### TABLE 5:2 Relative expression of the fusion proteins from pAS2 and pAS56

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Start codon</th>
<th>β-Gal.</th>
<th>Ratio&lt;sup&gt;a&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>pAS2</td>
<td>Wild type 'CTG' codon</td>
<td>30 ± 2</td>
<td></td>
</tr>
<tr>
<td>pAS56</td>
<td>Mutant 'ATG' codon</td>
<td>114 ± 5</td>
<td>3.8</td>
</tr>
</tbody>
</table>

<sup>a</sup> Relative expression of β-galactosidase activity.

Cultures (inoculated with a 1/100 dilution of an overnight culture) were grown to early-log phase before β-galactosidase activities were determined. Mean Miller units and standard errors are shown. The plasmids were maintained in DH5α with ampicillin.

### 5.4 Comment on the translational start codon used in the expression of repA

The RepA N-terminal sequence indicates that the translation of *repA* is unusual as the majority of prokaryote genes are expressed from either an ATG or GTG codon (this is discussed in depth in Chapter Eight, 8.3.2). A clear prediction resulting from the use of a CTG start codon is that the efficiency of translation should be lower compared to initiation from an ATG codon. This prediction has been shown to be correct by comparing the levels of expression of a RepA:β-galactosidase fusion protein from two plasmids which differ only in the start codon of the fusion gene. The N-terminal sequence information has been used to produce several RepA expression plasmids, each of which produce functional RepA protein (Chapter Three, Four and Six). RepA isolated from cells carrying one of these plasmids has been used to prepare RepA antiserum which recognises the RepA:β-galactosidase fusion protein, RepA expressed by a RepFlB mini-plasmid as well as RepA expressed by another over-expression plasmid (Chapter Six). These results suggest that the N-terminal sequence data reported here is correct.
6 RepA DNA-binding in vitro

6.1 Introduction

The RepA protein of RepFIB is predicted to be a DNA-binding protein which specifically recognises and binds to DNA containing one or more copies of the RepFIB repeat element. This prediction is based on the organisational homology seen between RepFIB and other Step function replicons, and in particular, the organisational homology with mini-P1 and the sequence homology seen between the RepA proteins of mini-P1 and RepFIB. This prediction has been supported by a number of in vivo experiments where RepA is able to repress three different promoters which are embedded within the BCDD'D" and EFG repeat groups (ori_p, repAp and EFp; see Chapters Three and Four). Additional in vivo genetic experiments suggest that RepA DNA-binding activates the origin of replication and helps determine RepFIB mini-plasmid copy numbers (Chapter Seven).

An in vitro demonstration of RepA DNA-binding is given in this chapter which complements the in vivo genetic evidence of RepA DNA-binding activity reported elsewhere in this Thesis. The in vitro demonstration required a suitable expression plasmid capable of producing large amounts of RepA for isolation and the development of a simple test for RepA DNA-binding.

6.2 The over-expression of RepA

6.2.1 RepA expression plasmid

A repA over-expression plasmid (pAS49) was constructed using the expression plasmid pJLA602 (Schauder et al., 1987) and PCR-amplification of the repA coding sequence from pNZ9451. pJLA602 contains a hybrid λ promoter (λpL,R) which is repressed by the c1857 (C1) protein carried by pJLA602 when grown at the permissive temperature of 30°C. On induction by transfer to 42°C, C1 is inactivated and λpL,R is de-repressed. The upstream primer used in the amplification of the repA coding sequence was designed to make two alterations to the repA sequence: i) the repA start codon was altered from CTG to ATG to increase the efficiency of repA translation; and ii) to form a Ncol site covering the start codon to allow the simple ligation of the PCR-amplified fragment into pJLA602.

1 Details of all plasmids constructed for this Thesis are supplied in Appendix Two.
Although the DNA sequence of the upstream region of repA was altered, the amino acid sequence of RepA expressed by the recombinant plasmid is identical to that of the wild type RepA.

An analysis of the expression of RepA by pAS49 in PB1905 and in DH5α has shown that RepA is rapidly expressed after induction. RepA is readily detected by SDS-PAGE analysis of whole cell lysates using coomassie blue stain within thirty minutes of induction. The expressed protein has an apparent molecular weight of 39kDa, which is in close agreement with the predicted size of RepA (38.87kDa). The identity of RepA expressed by pAS49 was confirmed by limited N-terminal sequence analysis of the 39kDa band removed from an in vivo 14C-labelling experiment (Figure 6:1).

**FIGURE 6:1**

*Photographic plate opposite, Top* RepA is rapidly expressed by cells containing pAS49 following an elevation of incubation temperature to 42°C. RepA was identified on the basis of the appearance of a labelled protein (after induction) with a relative molecular mass of 39kDa (→). The identification was confirmed by obtaining limited N-terminal sequence of the protein from the 120 minute sample (+). **Method:** An overnight culture of DH5α + pAS49 was used to inoculate 20ml of L-broth. The culture was grown to early-log phase at 28°C before the incubation temperature was shifted to 42°C. 50 μCi of 14C-labelled amino acids were added directly to the culture after the shift to 42°C. Samples of the culture were removed after 0, 15, 30, 60, 90, 120 and 180 minutes of incubation at 42°C. The cells were pelleted by centrifugation and washed once in L-broth before lysis in 2x SDS-PAGE sample buffer. The whole cell lysates were separated by 12.5% SDS-PAGE and the gel washed in Amplify before autoradiography.

**FIGURE 6:2**

*Photographic plate opposite, Bottom* Rabbit RepA antiserum can be used to identify the RepA:β-galactosidase fusion protein originally isolated in order to determine the translational start codon of repA (expressed by pAS7), as well as RepA naturally expressed by the copy mutant RepFIB mini-plasmid, pMA4322, and RepA from two different expression plasmids, pAS49 and pAS60. However, the antiserum does not detect RepA expressed from the wild type RepFIB mini-plasmid pSS3928, presumably because of the low copy number of this plasmid. The migration positions of the 39kDa RepA and the 145kDa RepA:β-galactosidase fusion protein are indicated (→). **Lanes:** 1, 8M urea extract containing RepA (from DH5α carrying pAS49, induced at 42°C); 2, DH5α; 3, DH5α carrying pSS3928; 4, DH5α carrying pMA4322; 5, DH5α carrying pAS7; 6, DH5α carrying pAS49 induced at 42°C for two hours; 7, DH5α carrying pAS60 induced with IPTG for two hours; 8, Cell lysate of DH5α; 9, Cell lysate of DH5α carrying pSS3928; 10, same as lane 1. Note that lanes 6 and 7 are over-loaded and the antiserum recognises a RepA degradation product of ~36kDa and some higher-molecular weight RepA-aggregates (unmarked). **Method:** Protein samples were separated by 12.5% SDS-PAGE and transferred to a nitrocellulose membrane. Cells were lysed directly in 2x SDS-PAGE sample buffer (lanes 1-7) or prepared in lysis buffer with lysozyme (lanes 7 and 8). RepA and RepA:β-galactosidase fusion protein were detected using rabbit RepA antiserum and an Amersham ECL kit.

**RepA DNA-binding *in vitro***
This experiment was designed to produce radioactively-labelled RepA which could then be used to evaluate the efficiency of various purification strategies such as ion exchange and gel-filtration chromatography. However, although RepA was readily labelled using $^{14}$C-amino acids, the majority of $^{14}$C-RepA remained with the cell debris after cell lysis and could not be solubilised without the use of denaturing buffers. The use of such buffers would presumably destroy the DNA-binding activity of RepA. As a result, that particular strategy to isolate functional RepA for the purposes of in vitro DNA-binding analysis was abandoned.

### Table 6:1 Extraction of RepA

<table>
<thead>
<tr>
<th>Conditions</th>
<th>Release</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cell lysis by treatment with lysozyme</td>
<td>&lt; 1%</td>
</tr>
<tr>
<td>Pellet washed with 1% Triton X-100</td>
<td>&lt; 5%</td>
</tr>
<tr>
<td>Pellet washed with 1% SDS</td>
<td>90-95%</td>
</tr>
<tr>
<td>Pellet washed with 375mM NaCl</td>
<td>&lt; 5%</td>
</tr>
<tr>
<td>Pellet washed with 0.25, 0.5, 1 and 2M urea</td>
<td>&lt; 5%</td>
</tr>
<tr>
<td>Pellet washed with 4M urea</td>
<td>50%</td>
</tr>
<tr>
<td>Pellet washed with 8M urea</td>
<td>90-95%</td>
</tr>
<tr>
<td>8M insoluble fraction</td>
<td>&lt; 5%</td>
</tr>
</tbody>
</table>

These values have been assessed from a number of different attempts to extract RepA from PB1905 and DH5α strains carrying pAS49. Each wash was for five minutes at room temperature. The pellet and supernatant fractions were separated by centrifugation and the amount of RepA protein released into the supernatant determined by coomassie blue staining after SDS-PAGE.

#### 6.2.2 Solubilisation of RepA

RepA expressed by pAS49 in PB1905 or DH5α strains appear to form insoluble aggregates which are associated with the cellular debris after cell lysis by sonication, freeze-thawing or lysozyme treatment. The presence of 1% Triton X-100 or 375 mM NaCl in the lysis buffer does not result in substantial quantities of soluble RepA. However, RepA solubilisation occurs in the presence of 1% SDS or high concentrations of urea (4-8M) (Table 6:1).

Some attempt has been made to purify RepA from the 8M urea samples since purified RepA was required to produce an antiserum for future work (not to be covered in this Thesis) and semi-purified

---

2 Figure 6:1 is an over-exposure designed to determine whether RepA was the most highly-labelled protein expressed by the cells and whether any host proteins were equally well-labelled. Although other proteins are labelled with $^{14}$C, preliminary chromatography trials suggested that the specific activity of $^{14}$C-RepA was sufficient to assess various purification protocols, especially ion exchange and gel-filtration chromatography.

3 8M urea samples containing RepA (and 100mM NaCl) can be dialysed down to 200mM NaCl before RepA precipitates in any significant quantities. If the 8M urea sample is first dialysed against 1M urea 500mM NaCl, then against 500mM...
(or purified) RepA was required for \textit{in vitro} DNA-binding experiments. However, this became unnecessary when a RepA antiserum was successfully prepared directly from the 8M urea sample (next section), and once RepA DNA-binding was demonstrated using the Western-DNA protocol described in Section 6.3.

### 6.2.3 RepA rabbit antiserum

RepA over-expressed by pAS49 in DH5\(_\alpha\) was isolated after SDS-PAGE and used to produce rabbit antiserum. The antiserum was then used in Western analysis to detected RepA expressed by a number of different plasmids used in this Thesis (Figure 6.2). In particular, the antiserum recognised the RepA:β-galactosidase fusion protein expressed by pAS7, as well as RepA expressed by the RepFlB mini-plasmid, pMA4322. The fusion protein expressed by pAS7 was isolated by immuno-affinity chromatography in order to determine the N-terminal sequence of RepA (described in Chapter 5). The detection of the fusion protein and a 39kDa protein expressed by pMA4322 by antiserum generated by the expected gene product (RepA) is a clear confirmation that repA is expressed and translated to produce a single mature protein.

The antiserum does not detect RepA expressed by the wild type RepFlB mini-plasmid, pSS3928. The most probable explanation for such a failure is that the levels of RepA expression from pSS3928 are so low that insufficient RepA is present on the membrane to bind the antiserum and produce a signal. In an attempt to run more RepA on the SDS-PA gel, a cell lysate was used rather than a sample prepared by lysing cells directly in SDS-PAGE sample buffer (Figure 6.2, lane 8). However, the antiserum was still unable to detect RepA expressed by pSS3928. In contrast, pMA4322 has a 6-8x higher copy number than pSS3928 (Spiers, MSc Thesis, 1989), and RepA expressed by pMA4322 can be detected using the antiserum.

### 6.2.4 Summary

The production of a RepA over-expression plasmid has made the rapid production of RepA protein possible. However, the high level expression of RepA from pAS49 results in the accumulation of insoluble RepA aggregates which require denaturing agents for further purification. This requirement posses some problems if active RepA is to be isolated for \textit{in vitro} demonstration and investigation of DNA binding activity.

The RepA antiserum has been used to confirm the relationship between the repA coding region, the fusion protein used to determine the N-terminal sequence of RepA and the expression of RepA by RepFlB mini-plasmids. However, the antiserum will probably be of more value as an aid to the future purification of RepA.

NaCl and finally against 10mM NaCl, approximately 50% of RepA remains in the supernatant after centrifugation at 16,000 xg for thirty minutes. Preliminary experiments suggest that RepA will bind to DEAE ion exchange columns when loaded after dialysis against 200mM NaCl or after dilution to 500mM urea, 100mM NaCl. However, when the column is washed with increasing concentrations of NaCl (with or without 100 or 200mM urea), RepA elutes over a wide range of NaCl concentrations along with the majority of other bound proteins.
6.3 Western-DNA

6.3.1 Introduction

*In vitro* DNA-binding activity of proteins is typically demonstrated using either gel retardation assays or filter binding assays. However, both techniques rely on fairly large amounts of semi-purified or purified protein which is both soluble and biologically active. Rather than attempt to purify RepA from the 8M urea extracts and then to develop a suitable renaturation protocol before testing the DNA-binding activity of the protein, I have developed a technique which has allowed me to demonstrate DNA-binding activity very simply. This 'Western-DNA' technique is based on a number of protocols which have used Western transfer of proteins ('protein blotting') to membranes which have then been incubated with DNA probes to demonstrate DNA-binding (Bowen *et al*., 1980; Herlt *et al*., 1988; and Lelong *et al*., 1989). The success of protein blotting varies according to the DNA-binding protein of interest and can be divided into three stages which affect the outcome of the experiment: i) electrophoresis of the protein mixture, where in some cases non-denaturing conditions are essential, but in others SDS-PAGE is acceptable; ii) transfer to membranes, where in some cases capillary transfer in urea buffers is more appropriate than electrotransfer; and iii) post-transfer incubation, where in some situations protein denaturation is required before renaturation can successfully occur to regain DNA-binding activity.

The Western-DNA technique described here arose fortuitously (Table 6:2). It has not been applied to other DNA-binding proteins in this laboratory and has not been modified to improve RepA DNA-binding, simply because it effectively demonstrates RepA DNA-binding without modification.

**Table 6:2** The western-DNA technique

1. 12.5% SDS-PAGE separation of 8M urea extract (samples are boiled for 2 minutes in SDS-PAGE sample buffer containing 8-mercaptoethanol).
2. Gel equilibration in transfer buffer (15 minutes).
3. Electrotransfer to nitrocellulose (2 hours at 50 Volts).
4. Post-transfer washes in TBS, Tween-20 to remove SDS and methanol (15 minutes).
5. Incubation in blocking solution containing TBS, Tween-20, non-fat milk powder, sonicated calf thymus DNA (2 hours).
6. Incubation in diluted blocking solution with end-labelled probe DNA (1-2 hours).
7. Membrane washes in TBS, Tween-20 to remove unbound probe DNA (15 minutes).
8. Autoradiography
**FIGURE 6:3 Western-DNA analysis of RepA DNA-binding**

**Photographic plate opposite** A number of different experiments using the Western-DNA technique demonstrate the nature of RepA DNA-binding. The technique involves SDS-PAGE separation of a protein sample which contains RepA, electrotransfer to a nitrocellulose membrane, a period of post-transfer washes to allow protein renaturation, followed by incubation with a 32P-labelled DNA probe. The membrane is washed to remove unbound probe before autoradiography. In this figure, the strips of membrane are arranged and cropped so that the SDS-PAGE origin is on the left and the dye-front (DF) on the right of the box surrounding each group of strips (the direction of electrophoresis is indicated by the arrow). The electrophoretic position of RepA is indicated at the top of the figure; note also that a 36kDa RepA-degradation product also binds probe DNA (→) and a faint RepA-band in D3 is marked by a '▼'. The methods are described on the page following the photographic plate.

**A** Post-transfer renaturation in Western-DNA. Several different methods were tested to determine whether RepA DNA-binding activity could be increased by changing the renaturation post-transfer washes before incubation with probe DNA. **Lanes:** 1, 15 min. wash in TBS followed by three 5 min. washes in TBS; 2, 15 min. wash in TBS followed by three 5 min. washes in TBS, 0.5% Tween-20; 3, 15 min. wash in TBS, 5 min. wash in TBS, 1% Triton X-100 followed by two 5 min. washes in TBS, 0.5% Tween-20; 4, 15 min. wash in TBS, 5 min. wash in TBS, 500mM urea followed by two 5 min. washes in TBS, 0.5% Tween-20; 5, 15 min. wash in TBS, 5 min. wash in TBS, 500mM urea, 5 min. wash in TBS, 250mM urea, 5 min. wash in TBS, 125mM urea followed by 5 min. wash in TBS, 0.5% Tween-20.

**B** Only RepA shows specific RepFIB DNA-binding activity. Various protein extracts from DH5α cultures were tested to determine whether host proteins were responsible for DNA-binding in the Western-DNA assay. **Lanes:** 1, 8M urea extract containing RepA (from a DH5α carrying pAS49 culture grown at 42°C); 2, 8M urea extract from a DH5α culture grown at 42°C; 3, cells from a DH5α culture grown at 42°C lysed in 8M urea extraction buffer.

**C** RepA binds to the BCDD'D", EFG and HIJ repeat groups. RepA is capable of specifically binding to RepFIB DNA containing groups of the A-K repeat elements. **Lanes:** (first gel) 1, incubated with pBS⁺ probe DNA (control for RepFIB-specific binding), 2, incubated with pAS20 probe DNA (containing the origin region); 3, incubated with pAS19 probe DNA (containing the origin region plus the BCDD'D" repeat elements); 4, incubated with pAS4 probe DNA (only BCDD'D") 5, incubated with pAS16 probe DNA (EFG); 6, incubated with pAS15 probe DNA (EFGHIJ); 7, incubated with pAS17 probe DNA (HIJ); (second gel) 8, incubated with pBS⁺ probe DNA, 9, incubated with pAS20 probe DNA; 10, incubated with pAS19 probe DNA.

**D** RepA DNA-binding requires one DNA repeat element. RepA DNA-binding requires the presence of at least one DNA repeat element. **Membranes containing lanes:** 1, 2 and 3, incubated with pAS2 probe DNA (no DNA repeat elements); 4, 5 and 6, incubated with pAS37 probe DNA (containing only the D" repeat element); 7, 8 and 9, incubated with pAS7 probe DNA (containing the BCDD'D" repeat elements). **Lanes:** 1, 4 and 7, Cells from a DH5α culture grown at 42°C lysed in 8M urea extraction buffer; 2, 5 and 8, 8M urea extract from a DH5α culture grown at 42°C; 3, 6 and 9, 8M urea extract containing RepA (from a DH5α/pAS49 culture grown at 42°C).
FIGURE 6.3 (continued)

Method (A) : 8M urea extract protein samples containing RepA were separated by 12.5% SDS-PAGE in this Western-DNA analysis. After transfer, the membrane was cut into strips and washed in a number of different ways. The membrane strips were finally incubated with BamHI-cut, 32P-end labelled pAS4 DNA for 60 minutes.

Method (B) : Each lane was loaded with 2-3μg total protein and the 8M urea RepA sample contained = 40-50% RepA. The protein extracts were adjusted so that each sample contained the same amount of host protein (determined by a comparison of coomassie blue staining after SDS-PAGE). The protein samples were separated by 12.5% SDS-PAGE in this Western-DNA analysis. The membrane was incubated with BamHI-cut, 32P-end labelled PCR-amplified DNA (oligonucleotide primers #8 and #2) for another hours.

Method (C, lanes 1-7) : 8M urea extract protein samples containing RepA were separated by 12.5% SDS-PAGE in this Western-DNA analysis. The membrane strips were incubated with Xbal-cut, 32P-end labelled plasmid DNA probes for two hours.

Method (C, lanes 8-10) : Same as for C 1-7, except a different preparation of probes were used and post-incubation washes were carried out with larger volumes of buffer.

Method (D) : Each lane was loaded with 2-3μg total protein and the 8M urea RepA sample contained = 40-50% RepA. The protein extracts were adjusted so that each sample contained the same amount of host protein (determined by a comparison of coomassie blue staining after SDS-PAGE). The protein samples were separated by 12.5% SDS-PAGE in this Western-DNA analysis. The membrane strips were incubated with BamHI-cut, 32P-end labelled plasmid DNA probes for two hours. The intensity of the bands should not be compared between sections of the membrane as the specific activity of each probe varied.

6.3.2 The western-DNA technique

The complex mixture of proteins present in the 8M urea extract containing RepA was separated by SDS-PAGE. 12.5% SDS polyacrylamide gels were used and the electrophoresis conditions chosen to optimise the separation of RepA and other similarly-sized host proteins present in the 8M urea extract. After SDS-PAGE, the gel was washed to remove unbound SDS and to equilibrate the buffer prior to electrotransfer to a nitrocellulose membrane. After transfer, the membrane-bound proteins were incubated in a manner which allows RepA to renature, recognise and specifically bind RepFIB probe DNA.

Presumably, the 8M urea-denatured RepA passes through a number of different denaturing environments before refolding to form a biologically active protein. Several different post-transfer

---

4 The isolation procedure used to obtain RepA using 8M urea would destroy the tertiary structure of the proteins present in the sample. During the early stages of electrophoresis as the proteins migrate into the stacking gel, the urea present in the sample will remain in the gel well. However, the proteins will be maintained in a denatured form during electrophoresis by the presence of SDS. The washing steps following electrophoretic separation and before electro-transfer will remove excess SDS from the polyacrylamide gel as well as some of the SDS bound to the proteins present in the gel. Unbound SDS will be further diluted during transfer to the nitrocellulose membrane. Once bound to the membrane, renaturation of the proteins is possible as bound SDS will be removed by competitive binding.
incubation protocols have been tested to determine whether the DNA-binding activity of renatured RepA can be increased. However, no appreciable increases of activity compared to simple washes with TBS, 0.5% Tween-20 were observed (Figure 6:3, A). These results imply that RepA readily refolds into an active configuration and that any cofactors required for refolding are provided by the basic blocking buffer containing Tris, Tween-20, non-fat milk powder (NFMP) and sonicated calf thymus DNA.

No attempt has been made to estimate what proportion of RepA bound to the membrane refolds into an active configuration. It is unlikely that all bound RepA is refolded and is capable of binding probe DNA. In addition, there is some evidence that a small proportion of bound RepA refolds and binds non-specifically to probe DNA (for example, RepA binding to pBS+ DNA; see Figure 6:3, C : lanes 1 and 3; and Figure 6:4, A). A = 36kDa RepA-degradation product seen in Western-DNA analysis when an old stock of 8M urea extract containing RepA is used binds probe DNA efficiently (for example, see Figure 6:3, D).

6.4 Specific in vitro RepA DNA-binding activity

Specific RepA binding to RepFIB DNA containing copies of the A-K repeat elements was demonstrated by incubating a variety of DNA probes with a number of different of Western-DNA membranes. These membranes have been presented photographically in a composite figure (Figure 6:3) and as a scanned image (Figure 6:4). The results have been summarised in Table 6:3.

Initial probes were constructed using pAS4 which contained the BCDD'D" repeat elements and 3kb of pBS+ sequence, or PCR-amplified RepFIB DNA which included the CDD'D" repeat elements as well as half of repA. Both probes where bound by RepA on Western-DNA membranes (Figure 6:3, A and B, respectively). Figure 6:3, B demonstrates that host proteins with the same electrophoretic mobility as RepA are not responsible for DNA-binding, and that the 8M urea extraction technique used to prepare the RepA sample has not enhanced or modified host proteins which would otherwise not bind DNA, or bind DNA very poorly compared with RepA. Figure 6:3 A and B indicate that RepA is a DNA-binding protein, and that the acquisition of DNA-binding ability from an initial denatured state is not affected by the range of post-transfer washing conditions tested. The information presented in Figure 6:3, C demonstrates that RepA will bind to DNA carrying the BCDD'D", EFG or HIJ repeat elements. However, RepA appears to bind very poorly probe DNA constructed using = 500bp of RepFIB DNA which includes part of ORF-2, half of the origin of replication and the 'A' repeat element (pAS20).

with a vast excess of milk proteins and by replacement with Tween-20 detergent. Renaturation and the restoration of DNA-binding ability of RepA may require various ions and co-factors which are introduced by the non-fat milk powder used in the blocking solution. Finally, the milk proteins and calf thymus DNA present in the final wash block non-specific protein and DNA binding sites present on the membrane and non-specific binding reactions presented by the membrane-bound proteins.

RepA DNA-binding in vitro
Comparative binding of pBS+, pAS20, pAS19 and pAS4 probes

RepA DNA-binding has been determined by densitometric measurement of the Western-DNA autoradiograph. DNA-binding is measured as a function of film opacity (relative units, vertical axis; the same scale has been used for all traces); strips of membrane were incubated with $^{32}$P-labelled probe DNA: pBS+ (A); pAS20 (B); pAS19 (C) and pAS4 (D). In this figure, the traces are arranged so that the SDS-PAGE origin is on the left and the dye-front (DF) on the right of the box surrounding each trace (the direction of electrophoresis is indicated by the arrow). The electrophoretic position of RepA is indicated at the top and bottom of the figure. A 36kDa RepA-degradation product (→) and several 'bands' caused by $^{32}$P-spots on the autoradiograph (▲) are marked. Method: 8M urea extract protein samples containing RepA were separated by 12.5% SDS-PAGE in this Western-DNA analysis. The membrane strips were incubated with Xbal-cut, $^{32}$P-end labelled plasmid DNA probes for two hours. Density scans were made directly from the autoradiograph.
The A repeat shows the least degree of sequence similarity to the A-K repeat consensus sequence and is not involved in the control of RepFIB replication.

All of the RepFIB probes used in Figure 6:3, C were produced from pBS+-derived plasmids with similar specific activities and concentrations. The degree of RepA DNA-binding to pBS+ or pAS20 probes could be reduced compared to binding to a pAS19 or pAS4 probe by more extensive washing before autoradiography (Figure 6:3, C: lanes 8, 9 and 10). A pBS+ probe is bound to RepA as efficiently as a pAS20 probe (containing only the A repeat element), but is bound very poorly compared with either a pAS19 probe (ABCDD‘D* repeats) or a pAS4 probe (BCDD‘D* repeats) (Figure 6:4). These results imply that RepA DNA-binding is specific for DNA containing copies of the repeat elements which share a good degree of homology with the A-K repeat element consensus sequence.

The small degree of binding to pBS+ and pAS20 by RepA is probably due to non-specific DNA binding resulting from either i) a small portion of the total RepA bound to the membrane refolding incorrectly, or ii) a very high local concentration of protein bound to the membrane at that position.

<table>
<thead>
<tr>
<th>Figures</th>
<th>Probe DNA</th>
<th>Observations</th>
</tr>
</thead>
<tbody>
<tr>
<td>6:3, A</td>
<td>pAS4</td>
<td>BCDD‘D*</td>
</tr>
<tr>
<td>6:3, B</td>
<td>PCR DNA</td>
<td>CDD‘D*, repA</td>
</tr>
<tr>
<td>6:3, C and 6:4</td>
<td>pBS+</td>
<td>no RepFIB DNA</td>
</tr>
<tr>
<td></td>
<td>pAS20</td>
<td>origin regionA, no repeats</td>
</tr>
<tr>
<td></td>
<td>pAS19</td>
<td>origin regionA, BCDD‘D*</td>
</tr>
<tr>
<td></td>
<td>pAS4</td>
<td>BCDD‘D*</td>
</tr>
<tr>
<td></td>
<td>pAS16</td>
<td>C-terminus of repA, EFG repeats</td>
</tr>
<tr>
<td></td>
<td>pAS15</td>
<td>C-terminus of repA, EFGHJ repeats</td>
</tr>
<tr>
<td></td>
<td>pAS17</td>
<td>HIJ repeats</td>
</tr>
<tr>
<td>6:3, D</td>
<td>pAS2</td>
<td>no repeat elements, N-terminus of repA</td>
</tr>
<tr>
<td></td>
<td>pAS37</td>
<td>D*, N-terminus of repA</td>
</tr>
<tr>
<td></td>
<td>pAS37</td>
<td>BCDD‘D*, N-terminus of repA</td>
</tr>
</tbody>
</table>

a The ‘origin region’ in pAS19 and pAS20 contains ~500bp of RepFIB DNA including the A repeat element (which shares less homology to the A-K repeat consensus sequence than any of the other repeat elements), the ssi sequence and the dnaA box.

RepA does not require multiple copies of the repeat elements in order to bind probe DNA. This fact is demonstrated in Figure 6:3, D where RepA binds to pAS37 probe DNA carrying only the D* repeat element. The D* repeat element shows a higher degree of sequence homology to the A-K repeat.

5 Both explanations are equally valid, and are supported in part by early Western-DNA attempts (using larger RepA sample loadings than the experiments reported here) which resulted in more intense binding of pBS+ probe DNA compared to pAS4 probe DNA.
consensus sequence than repeat A (14/21 cf 9/21). This observation may indicate that the strength of RepA DNA-binding varies according to the sequence of the conserved repeat elements.

6.5 Summary of RepA DNA-binding in vitro

A number of experiments described in this Thesis demonstrate in vivo that RepA is a DNA-binding protein which recognises and binds to DNA containing copies of the A-K repeat elements (Chapters Three and Four). These in vivo observations have been confirmed using an adaptation of the Western transfer technique, where RepA protein has been separated from other proteins present in a sample by SDS-PAGE, transferred to a nitrocellulose membrane and permitted to renature in such a way which has then allowed the protein to bind specifically to RepFIB probe DNA ('Western-DNA').

This technique has allowed an analysis of RepA-DNA binding activity in vitro, despite the fact that the only source of soluble RepA was as a denatured 8M urea extract. RepA appears to readily renature using the Western-DNA technique, although a small percentage of the protein appears to refold into a form which binds non-specifically to probe DNA. However, proteolytic cleavage of RepA by cyanogen bromide or by incubation at pH 2.5 completely destroys the ability of RepA or RepA fragments to renature and bind probe DNA, and this has prevented the identification of which of the three potential DNA-binding domains is responsible for RepA DNA-binding (described in Appendix Four).6

Finally, the fact that RepA is capable of binding to DNA in the Western-DNA procedure may indicate that RepA monomers are DNA-binding. Presumably, any multimeric RepA complexes would be destroyed during the sample preparation and would be unable to reform during SDS-PAGE. However, it is possible that RepA dimerisation could occur during electrotransfer to the membrane, or between adjacent proteins after transfer during the renaturation washes7.

---

6 Appendix Four also describes an experiment attempting to inhibit RepA-DNA binding by preincubation with RepA antiserum. However, DNA binding was not inhibited.

7 Due to the similarities between RepFIB and other Step function replicons, and in particular, the homology between RepFIB RepA and mini-P1 RepA, there is a strong prediction that RepA DNA-binding requires dimerisation of the protein.
The functional relevance of the DNA repeat groups

7.1 Functional significance of the repeat groups

7.1.1 Predictions based on homology with step function replicons
The organisational homology seen between RepFlB and other replicons of the Step function class strongly suggests roles for RepA and the two groups of repeat elements which flank the repA gene. The shared homology predicts that RepA is the activator of the RepFlB origin of replication and that this initiator is absolutely required for RepFlB replication. This observation has been partially confirmed by showing that transposon disruption or deletion of the repA coding region destroys RepFlB-directed replication (Saul et al., 1989). The organisational homology also suggests that the BCDD'D'' repeats would be involved in origin function, and that the EFGHIJ repeat elements be involved in the sensing and maintenance of the correct replicon copy number via titration (hence, the BCDD'D'' group is referred to as the ‘origin’ group, and the EFGHIJ group referred to as the ‘titration’ group). It is presumed that the functional significance of both groups of repeats relies on the interaction between RepA and the repeat element DNA.

The interaction of RepA with the repeat elements also suggests a mechanism for the expression of incE incompatibility associated with the RepFlB replicon. Previous analysis had identified a ∼450bp region including the EFGHIJ repeats which represented the cis-acting factor of the incE phenomenon\(^1\). The trans-acting factor therefore would be RepA, which through the organisational homology with other Step function plasmids, is thought to bind to the repeat elements (confirmed by work reported in Chapters Three, Four and Six). Theoretically, the BCDD'D'' repeats should also act in cis to express incE incompatibility, as the sequences of the repeat elements in that group are not significantly different from the sequences of the other repeat elements in the titration group.

The work reported in this chapter defines the origin nature of the BCDD'D'' repeat elements and the involvement of the titration group in the copy number control of RepFlB. A brief description of incompatibility reactions between RepFlB mini-plasmids and plasmids carrying copies of the repeat elements is given below.

---

1 For this reason the EFGHIJ repeats have also been referred to as the ‘incE’ repeats, or simply as ‘incE’.
FIGURE 7:1  RepFiB repeat groups

The basic RepFiB replicon contains eleven repeat elements sharing a consensus sequence of 5'-ATATAAGCTGTAGTAAGTAAA-3'. The origin group consists of five elements forming a direct repeat located upstream of the repA gene (BCDD'D*). The remainder of the repeat elements are located downstream of repA and have a more complex arrangement. The E and F elements form an inverted repeat whilst FG and HIJ form direct repeats. The EFGHIJ repeat elements form the titration group. The origin group can be isolated as a = 200bp BamHI-PstI fragment, the titration group as a = 480bp HindIII-BamHI fragment (EFG: = 210bp HindIII-SnaBl, HIJ: = 270bp SnaBl-BamHI). This figure is not to scale.

7.1.2  A summary of observations from incompatibility reactions

In a variety of experiments described in this Thesis, incompatibility reactions between RepFiB mini-plasmids and other plasmids carrying copies of the repeat elements have been noted and lead to three broad observations.

The first observation is that the origin group of repeat elements can elicit an incompatibility reaction from a RepFiB mini-plasmid as readily as the titration group of repeats (EFG, HIJ or EFGHIJ). This observation implies that the origin group reacts in the same way with RepA as do the other repeat elements within the context of incompatibility reactions. The second observation is that the incompatibility reaction of the RepFiB mini-plasmid pSS3928 to the presence of repeat elements carried by high copy number plasmids such as pBS+ is far greater than the reaction seen with pMA4322, a copy number mutant RepFiB mini-plasmid. In particular, pSS3928 is intolerant to the origin group in trans, where as pMA4322 can be maintained with selection in the presence of a plasmid carrying the origin repeat elements. This difference has allowed a demonstration of repAp autoregulation and oriP regulation by RepA in Chapter Four.

The third observation is that the demonstration of incompatibility reactions using RepFiB mini-plasmids and plasmids carrying copies of the repeat elements can only be carried out in one direction. Classically, incompatibility reactions should be observable independently of which of the two plasmids (mini-plasmid or the inc-determinant plasmid) is maintained with selection. In the case of RepFiB, a resident RepFiB mini-plasmid can not be displaced by a challenging inc-determinant plasmid, although a resident inc-determinant plasmid can be displaced by a challenging mini-plasmid (Table 7:1). If no
selection is used for either plasmid, the inc-determinant plasmid will be lost after a period of incubation despite a higher copy number replicon.

The asymmetric outcome of RepFIB incompatibility experiments has been explained by a test to see whether pSS3928 required DnaA for replication (Section 7.2.2). The results of this experiment strongly suggest that the E11 fragment carrying the RepFIB replicon used to make RepFIB mini-plasmids contains a coupled cell death system (ccd) similar to that carried by mini-F and R1 (see Appendix Five). The ccd function ensures that any cell which has lost the mini-plasmid is killed within 2-3 generations.

### Table 7:1

<table>
<thead>
<tr>
<th>Plasmids</th>
<th>Selection</th>
<th>Expected Outcome</th>
<th>Observed Outcome</th>
<th>Outcome influenced by ccd function</th>
</tr>
</thead>
<tbody>
<tr>
<td>Resident</td>
<td>Challenging&lt;sup&gt;a&lt;/sup&gt;</td>
<td>RepFIB</td>
<td>inc-plasmid lost</td>
<td>inc-plasmid lost</td>
</tr>
<tr>
<td>RepFIB inc-plasmid</td>
<td>inc-plasmid</td>
<td>RepFIB lost</td>
<td>no viable transformants obtained</td>
<td></td>
</tr>
<tr>
<td>Co-transformants&lt;sup&gt;b&lt;/sup&gt;</td>
<td>RepFIB</td>
<td>inc-plasmid lost</td>
<td>inc-plasmid lost</td>
<td>RepFIB retained</td>
</tr>
<tr>
<td>RepFIB and inc-plasmid</td>
<td>inc-plasmid</td>
<td>RepFIB lost</td>
<td>RepFIB retained</td>
<td>yes</td>
</tr>
</tbody>
</table>

<sup>a</sup> the inc-determinant plasmid is used for either plasmid, the inc-determinant plasmid will be lost after a period of incubation despite a higher copy number replicon.

<sup>b</sup> a high copy number plasmid carrying repeat elements, for example: BCDD'D' (pAS4, 19, 7), CDD'D' (pAS39), D' (pAS37), EFG (pAS16), HIJ (pAS17), EFGHJ (pAS15), BCDD'D'EFGHJ (pNZ945).

The challenging plasmid was used to transform a strain containing the resident plasmid. Two plasmids were used in transformation and both plasmids were selected. The outcomes shown here are the result of transfer to media containing selection for one of the plasmids.

## 7.2 The RepFIB origin of replication

### 7.2.1 Location of the origin of replication

The RepFIB origin of replication was located initially on the basis of DNA sequence features and by organisational homology with the mini-P1 replicon. The origins of many Step function replicons consist of two elements: i) a set of DNA repeat elements located upstream of the initiator gene which bind the initiator protein; and ii) a group of sequence elements which are important with regard to the physical structure of the origin DNA or are recognised by host factors involved in the initiation of replication.

In the case of RepFIB, the BCDD'D' repeats are appropriately located to form part of the origin of replication. Immediately to the left of the B repeat element is a = 100bp region of DNA which includes origin sequences and recognition sites for host factors (one dnaA box, seven GATC repeat elements,
a 13mer element and GC-spacer). Finally, a single strand initiation sequence (ssi) required to initiate lagging-strand synthesis is located $=100$bp to the left of the dnaA box (Figure 7:2).

![Diagram](image.png)

**FIGURE 7:2** Sequence features of the RepFIB origin

The region between 1800-2200cbp of the E11 fragment containing RepFIB contains a number of sequence features which suggests that it represents the origin of replication. The region includes five direct repeat elements BCDD'D'' (large triangles), seven small repeat elements associated with the sequence 'GATC' recognised by Dam methylase (■), a single dnaA box (○), and a single strand initiation sequence (ssi). The best of four potential 13mer sequences is located near the dnaA box and a 'GC-spacer' homologous to the mini-P1 spacer is located to the right of the GATC repeats and overlaps the B repeat element. The region also contains two transcriptionally active promoters located within the BCDD'D'' repeats (→), oriP (transcription to the left) and repAp (transcription to the right). Scale is shown in base pairs.

The presence of a number of origin-specific sequence elements strongly suggests that the RepFIB origin of replication is located in the left hand portion of the minimal replicon (=1900-2200cbp). The right hand limit of the origin sequence can be determined using a pot 'rescue' assay. Such an assay relies on the fact that some replicons have an absolute dependence on the host DNA polymerase I for replication (such as the replicon in the plasmid pBS+), where as other replicons (such as RepFIB) can replicate independently of Pol I (Figure 7:3). However, pBS+ can be maintained in a pot host by forming a hybrid plasmid with a Pol I-independent replicon. Such 'rescue' can also be achieved if pBS+ carries only the origin of a Pol I-independent replicon if the cell is also supplied with the specific activator of the Pol I-independent origin. However, if the trans-activator is not supplied, the plasmid will not be able to replicate within the host.

In the case of RepFIB, the trans-activator is presumed to be the RepA protein. RepA could be provided in a pot host by a RepFIB mini-plasmid such as pMA4322. In this assay, a pBS+ derived

---

2 Both the wild type mini-plasmid pSS3928 and the copy mutant pMA4322 were tested in the development of this assay. However, incompatibility reactions between pSS3928 and pBS+ plasmids carrying the BCDD'D'' repeats was so strong that no co-transformants could be isolated in either pot+ (PB1965) or pot− (PB1966) strains.
plasmid which carried a functional RepFIB origin of replication would be expected to be capable of transforming pot competent cells containing pMA4322. However, if the plasmid carried a partial origin, it would not be rescued by the resident RepFIB mini-plasmid and would thus be unable to transform the cells.

![Diagram of replication origins (A, B, C)]

**Figure 7:3**  
*pot*/pot* rescue assay*

The origin of replication of a replicon which does not require host DNA polymerase I for replication can be functionally located using a 'rescue' assay. Plasmids such as pBS+ depend on Pol I for replication and as a result, can not replicate in a pot host (A). In contrast, some replicons do not have a Pol I dependence. If such a replicon is cloned into pBS+, the hybrid plasmid is 'rescued' and will be able to replicate in a pot host (B). pBS+ could also replicate in a pot host if it carries the origin of replication of the Pol I-independent replicon, and if the origin-activator protein (O) is supplied in trans from a second plasmid (C).

In order to confirm the sequence-based prediction that the 1800-2200bp region of RepFIB represents the origin of replication, a number of Xho31 derivative plasmids (Figure 7:4) were constructed with the aim of determining whether they could be rescued by pMA4322 in a pot strain. An isogenic pair of strains (PB1965: pol+ and PB1966: pot-) were first transformed with pMA4322 and then used to prepare competent cells. In this assay, pMA4322 was maintained with 50μg/ml spectinomycin, the pBS+ derived plasmids were selected for using 50μg/ml ampicillin and pHSG576 was selected for using 25μg/ml chloramphenicol.

The transformation frequencies of both pol+ and pot strains for each pBS+ derived plasmid was adjusted to account for the difference in the competency of the cells (determined by comparing the frequency of transformation of the Pol I-independent plasmid pHSG576 into the two strains) resulting in a transformation frequency of 5-65 times higher PB1965/pMA4322 transformation frequencies. This

---

3 Details of all plasmids constructed for this Thesis are supplied in Appendix Five.

4 DNA quality and concentration will also affect transformation frequencies. However, since all plasmid DNA was capable of transforming the pol+ strain one can assume that the DNA is capable of transforming the pot strain if the plasmid itself does not depend on Pol I for replication. The concentrations of the test plasmid DNAs were greater than the pHSG576 DNA, resulting in 5-65 times higher PB1965/pMA4322 transformation frequencies. This
in an 'efficiency' of transformation of 100% for pHSG576 and an 'efficiency' of less than 0.3% for pBS+ (Table 7:2).

Of the five Xho31 derivative plasmids tested in the rescue assay, only pAS19 and pAS36 appeared to transform the pol strain with any efficiency. Plasmids which lack either the BCDD'D' repeats which bind RepA or the DNA region containing the dnaA box are unable to replicate in a pol host even if RepA is supplied in trans. Since pAS36 is capable of transformation, it appears that the orientation of the two origin portions does not affect replication. However, if the two fragments are separated by increases the sensitivity of the assay (since all test transformations are at low frequencies) but does not alter the 'efficiency' ratio.
2kb (as in pAS28) origin function is destroyed and the plasmid can not be rescued in this type of assay.

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Efficiencya</th>
<th>Origin function is destroyed and the plasmid can not be rescued in this type of assay.</th>
</tr>
</thead>
<tbody>
<tr>
<td>pHSG576</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>PBS*</td>
<td>&lt; 0.003</td>
<td></td>
</tr>
<tr>
<td>pAS20</td>
<td>&lt; 0.04</td>
<td></td>
</tr>
<tr>
<td>pAS4</td>
<td>&lt; 0.01</td>
<td></td>
</tr>
<tr>
<td>pAS19</td>
<td>0.34</td>
<td></td>
</tr>
<tr>
<td>pAS36</td>
<td>0.27</td>
<td></td>
</tr>
<tr>
<td>pAS28</td>
<td>&lt; 0.006</td>
<td></td>
</tr>
</tbody>
</table>

Table 7:2

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Efficiencya</th>
</tr>
</thead>
<tbody>
<tr>
<td>pHSG576</td>
<td>1.00</td>
</tr>
<tr>
<td>PBS*</td>
<td>&lt; 0.003</td>
</tr>
<tr>
<td>pAS20</td>
<td>&lt; 0.04</td>
</tr>
<tr>
<td>pAS4</td>
<td>&lt; 0.01</td>
</tr>
<tr>
<td>pAS19</td>
<td>0.34</td>
</tr>
<tr>
<td>pAS36</td>
<td>0.27</td>
</tr>
<tr>
<td>pAS28</td>
<td>&lt; 0.006</td>
</tr>
</tbody>
</table>

a Efficiency is the ratio of transformation frequencies of PB1966/pMA4322 and PB1965/pMA4322 (pol+/pol+) where the relative competency of the two strains have been corrected so that pHSG576 gives 1.00.

PB1965 and PB1966 were transformed with pMA4322. Transformant colonies were tested for the pol' allele with MMS, and a suitable colony used to prepare competent cells of the two strains. CsCl plasmid DNA was used to transform the cells and the bacteria were spread onto plates containing 50μg/ml spectinomycin (to maintain pMA4322) and either 50μg/ml ampicillin or 25μg/ml chloramphenicol (to select for the incoming plasmids). Colonies were counted after incubation overnight.

The < 0.3% efficiency of PBS+ transformation is accounted for by the relatively high reversion frequency of the polA1 allele present in PB1966. pAS4 and pAS20 transformation efficiencies are greater than that of PBS+ and may not be accounted for by pol' reversions. However, they may be explained as the result of homologous recombination events occurring between pMA4322 and the PBS+ plasmid which form Pol I-independent, ampicillin and spectinomycin-resistant cointegrate plasmids. This is certainly possible as PB1966 is a rec strain and recombination between identical sequences in both plasmids would offer the host a selective advantage. At most, pol→polh reversion and co-integrate formation could account for ~ 4% of the efficiency of transformation in the case of pAS19 and pAS36. This suggests that plasmid 'rescue' is responsible for the extra 22-29% efficiency of transformation with pAS19 and pAS36, and is the result of the trans activation of the PBS+-RepFIB origin by RepA expressed from pMA4322.

The low efficiency of transformation by both pAS19 and pAS36 compared to pHSG576 might be explained by competition between the PBS+ plasmid origins and the origin carried by pMA4322. If pMA4322 does not express sufficient RepA to activate the origins present in both

5 Several pMA4322 + PBS+ PB1966 colonies were restreaked onto plates containing MMS and shown to be MMS-sensitive. This results suggest that either i) a secondary mutation has occurred which has increased the cellular supply of Pol I but has retained MMS-sensitivity; or ii) that PBS+ has integrated into the host chromosome and is now longer dependent on Pol I for replication. In this and other PB1965/PB1966 experiments, the reversion frequency of PB1966 is usually less than 2%.
plasmids, RepA should 'preferentially' be used by pMA4322 rather than by the pBS+ plasmids as pMA4322 is maintained by selection with spectinomycin. This 'preferential' use of RepA is simply the result of the selection used to isolate the correct transformant colonies in this assay: if RepA was used to rescue the pBS+ plasmid instead of pMA4322, pMA4322 would be lost and colony growth prevented by spectinomycin; and in addition, pBS+ plasmid replication would fail once available RepA had been used.

If RepA is a limiting factor, then developing colonies resulting from transformation with pAS19 or pAS36 presumably contain cells which have recently lost one of the two plasmids, and that the rate of loss has slowed the growth of the colonies and resulted in a lower 'transformation' efficiency.

7.2.2 Host factors required for RepFIB replication

The RepFIB requirement for the host Dam methylase or DnaA protein was determined using transformation assays of appropriate pairs of *E. coli* strains. The involvement of Dam in RepFIB replication was tested using PB2946 (dam+) and PB2989 (dam-), and the involvement of DnaA was tested using DK249 (dnaA+) and AQ699 (dnaA-).

**TABLE 7:3** Requirement for Dam methylase

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Efficiencya</th>
</tr>
</thead>
<tbody>
<tr>
<td>pBS+</td>
<td>Dam-independent replication</td>
</tr>
<tr>
<td>pSS3928</td>
<td>RepFIB mini-plasmid</td>
</tr>
<tr>
<td>pMA4322</td>
<td>RepFIB mini-plasmid copy number mutant</td>
</tr>
<tr>
<td>pNZ955</td>
<td>RepFIB/pUC19 hybrid</td>
</tr>
<tr>
<td>pNZ956</td>
<td>RepFIB copy number mutant/pUC19 hybrid</td>
</tr>
</tbody>
</table>

a Efficiency is the ratio of transformation frequencies of PB2989 (dam-) and PB2946 (dam+) with pBS+ where the relative competency of the two strains have been corrected so that pBS+ gives 1.00.

PB2946 and PB2989 were transformed with CsCl plasmid DNA and the bacteria were spread onto plates containing appropriate antibiotics (50μg/ml spectinomycin or 50μg/ml ampicillin). Colonies were counted after incubation overnight.

Plasmids with replicons derived from pMB1 such as pBS+ and pUC19 do not require Dam for replication, although the transformation efficiency of these plasmids into a dam- strain will be lower compared to a dam+ strain. The transformation efficiency of both the wild type RepFIB mini-plasmid pSS3928 and the copy mutant mini-plasmid pMA4322 are very low compared to hybrid mini-plasmids which also include pUC19 and a Dam-independent replicon (Table 7:3). These results imply that

---

6 This assay measures the ability of plasmid DNA to transform strains and is determined by counting the number of transformant colonies after incubation on agar plates. However, in reality the transformation frequencies reflect both the ability of plasmid transformation, the replication efficiency of the plasmid within the cell and the plasmid segregation rate since only colonies which are sufficiently developed to see are counted.
although RepFlB may not have an absolute requirement for Dam, RepFlB replication is severely affected by a lack of Dam methylase in the host.

The DnaA-replication dependence of RepFlB was tested using a pair of isogenic strains, DK249 and AQ699 (Table 7:4). pMB1-derived replicons do not require DnaA for replication, and pUC19 has been used as a control for DnaA-independent replication, where as pSC101-derived replicons have an absolute requirement for replication and pHSG576 has been used as a control for DnaA-dependent replication. From the transformation efficiencies obtained for this experiment it is apparent that a RepFlB mini-plasmid has an absolute requirement for DnaA, whilst a hybrid plasmid including a DnaA-independent replicon (pWM114) has no requirement for host DnaA.

<table>
<thead>
<tr>
<th>Table 7:4</th>
<th>Requirement for DnaA</th>
</tr>
</thead>
<tbody>
<tr>
<td>pUC19</td>
<td>DnaA-independent replication</td>
</tr>
<tr>
<td>pHSG576</td>
<td>DnaA-dependent replicon</td>
</tr>
<tr>
<td>pMA4322</td>
<td>RepFlB mini-plasmid copy number mutant</td>
</tr>
<tr>
<td>pWM114</td>
<td>RepFlB/pBR322 hybrid</td>
</tr>
</tbody>
</table>

a Efficiency is the ratio of transformation frequencies of AQ699 (dnaA*) and DK249 (dnaA*) with pUC19 where the relative competency of the two strains have been corrected so that pUC19 gives 1.00.

AQ699 and DK249 were transformed with CsCl plasmid DNA and the bacteria were spread onto minimal plates containing appropriate antibiotics (50μg/ml spectinomycin or 50μg/ml ampicillin). Colonies were counted after incubation overnight.

Although the transformation efficiencies shown in Table 7:4 are both convincing and replicable, the actual transformation efficiencies of both AQ699 and DK249 are very low compared to such strains as PB2946 or DH5α. Both AQ699 and DK249 are sensitive to rich medium and as a result, the preparation of competent cells and the transformation assay was carried out using minimal media supplemented with the appropriate amino acids. In order to confirm the RepFlB replication requirement for DnaA, a second assay was made using a temperature sensitive dnaA strain, PB1849 (dnaA46ts).

A RepFlB mini-plasmid (pSS3928) and a RepFlB-pBS+ hybrid plasmid (pNZ945) were used to transform PB1849. Test cultures were grown at the non-permissive temperature (42°C) with or without selection for the resident plasmid. Since RepFlB requires DnaA, cultures of PB1849 + pSS3928 and spectinomycin should not grow. In contrast, PB1849 + pNZ945 and ampicillin should grow, since replication of the plasmid will be from the DnaA-independent pBS+ replicon and not from RepFlB. These predictions were found to be true, supporting the dnaA+/dnaA- transformation results which indicate that RepFlB replication has an absolute dependence for host DnaA. However, the dnaA46ts experimental results were more complex than expected and required further work to
resolve the experimental observations. When completed, the results indicated that the E11 fragment carrying RepFIB used to make plasmids such as pSS3928 and pNZ945 also contained a ccd-like system (coupled cell death). The investigation of the ccd-system, including the dnaA461s results, are reported in Appendix Five.

7.3 Control function of the EFGHIJ repeats

7.3.1 Introduction

The organisational homology seen between RepFIB and a number of Step function replicons strongly suggests that the titration repeat elements are involved in the sensing and setting of plasmid copy number. In order to test this prediction, the copy number of RepFIB mini-plasmids containing various deletions was determined. The deletions were derived from pNZ945 for the purpose of sequencing the E11 fragment containing RepFIB (Saul et al., 1989). Since the deletions also contained the pBS+ replicon, RepFIB copy numbers were determined in a DNA polymerase I mutant strain where the pBS+ replicon is inactive.

7.3.1 Determination of the copy number of RepFIB mini-plasmids

RepFIB mini-plasmids were transformed into the pol+ strain PB1966. Plasmid copy number was determined relative to the copy number of a wild type RepFIB mini-plasmid using the PstI 1.2kb fragment labelled with 32P-dCTP (Table 7:5).

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Repeats present</th>
<th>N²</th>
</tr>
</thead>
<tbody>
<tr>
<td>pSS3928, pNZ945</td>
<td>BCDD'D*EFGHU</td>
<td>1</td>
</tr>
<tr>
<td>pMA4322</td>
<td>BCDD'D*EFGHU</td>
<td>2-3</td>
</tr>
<tr>
<td>pAS43</td>
<td>BCDD'D*EFG(Ω)HIJ</td>
<td>6-12</td>
</tr>
<tr>
<td>XS61, SnaBI-BaII Δ</td>
<td>BCDD'D*EFG</td>
<td>2-4</td>
</tr>
<tr>
<td>Xho26, Xho36</td>
<td>BCDD'D*EFGHU</td>
<td>1</td>
</tr>
</tbody>
</table>

a Copy number relative to a RepFIB mini-plasmid (pSS3928).

The deletion of E11 fragment DNA on the left hand side of RepFIB up to 1726cbp does not alter the copy number of the mini-plasmid. Xho26 and Xho36 have the same copy number, despite the fact that Xho36 has a more extensive deletion of DNA than Xho26. Xho26 retains the 'A' repeat element which has been lost in Xho36. Since the presence or absence of this repeat element does not correspond to a
change in copy number, it appears that repeat element A is not involved significantly in the control of plasmid copy number. In contrast to the left hand deletions of RepFIB DNA, deletion of DNA on the right hand side of the replicon which removes the HIJ repeat elements results in an increase in copy number. The copy number of RepFIB can also be increased by inserting the 2.0kb Ω fragment into the SnaBI site located between the EFG and HIJ repeat elements. This observation suggests that the right hand border of the minimal replicon is = 200bp to the left of the basic replicon border7.

A second copy number experiment was conducted in order to determine what affect a second plasmid (pAS18) carrying a copy of the HIJ repeats might have on RepFIB copy number (Table 7:6). The HIJ repeats were carried by pHSG576 which has a pSC101-derived Pol I independent replicon and RepFIB mini-plasmid copy number was determined using the BamHI 0.97kb fragment labelled with 32P-dCTP (this fragment does not include the HIJ repeats). When pAS18 is present in PB1966, the copy numbers of pNZ945, pMA4322 and SnaBI-Ball Δ increase. However, the increase is more dramatic for SnaBI-Ball Δ than for either pNZ945 or pMA4322.

<table>
<thead>
<tr>
<th>Plasmids</th>
<th>Repeats</th>
<th>Na</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>pNZ945 + pAS18</td>
<td>BCDD'D'EFG HIJ</td>
<td>1 → 1.25</td>
<td>Copy number of pNZ945 increases when pAS18 is present.</td>
</tr>
<tr>
<td>pMA4322 + pAS18</td>
<td>BCDD'D'EFGHIJ</td>
<td>1 → 1.25</td>
<td>Copy number of pMA4322 increases when pAS18 is present.</td>
</tr>
<tr>
<td>SnaBI-Ball Δ + pAS18</td>
<td>BCDD'D'EFG</td>
<td>1 → 10</td>
<td>Copy number of SnaBI-Ball Δ increases when pAS18 is present</td>
</tr>
</tbody>
</table>

a Copy number change relative to RepFIB without pAS18.

7.3.1 Comment on copy number determinations
The results of the copy number determinations of RepFIB deletions suggest that the HIJ repeat elements serve to sense and limit RepFIB copy number to 1-2 copies per chromosome. If the HIJ repeats are removed, the copy number increases, suggesting that the repeat elements titrate RepA. These observations can be explained in the context of this control mechanism, if RepA is thought of as a replication initiation factor with a positive influence on plasmid copy number. By removing some of the RepA binding sites from RepFIB, RepA binding at the origin repeats (BCDD'D") is enhanced leading to more frequent initiation of replication and a higher copy number.

7 The minimal replicon is defined by the minimal amount of DNA required for autonomous replication. The basic replicon is defined by the minimal amount of DNA required for autonomous replication with the same control behaviour as the parental replicon.
When additional copies of the HIJ repeat elements are provided in trans, the copy number of RepFIB mini-plasmids increases. This observation appears to be in contradiction to the results described above. However, if in this context RepA is thought of as a repressor of initiation, the copy number response of RepFIB to extra sets of the HIJ repeats in trans can be explained. In this instance, RepA binding to the origin repeats (as well as to the titration repeats) in RepFIB will be lowered due to competition by the HIJ repeats in trans.

The two explanations given here appear to rely on the mutually exclusive assumptions that RepA is either an initiator of replication or a repressor of initiation. This problem has been addressed in the Introduction (Chapter One) where it has posed a major problem in the development of control models for both mini-F and mini-P1.

7.4 Summary of the functional relevance of the DNA repeat groups

A number of in vivo genetic experiments reported in this chapter have investigated the functional relevance of the DNA repeat groups which flank repA in the RepFIB. The BCDD'D" repeat elements located upstream of repA are important components in the autoregulation of repA (Chapter Three) and in the origin of replication. The origin itself appears to be divided into two separate regions, one of which contains the BCDD'D" repeat elements, and the second which contains a single dnaA box and the majority of Dam methylation sites. Both DnaA and Dam are required for RepFIB replication, and the origin region can replicate by itself if a RepFIB mini-plasmid is present in trans.

The EFGHIJ repeats located downstream of repA appear to be involved in the sensing and setting of plasmid copy number. A deletion of the HIJ repeat elements results in an increase in copy number, and copies of the HIJ repeats in trans will further increase the copy number of a RepFIB ΔHIJ mini-plasmid. If the EFGHIJ repeats are separated by the insertion of Ω, the copy number increases suggesting that the titration ability of the repeats is not simply the sum of the separate titration abilities of the EFG and HIJ repeat elements alone. Presumably, it is the interaction between RepA and these repeats which senses and sets RepFIB mini-plasmid copy numbers.

---

8 The activation of the origin of replication presumably only requires RepA in trans, and no other factor expressed by the RepFIB mini-plasmid. This assumption is as yet untested, but is supported by the homology between RepFIB and other Step function plasmids (Chapter One, 1.2; See also Appendix Six, A6.4 and A6.5), deletion and transposon-disruption experiments on P307 RepFIB (Chapter One, 1.3.2) and the RepA DNA-binding studies reported in this Thesis (Chapters Three, Four, and Six).
CHAPTER 8 Review and discussion

This chapter has been divided into a number of sections including a review of the experimental results described in this Thesis (Section 8.1); a summary of the research work (Section 8.2); a discussion of some of the unusual, contentious or conflicting results of this work (Section 8.3); and finally, a description of some future research of potential value for the understanding of the replication strategy of RepFIB (Section 8.4). Page 103 contains a fold-out schematic figure of RepFIB plus a table listing the various areas of research which may be of use when reading this chapter.

8.1 Review of experimental results

The research presented in this thesis has investigated a number of elements or features of the copy number control system of RepFIB. A review of the research is given below in a manner which relates all of the different elements to one another, despite the fact that they have been described in separate research chapters.

8.1.1 Promoter sequences in the E11 fragment

A number of potential promoter sequences in the E11 fragment were located by computer analysis utilising an E. coli consensus promoter sequence (Mulligan et al., 1984). The potential sequences were then examined to determine whether they represented transcriptionally active promoters through the use of promoter-probe plasmids. The formation of protein or gene fusions under the transcriptional control of these sequences was chosen over the more commonly used Northern or primer extension assays since earlier Northern analysis of the E11 fragment were inconclusive (W. Maas, personal communication to P. Bergquist), despite the fact that at least three proteins are known to be expressed by the E11 fragment (Saul et al., 1989).

A computer analysis of the E11 sequence identified eighteen potential promoters. Of these, two produce transcripts which express repA and ORF-7 (repAp and ORF-7p), whilst another two appear to produce transcripts which are not translated (oriP and EFp). repAp, oriP and EFp have been examined in more detail as they are located within the RepFIB minimal replicon portion of the E11 fragment. Each of the three promoters are located within repeat elements which flank the repA gene. repAp is responsible for the expression of repA and is partially covered by repeat element D", oriP is embedded in the BC repeat elements whilst EFp is located in the EF inverted repeat group.
Of the three promoters, only repA<sub>p</sub> appears to have a clear role in copy number control. Although the functional relevance of ori<sub>p</sub> and EF<sub>p</sub> has not been examined, the presence of promoters in the repeat groups has allowed the development of an *in vivo* assay for RepA DNA-binding. In this genetic assay, RepA DNA-binding to the repeat elements represses promoter transcription which results in the reduction of the expression of a readily assayed hybrid β-galactosidase fusion protein.

8.1.2 RepA repression of three distinct promoters

The results of a series of repression tests demonstrate clearly that RepA binds to the BCDD'D'' and EFG repeat groups, and in doing so, represses the transcriptional activity of repA<sub>p</sub>, ori<sub>p</sub> and EF<sub>p</sub>. The individual repression assays do not prove that the repression is due to RepA binding to the repeat elements rather than somewhere else near the promoter sequence *per se*. However, the repression of all three promoters indicates that RepA must bind to DNA common to all three promoter-probe plasmids and the only common sequences are copies of the A-K repeat elements.

A comparison of RepA repression of ori<sub>p</sub> and repA<sub>p</sub> suggests that RepA may recognise and bind to individual repeat elements within the BCDD'D'' group with differing affinities. The repression of repA<sub>p</sub> results in a gradual decrease of transcriptional activity compared to a more rapid decrease of ori<sub>p</sub> promoter activity. This difference implies that there is a significant difference between RepA binding to the BC repeat elements (which cover ori<sub>p</sub>) and binding to the D'D'' repeats which cover repA<sub>p</sub>. One plausible explanation might be that the differential binding may be due to sequence differences between repeat elements within the BCDD'D'' group. However, sequence analysis has not demonstrated any significant difference between the BC and D'D'' repeat elements.

Specialised over-expression plasmids were used to supply RepA *in trans* for most of the repression assays. However, in order to produce such plasmids, the N-terminal sequence of RepA needed to be determined since the sequence analysis of repA had not been able to indicate with confidence which of several start codons was used in the translation of RepA.

8.1.3 The initiation codon used in the translation of repA

The true RepA start codon could not be successfully identified by comparing predicted RepA molecular weights with that determined experimentally since RepA had only been identified in maxicells containing a copy mutant RepFIB mini-plasmid (pMA4322), and not in maxicells containing the wild type mini-plasmid (Saul *et al.*, 1989). Presumably, the increased copy number of pMA4322 resulted in sufficiently high levels of RepA to allow detection by maxicell analysis. However, at that time the sequence of pMA4322 was not available and it was thought possible that the copy mutation was the result of a truncation or extension of the repA coding region producing a protein with an altered molecular weight<sup>1</sup>. Because such an alteration might have occurred through either a N-terminal or a

---

<sup>1</sup> In fact, the copy mutation is the result of a single residue difference in RepA.
C-terminal mutation, the molecular weight of RepA expressed by pMA4322 could not be safely used to determine which of the wild type start codons was used in the translation of RepA.

With these problems in mind, the true start codon used in the translation of RepA was determined directly by limited N-terminal sequence analysis obtained from a RepA:ß-galactosidase fusion protein expressed by a promoter-probe plasmid. The fusion protein was partially-purified by immuno-affinity chromatography and identified using monoclonal anti-ß-galactosidase antibodies. Sequence analysis of the fusion protein demonstrated that the initiation of translation did not begin at either of the start codons predicted by Saul et al. (1989) or by Perez-Casal et al. (1989). The N-terminal residue of the fusion protein is a methionine which corresponds to the initiation of translation from a CTG codon at 22796bp (resulting in the expression of a 38.87kDa protein). The sequence information has been used to produce several RepA expression plasmids and RepA expressed from one such plasmid was used to generate antiserum which confirmed the identity of the RepA:ß-galactosidase fusion protein used to determine the N-terminal sequence of RepA.

8.1.4 RepA-DNA binding in vitro
RepA DNA-binding has been demonstrated in an in vitro system making use of the 'Western-DNA' protocol. Western-DNA has demonstrated that RepA specifically binds to DNA containing copies of the BCDD'D", EFG and HIJ repeat elements. The relatively low degree of binding to DNA lacking repeat elements suggests that a small proportion of RepA present on the membrane refolds into a form which will bind DNA non-specifically.

8.1.5 Three-plasmid assay
An attempt was made to determine whether RepA bound with different affinities to DNA which carried different sets of repeat groups. The three-plasmid assay used to test this proposal involved a RepA expression plasmid which could be induced with IPTG, a 'reporter' plasmid which carried a repA: lacZ protein fusion under the control of repAp (and the BCDD'D" repeats), and a 'third' plasmid which carried a copy of the BCDD'D", EFG, HIJ or EFGHIJ repeats.

The experimental results of this assay lead to two important observations. The first was that RepA bound to DNA repeat elements present on the 'third' plasmid is able to repress repAp present in the reporter plasmid. Such in trans repression occurred when the third plasmid carried a partial or complete copy of the BCDD'D" repeats or the EFG repeats. The second observation was that the repression did not occur when the third plasmid carried the HIJ repeats or the EFGHIJ repeats. These results imply that RepA binding to the HIJ repeats is functionally different from binding to the BCDD'D" or EFG repeats, and that the repressor-effect of RepA bound to the EFG repeats is 'neutralised' by the presence of the HIJ repeat elements. The neutralising effect is distance-dependent, since the separation of the EFG and HIJ repeats by the insertion of Ô resulted in an active repressor element. A cladistic analysis of the sequences of the repeat elements does not reveal a relationship
between the spatial groupings and the sequence conservation of the repeat elements. In addition, the HIJ repeat elements are not significantly different from the other repeat elements in terms of sequence. Therefore, the neutralising effect of the HIJ repeats must be the result of the orientation and spacing of the individual repeat elements, or perhaps, the ability of one or more of the repeats to interact with the EFG repeat elements in cis.

8.1.6 The functional significance of RepA binding to the repeat elements
RepA binding to the repeat elements arrayed on either side of repA are functionally significant in three respects: binding to the BCDD'D'' repeat elements autoregulates repAp, activates the origin of replication, and binding to the HIJ repeats affects RepFIB mini-plasmid copy number.

RepA binding to the BCDD'D'' repeats controls the expression of repA by autoregulation of repAp. Depending on the choice of repAp, reporter plasmid and RepA expression plasmid used in a repression assay, the degree of repA repression varies between 1-50% of the non-repressed repAp transcriptional activity. The lower value corresponds to a test situation in which the copy number of the reporter plasmid is approximately ten times that of a RepFIB mini-plasmid, and where RepA is supplied in vast excess over wild type levels (and in excess with respect to the number of copies of repAp). When the reporter plasmid copy number is very high in comparison to RepFIB and where RepA is supplied at lower levels with respect to repAp, repAp repression is not complete. Although both of these test situations have no natural parallels, these experiments show quite clearly that repA is autoregulated.

The BCDD'D'' repeats also form one half of the RepFIB origin of replication. The origin region has been identified using a rescue assay where a Pol I-dependent plasmid carrying the origin ('origin-plasmid') region is able to replicate in a Pol- host if a RepFIB mini-plasmid is present in trans. Using a similar rescue assay, Gammie and Crosa (1991a) have also mapped the origin of replication to this region of RepFIB. The origin includes the BCDD'D'' repeat elements plus a further ~150bp of DNA which contains a number of sequence features associated with origins of replication. One of these features is a single dnaA box capable of binding DnaA, and DnaA has been shown to be essential for RepFIB replication. The dnaA box and the BCDD'D'' repeats are separated by seven 13bp repeat elements, five of which include the 5'-GATC-3' Dam methylation sequence (all seven are referred to as the GATC-repeats). Although RepFIB mini-plasmids are unable to transform a Dam- strain, it is unclear how the methylation status of the origin GATC-repeats affects replication.

A functional origin of replication requires both the BCDD'D'' repeats and the region containing the dnaA box and GATC-repeats. However, the BCDD'D'' repeats can be inverted with respect to the dnaA box-GATC-repeat region, suggesting that the two regions play functionally different roles during origin activation. This suggestion is further supported by the fact that if the two regions are

separated by 2kb, origin function is lost. In the case of RepFIB, the distance from the single dnaA box to the B repeat element is the same as the distance between the dnaA box and the C repeat element in the BCDD'D"-inversion origin. If the distance between the dnaA box and the RepA binding sites is critical for origin activation, the fact that a BCDD'D"-inversion origin-plasmid replicates almost as efficiently as a BCDD'D" origin-plasmid suggests that only the BCD elements are essential for origin function.

The EFGHIJ repeats also bind RepA. In doing so, RepA represses the activity of EFp located in the EFG repeat group. However, the functional significance of EFp is unknown. In contrast, the HIJ repeats are certainly involved in the setting of RepFIB mini-plasmid copy number as if the repeats are deleted, the copy number increases. If additional HIJ repeats are provided in trans, the copy number of a RepFIB ΔHIJ mini-plasmid or a normal mini-plasmid increases. The most likely explanation of this behaviour is that the HIJ repeats are binding a repressor of replication. The more repressor which is removed from the mini-plasmid by competitive binding to another plasmid carrying the HIJ repeats, the more the mini-plasmid copy number will increase.

Although the copy number experiments do not demonstrate that RepA binding to the EFGHIJ repeats affects RepFIB copy number, observations from other experiments reported in this Thesis demonstrate that RepA can bind to the EFGHIJ repeats. The pattern of copy number changes with different combinations of repeats in cis and in trans is very similar to that seen in other Step function replicons where binding of the initiator protein to DNA repeats sets the plasmid copy number.

8.2 Summary of the research work

The aim of this Thesis has been to determine the fundamental features of the replication and copy number control system of RepFIB. These features have been identified by direct sequence and genetic analysis of the RepFIB basic replicon, as well as through the testing of predictions generated by comparing RepFIB with the related mini-P1 replicon.

A substantial portion of the research work described here investigated the expression and control of the single gene required for replication (repA). This research has shown that RepA translation does not begin from either of the two the predicted start codons, but from a CTG codon further upstream to express a 39kDa protein. In addition, autoregulation of repA by RepA binding to the BCDD'D" repeat elements which overlap the repA promoter sequence (repAp) has been demonstrated. RepA has been shown to bind to the EFGHIJ repeat elements, and these elements appear to be involved in the copy number determination of RepFIB mini-plasmids.

The experimental results reported here suggest that RepA binding to the BCDD'D" repeat elements activates the origin of replication and that RepA autoregulation and titration affect copy number.
These three suggestions indicate in terms of control functions, that RepFIB is very similar to replicons such as mini-P1 and mini-F. The results presented here elevates the level of understanding of RepFIB so that a useful comparison can be made with other Step function replicons. Such a comparison may prove valuable, as there are currently no universally acceptable models to describe the replication control systems of either mini-P1 or mini-F, due mainly to the apparently inescapable paradox posed by concurrent autoregulation and titration of the initiator proteins of these replicons. Having identified RepA autoregulation and titration in RepFIB, any RepFIB control model must also overcome the control paradox.

Since replicons such as RepFIB, mini-P1, Rts1, mini-F, and R6K are so similar in terms of genetic organisation, control elements and control pathways, a successful model for one replicon might be further developed to provide a general model for all step function replicons.

8.3 Discussion

Some of the work presented in this Thesis deserves further discussion due to the unusual or conflicting results they have produced. The following broad areas are discussed in greater depth: the location of the repA promoter, the RepA initiation codon, the multiple functions of RepA and the similarities between RepFIB and mini-P1.

8.3.1 The repA promoter

The repA promoter (repAp) has been located by in vivo genetic analysis using a set of promoter-probe plasmids in which a number of potential repAp sequences were separated by restriction enzyme digestion or PCR amplification. An analysis of the expression of the repA: lacZ protein fusion from these plasmids indicated that one promoter sequence is responsible for at least 86% of the expression of the gene. This promoter sequence has been designated as repAp and is partially embedded within the D repeat element. Although the experimental designation of repAp here appears to be quite straightforward, a second group has designated a completely different repAp sequence based on an analysis of the ColV3-K30 RepFIB repA gene. Gammie and Crosa (1991b) have used primer extension to locate the transcription initiation point (+1 nucleotide) and have then used this information to define the promoter sequence of repAp. The +1 nucleotide corresponds to the cytosine residue at 2163cbp in the P307 RepFIB replicon. Since the sequence of a number of RepFIB replicons is highly conserved in this region (Gibbs, MSc. Thesis, 1991), it is possible that P307 RepFIB repA transcripts originate from this position. However, this assumption would place the +1 nucleotide within the -35 sequence of the promoter I have determined on functional grounds to be repAp.

Gammie and Crosa (1991b) have defined repAp by locating a promoter sequence a suitable distance upstream of the +1 nucleotide determined by primer extension. I do not accept this definition for two reasons. The first is that although the position of the +1 nucleotide is defined primarily by the distance
from the promoter -10 sequence, determination of the +1 site does not define the -35 and -10 sequences per se. Conversely, genetic deletion analysis of a region containing an active promoter sequence does not depend on a specific +1 nucleotide associated with the promoter sequence. The second reason is that although Gammie and Crosa claim that although Gammie and Crosa claim to have identified the -35 and -10 promoter sequences based on the Hawley and McClure (1983) consensus sequence, my examination of the area (using the same consensus sequence) failed to identify a promoter sequence in an equivalent position, despite the fact that P307 and ColV3-K30 RepFIB sequences are 98.7% identical in that region (Gibbs, MSc Thesis, 1991). The designation of two different repAp sequences will need to be resolved by S1 nuclease mapping of transcripts from P307 RepFIB.

8.3.2 Initiation of translation from a CTG codon
Several non-ATG triplets are capable of acting as initiation codons in protein synthesis in prokaryotes. Seven of the possible nine triplets with a single variant nucleotide from the universal 'ATG' codon have been reported as functional start codons (for references, see Romero and García, 1991). The moderate efficiency of initiation from non-ATG codons recalls Crick's wobble hypothesis, which suggests that the pairing between codon and anticodon at the first two positions always follows the usual rules, but exceptional 'wobbles' might occur at the third position (Crick, cited in Romero and García, 1991). These unconventional base pairings occur due to the conformation of the tRNA anticodon loop that allows unusual flexibility at the first base of the loop. However, in the case of initiation codons, it appears that the 'wobble' position need not be at the third position of the codon. Indeed, the initiator tRNA must allow P-site (5') 'wobbles' with the 5' mRNA nucleotide since 'GTG' is a frequently used start codon (Gold, 1988).

During the initiation phase of translation (reviewed by Bosch and Hofstad, 1979; and by Gold, 1988), the 30S small ribosomal unit with associated initiation factors (IF) will bind to mRNA and identify the translational start codon. Once this has occurred, the 50S large ribosomal subunit will bind to the 30S subunit to form a functional ribosome. The initiation factors are no longer required (Bosch and van der Hofstad, 1979); a variety of elongation factors become associated with the ribosome and the first peptide bond is formed.

Initiation factors are used by E. coli to select initiator tRNAs over elongator tRNAs during translation initiation. Initiation Factor Three (IF3) catalyses the formation of 30S initiation complexes (Winterneyer and Gualerzi, 1983) and selects the initiator tRNA complex used to begin translation (Hartz et al., 1989). In effect, IF3 selects specific, charged Met-tRNA^{Met}-30S complexes from a pool of complexes which also contain Met-tRNA^{Met}-30S apparently on the basis of unique sequences located in the anticodon stem and loop of the tRNA^{Met} (Hartz et al., 1990). These unique sequences

3 The programme TargSearch was used to locate promoter sequences (Mulligan et al., 1984). TargSearch uses the Hawley and McClure (1983) consensus sequence to detect and score potential promoter elements.

4 In a 154bp region which covers the BCOVD2 repeats (up to the Psfi site at 2194cbp), P307 and ColV3-K30 differ at only two nucleotide positions, none of which are near either 'repAp' sequences.
involve the anticodon sequence (Berkhout et al., 1986), as well as three GC base pairs adjacent to the anticodon which are also important for in vitro translation (Seong and RajBhandary, 1987).

Hartz et al. (1990) have shown that IF3 appears to 'inspect' the anticodon end of the initiator tRNA^Met, probably along with the initiation codon on the mRNA. In vitro analysis of the IF3 selection mechanism suggests that the tRNA^Met and the mRNA codon can be thought of as a single unit. IF3 allows translation initiation to occur if the anticodon stem and loop sequences are correct, and if the tRNA^Met is associated with either an AUG, GUG or UUG mRNA codon. The 5' nucleotide of the mRNA codon has little influence on IF3 selection and GUG or UUG codons can be used quite efficiently as translation start sites in vivo (Gold, 1988). In vivo translation initiation from CTG has not been reported widely, but Gold suggests that initiation is quite possible from this codon with an efficiency approaching that of initiation from a TTG codon. The 3' nucleotide of the mRNA codon appears to be inspected by the action of IF3; if the nucleotide is not a guanosine, IF3 destabilises the tRNA:mRNA complex and translation initiation does not occur in vitro (Hartz et al., 1990) and very poorly in vivo (Shinedling et al., 1987). However, Köpke and Leggatt (1991) have demonstrated that initiation from an ATA codon is possible using an over-expression plasmid, a result that suggests that IF3 destabilisation does not always prevent gene expression.

Despite the fact that so few reports have been made about natural gene expression initiating from codons other than ATG or GTG, there appears to be no reason why genes starting with a TTG or CTG codon could not be expressed. It appears that the only result of initiation from TTG or CTG is a reduction in the efficiency of the initiation of translation compared with initiation from an ATG or GTG codon. Whilst efficient gene expression is affected by the choice of start codon, the initiation of translation is not prevented so long as the start codon passes IF3 inspection. Although RepA expression was increased experimentally by altering the start codon to ATG, the low-level expression of RepA by RepFIB is clearly sufficient to allow normal copy control and the initiation of replication. In the case of mini-P1, only 20 RepA dimers are found per plasmid copy in vivo (Swack et al., 1987). Since mini-P1 and RepFIB are so similar, I conclude that RepFIB is able to produce equivalent amounts of RepA despite the inefficient initiation of translation without requiring a CTG → ATG mutation to enhance expression.

8.3.3 RepA DNA-binding

RepA binds to the BCDD'D", EFG and HIJ repeat elements in both in vivo and in vitro experiments. In doing so, RepA autoregulates repAp, and represses the transcriptional activity of oriP and EFp. RepA binding to these repeats is also important for origin activation and the sensing and setting of plasmid copy number. The fact that RepA DNA-binding has so many different consequences poses two major problems for any control model which might explain the regulation of RepFIB replication. The first problem is that repA autoregulation and RepA titration can not be concurrent without facing the same

---

5 Personal communication to P.L. Bergquist in a discussion of this work.
control paradox identified in both mini-F and mini-P1. The second problem is that despite the fact that there are only two groups of repeat elements (BCDD'D' and EFGHIJ), RepA binding results in at least three different important control functions (autoregulation, titration and origin activation). This observation implies that RepA binding to one set of repeat elements can result in one of two different outcomes, since all three functions are essentially incompatible.

However, both of these problems may be resolved by the experimental evidence which suggests that RepA binding to the BC and the D'D' elements in the BCDD'D' repeat group is slightly different, and that once bound to the HIJ repeats, a RepA-HIJ complex is significantly different in terms of repressor activity than either a RepA-BCDD'D' or RepA-EFG DNA complex.

These differences in RepA DNA-binding suggest that RepA is able to distinguish between different repeat elements and between the different groups of repeats. This ability may then allow RepA to carry out different control functions at the same time which might be achieved in a manner similar to that described by either of the Hierarchical Binding, Steric Hindrance or Anti-parallel Pairing models (Trawick and Kline, 1985; McEachern et al., 1989; and Abeles and Austin, 1991). However, it is not clear how RepA is able to distinguish between repeat groups or elements. Although the repeat sequences all differ slightly from the A-K repeat consensus, and although the organisation of repeats within the groups differ, there is no obvious correlation between repeat sequence or organisation and the different control functions of the repeat groups themselves.

The lack of such obvious correlations implies that the criteria used to judge the significance of the difference between repeat elements and groups is either faulty or insufficiently resolved. If RepA has different binding affinities for each of the repeat elements, then element placement (with respect to adjacent elements) and sequence must be important. If this is the case, then the A-K repeat consensus sequence fails to identify which of the nucleotides within the sequence are important for RepA recognition and binding. If RepA DNA-binding does differentiate between the different repeat groups, then element sequence, placement, orientation and number must be important. If RepA DNA-binding is co-operative or if RepA DNA-binding involves a dimeric RepA molecule, then these four factors may also affect the RepA DNA-binding affinity for each repeat group.

The fact that RepA bound to the BCDD'D' repeats (or even simply to the D' repeat element) is able to repress in trans a repAP promoter-probe plasmid which contains a second copy of the BCDD'D' repeats.

---

6 See Chapter One, 1.2.4-1.2.10, and Appendix Six, A6.4.6-A6.4.11 (mini-F) and A6.5.8 (mini-P1) for a description of the control paradox and the various models which have been proposed to resolve the problem.

7 Autoregulation and titration are incompatible and result in the control paradox; autoregulation and origin activation must also be incompatible, since both result from RepA binding to the BCDD'D' repeat elements.

8 Claimed by Gammie and Crosa (1991b) on the basis of a filter-binding assay and in vivo repression assay involving deletions of the BCDD'D' repeats. The claim is based on the non-linear relationship between the number of repeat elements present in the DNA and the ability of RepA to bind to the DNA or the ability of the DNA to repress a repAP promoter-probe plasmid (both assays give the same relationship). The contentious issue from these assays is whether or not the relationship is linear or non-linear, since the non-linearity results from the placement of one value (only three deletions of the BCDD'D' repeats plus the BCDD'D' repeats were used).
repeats allows the possibility that RepFIB DNA-looping or inter-replicon pairing may occur in vivo. Although these observations do not imply that such inter or intra-molecular pairing occur per se, the ability of the initiator protein (RepA) to first bind to one set of repeat elements and then bind to a second set is a fundamental requirement of the Anti-parallel Pairing and Steric Hindrance models, both of which have been proposed for Step function replicons (Abeles and Austin, 1991; McEachern et al., 1989) of which RepFIB is an example. The ability of RepA to bind to two sets of repeat elements at the same time would then allow RepA to carry out two functions concurrently. If this was the case, then the control paradox would be resolved, since autoregulation and titration be carried out at the same time.

The different repressor activity of a RepA-HIJ repeat complex compared to either a RepA-BCDD'D' or RepA-EFG complex also raises the possibility that the EFGHIJ repeat group is functionally divided into two different sets of repeat elements. Whilst it is fairly clear that the HIJ repeats are involved in the sensing and setting of copy number, direct evidence that the EFG repeats are similarly involved and are not essential for replication is not yet available.

8.3.5 Comparison of RepFIB and mini-P1
RepFIB and mini-P1 appear to represent two classes of related replicons. Although the two classes differ in cis control sequences (RepA-binding repeat elements), the similarity of the trans control elements (RepA) and the genetic organisation of the replicons support the hypothesis that the two classes are evolutionarily related (homologous) rather than merely functionally similar (analogous). The two replicons are very similar in terms of the organisation of the origin of replication, the initiator gene repA, and the repeat elements which flank repA and form part of the origin (Table 8.1). Although the repA promoter location differs slightly between the two replicons, in both cases the promoter (repAp) is located in the repeat elements found upstream of the gene. In both replicons the gene is autoregulated by the direct binding of RepA to the repeat elements.

Binding to the upstream repeats represses a second promoter element found in both replicons. In the case of RepFIB, oriP is of similar transcriptional strength as repAp. However, the mini-P1 homologue (pl) is approximately 5% of the strength of mini-P1 repAp (Chattoraj et al., 1985a). The upstream repeats form part of the ~300bp origin of replication of both replicons which are bounded by one or two dnaA boxes on the left hand side, and by the five RepA binding elements on the right. A series of Dam methylation sequences are located between the dnaA boxes and the RepA binding sites, and replication of both replicons requires DnaA and Dam methylase.

In the case of RepFIB from ColV3-K30, a deletion or duplication of one of the GATC repeat elements deactivates the origin (Gammie and Crosa, 1991a). Such an alteration also changes the spacing between the dnaA box and the BCDD'D' repeat elements. In the case of mini-P1, alteration of the

---

9 The mini-P1 class includes P7 (Froehlich and Scott, 1988), Rts1 (Kamio and Terawaki, 1983; Kamio et al., 1984; and Nozue et al., 1988) and R401 (Tabuchi, 1985). See Appendix Three, A3.7.
DNA box-RepA binding site distance by as much as one nucleotide deactivates the origin (Brendler et al., 1991b). However, it is not known whether the ColV3-K30 RepFIB origin inactivation is the result of a change in the DNA box-RepA binding site distance per se, or whether deactivation is the result of the enhancement or reduction of some other regulatory system acting at the origin.

Gammie and Crossa (1991a) suggest that Dam methylation of the origin GATC-repeats causes local destabilization of the DNA helix as well as marking the origin DNA for protein recognition. In particular, they suggest that RepA can recognize methylated DNA and that alterations in RepA DNA-binding to the GATC-repeats affect origin activation. Although mini-P1 RepA and RepFIB RepA share the same putative 'GATC-repeat-binding' domain, the variety of in vitro footprinting assays reported in the literature have not demonstrated that RepA binds to the GATC-repeats of mini-P1.

**Table 8:1**

<table>
<thead>
<tr>
<th>Similarities</th>
<th>Differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>RepA protein</td>
<td>Three potential DNA-binding domains and a conserved region not associated with any known functional domain.</td>
</tr>
<tr>
<td>RepA function</td>
<td>DNA-binding, autoregulation, titration, origin activation.</td>
</tr>
<tr>
<td>Origin</td>
<td>Five repeats, Dam methylation sequences, DNA box, contains repA, oriP/pl.</td>
</tr>
<tr>
<td>Repeat elements</td>
<td>Five repeats upstream which form part of the origin, downstream repeats influence copy number.</td>
</tr>
</tbody>
</table>

RepA from RepFIB and mini-P1 appear to have the same roles in the control of replication in each of the replicons. Despite these differences, the two proteins share 66% overall amino acid sequence homology, although the homology is not randomly spread along the lengths of the two proteins and the most highly homologous regions correspond to three very conserved putative DNA-binding domains. Despite the fact that the two proteins have the same potential DNA-binding sites, the RepA binding sites recognised by each protein share no homology.

10 This claim has been made on the basis that RepA shares a potential DNA-binding domain found amongst several bacterial methylases known to recognise GATC sequences. However, the putative 'GATC'-binding domain in RepA overlies a second larger potential DNA-binding domain (domain II) which has been identified using a consensus sequence derived from over fifty different prokaryotic DNA-binding proteins (Nakata and Maizel, 1989). The overlap between the GATC-binding domain and domain II suggests that the GATC-binding domain may simply be part of a general DNA-binding domain, and may in fact not contain the elements necessary to bind specifically to a GATC sequence.

11 The two proteins share a fourth highly homologous region located between residues 25-46 of RepFIB RepA. This region (which is also highly conserved in Rts1 RepA) shares no homology with any known functional protein domains.
8.4 Future RepFIB research

Previous research on RepFIB had suggested that it was a member of the Step function group of replicons. This suggestion was based primarily on the genetic organisation of RepFIB and lead to a number of predictions about how replication and copy number would be controlled by the replicon. This Thesis has examined a number of these predictions and has identified the main elements involved in the control of replication. The work reported here can be further extended in a number of areas, each of which has the potential to answer interesting questions which have relevance not only to an understanding of RepFIB, but to the Step function replicons in general. These extensions may resolve the control paradox faced by Step function replicons generated by the apparently multiple and incompatible roles of the initiation protein of each replicon. Such a resolution would represent a major advance in the understanding of the control and initiation of replication in Step function replicons.

8.4.1 Examination of the promoters

The designation of different repA\textsubscript{p} sequences for two essentially identical RepFIB replicons has meant that the repA\textsubscript{p} of P307 RepFIB should be confirmed by further work. Such confirmation could be obtained by S1 nuclease mapping of transcripts expressed at high copy number from a plasmid such as pAS7, or from a RepFIB mini-plasmid such as pSS3928 at a lower copy number. A genetic examination of the functional role of orip\textsubscript{p} and EF\textsubscript{p} (including nuclease mapping of the transcripts) may also prove valuable, since these promoters may be involved in origin function or the control of replication of RepFIB. Transcription from orip\textsubscript{p} might affect the helix stability of the origin region, alter the binding rates of host factors or provide a RNA molecule involved in the initiation of DNA synthesis. An EF\textsubscript{p} RNA molecule might interact with RepA mRNA in a counter-transcript manner to inhibit translation, or the transcription from EF\textsubscript{p} itself may limit transcription from repA\textsubscript{p}.

The repression of orip\textsubscript{p} and repA\textsubscript{p} by RepA has suggested that there is a difference in the repression of each promoter, despite the fact that repression is achieved by RepA binding to the same repeat elements. Since the differences are subtle and possibly the result of experimental design\textsuperscript{12}, the differences in repression should be investigated further using a divergent promoter-probe plasmid (such as pCB267, Schneider and Beck, 1986). By presenting RepA \textit{in trans}, the simultaneous effect on each promoter could be established, which might in turn demonstrate whether RepA DNA-binding to the BCDD'D\textsuperscript{r} repeat elements is sequential or random.

8.4.2 RepA DNA-binding

The repression of orip\textsubscript{p} and repA\textsubscript{p} by RepA and the Three-plasmid assay suggest that RepA may recognise and bind to different repeat groups or repeat elements with different affinities and results. This suggestion deserves further investigation because differential RepA DNA-binding could solve the

\textsuperscript{12} In particular, any copy number differences between the orip\textsubscript{p} and repA\textsubscript{p} promoter-probe plasmids used in the assay might be responsible for the differences in the RepA repression characteristics of the two promoters.
control paradox faced by RepF1B and other Step function replicons. A variety of double-stranded oligonucleotide probes based on the A-K repeat consensus sequence could be used to determine the relative RepA binding affinities for particular repeat sequences. Although such an analysis may be expensive, an initial comparison of the B and D* repeats alone may be valuable. Such a comparison would require the synthesis of two 50mer oligonucleotides capable of forming hairpin loops, or two pairs of complementary 25mer oligonucleotides which, when annealed, would produce double-stranded DNA suitable for RepA recognition and binding. RepA-probe binding affinities could be determined by Western-DNA or by using a filter-binding approach where RepA-probe complexes were separated from unbound probe DNA by filtration. A filter-binding assay may prove to be more sensitive to differences in RepA DNA-binding affinities than Western-DNA and would be amenable to modifications which might examine DNA-binding directly.

If RepA can refold successfully in solution, the isolation of soluble RepA-DNA complexes may then allow crystallographic analysis of the interaction of RepA and the DNA sequence. Although this type of experiment may prove to be technically difficult, the information obtained could be of considerable value. A comparison of the secondary and tertiary structure adopted of free RepA and RepA bound to a variety of different repeat element sequences might identify the basis of RepA discrimination and DNA-binding ability.

Finally, an examination of RepA-DNA complexes by electron microscopy might indicate whether RepF1B is capable of forming inter and intra-molecular linkages similar to those seen with mini-P1, mini-F and R6k. The microscopy could be extended by examining cell lysates to determine whether similar DNA-linkages occur naturally in the cell, whilst the use of proteolytic cleavage of RepA and a variety of DNA-probes could be used to examine DNA-binding directly.

8.4.3 Similarities with mini-P1

Although the RepA proteins of RepF1B and mini-P1 are very similar in terms of function and the conservation of potential DNA-binding domains, the two proteins differ in terms of protein lengths, and perhaps more significantly, in terms of the repeat element DNA sequences they recognise and bind. An examination of the similarities and differences of the two proteins may prove useful in the understanding of both RepF1B and mini-P1. Relatively simple experiments such as PCR-directed truncation of RepF1B RepA could be used to determine whether the N-terminal or C-terminal sequence extensions compared to mini-P1 RepA are required for activity, and PCR-directed deletion of potential DNA-binding domains could be used to identify which portion of the protein is responsible for DNA-binding. A series of ordered deletions of RepA may also identify which regions of the protein (if any) are required for protein-protein interactions which may affect aspects of replication control. If similar deletion experiments with mini-P1 RepA were carried out, a comparison of the two proteins might provide an insight to how two homologous proteins are able to perform similar control functions by the recognition and binding of quite different DNA repeat sequences.

---

**Figure 8.1** Significance of the repeat groups and differential RepA binding

The basic RepF1B replicon is located in a 1.6kb segment of the EcoR1 4.3kb E11 fragment from P307. The replicon contains a single gene (repA) which expresses a 39kd RepA protein from repAP. The replicon consists of three active promoters (oriP, repAP and EFp) which are embedded within the groups of repeat elements located on either side of repA (BCDD'd*, EFG and HIJ). The origin of replication consists of the BCDD'd* repeats and a number of host-specific binding sites (white box). The BCDD'd* repeats are involved in the autoregulation of repA whilst the HIJ repeats (and probably also the EFG repeats) are involved in the copy number determination. RepA binds to both sets of repeat groups, but with differing affinities and outcomes.

**Table 8.1** Areas of research in this thesis

<table>
<thead>
<tr>
<th>Areas of research in this thesis</th>
<th>Chapter 2</th>
<th>Chapter 3</th>
<th>Chapter 4</th>
<th>Chapter 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Repeat Elements</td>
<td>Three plasmid assay Copy number determination Origin of replication</td>
<td>List of repeat sequences Cladistic analysis of repeat sequences</td>
<td>Three plasmid assay Copy number determination Origin of replication</td>
<td>List of repeat sequences Cladistic analysis of repeat sequences</td>
</tr>
</tbody>
</table>

---

**Chapter Eight** 103
One of the early replication control models for a Step function plasmid incorporated the Clepsydra analogy proposed by Chattoraj et al. (1985a). A clepsydra is a water clock in which time is kept by regulating the supply of water into a reservoir of predetermined volume (a schematic diagram of a clepsydra is given in the Frontispiece). The time is read using a pointer floating in the reservoir and every twelve (or twenty-four) hours the reservoir automatically empties to reset the clock. In the control model, the initiation of replication is analogous to the emptying of the reservoir and occurs when sufficient initiator protein has accumulated within the cell. Although subsequent examination of the mini-P1 and mini-F replicons suggested that the Clepsydra analogy was inadequate, the analogy can be modernised to include recent discoveries and represent accurately the control systems of both replicons as well as RepFIB.

The necessary modifications (shown in the Tailpiece) provide a clock mechanism which, in the corresponding modernised analogy, is able to regulate the accumulation of initiator protein by allowing for variations in copy number, degradation of the initiator and the involvement of specific host replication factors. A feed-back system involving a third reservoir linked to the over-fill pipe can mechanistically account for recent observations such as steric hindrance, whilst the drip tap and spring components can be used to describe cellular degradation of the initiator protein which may occur through modification, proteolysis or dilution by cell division. Finally, a weight system has been added to represent cellular factors which may be involved in the timing of replication. The presence of initiator co-factors may increase the frequency of replication, whilst the presence of inhibitors (as depicted in the Tailpiece) may reduce the frequency of replication. Although the modified clepsydra can account for such recent observations as steric hindrance and allow the involvement of host factors in the timing of replication, the number of containers, joints and pipes would probably be failed by Occum but delight Murphy.
Appendices and References
A1.1 Bacterial strains

The *Escherichia coli* strains which have been used in this research are listed with their relevant genotypes in Table A1.1.

**Table A1.1: Escherichia coli strains**

<table>
<thead>
<tr>
<th>Strain/Synonym</th>
<th>Genotype</th>
<th>Source/Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>PB1849</td>
<td>F-, leu, thi, thy, Δ(proB-lac), dra/drm, dnaA46&lt;sub&gt;ls&lt;/sub&gt;, Str-&lt;sub&gt;r&lt;/sub&gt;</td>
<td>D. Lane</td>
</tr>
<tr>
<td>PB1905</td>
<td>ara, thi, Δ(lac-pro)</td>
<td>B. Kline</td>
</tr>
<tr>
<td>PB1965</td>
<td>Tn10 (Tet-r); isogenic pair with PB1966</td>
<td>P. Bergquist</td>
</tr>
<tr>
<td>PB1966</td>
<td>polA1::Tn10 (Tet-r); isogenic pair with PB1965</td>
<td>P. Bergquist</td>
</tr>
<tr>
<td>PB2946</td>
<td>ara, thi, Δ(lac-proAB), rpsL, ϕ80d1acΔM15</td>
<td>Viera and Messing (1982)</td>
</tr>
<tr>
<td>PB2989</td>
<td>thi-1, dam-3, lacU169, λ(tyrP-lac)</td>
<td>P. Bird</td>
</tr>
<tr>
<td>PB4955</td>
<td>F-, ϕ80d1acΔM15, Δ(lacZYA-argF)U169, recA1, endA1, gyrA96, thi-1, relA1, supE44, hsdR17(rk&lt;sup&gt;-&lt;/sup&gt;, m&lt;sup&gt;+&lt;/sup&gt;k&lt;sup&gt;-&lt;/sup&gt;), λ&lt;sup&gt;+&lt;/sup&gt;</td>
<td>Hanahan (1983)</td>
</tr>
<tr>
<td>AQ699</td>
<td>sdrA224, metD88, ilv, metB, his-29, trpA9605, thyA, deoB (or C), rpoB</td>
<td>Kogoma and von Meyenburg (1983)</td>
</tr>
<tr>
<td>DK249</td>
<td>sdrA244, dnaA50::Tn10, ilv, metB, his-29, trpA9605, thyA, deoB (or C), rpoB</td>
<td>Kogoma and von Meyenburg (1983)</td>
</tr>
</tbody>
</table>

| Str-<sub>r</sub> Streptomycin resistance | Tet-<sub>r</sub> Tetacycline resistance |

A1.2 Growth media

A1.2.1 Media

Bacterial strains were typically grown in either L medium (sometimes referred to as L-broth) cited in Sambrook *et al.*, 1989, as 'LB medium') or minimal (defined) medium using 56/2 Salts (Adelberg and Burns, 1960) supplemented with thiamine (0.5μg/ml), 0.2% glucose, and when required, amino acids.
up to 50μg/ml. Solid media was derived from either L or minimal media by the addition of 2% Davis (NZ) agar. Long term storage of bacterial strains was at -70°C in revco medium.

A1.2.2 Antibiotics
Antibiotics for the selection and maintenance of plasmids in strains were used at the following concentrations in either solid or liquid media: ampicillin (50 or 100μg/ml), chloramphenicol (25μg/ml), kanamycin (40 or 50μg/ml), spectinomycin (50μg/ml), tetracycline (50μg/ml) and trimethoprim (10μg/ml). The lower concentrations were used when the antibiotic resistance was carried by a low copy number plasmid such as a RepFIB mini-plasmid. Chloramphenicol was dissolved in ethanol, tetracycline in 1:1 methanol:ethanol and the other antibiotics were made up in water. Antibiotics were filter sterilised and stored at 4°C or at -20°C (ampicillin).

A1.2.3 β-Galactosidase indicators
The expression of β-galactosidase by colonies on agar plates was detected using plates containing 0.003% Xgal and 0.006% IPTG. Xgal was made up as a 2% solution in dimethylformamide and IPTG as a 4% solution in water. Both were stored at -20°C.

A1.2.4 Methylmethane sulphonate (MMS)
Strains carrying a DNA Pol I mutation (polA1) were checked using L plates containing methylmethane sulphonate at a final concentration of 0.04% (= 10μl spread per plate). The polA1 mutation is sensitive to MMS and polA1 strains will not grow on MMS plates.

A1.3 Transformation

Two procedures were used to transform E. coli strains with plasmid DNA. A two-stage heat shock transformation method modified from the procedure of Lederberg and Cohen (1974) was used preferentially when cells were to be transformed with ligated DNA. The DMSO transformation protocol of Chung et al. (1989) was a very fast and convenient single-step protocol ideally suited to the transfer of plasmid DNA between strains. Although the highest transformation frequencies were obtained using mid-log phase cultures, stationary, early or late-log phase cultures could be made competent using this protocol.

A1.3.1 Heat shock transformation
Mid-log phase cells were resuspended in 0.05-0.1 volumes of cold 100mM MgCl₂ and left on ice for fifteen minutes. The cells were harvested, resuspended in 1ml of 100mM KCl and then left on ice for one hour. The cells were again harvested and resuspended in 1ml of 50mM CaCl₂, 15% glycerol. Cells at this stage could be stored frozen at -70°C for several months or transformed immediately.

1 Revco Medium 42mM K$_2$HPO₄, 22mM KH₂PO₄, 1.7mM sodium citrate, 0.4mM MgSO₄, 30% glycerol.
Plasmid DNA was added to 100-200μl competent cells and the mixture left on ice for thirty minutes. The mixture was heat-shocked in a 42°C water bath for three minutes. Nine volumes of growth medium (typically L-broth) was then added to the cells, and the culture incubated with shaking for thirty to forty-five minutes at 37°C to allow expression of the plasmid-borne antibiotic resistance gene. The culture was then spread on selective agar plates at appropriate dilutions to obtain single colonies. Suitable transformant colonies were restreaked to new selective plates before further use.

**A1.3.2 DMSO transformation**

Mid-log phase cells were harvested and resuspended in 0.5-1ml of TSS medium. The cells were made competent by the addition of 0.05 volumes of dimethyl sulfoxide (DMSO). Plasmid DNA was added to the competent cells and the mixture left on ice for thirty minutes. Nine volumes of growth medium was then added to the cells and the culture incubated at 37°C with shaking for thirty to forty-five minutes before plating out. DMSO was stored frozen under nitrogen.

Satellite-colony contamination was often a problem with this procedure, and was minimised by centrifugation of the transformation mixture after the 37°C incubation and resuspending the cells in fresh L broth before spreading onto the appropriate selective plates.

**A1.4 Isolation of DNA from bacterial cultures**

A modification of the procedure of Holmes and Quigley (1981) was used to prepare plasmid DNA suitable for restriction enzyme digestion on a small scale, or for caesium chloride purification on a larger scale.

**A1.4.1 Mini-preparations of plasmid DNA**

Cells were harvested from a 1-2 ml overnight culture and resuspended in 200-300μl of STET buffer. 2μl of fresh 10mg/ml lysozyme (in STET) was added and the mixture left on ice for ten minutes. The mixture was then centrifuged at 13,000xg for eight minutes and the pelleted cell debris removed with a toothpick. A equal volume of cold isopropanol was added to the supernatant and the DNA left to precipitate on ice or at -20°C for at least thirty minutes. The DNA was pelleted by ten minutes centrifugation at 13,000xg and dried under a vacuum before resuspension in TE buffer.

**A1.4.2 Large-scale preparation of plasmid DNA**

An overnight 600ml culture was harvested and resuspended in 20-25ml of STET buffer. 2ml of 10mg/ml lysozyme was added, and after fifteen minutes incubation on ice the mixture was boiled for

<table>
<thead>
<tr>
<th></th>
<th>Medium/Buffer</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>TSS Medium</td>
<td>1% Difco Bacto-tryptone, 0.5% Difco Yeast extract, 10% PEG 4000, 170mM NaCl, 20mM MgCl₂; pH 6.5.</td>
</tr>
<tr>
<td>3</td>
<td>STET Buffer</td>
<td>8% Sucrose, 5% Triton X-100, 50mM EDTA, 50mM Tris; pH 8.0.</td>
</tr>
<tr>
<td>4</td>
<td>TE Buffer</td>
<td>10mM Tris, 10mM EDTA; pH 8.0.</td>
</tr>
</tbody>
</table>
five minutes. The mixture was centrifuged at 16,000xg for thirty minutes and the DNA in the supernatant precipitated with an equal volume of isopropanol. The DNA pellet was routinely resuspended in TE buffer for cesium chloride purification.

A1.4.3 CsCl purification of plasmid DNA
Plasmid DNA was purified by banding the DNA in cesium chloride in the presence of ethidium bromide (Maniatis et al., 1982) overnight at 50,000rpm in a Sorvall TV-865 rotor, or for forty hours at 40,000rpm in a Beckman 50 Ti rotor. The ethidium bromide was removed from the DNA preparation by repeated washes with butanol, and the cesium chloride removed by dialysis against 1,000-2,000 volumes of cold TE buffer.

A1.5 Isolation of DNA after electrophoresis

A number of methods were used to isolate DNA from agarose gels following electrophoresis. PCR DNA or restriction fragments were isolated from agarose gels using either Geneclean (Bio 101), GELase (Epicentre Technologies), by diffusion or melting of the gel slice followed by phenol extraction and alcohol precipitation.

A1.5.1 DNA isolation by diffusion and phenol-extraction
This method was used to isolate small DNA fragments of 100-200bp from high concentration LMP-TA gels. The gel slice was diced and left in shaking in 500-600μl TE buffer overnight. The gel fragments were removed by centrifugation, the solution extracted with phenol until the interface was clear and the DNA precipitated with alcohol.

A1.5.2 DNA isolation by melting and phenol-extraction
Phenol purification was commonly used to clean DNA of contaminating agarose by adding an equal volume of TE buffer to the gel slice, melting the agarose and repeated extraction with phenol until the interface was clear (Sambrook et al., 1989). The DNA in the aqueous phase was precipitated with either isopropanol or by ethanol and sodium acetate.

A1.6 Electrophoresis of DNA

The examination of DNA by gel electrophoresis was routinely carried out using submerged agarose gels with a Tris-borate running buffer5 (TBE, Maniatis et al., 1982). Agarose gel concentrations of between 0.7% and 2% were used to separate DNA fragments from up to 5-7kb and down to 100-200bp. Ethidium bromide was included in the running buffer and following electrophoresis, DNA bands

---

5 TBE Buffer: 40mM Tris, 0.4mM EDTA, 5mM sodium citrate, pH 7.8. 0.5mg/mL EtBr in gel running buffer.
were visualised using a UV transilluminator. Low melting point (LMP) agarose and a Tris-acetate buffer\(^6\) (TA) were used whenever DNA fragments were to be isolated from gels after electrophoresis.

A1.7 Manipulation of DNA

A1.7.1 Enzymes
Restriction enzyme digestion, DNA ligation and the fill-in or blunting of sticky ends with DNA Polymerase or Klenow fragment followed general protocols such as those described in Maniatis et al. (1982), Sambrook et al. (1989) or Ausubel et al. (1989). Most restriction enzymes were supplied by Bethesda Research Laboratories (BRL), with the exception of PstI from New England Biolabs, PvuI from Pharmacia and SnaB1 from Boehringer Mannheim (BM). BM T4 Ligase, T4 DNA Polymerase and Klenow fragment were used for DNA modification. The dephosphorylation of DNA was achieved using either BM Alkaline Phosphatase or the heat sensitive HK Phosphatase supplied by Epicentre Technologies.

A1.7.2 Polishing reactions
The 'sticky ends' produced by some restriction enzymes and by Taq polymerase during PCR amplification were made flush (polished) using Klenow or DNA Polymerase I with 20\(\mu\)M dNTPs in BRL REact I buffer at 37\(\degree\)C for 15-20 minutes. In some instances, NEN \(^{32}\)P-dCTP was used instead of dCTP to produce labelled DNA probes.

A1.7.3 Labelling DNA by nick-translation
A Bethesda Research Laboratories' Nick Translation kit was used to label DNA fragments and whole plasmid DNA with NEN \(^{32}\)P-dCTP or \(^{35}\)S-dCTP for use as probes in Southern analysis, colony hybridisations and in copy number experiments. Unincorporated label was removed from the probe by successive alcohol precipitations of the probe DNA.

A1.7.3 Labelling DNA by random-priming
Radioactively labelled DNA probes were also produced using Amershams' Multiprime DNA Labelling System. The kit was used to label whole plasmid DNA with NEN \(^{32}\)P-dCTP, and unincorporated label was removed from the probe by alcohol precipitation.

A1.7.4 PCR synthesis of DNA
The Polymerase Chain Reaction (PCR) was used to synthesize defined DNA fragments as described by Sambrook et al. (1989). Oligonucleotide primers were made according to an analysis of the RepFlB sequence and to computer predicted-oligonucleotide melting temperatures (Table A1:2). Some

---

6 TA Buffer 90mM Tris, 1.25mM EDTA, 90mM boric acid; pH 8.4. 0.5mg/ml EtBr in gel running buffer.
primers were altered to include restriction sites not present in the RepFlB sequence. These modifications were usually at the 5' end of the primer, and did not change the ability of the primer to correctly recognise the appropriate template sequence during PCR amplification. Additional nucleotides were added 5' to the restriction sites to allow efficient enzymatic cleavage of the PCR product (1 nucleotide for EcoRI, ≥5 for BgIII and NcoI).

<table>
<thead>
<tr>
<th>Table A1:2</th>
<th>PCR oligonucleotide primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oligo</td>
<td>RepFlB bp</td>
</tr>
<tr>
<td>#2 3306-3282 Oligonucleotide GGAATTCATCAGCGGGATTTGAAGA RepFlB TTTA GT</td>
<td>C-terminus of repA two stop codons followed by an EcoRI site.</td>
</tr>
<tr>
<td>#6 2736-2716 Oligonucleotide GATCACGGTACCGGCAATGCGC RepFlB</td>
<td>Central portion of repA covering the Kdel site.</td>
</tr>
<tr>
<td>#8 2091-2114 Oligonucleotide CTGTAACATGCTAATGATAAGCTG RepFlB</td>
<td>Located inside the origin repeat elements (C and D).</td>
</tr>
<tr>
<td>#9 2142-2165 Oligonucleotide GTTATGACATAAATGATAAGCTG RepFlB</td>
<td>Located inside the origin repeat elements (O' and D').</td>
</tr>
<tr>
<td>#17 2270-2302 Oligonucleotide GAGGCCATCCATGGAAACGGAATTCAGACATC RepFlB TTC</td>
<td>N-terminus of repA start codon has been altered 'ATG' and includes a NcoI site.</td>
</tr>
<tr>
<td>#21 1720-1750 Oligonucleotide ATATGAAATTCTACCTCAGGATATGACGTACG RepFlB GTG</td>
<td>Equivalent to the XhoI left hand border and includes an EcoRI site.</td>
</tr>
<tr>
<td>#24 3315-3283 Oligonucleotide CGACTACCCATGGAAATTCAGGATATTTGAAG RepFlB AGT TATG</td>
<td>C-terminus of repA two stop codons followed by EcoRI and NcoI sites.</td>
</tr>
<tr>
<td>#26 2290-2265 Oligonucleotide TCGTTTCCATAAGCCCTCAGCCT RepFlB G</td>
<td>Alteration of the repA translation start codon. Complement of #25.</td>
</tr>
</tbody>
</table>

Restriction sites are **underlined**, alterations to the RepFlB sequence are shown in red.
DNA synthesis used either Cetus Corporation AmpliTaq or BRL TaqI polymerase. Pharmacia Ultrapure dNTP's were used to make 50x dNTP mixes buffered in 20mM Tris, pH8.8 (12.5mM of each dNTP). 10x Optaq buffers were used to give a range of Mg concentrations from 5-50mM per reaction. The 50-100μl reaction mixture typically contained 0.25 units of polymerase, oligonucleotide primers diluted to 0.1μM, and between 20-100ng plasmid (template) DNA. A further 30μl of mineral oil was placed on top of the reaction mixture to prevent evaporation during amplification. DNA amplification was controlled automatically and involved 10-30 amplification cycles. The reaction temperature was maintained at 4°C until the samples were collected to avoid degradation of the PCR products.

A1.8 Colony hybridisation

Colony hybridisations were carried out according to the protocol of Bergquist (1987). Colonies were replica plated onto NEN Genescreen-plus membranes, and the cells lysed by floating the membranes on 750μl puddles of 0.5M NaOH for ten minutes. The membranes were neutralised with 1M Tris (pH 7.5) for one minute, and then transferred to 1.5M NaCl, 0.5M Tris (pH 7.5) for five minutes. The membranes were transferred to two new 1.5M NaCl, 0.5M Tris puddles for five minutes each before drying at 42°C for forty-five minutes. The membranes were washed in ethanol and dried for a further ten minutes at 42°C before prehybridisation. The membrane was prehybridised for six hours at 42°C. The radioactively-labelled probe was boiled for five minutes and then added to prewarmed hybridisation buffer at 42°C. The membrane was then transferred to the hybridisation buffer and left shaking at 42°C for 12-16 hours. After hybridisation, the membrane was placed in 100-200ml of prewarmed solution of 2xSSC, 0.1% SDS and incubated for twenty minutes at 42°C. The wash was repeated and then the membrane transferred to prewarmed 0.1xSSC, 0.1% SDS and washed twice at 42°C. The membrane was then rinsed with water, blotted dry and autoradiographed. The stringency of this wash regime allows a mismatch of 3.2%.

A1.9 Copy number determination

The protocol for this procedure is described by Maas et al. (1987). Log phase cultures were harvested and resuspended in 0.8ml of 0.5M NaOH, 1.5M NaCl and boiled for three minutes. The mixture was plunged into ice and neutralised with 0.8ml of 1M Tris (pH 7.5), 2M NaCl. Each sample was diluted and aliquots from each dilution filtered on to a NEN Genescreen-plus membrane. Each well

<table>
<thead>
<tr>
<th>Buffer Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>7 OpTaq Buffer</td>
<td>(10x) 0.5-50mM MgCl₂, 500mM KCl, 100mM Tris (pH 8.8), 0.01% gelatin.</td>
</tr>
<tr>
<td>8 Prehybridisation Buffer</td>
<td>6x SSC, 1% SDS, 50 mg/ml sonicated Calf Thymus DNA, 50% formamide, 5x</td>
</tr>
<tr>
<td></td>
<td>Denhardt's solution.</td>
</tr>
<tr>
<td>9 Hybridisation Buffer</td>
<td>6x SSC, 1% SDS, 50 mg/ml sonicated Calf Thymus DNA, 50% formamide, 5x</td>
</tr>
<tr>
<td></td>
<td>Denhardt's solution, 10% Dextran sulphate.</td>
</tr>
<tr>
<td>10 SSC</td>
<td>(20x) 3M NaCl, 0.3M tri-sodium citrate.</td>
</tr>
</tbody>
</table>
was washed with 100μl 1M Tris (pH 7.5), dried and then hybridised as described for colony hybridisations.

DNA binding to the membrane was affected by the amount of cell debris contained within the sample. To avoid problems associated with DNA binding due to varied culture densities or plasmid copy numbers, each sample was made up to a standard cell number of 10⁸ cells by the addition of plasmid-free cells before resuspension after culture harvesting. Culture densities were determined by the optical density of the culture at 600nm (OD₆₀₀) and were confirmed by colony counts of culture dilutions spread on appropriate selective agar plates.

### A1.10 Preparation of cell extracts

#### A1.10.1 8M urea extracts containing RepA

A 40ml overnight culture was used to inoculate 1.2 litres of L-broth containing 100μg/ml ampicillin. The culture was incubated at 30°C with vigorous shaking for two hours before transfer to 42°C to induce expression and incubated for a further three hours. The cells were harvested and resuspended in 10ml lysis buffer¹¹ (Ausubel et al., 1989). 8μl of 50mM PMSF and 200μl of 10mg/ml lysozyme were added and the cells left on ice for twenty minutes. 250μl of 40mg/ml deoxycholic acid and 50μl of DNase I were added and the cells left at room temperature for twenty minutes. The viscosity of the solution was reduced by passage through a 20 gauge needle before the cell debris were collected by centrifugation at 12,000 xg for ten minutes. The debris was resuspended in 10ml lysis buffer, 0.5% Triton X-100, vortexed and left at room temperature for ten minutes. The insoluble material was collected by centrifugation at 12,000 xg for ten minutes and then resuspended in 10ml 8M urea, 1M NaCl, 50mM Tris (pH 8.7). The sample was vortexed and left at room temperature for ten minutes before centrifugation at 12,000 xg for ten minutes. The supernatant was collected and stored at -20°C until required.

#### A1.10.2 Extracts containing the RepA:β-galactosidase fusion protein

The disintegration of cells followed a modification of the procedures described by Germino et al., (1983), Carrol and Laughon (1987) and Erickson and Steers (1970). Cells were harvested from overnight cultures and resuspended in 10 volumes (of the cell pellet) of B⁺ buffer¹², 20mg/ml fresh lysozyme was added to the suspension and left to stir on ice for thirty minutes. The cells were then sonicated with three ten second bursts with a Soniprep sonicator (MSE Scientific Instruments) before centrifugation at 12,000 xg for thirty minutes. The supernatant was poured off the pellet of cell debris and kept on ice before use.

<table>
<thead>
<tr>
<th></th>
<th>Lysis Buffer</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>¹¹</td>
<td>50mM Tris (pH 8.0), 1mM EDTA, 100mM NaCl.</td>
<td></td>
</tr>
<tr>
<td>¹²</td>
<td>0.25M NaCl, 10mM, magnesium acetate, 10mM β-me, 5% glycerol, 20 mM Tris, 1mM PMSF, 2mM iodoacetemate, 10mM EDTA; pH 7.6.</td>
<td></td>
</tr>
</tbody>
</table>
A1.10.3 Preparation of tubing for the dialysis of protein samples

Dialysis tubing was prepared by boiling 2-3 metres of tubing in 2 litres of deionised water for five minutes. The water was then drained off, and the tubing boiled for five minutes in 0.1% Na₂CO₃. The tubing was then boiled again in deionised water for five minutes, then in 0.1% EDTA for another five minutes. The tubing was boiled once more in deionised water and then stored at 4°C until required.

A1.11 Cleavage of protein samples

8M protein extracts were cleaved by incubation in formic acid and in a formic acid/cyanogen bromide solution. 500µl of water was added to 25µl portions of the protein extract and freeze-dried. The lyophilised material was resuspended in 500µl water before freeze-drying for a second time. Finally, the material was resuspended in 75µl 70% formic acid (pH 2.5 digestion) or 75µl 70% formic acid, 130mg/ml cyanogen bromide (cyanogen bromide digestion). The pH 2.5 digestion was carried out at 37°C for 24 hours, whilst the cyanogen bromide digestion took place at room temperature overnight. After digestion, the samples were diluted with 500µl water and freeze-dried twice before resuspension in 100µl TBS buffer. The digests were immediately used in Western-DNA analysis.

A1.12 Protein concentration

A1.12.1 Precipitation with ammonium sulphate

Crude cell extracts were differentially precipitated by the addition of solid ammonium sulphate at 0°C (England and Seifer, 1990). Ammonium sulphate was added to protein samples to the appropriate concentration (calculated from Dawson et al., 1986), and the solution gently stirred at 0°C for two hours before centrifugation at 12,000xg for thirty minutes. The ammonium sulphate was removed from the sample by dialysis.

A1.12.2 Precipitation with acetone

Nine volumes of ice-cold acetone was added to the protein sample and left on ice or at -20°C for thirty minutes. The solution was centrifuged at 12,000xg for twenty minutes and the acetone drained from the pellet. The sample was dried under vacuum and resuspended in a suitable buffer. It is important to note that acetone precipitation is inefficient with low protein concentrations, and that other components of the sample may co-precipitate with the proteins (for example, salts and SDS).

A1.12.3 Freeze drying

Samples were concentrated by freeze drying (lyophilise) and resuspension in an appropriate buffer.

TBS Buffer

20mM Tris, 137mM NaCl; pH 7.6.
A1.13 Electrophoresis of proteins

A1.13.1 SDS-PAGE
Protein samples were analysed by SDS-PAGE according to the method of Laemmli (1970) described by Hames and Rickwood (1991). Small 70 x 100 x 0.75 mm gels were used, and could be cast and stored with the stacking gel poured for one to two weeks without deterioration of the performance of the gel. Samples were boiled for two minutes with an equal volume of 2x sample buffer, centrifuged briefly and loaded onto the gel. Typically the gels were run at 90-150 Volts, and electrophoresis completed in one to two hours. SDS polyacrylamide gels were electrotransferred onto a membrane, or stained with coomassie blue and dried onto filter paper.

A1.13.2 Staining SDS-PA gels with coomassie blue
Polyacrylamide gels were stained with 0.1% coomassie blue, 50% methanol, 10% acetic acid for thirty minutes, then washed several times with 10% methanol, 7% acetic acid until adequate destaining was achieved.

A1.14 Western analysis
Proteins were transferred from SDS polyacrylamide gels following electrophoresis onto nitrocellulose or polyvinylidene difluoride (PVDF) membranes after Timmans and Dunbar (1990).

A1.14.1 A general protocol for the transfer of proteins to membranes
SDS-PA gels were first washed for ten minutes in transfer buffer following electrophoresis. The separated proteins in the gel were electrotransferred to nitrocellulose, Amersham Hybond-ECL or Millipore Immobilon membranes over two hours at 50 Volts. After transfer proteins could be visualised on the membrane with amido black stain or proteins identified by immunodetection using appropriate antibodies or antiserum.

A1.14.2 Transfer to Immobilon for sequencing purposes
A PVDF membrane (Immobilon, Millipore) was used to bind proteins in a Western transfer in order to allow amino-terminal protein sequencing (Matsudaira, 1990). This procedure differs from the general protocol used for Western transfer of proteins following SDS-PAGE: i) polyacrylamide gels were aged overnight before use; ii) samples were boiled for one minute in a modified sample buffer lacking coomassie blue; and iii) the gel was not washed in transfer buffer prior to electrotransfer.

14 Sample Buffer (2x) 2% SDS, 10% glycerol, 0.005% bromophenol blue, 10mM Tris, 0.3M 8-me; pH 6.8.
15 Transfer Buffer 20% methanol, 192mM Tris, 89mM glycine.
16 Sample Buffer (2x) 2% SDS, 10 glycerol, 10mM Tris; pH 6.8.
17 Transfer Buffer (for PVDF membranes) 10% methanol (HPLC grade), 10mM CAPS (pH 11), 0.05% SDS.
Optimal transfer of the β-galactosidase fusion protein involved transfer for thirty-five minutes at 90 Volts and the inclusion of 0.05% SDS in the transfer buffer (Lissilour and Godinot, 1990). Following transfer, the membrane could be stained with coomassie blue for sequencing or used for immunodetection.

A1.14.3 Staining PVDF membranes with coomassie blue
Coomassie blue staining of PVDF membranes followed the protocol of Matsudaira (1990). The membrane was washed in two changes of transfer buffer then stained in a fresh solution of 0.1% coomassie blue, 50% methanol (HPLC grade) for two minutes. The membrane was destained by washing in five 10ml aliquots of 50% methanol, 10% acetic acid (both HPLC grade) until the background had cleared. The membrane was finally washed in water twice, blotted dry and stored at -20°C until required for sequencing.

A1.14.4 Staining Hybond-ECL membranes with amido black
Hybond-ECL membranes were washed in 0.2% amido black, 7% acetic acid for two to three minutes, then washed thoroughly in water. Sigma Amido Black solution (2x) was also used for detection.

A1.14.5 General incubation protocol for immunodetection
A number of different immunodetection systems were used to identify proteins on Western or dot-blotted membranes. However, the same system of incubation times and washing steps were used before the final chemical reaction.

The membrane was blocked by soaking overnight at 4°C or for at least two hours at room temperature in TBS18, 0.05% Tween-20, 3% Anchor non-fat milk powder (NFMP). The blocking solution was diluted by adding four parts of TBS, 0.05% Tween-20 to form the buffers for primary and secondary antibody incubations. The membrane was incubated for two hours with the primary antibody (or antiserum) and then the membrane was washed. Membranes were washed four times with water for five minutes each, then twice with TBS, 0.05% Tween-20 for ten minutes each. Incubation with the secondary antibody in diluted blocking solution was for one hour before the membrane was washed again. The membrane was washed once in TBS for ten minutes before the chemical colour or light reaction was started.

A1.14.6 Immunodetection using anti-β-galactosidase antibodies and an anti-mouse-HRP conjugate
Promega monoclonal mouse anti-β-galactosidase antibody was used as primary antibody to detect β-galactosidase fusion proteins. The antibody was used at a 1/1000 dilution and was recognised by an anti-mouse rabbit immunoglobulin-HRP (Horse radish peroxidase) conjugate (Promega) at a 1/2500 dilution. After the final round of membrane washing, the detection reaction was carried out in 50ml of

---

18 TBS Buffer 20mM Tris, 137mM NaCl; pH 7.6.
TBS, 15μl 30% H₂O₂ and 10mg 3,3’-diaminobenzidine (DAB). The colour developed within five minutes and the reaction was terminated by thorough washing in distilled water. The membrane was stored in the dark to prevent fading of the signal and the browning of the membrane.

A1.14.7 Immunodetection using RepA antiserum, anti-rabbit-Ig-biotin and a streptavidin-HRP complex

RepA rabbit antiserum bound to RepA was detected using anti-rabbit biotinylated-immunoglobin (Amersham) at a 1/500 dilution in the second incubation, followed by a streptavidin-HRP (Horse radish peroxidase) complex in a third incubation for 30 minutes in TBS at a 1/600 dilution. After the final round of washing, the detection reaction was carried out in 20ml of TBS, 5ml 3mg/ml 4 chloro-napthiol (in methanol) and 12μl 30% H₂O₂. The colour developed within 15 minutes and the reaction was terminated by thorough washing in distilled water. The membrane was stored in the dark to prevent fading of the signal and the browning of the membrane.

A1.14.8 Immunodetection using RepA antiserum and an ECL kit

An Amersham ECL (Enhanced chemiluminescent) detection kit was used to detect RepA antiserum bound to RepA. After electro-transfer, the membrane was processed according to the kit instructions. RepA antiserum was used at a 1/20000 dilution, and the secondary anti-rabbit immunoglobin-HRP conjugate at a 1/10000 dilution. The chemiluminescence was detected using Hyperfilm-ECL sheets (Amersham).

A1.14.9 Limited N-Terminal sequencing from PVDF membranes

Limited N-terminal amino acid sequence was obtained from PVDF-blotted, coomassie blue stained protein samples after the method of Matsudaira (1990).

A1.15 Western-DNA analysis of RepA DNA-binding

The Western transfer of proteins separated by SDS-PAGE has been adapted in this Thesis to allow the analysis of RepA DNA-binding (described in detail in Chapter Six, 6.3). Protein samples were separated by SDS-PAGE using 12.5% gels. After electrophoresis, the gel was washed three times in transfer buffer¹⁹ for five minutes each before electrotransfer to Amersham Hybond-ECL membrane over two hours at 50 Volts. The membrane was then washed three times in TBS, 0.5% Tween-20 for five minutes each and then blocked in TBS, 0.5% Tween-20, 3% non-fat milk powder (NFMP), 0.01mg/ml sonicated calf thymus DNA (CTD) for two hours at room temperature. The membrane was incubated for a further two hours with double stranded ³²P-end labelled probe DNA in TBS, 0.5% Tween-20, 0.6% NFMP, 0.01mg/ml sonicated CTD. The membrane was washed three times with TBS, 0.5% Tween-20 for five minutes each, before blotting dry and autoradiography at -70°C.

¹⁹ Transfer Buffer: 20% methanol, 192mM Tris, 89mM glycine.
A1.16 Protein assays

A1.16.1 Estimation of the amount of total protein in samples
An indication of total protein concentrations in ammonium sulphate supernatants and precipitants was gained using the dye-binding methods of Bradford (1976) and Spector (1978). Protein concentrations were also estimated in gel filtration column fractions by determining the optical absorbance of the sample at 280nm.

A1.16.2 Assay for β-galactosidase activity
Quantitative β-galactosidase assays of bacterial cultures was by a modification of the method of Miller (1972). Up to 400µl of bacterial culture could be assayed, and the culture was made up to 800µl with Z buffer. The cells were lysed by the addition of 10µl toluene and vigorous vortexing, and the sample was kept on ice until required. Immediately prior to use, the sample was equilibrated at 37°C for two and a half minutes before the addition of 160µl of 4mg/ml ONPG initiated the reaction. The reaction was terminated by the addition of 400µl 1M sodium carbonate after appropriate colour development of the sample. Each sample was centrifuged at 13,000xg for one minute before absorbance readings were made at OD420 and OD550. β-galactosidase activity in Miller Units is determined by the equation:

\[
\frac{800 \times (OD_{420} - 1.75 \times OD_{550})}{T \times C \times OD_{600}}
\]

where time (T) is measured in minutes, C is the culture volume in millilitres and the cell density of the culture is read at OD600. The same protocol was used for the qualitative determination of β-galactosidase activity in crude cell lysates, ammonium sulphate suspensions and precipitates, and in some column eluent fractions.

A1.16.3 In vivo 14C-labelling of proteins
A bacterial culture was labelled with a mixture of 14C-labelled amino acids in order to determine changes in the in vivo translation of proteins following a temperature shift from 30°C to 42°C. An overnight culture (160µl) was used to inoculate 20ml of L-broth which was then grown for two hours at 30°C. The culture was shifted to 42°C, and 50µCi 14C-labelled amino acids (Amersham High Specific Activity Amino Acid Mixture, RD4656) added to the culture. Samples were removed from the culture after various time intervals after the shift in temperature in order to analyse the radiolabelling of newly synthesised proteins. The samples were analysed by SDS-PAGE and after staining with coomassie blue, the gels were washed in Amplify (Amersham), dried onto pieces of filter paper and autoradiographed.

20 Z Buffer 60mM NaHPO4, 40mM NaH2PO4, 10mM KCl, 1mM MgSO4, 50mM β-me. This buffer must be freshly made.
A1.17 RepA rabbit antiserum

A1.17.1 Preparation of RepA antigen
RepA antigen was obtained from partially purified RepA samples by preparative SDS-PAGE. Cells containing pAS49 were harvested after a period of growth at 42°C and a crude preparation of RepA was obtained in 8M urea. The urea extract was then dialysed overnight against 1M urea, 50mM Tris (pH 8.7). The sample was centrifuged at 12,000xg for thirty minutes and aliquots of the supernatant run on preparative (1.5 x 70 x 100 mm) 12.5% SDS-PAGE gels to separate the proteins. After electrophoresis, the gels were stained with coomassie blue for 5 minutes, washed with destain for two minutes and then with water for a further two minutes. The lightly-stained band corresponding to RepA was cut from the gel and the fragment washed five times with water before storage at -20°C until required.

A1.17.2 Immunisation with RepA
Acrylamide gel slices containing RepA were crushed in an eppendorf tube and then passed through a 25 gauge needle. The crushed gel was mixed with an equal volume of Freund’s incomplete adjuvent and used to immunise a Chinchilla rabbit. Twenty small dorsal intradermal injections were made according to Sambrook et al. (1989). The rabbit was given a RepA booster shot prepared by electroelution (D. Christie) on Day 40. A small sample of blood was tested on Day 48, boosted again on Day 71 and bled on Day 78.

A1.17.3 Preparation of the antiserum
The rabbit blood was allowed to clot overnight at 4°C and the clear serum removed using a pippette. The serum was briefly centrifuged at 5,000xg and the supernatant aliquoted for storage. Some of the antiserum was stored at -70°C whilst the rest was lyophilised in 1ml units.

A1.18 Immunoaffinity chromatography
A 1ml bed-volume anti-β-galactosidase immunoaffinity column (Promega) was used to isolate the β-galactosidase fusion protein expressed from pAS7 in PB1905. The sample was first dialysed against TEP buffer21 and cleared by centrifugation before passage through the column according to the manufacturer’s instructions. The column was washed with 25ml TN buffer22 before elution with three 1ml aliquots of 0.1M sodium carbonate (pH 10.8), immediately followed by 1ml of TBS buffer23. The resulting 4ml eluate was collected, dialysed against 5mM Tris (pH7.5), and then lyophilised in several aliquots. It was necessary to immediately re-equilibrate the column after use with TBS.

<table>
<thead>
<tr>
<th>Buffer</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>TEP</td>
<td>100mM Tris (pH 7.4), 10mM EDTA, 1mM PMSF.</td>
</tr>
<tr>
<td>TN</td>
<td>50mM Tris (pH 7.3), 0.2% NP-40.</td>
</tr>
<tr>
<td>TBS</td>
<td>20mM Tris, 137mM NaCl; pH 7.6.</td>
</tr>
</tbody>
</table>
A1.19 DNA sequence analysis

University of Wisconsin Genetics Computer Group (UWGCG) programmes were used to analyse DNA sequences, compare DNA sequences obtained from GenBank and to predict protein secondary structures. DNA sequences showing significant similarity to the consensus Escherichia coli promoter sequence were located using TargSearch (Mulligan et al., 1984). Statistical examination of β-galactosidase assays involved both CricketGraph (Cricket Software, 1987) and StatView 512+ (Abacus Concepts, 1986).

A1.20 Imaging techniques

Autoradiographs were made using Cronex 8 Medical X-Ray film or Amersham Hyperfilm-MP. Western-ECL membranes were exposed to Amersham Hyperfilm-ECL. A 35mm camera was used to take photographs of ethidium bromide-stained DNA agarose gels, coomassie blue-stained SDS-polyacrylamide gels, radioautographs and ECL films. The Professional Image Processing System (Digital Optics Ltd, 1990) was used to analyse autoradiographs.
A2.1 Plasmids used for cloning

Plasmids used in cloning for this thesis are listed and described briefly in Table A2.1. A number of RepFIB plasmids were also used both in experimentation and as a source of DNA for cloning and PCR amplification. Those plasmids are listed in Table A2.2.

A2.1.1 General plasmids used in this thesis
The general high copy-number plasmid pBS+ (3.0kb; Stratagene) was used in the majority of cloning work for this Thesis. pACYC184 (4.24kb; Chang and Cohen, 1978) and pCGN565 (2.48kb; Stalker et al., 1988) have also been used for general cloning purposes. Several plasmids generated for this thesis are deletion derivatives of pNZ945 or Xho31.

pHSG576 (3.61kb; Takeshita et al., 1987) and pUC19 (2.69kb; Yanisch-Perron et al., 1985) have been used as control plasmids for some transformation assays. pHSG576 replication is independent of DNA polymerase I (Pol I) whilst the replication of pUC19 (and pBS+) is dependent on host Pol I. The replication of plasmids such as pUC19 and pBS+ is also independent of host DnaA.

A2.1.2 Promoter-probe plasmids used in this thesis
The promoter-probe plasmids pMLB1034 (6.259kb; Shultz et al., 1982), pNM480, pNM481, pNM482 (8.6kb; Minton, 1984) and pNZ3381 have been used to produce fusion proteins to determine whether particular open reading frames (ORFs) are capable of expression. Where successful fusions were formed, the expression of the β-galactosidase fusion protein was readily detected using agar plates containing Xgal. Some plasmids were further derivatised by the insertion of the 2.0kb Omega spectinomycin resistance fragment (ΩSpec; Prentki and Krisch, 1984) upstream of the protein fusion in order to prevent vector-initiated transcription affecting the expression of the gene.

pKK232-8 (5.1kb; Brosius, 1984) and pMU575 (15kb; Yang and Pittard, 1987) have been used to construct gene fusions where protein fusions were not possible. Successful gene fusions in pKK232-8 which expressed the plasmid cml gene (chloramphenicol acetyltransferase) were detected by the

---

1 pNZ388 is a derivative of pMLB1034 (R. de Feyter, this laboratory). The lacZ fusion gene segment of pMLB1034 (including the EcoRI, Smal and BamHI sites) was transferred into pLG338 to produce a kanamycin resistant, low-copy number promoter-probe vector.
growth of colonies on agar plates containing chloramphenicol. Successful fusions in pMU575 expressed a hybrid β-galactosidase which was readily detected using agar plates containing Xgal.

A2.1.3 Expression plasmids used in this thesis

pJLA602 is a high copy-number over-expression plasmid utilising a hybrid λP LR promoter for expression and the c1857 gene for regulation (4.9kb; Schauder et al., 1987). Expression of the cloned gene is induced by incubation at 42°C. pKK233-2 is second high copy-number expression plasmid where the expression from a hybrid trp/lac promoter (pLcT) is induced with IPTG (4.6kb; Amann and Brosius, 1985). Both plasmids contain transcription terminator sequences downstream of the cloning site.

### Table A2:1

<table>
<thead>
<tr>
<th>Vector</th>
<th>Resistance</th>
<th>Cloning plasmids used in this thesis</th>
<th>Source/Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>pACYC184</td>
<td>Chloramphenicol, Tetracycline</td>
<td>Cloning plasmid, compatible with plasmids containing pMB1 or pMB1-derivative replicons.</td>
<td>Chang and Cohen, 1978</td>
</tr>
<tr>
<td>pBS+</td>
<td>Ampicillin</td>
<td>General cloning plasmid, DNA Pol I-dependent replication.</td>
<td>Stratagene</td>
</tr>
<tr>
<td>pCGN565</td>
<td>Chloramphenicol</td>
<td>General cloning plasmid, multiple cloning site from pUC18.</td>
<td>Stalker et al., 1988</td>
</tr>
<tr>
<td>pH45</td>
<td>Ampicillin, Spectinomycin</td>
<td>Source of the Omega (ΩSpec) fragment. Excised as an EcoRI, BamHI or Smal fragment.</td>
<td>Pretenki and Krisch, 1984</td>
</tr>
<tr>
<td>pH45Ω-Tc</td>
<td>Ampicillin, Tetracycline</td>
<td>Source of a modified Omega (ΩTel) fragment. Excised as an EcoRI, BamHI or Smal fragment.</td>
<td>Fellay et al., 1987</td>
</tr>
<tr>
<td>pHSG576</td>
<td>Chloramphenicol</td>
<td>Cloning plasmid, DNA Pol I-independent replication, pUC8-type multiple cloning sites.</td>
<td>Takeshita et al., 1987</td>
</tr>
<tr>
<td>pJLA602</td>
<td>Ampicillin</td>
<td>Over-expression plasmid, λP LR hybrid promoter, c1857 repressor, activated at 42°C.</td>
<td>Schauder et al., 1987</td>
</tr>
<tr>
<td>pKK233.2</td>
<td>Ampicillin</td>
<td>Over-expression plasmid, tcr promoter, induced with 1μM IPTG.</td>
<td>Amann and Brosius, 1985</td>
</tr>
<tr>
<td>pKK232-8</td>
<td>Ampicillin</td>
<td>Promoter-probe plasmid, gene fusion with chloramphenicol acetyltransferase.</td>
<td>Brosius, 1984</td>
</tr>
<tr>
<td>pMC9</td>
<td>Ampicillin</td>
<td>Source of the lacR fragment, excised as an EcoRI fragment.</td>
<td>Calos et al., 1983</td>
</tr>
<tr>
<td>pMLB1034</td>
<td>Ampicillin</td>
<td>Promoter-probe plasmid, protein fusion with β-galactosidase, EcoRI, Smal and BamHI cloning sites.</td>
<td>Shultz et al., 1982</td>
</tr>
<tr>
<td>pNM480</td>
<td>Ampicillin</td>
<td>Promoter-probe plasmid, protein fusion with β-galactosidase, pUC8-type multiple cloning sites, all three reading frames are supplied by these three vectors.</td>
<td>Minton, 1984</td>
</tr>
<tr>
<td>pNM481</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>pNM482</td>
<td>*</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td>pNZ338</td>
<td>Kanamycin</td>
<td>Promoter-probe plasmid, protein fusion with β-galactosidase. Derived from pMLB1034 and pLG338.</td>
<td>R. de Feyter</td>
</tr>
</tbody>
</table>
A2.2 Plasmids made for this thesis

The plasmids made for this thesis are listed numerically in the following pages. Schematic maps of most can be found in Figures A2:1 and A2:2 (a fold-out from page 131), and references to figures found in the main body of this Thesis are given where appropriate. Table A2:3 lists the replicon, incompatibility determinants and antibiotic resistance genes for each plasmid. Incompatible plasmids are essentially those sharing the same plasmid replicons or resistances. RepFIB incompatibility reactions may arise between a RepFIB mini-plasmid and a second plasmid if the second expresses RepA or carries copies of the A-K repeat elements.

---

### Table A2:2

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Resistance</th>
<th>Descriptiona</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>pNZ945</td>
<td>Ampicillin</td>
<td>RepFIB ligated to pBS+</td>
<td>This laboratory</td>
</tr>
<tr>
<td>pMA4322</td>
<td>Spectinomycin</td>
<td>A spontaneous copy mutant of pSS3928.</td>
<td>W. Maas</td>
</tr>
<tr>
<td>SnaBl-Bal △</td>
<td>Ampicillin</td>
<td>SnaBl-Bal deletion of pNZ945.</td>
<td>This laboratory</td>
</tr>
<tr>
<td>pSS3928</td>
<td>Spectinomycin</td>
<td>RepFIB mini-plasmid with the (\Omega^\text{Spec})-resistance fragment.</td>
<td>S. Saadi</td>
</tr>
<tr>
<td>pWM114</td>
<td>Ampicillin, Tetracycline</td>
<td>RepFIB ligated to pBR325.</td>
<td>W. Maas</td>
</tr>
<tr>
<td>Xho26</td>
<td>Ampicillin</td>
<td>A Tn10 deletion-derivative of pNZ945, lacking DNA from 0-1317bp.</td>
<td>This laboratory</td>
</tr>
<tr>
<td>Xho31</td>
<td>Ampicillin</td>
<td>A Tn10 deletion-derivative of pNZ945, lacking DNA from 0-1252bp.</td>
<td>This laboratory</td>
</tr>
<tr>
<td>Xho36</td>
<td>Ampicillin</td>
<td>A Tn10 deletion-derivative of pNZ945, lacking DNA from 0-1726bp.</td>
<td>This laboratory</td>
</tr>
<tr>
<td>XS61</td>
<td>Ampicillin</td>
<td>A Exo III-Mung Bean nuclease deletion-derivative of pNZ945, lacking DNA from 3393-4298bp.</td>
<td>This laboratory</td>
</tr>
<tr>
<td>#285</td>
<td>Chloramphenicol</td>
<td>PstI 1.2kb fragment ligated to pCGN565.</td>
<td>This laboratory</td>
</tr>
<tr>
<td>#288</td>
<td>Chloramphenicol</td>
<td>EcoRI-PstI 2.26kb ligated to pCGN565.</td>
<td>This laboratory</td>
</tr>
</tbody>
</table>

a Unless otherwise stated, all RepFIB DNA has been obtained from the P307 E11 4.3kb EcoRI fragment carrying RepFIB (rather than other examples of RepFIB).
<table>
<thead>
<tr>
<th>Plasmid (rep)</th>
<th>inc</th>
<th>Resistance to:</th>
<th>Plasmid (rep)</th>
<th>inc</th>
<th>Resistance to:</th>
</tr>
</thead>
<tbody>
<tr>
<td>pACYC184 (p15A)</td>
<td></td>
<td>A C S T K Tp</td>
<td>pBS+ (pMB1)</td>
<td></td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>pCGN565 (pMB1)</td>
<td></td>
<td>A C S T K Tp</td>
<td>pHPl45 (pMB1)</td>
<td></td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>pMC9 (pMB1)</td>
<td></td>
<td>A C S T K Tp</td>
<td>pMLB1034 (pMB1)</td>
<td></td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>pNM480 (pMB1)</td>
<td></td>
<td>A C S T K Tp</td>
<td>pNM481 (pMB1)</td>
<td></td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>pNM482 (pMB1)</td>
<td></td>
<td>A C S T K Tp</td>
<td>pMU575 (pMU575)</td>
<td></td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>pNZ945 (●, pMB1)</td>
<td>p,r</td>
<td>A C S T K Tp</td>
<td>pNZ955 (●, pMB1)</td>
<td>p,r</td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>pNZ956 (●, pMB1)</td>
<td>p,r</td>
<td>A C S T K Tp</td>
<td>SnaBi-BalIΔ(●, pMB1)</td>
<td>p,r</td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>#285 (pMB1)</td>
<td>r</td>
<td>A C S T K Tp</td>
<td>#288 (pMB1)</td>
<td>r</td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>pMA4322 (●)</td>
<td>p,r</td>
<td>A C S T K Tp</td>
<td>pSS3928 (●)</td>
<td>p,r</td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>pWM114 (●, pMB1)</td>
<td>p,r</td>
<td>A C S T K Tp</td>
<td>Xho26 (●, pMB1)</td>
<td>p,r</td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>Xho31 (●, pMB1)</td>
<td>p,r</td>
<td>A C S T K Tp</td>
<td>Xho36 (●, pMB1)</td>
<td>p,r</td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>X561 (●, pMB1)</td>
<td>p,r</td>
<td>A C S T K Tp</td>
<td>A C S T K Tp</td>
<td>A C S T K Tp</td>
<td></td>
</tr>
<tr>
<td>pAS1 (pMB1)</td>
<td></td>
<td>A C S T K Tp</td>
<td>pAS37 (pMB1)</td>
<td>r</td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>pAS7 (pMB1)</td>
<td></td>
<td>A C S T K Tp</td>
<td>pAS40 (pMB1)</td>
<td>r</td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>pAS9 (pSC101)</td>
<td>r</td>
<td>A C S T K Tp</td>
<td>pAS43 (●, pMB1)</td>
<td>p,r</td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>pAS10 (pMB1)</td>
<td>-</td>
<td>A C S T K Tp</td>
<td>pAS44 (pMB1)</td>
<td>-</td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>pAS15 (pMB1)</td>
<td>r</td>
<td>A C S T K Tp</td>
<td>pAS46 (pMB1)</td>
<td>r</td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>pAS16 (pMB1)</td>
<td>r</td>
<td>A C S T K Tp</td>
<td>pAS48 (pMB1)</td>
<td>-</td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>pAS17 (pMB1)</td>
<td>r</td>
<td>A C S T K Tp</td>
<td>pAS49 (pMB1)</td>
<td>p</td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>pAS18 (pMB1)</td>
<td>r</td>
<td>A C S T K Tp</td>
<td>pAS50 (pMB1)</td>
<td>r</td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>pAS19 (pMB1)</td>
<td>-</td>
<td>A C S T K Tp</td>
<td>pAS52 (pMB1)</td>
<td>-</td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>pAS20 (pMB1)</td>
<td>r</td>
<td>A C S T K Tp</td>
<td>pAS53 (pMB1)</td>
<td>-</td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>pAS21 (pMB1)</td>
<td>r</td>
<td>A C S T K Tp</td>
<td>pAS54 (pMB1)</td>
<td>-</td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>pAS22 (pMB1)</td>
<td>r</td>
<td>A C S T K Tp</td>
<td>pAS56 (pMB1)</td>
<td>-</td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>pAS24 (pMB1)</td>
<td>-</td>
<td>A C S T K Tp</td>
<td>pAS58 (pMB1)</td>
<td>r</td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>pAS25 (pMB1)</td>
<td>r</td>
<td>A C S T K Tp</td>
<td>pAS59 (pMB1)</td>
<td>r</td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>pAS28 (pMB1)</td>
<td>r</td>
<td>A C S T K Tp</td>
<td>pAS60 (pMB1)</td>
<td>p</td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>pAS29 (pMB1)</td>
<td>r</td>
<td>A C S T K Tp</td>
<td>pAS62 (pMB1)</td>
<td>p</td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>pAS31 (pMU575)</td>
<td>r</td>
<td>A C S T K Tp</td>
<td>pAS63 (pMB1)</td>
<td>r</td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>pAS32 (pMU575)</td>
<td>r</td>
<td>A C S T K Tp</td>
<td>pAS64 (p15A)</td>
<td>p</td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>pAS33 (pMU575)</td>
<td>r</td>
<td>A C S T K Tp</td>
<td>pAS65 (pMB1)</td>
<td>r</td>
<td>A C S T K Tp</td>
</tr>
<tr>
<td>pAS36 (pMB1)</td>
<td>r</td>
<td>A C S T K Tp</td>
<td>pAS69 (pMB1)</td>
<td>r</td>
<td>A C S T K Tp</td>
</tr>
</tbody>
</table>

rep The replicon (or replicon derivative such as a copy mutant or deletion) are shown in brackets; some also contain functional RepFlB replicons ('●': wild type RepFlB, '●': copy mutant RepFlB).

inc Some plasmids contain RepFlB incompatibility determinants ('p' : expresses RepA; 'r': contains some of the A-K repeat elements; '!': no significant incompatibility reactions observed; '!': unknown).

RepFlB promoter expresses the plasmid cml gene (CAT).

A  Ampicillin C  Chloramphenicol  S  Spectinomycin
T  Tetracycline K  Kanamycin  Tp  Trimethoprim
pAS1
A promoter-probe plasmid. The *Pst*I-*Bam*HI 0.70kb fragment from pNZ945 was ligated to *Sma*I-*Bam*HI cut pMLB1034 to give a *repA*:lacZ fusion. The RepFlB *Pst*I site was treated with Klenow, and the *Sma*I site in pMLB1034 lost. In this plasmid, the RepA:β-galactosidase fusion protein is not expressed although colonies containing pAS1 on plates with Xgal develop a faint blue colour. The plasmid confers resistance to 100μg/ml ampicillin. See Figures 3:1 and A2:2.

pAS2
A promoter-probe plasmid. The *Pst*I-*Bam*HI 0.77kb fragment from pNZ945 was ligated to *Sma*I-*Bam*HI cut pMLB1034 to give a *repA*:lacZ fusion. The RepFlB *Pst*I site was treated with Klenow, and the *Sma*I site in pMLB1034 lost. In this plasmid, the RepA:β-galactosidase fusion protein is not expressed although colonies containing pAS2 on plates with Xgal develop a faint blue colour. The plasmid confers resistance to 100μg/ml ampicillin. See Figures 3:1 and A2:2.

pAS4
The *Bam*HI-*Pst*I 0.20kb fragment from pNZ945 was ligated to *Bam*HI-*Pst*I cut pBS+. The plasmid confers resistance to 100μg/ml ampicillin. See Figures 4:6, 7:4 and A2:2.

pAS7
A promoter-probe plasmid. The *Bam*HI 0.97 kb fragment from pNZ945 was ligated to *Bam*HI cut pMLB1034 to give a *repA*:lacZ fusion. In this plasmid, the RepA:β-galactosidase fusion protein is expressed from *repA* and colonies containing pAS7 on plates with Xgal develop a strong blue colour. The plasmid confers resistance to 100μg/ml ampicillin. See Figures 3:1, 4:1, 5:1 and A2:2.

pAS9
A promoter-probe plasmid. The *Bam*HI 0.97kb fragment from pMA4322 was ligated to *Bam*HI cut pNZ338 to give a *repA*:lacZ fusion. The *Bam*HI fragment includes the single base change responsible for the mutant copy number of pMA4322. In this plasmid, the RepA:β-galactosidase fusion protein is expressed from *repA* and colonies containing pAS9 on plates with Xgal develop a strong blue colour. The plasmid confers resistance to 40μg/ml kanamycin. See Figures 4:1, 4:6 and A2:2.

pAS10
The *Bgl*II 0.67kb fragment from pNZ945 was ligated to *Bam*HI cut pBS+. As a result of this ligation, both the flanking *Bgl*II and *Bam*HI sites have been lost. The plasmid confers resistance to 100μg/ml ampicillin. See Figure A2:2.

pAS15
The *Hind*III-*Bam*HI 0.51kb fragment from pNZ945 was ligated to *Hind*III-*Bam*HI cut pBS+. The plasmid confers resistance to 100μg/ml ampicillin. See Figures 4:6 and A2:2.
pAS16
The HindIII-SnaBl 210bp fragment from pNZ945 was ligated to HindIII-Smal cut pBS+. As a result of this ligation, both the flanking SnaBl and Smal sites have been lost. The plasmid confers resistance to 100μg/ml ampicillin. See Figures 4:6 and A2:2.

pAS17
A SnaBl-Sal deletion of pAS15. As a result of this deletion, both the SnaBl and Sal sites have been lost. The plasmid confers resistance to 100μg/ml ampicillin. See Figures 4:6 and A2:2.

pAS18
The SnaBl-BamHl 270bp fragment from pNZ945 was ligated to Smal-BamHl cut pHSG576. The SnaBl and Smal sites have been lost as a result of this ligation. The plasmid confers resistance to 25μg/ml chloramphenicol. pAS18 is similar to pAS17 (Figure 4:6). See Figure A2:2.

pAS19
This plasmid is a PstI deletion derivative of Xho31. The plasmid confers resistance to 100μg/ml ampicillin. See Figures 7:4 and A2:2.

pAS20
This plasmid is a BamHl deletion derivative of Xho31. The plasmid confers resistance to 100μg/ml ampicillin. See Figures 7:4 and A2:2.

pAS21
A promoter-probe plasmid. The BamHl fragment from pAS19 was ligated to BamHl cut pKK232-8. Digestion of pAS21 with PstI will not release a 200bp fragment and the fragment is in the opposite orientation in pAS22. In this plasmid, CAT is expressed as a result of promoter activity (repAp) in the cloned insert. Colonies containing pAS21 are resistant to chloramphenicol. This plasmid confers resistance to 100μg/ml ampicillin. See Figures 4:1 and A2:2.

pAS22
A promoter-probe plasmid. The BamHl fragment from pAS19 was ligated to BamHl cut pKK232-8. Digestion of pAS22 with PstI will release a 200bp fragment and the fragment is in the opposite orientation in pAS21. In this plasmid, CAT is expressed as a result of promoter activity (oriP) in the cloned insert. Colonies containing pAS22 are resistant to chloramphenicol. This plasmid confers resistance to 100μg/ml ampicillin. See Figures 4:1 and A2:2.

pAS24
A modified promoter-probe plasmid. The EcoRI cut $\Omega^{Spec}$ resistance fragment was inserted at the EcoRI site of pAS1. In this plasmid, the RepA:β-galactosidase fusion protein is expressed and colonies
containing pAS24 on plates with Xgal develop a faint blue colour. The plasmid confers resistance to 100μg/ml ampicillin and 50μg/ml spectinomycin. See Figures 3:1, 3:2 and A2:2.

pAS25
A modified promoter-probe plasmid. The EcoRI cut $\Omega^{Spec}$ resistance fragment was inserted at the EcoRI site of pAS7. In this plasmid, the RepA:β-galactosidase fusion protein is expressed from repA_p and colonies containing pAS25 on plates with Xgal develop a strong blue colour. The plasmid confers resistance to 100μg/ml ampicillin and 50μg/ml spectinomycin. See Figures 3:1, 3:2 and A2:2.

pAS28
The $\Omega^{Spec}$ resistance fragment was inserted at the RepFIB BamHI site of pAS19. The plasmid confers resistance to 100μg/ml ampicillin and to 50μg/ml spectinomycin. See Figures 7:4 and A2:2.

pAS29
A modified promoter-probe plasmid. The EcoRI cut $\Omega^{Spec}$ resistance fragment was inserted at the EcoRI site of pAS2. In this plasmid, the RepA:β-galactosidase fusion protein is not expressed although colonies containing pAS29 on plates with Xgal develop a faint blue colour. The plasmid confers resistance to 100μg/ml ampicillin and to 50μg/ml spectinomycin. See Figures 3:1, 3:2 and A2:2.

pAS31
A promoter-probe plasmid. The HindIII-BamHI 0.51kb fragment from pNZ945 was ligated to HindIII-BamHI cut pMU575. In this plasmid, the hybrid β-galactosidase gene is expressed as a result of promoter activity (EF_p) in the cloned insert. Colonies containing pAS31 on plates with Xgal develop a strong blue colour. The plasmid confers resistance to 10μg/ml trimethoprim. See Figures 4:1 and A2:2.

pAS32
A promoter-probe plasmid. The BamHI 0.20kb fragment from pAS19 was ligated to BamHI cut pMU575. Digestion of pAS32 with Psfl will release a 0.20kb fragment. In this plasmid, the hybrid β-galactosidase gene is expressed as a result of promoter activity (ori_p) in the cloned insert. Colonies containing pAS32 on plates with Xgal develop a strong blue colour. The plasmid confers resistance to 10μg/ml trimethoprim. See Figures 4:1 and A2:2.

pAS33
A promoter-probe plasmid. The BamHI 0.20kb fragment from pAS19 was ligated to BamHI cut pMU575. The insert is in the opposite direction to that of pAS32 and digestion of pAS33 with Psfl does not release a 0.20kb fragment. In this plasmid, the hybrid β-galactosidase gene is expressed as a result of promoter activity (repA_p) in the cloned insert. Colonies containing pAS33 on plates with Xgal develop a strong blue colour. The plasmid confers resistance to 10μg/ml trimethoprim. See Figures 4:1 and A2:2.
pAS36
This plasmid is an inversion derivative of pAS19. The BamHI fragment has been religated to the plasmid in the reverse orientation, such that PstI digestion of pAS36 will release a small 200bp fragment. The plasmid confers resistance to 100µg/ml ampicillin. See Figures 7:4 and A2:2.

pAS37
A promoter-probe plasmid. A PCR fragment generated from pNZ945 using oligonucleotide primers #2 and #9 was ligated to Smal-BamHI cut pMLB1034 to give a repA::lacZ fusion. The polymerase chain reaction used pNZ945 DNA as a template and involved 25 cycles of [1 min. at 95°C, 2 min. at 50°C and 2 min. at 70°C]. After the reaction, the PCR fragment was treated with Klenow and cut with BamHI (producing a = 820bp fragment) before ligation into pMLB1034. As a result of this cloning the Smal site in pMLB1034 has been lost. In this plasmid, the RepA::β-galactosidase fusion protein is expressed from repAp and colonies containing pAS37 on plates with Xgal develop a strong blue colour. The plasmid confers resistance to 100µg/ml ampicillin. See Figures 3:1 and A2:2.

pAS38
A modified promoter-probe plasmid. The EcoRI cut ΩSpec resistance fragment was inserted at the EcoRI site of pAS37. In this plasmid, the RepA::β-galactosidase fusion protein is expressed from repAp and colonies containing pAS38 on plates with Xgal develop a strong blue colour. The plasmid confers resistance to 100µg/ml ampicillin and to 50µg/ml spectinomycin. See Figures 3:1, 3:2 and A2:2.

pAS39
A promoter-probe plasmid. A PCR fragment generated from pNZ945 using oligonucleotide primers #2 and #8 was ligated to Smal-BamHI cut pMLB1034 to give a repA::lacZ fusion. The polymerase chain reaction used pNZ945 DNA as a template and involved 25 cycles of [1 min. at 95°C, 2 min. at 50°C and 2 min. at 70°C]. After the reaction, the PCR fragment was treated with Klenow and cut with BamHI (producing a = 870bp fragment) before ligation into pMLB1034. As a result of this cloning the Smal site in pMLB1034 has been lost. In this plasmid, the RepA::β-galactosidase fusion protein is expressed from repAp and colonies containing pAS39 on plates with Xgal develop a strong blue colour. The plasmid confers resistance to 100µg/ml ampicillin. See Figures 3:1 and A2:2.

pAS40
A modified promoter-probe plasmid. The EcoRI cut ΩSpec resistance fragment was inserted at the EcoRI site of pAS39. In this plasmid, the RepA::β-galactosidase fusion protein is expressed from repAp and colonies containing pAS40 on plates with Xgal develop a strong blue colour. The plasmid confers resistance to 100µg/ml ampicillin and to 50µg/ml spectinomycin. See Figures 3:1, 3:2 and A2:2.

pAS43
The ΩSpec resistance fragment was inserted at the SnaBI site of pNZ945. The plasmid confers resistance to 100µg/ml ampicillin and to 50µg/ml spectinomycin. See Figure A2:2.
pAS44
A promoter-probe plasmid. The *Bgl*II 0.67kb fragment from pNZ945 was ligated to *BamHI* cut pMN480 to give an ORF-7:*lacZ* fusion. As a result of this ligation, both the *Bgl*II and *BamHI* sites have been lost. In this plasmid, the ORF-7:*β*-galactosidase fusion protein is expressed from ORF-7p and colonies containing pAS44 on plates with Xgal develop a strong blue colour. The plasmid confers resistance to 100μg/ml ampicillin. See Figure A2:2.

pAS46
A promoter-probe plasmid. The *BamHI* 0.97kb fragment from pNZ945 was ligated to *BamHI* cut pMN480 to give a repA:*lacZ* fusion. The *EcoRI* cut *Ω*Spec resistance fragment was then inserted into the *EcoRI* site of the intermediary plasmid to give pAS46 (which has since been lost). In this plasmid, the RepA:*β*-galactosidase fusion protein is expressed from repAp and colonies containing pAS46 on plates with Xgal develop a strong blue colour. The plasmid confers resistance to 100μg/ml ampicillin and 50μg/ml spectinomycin. This plasmid is very similar to pAS25 (See Figures 3:1 and 3:2). See Figure A2:2.

pAS48
The *lacI* gene *EcoRI* fragment (~ 1.7kb) from pMC9 was ligated to the *EcoRI* site of pKK233.2. The plasmid confers resistance to 100μg/ml ampicillin.

pAS49
A RepA expression vector. A PCR fragment generated from pNZ945 using oligonucleotide primers #2 and #17 was digested with *NcoI* and *EcoRI*, and the fragment ligated to *NcoI-EcoRI* cut pJLA602. The PCR cycle [1 min. at 95°C, 1.5 min. at 72°C, 2 min. at 72°C] was repeated 25 times. The strain carrying this plasmid is grown at 30°C, and expression is induced by incubation at 42°C. Maximal expression of RepA is seen after two hours at 42°C. The plasmid confers resistance to 50μg/ml ampicillin. See Figure A2:1.

pAS50
A promoter-probe plasmid. The *EcoRI-BamHI* fragment from pAS18 was ligated to *EcoRI-BamHI* cut pNM481 to give an ORF-5:*lacZ* fusion. In this plasmid, the ORF-5:*β*-galactosidase fusion protein is not expressed although colonies containing pAS50 on plates with Xgal develop a faint blue colour. The plasmid confers resistance to 100μg/ml ampicillin. See Figure A2:2.

pAS52
A promoter probe plasmid. The *BamHI* 0.28kb fragment from pNZ945 was ligated to *BamHI* cut pNM480 to give an ORF-3:*lacZ* fusion (note that the ORF-3 component is an internal portion of ORF-3). Digestion of pAS52 with *EcoRI* does not release the inserted fragment. In this plasmid, the ORF-3:*β*-galactosidase fusion protein is not expressed although colonies containing pAS52 on plates with...
Xgal develops a faint blue colour. The plasmid confers resistance to 100μg/ml ampicillin. See Figure A2:2.

pAS53
A promoter probe plasmid. The *BamHI* 0.28kb fragment from *pNZ945* was ligated to *BamHI* cut *pNM481* to give a *lacZ* fusion (the N-terminal portion of the fusion is provided by an unnamed ORF). Digestion of pAS53 with *EcoRI* releases the inserted fragment which is in the opposite orientation in pAS52. In this plasmid, the β-galactosidase fusion protein is not expressed although colonies containing pAS53 on plates with Xgal develop a faint blue colour. The plasmid confers resistance to 100μg/ml ampicillin. See Figure A2:2.

pAS54
A promoter probe plasmid. The *BamHI* 0.33kb fragment from *pNZ945* was ligated to *BamHI* cut *pNM480* to give an ORF-3:*lacZ* fusion (note that the ORF-3 component is an internal portion of ORF-3). In this plasmid, the ORF-3:β-galactosidase fusion protein is not expressed although colonies containing pAS54 on plates with Xgal develop a faint blue colour. The plasmid confers resistance to 100μg/ml ampicillin. See Figure A2:2.

pAS56
A modified promoter probe plasmid. The *PstI*-BamHI 770kb fragment of pAS63 was ligated to Smal-BamHI cut *pMLB1034* to give a *repA*:lacZ fusion. The RepFIB *PstI* site was treated with Klenow, and as a result of this ligation, the *PstI* and Smal sites have been lost. The cloned fragment does not include *repA* and has a PCR-modified *repA* translation start codon (an ATG codon). In this plasmid, the RepA:β-galactosidase fusion protein is weakly expressed and colonies containing pAS56 on plates with Xgal develop a light blue colour. The plasmid confers resistance to 100μg/ml ampicillin. See Figure A2:2.

**Figure A2:1**

**RepA expression plasmids**

Three different RepA expression plasmids have been produced in this Thesis. In pAS49, *λpLR* is repressed at 30°C by the c1857 gene product and RepA expression is induced by growth at 42°C. In pAS60, LacI represses *pLcr* and expression is induced by IPTG. pAS60 was modified by the insertion of Ω to give pAS62. The expression cassette containing Ω^Spec^, lacI^Q^, pLcr, and repA was then transferred to pACYC184 to give pAS64.
This page contains a detailed representation of the DNA fragments used in two different studies. The DNA fragments are labeled with their respective positions and are illustrated in a schematic format.

Figure A.2.1: The primer is hybridized with the target sequence. The primer hybridization is followed by polymerase chain reaction (PCR). The PCR reaction is performed on the DNA fragment, and the resulting PCR products are analyzed. The DNA fragment is labeled with a specific primer sequence.

Figure A.2.2: The DNA fragment is hybridized with a specific probe. The probe hybridization is followed by hybridization with a specific target sequence. The hybridization is then detected using a specific detection method.

Opposite page (top) A schematic representation of the P3013 fragment. (bottom) A schematic representation of the P3014 fragment. (center) A schematic representation of the P3015 fragment. (right) A schematic representation of the P3016 fragment.
fragments were then combined to form the template for a second round of PCR synthesis (Figure 5:3).

The first round of PCR synthesis involved two reactions using primers #21 and #26; and #2 and #25. The PCR cycle of [1 min. at 94°C, 1.5 min. at 55°C, 1.5 min. at 72°C] was repeated 25 times. The PCR fragment resulting from each of the two reactions were isolated after LMP-TA gel electrophoresis using Geneclean. The two DNA fragments were combined to form the template DNA for the second round of PCR synthesis. This reaction used primers #2 and #21 and involved ten cycles of [1 min. at 94°C, 1.5 min. at 35°C, 1.5 min. at 72°C] followed by ten cycles of [1 min. at 94°C, 1.5 min. at 55°C, 1.5 min. at 72°C]. The resultant DNA (a single fragment) was digested with BamHI and ligated to BamHI cut pCGN565. The plasmid confers resistance to 25\(\mu\)g/ml chloramphenicol. See Figure A2:2.

pAS64
A RepA expression vector. The SacI-Scal fragment from pAS62 was ligated to the BamHI site of pACYC184 (disrupting the tetracycline resistance gene). The SacI-Scal fragment contains the \(\Omega^{\text{Spec}}\) resistance fragment, the lacI\(^q\) gene fragment (from pMC9), the \(\text{p}_{\text{cr}}\) promoter, the repA gene and RNA terminator genes. Both the fragment and the plasmid DNA were treated with Klenow before ligation, and as a result, the SacI, Scal and BamHI sites have been lost. The \(\text{p}_{\text{cr}}\) promoter is induced with IPTG at 37°C. Although the plasmid confers resistance to 50\(\mu\)g/ml spectinomycin and 25\(\mu\)g/ml chloramphenicol, the plasmid should only be maintained with chloramphenicol (spectinomycin may lead to plasmid amplification). See Figure A2:1.

pAS65
Smal-cuts \(\Omega^{\text{Tdl}}\) was inserted at the SnaBI site of pAS15 (N. Bhana). As a result of this ligation, both the Smal and SnaBI sites have been lost. The plasmid confers resistance to 50\(\mu\)g/ml tetracycline and 100\(\mu\)g/ml ampicillin. See Figures 4:6 and A2:2.

pAS69
A promoter-probe plasmid. The P
\(\text{III}\)-EcoRI fragment from #285 was ligated to EcoRI-Smal cut pAS7 to give an ORF-4:repA:laCZ fusion. The RepFIB P
\(\text{III}\) site was treated with Klenow, and as a result of this ligation the P
\(\text{III}\) and Smal sites have been lost. In this plasmid, the fusion protein is not expressed although colonies containing pAS69 on plates with Xgal develop a faint blue colour. The plasmid confers resistance to 100\(\mu\)g/ml ampicillin. See Figure A2:2.
Sequence analysis of the E11 fragment from P307

A3.1 General features of the E11 EcoRI fragment

The EcoRI fragment (E11) from P307 carrying the RepFlB replicons was the first example of RepFlB to be sequenced by this group (Saul et al., 1989). The sequence information was analysed using a number of UWCG computer programmes (Devereux et al., 1984) containing algorithms useful for determining regions likely to express proteins, locating regions with DNA repeat elements and locating sequences sharing homology with consensus sequences of DNA-binding factors. A schematic diagram of the sequence features of the E11 fragment is provided in Figure A3:1. A more detailed diagram of the origin of replication can be found in Chapter Seven, Figure 7:2. The sequence analysis presented here is a composite of work carried out for this Thesis and of other analyses (D. Saul, personal communication; Spiers, MSc. Thesis, 1988; Saul et al., 1989; and Gibbs, MSc. Thesis, 1991).

A3.2 Open reading frames

By definition, an open reading frame is a sequence bounded by two in-frame stop codons containing at least one codon which can be used for the initiation of translation. It is a simple matter to locate such regions in a DNA sequence, but a statistical analysis of the intervening sequence between the start and 3' stop codon is required to determine whether the open reading frame is likely to be expressed. DNA sequences which contain information necessary for the expression of a protein (coding regions) differ from non-coding regions in a number of ways. In degrees of increasing complexity, the composition of nucleotides differs between coding and non-coding regions; the nucleotides are not found in the same frequency in each of the three positions in a codon within a coding region; nucleotides are not randomly assorted within codons in coding regions; and the frequencies of particular codons vary between coding and non-coding regions. Each of these features can be used to indicate the likelihood of a particular region expressing a protein. These features have been exploited in a number of UWGCG programmes used to analyse the E11 fragment sequence and are summarised in Table A3:1. Only those open reading frames which might express proteins of 50 or more residues (> 5.5kDa) have been analysed.

Fickett's Testcode determines the frequency of each of the four nucleotides found at each position in the three reading frames of an ORF (Fickett, 1982). The results are expressed as a plot divided into
three areas: coding region, 'no opinion' and non-coding regions. Base Positional Preference determines the preference of bases in each of the three positions of a codon (Staden, 1984). If the plot is above the line, the probability of the region being a coding region is 76%, whilst the probability that it is in fact a non-coding region is only 24%. The frequency of bases within a codon is also tested by Shepherd's RNY Base Positional Preference (Shepherd, 1981), where the codons of coding regions are preferentially RNY; purine, any nucleotide, pyrimidine.

Finally, codon usage is used to locate coding regions; Codon Preference (Staden and McLachlan, 1982; and Gribskov et al., 1984) compare suspected coding regions with a codon frequency table generated from a number of E. coli genes which are expressed at high levels, and Codon Improbability (McLachlan et al., 1984) looks for regions showing a strong bias for particular combinations of codons. In the latter case coding regions will show peaks in all three reading frames of a coding region.

The probability of a coding region expressing a protein is enhanced by the presence of a Shine and Dalgarno (SD) sequence (5'-TAAGGAGG-3') or a SD-like sequence (5'-AGGA-3') between 4-15bp upstream of the start codon (Shine and Dalgarno, 1974; and Gold et al., 1981), and by the presence of a promoter sequence upstream of the gene.

**FIGURE A3:1** Features of the P307 E11 EcoRI fragment

*Opposite page (fold-out)* Sequence analysis of the E11 4.3kb EcoRI fragment from P307 which carries the RepFIB replicon has identified a number of sequence features.

A The RepFIB replicon is in the central region of the E11 fragment. The region to the right of the replicon is highly conserved amongst EcoRI fragments from other plasmids which contain RepFIB, whilst the region to the left of the replicon appears to be unique to P307. The origin of replication, minimal region required for replication and the incE regions are shown.

B Sequence analysis of the E11 fragment has identified a number of sequence features. Open reading frames are represented by half-arrows: repA (ORF-1), ORF-5 and ORF-7 have high probabilities of expression whilst ORF-2, ORF-3, ORF-4 and ORF-6 are unlikely to be expressed. The fragment includes two sets of DNA repeat elements (BCDD'D", EFGHIJ and the unrelated 123 repeats). ORF-7p, or1p, repAp and EFp promoters and potential DNA-binding domains in RepA are marked. The minimal region required for replication includes repA, the BCDD'D" repeats, origin factor binding sites (small box) and the EF repeat elements. Note that the G'G'HlJ repeats are not required for replication. The incE region includes the EFGG"HIJ repeat elements.

C A simplified restriction map of the E11 fragment is given. Not all PfMl sites are not shown and the coordinates for PfMl refer to the cut site of the enzyme, not the recognition site.

D Single strand annotated sequence of the E11 fragment. Restriction sites, open reading frame start and stop codons, active promoter sequences, A-K and 123 repeat elements are high-lighted. Only one PfMl cut site is shown.
Origin

Minimal region required for replication

B

ORF-7  ORF-6  ORF-2  repA  ORF-4  ORF-5

Co-ordinate base pair positions (cbp)

Annotated sequence of the E11 EcoRI fragment
A3.2.1  repA (ORF-1)

The repA gene is the largest open reading frame in the E11 fragment. The ORF stretches from a TGA stop codon at 2228cbp to a TGA stop codon at 3296cbp and contains a number of potential start codons for the initiation of translation. These start sites are listed in Table A3:2, along with the estimated sizes of the proteins each would express. Maxicell analysis has indicated that the repA gene product is between 35-40kDa, suggesting that of the eight start codons listed, initiation is most likely to occur from one of the four codons located between 2321-2393cbp. None of the codons are preceded by a suitably placed Shine and Dalgarno (SD) or SD-like sequence. However, codon a is located within a SD-like sequence which suggests that the initiation of translation might occur at that codon, or perhaps from the b codon 24bp downstream.

**TABLE A3:1** Statistical analysis of the open reading frames in the E11 fragment

<table>
<thead>
<tr>
<th>ORF</th>
<th>Stop&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Start&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Stop&lt;sup&gt;c&lt;/sup&gt;</th>
<th>a.d&lt;sup&gt;d&lt;/sup&gt;</th>
<th>kDa&lt;sup&gt;e&lt;/sup&gt;</th>
<th>Stat&lt;sup&gt;f&lt;/sup&gt;</th>
<th>SD&lt;sup&gt;g&lt;/sup&gt;</th>
<th>Score&lt;sup&gt;h&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>1&lt;sup&gt;i&lt;/sup&gt;</td>
<td>TGA (2228)</td>
<td>GTG (2321)</td>
<td>TGA (3296)</td>
<td>325</td>
<td>37.25</td>
<td>High</td>
<td>No</td>
<td>Good</td>
</tr>
<tr>
<td>2</td>
<td>TGA (1445)</td>
<td>GTG (1508)</td>
<td>TAA (2009)</td>
<td>167</td>
<td>19.44</td>
<td>Poor</td>
<td>Yes</td>
<td>Poor</td>
</tr>
<tr>
<td></td>
<td>ATG (1532)</td>
<td></td>
<td></td>
<td>159</td>
<td>18.54</td>
<td></td>
<td>Yes</td>
<td>Poor</td>
</tr>
<tr>
<td></td>
<td>GTG (1535)</td>
<td></td>
<td></td>
<td>158</td>
<td>18.41</td>
<td></td>
<td>Possible</td>
<td>Very Poor</td>
</tr>
<tr>
<td>3</td>
<td>TAA (4065)</td>
<td>GTG (3981)</td>
<td>TGA (3585)</td>
<td>118</td>
<td>13.81</td>
<td>Good</td>
<td>No</td>
<td>Poor</td>
</tr>
<tr>
<td></td>
<td>ATG (3939)</td>
<td></td>
<td></td>
<td>118</td>
<td>13.81</td>
<td></td>
<td>No</td>
<td>Poor</td>
</tr>
<tr>
<td></td>
<td>ATG (3927)</td>
<td></td>
<td></td>
<td>114</td>
<td>13.27</td>
<td></td>
<td>Yes</td>
<td>Good</td>
</tr>
<tr>
<td>4</td>
<td>TGA (2586)</td>
<td>ATG (2583)</td>
<td>TAA (2256)</td>
<td>109</td>
<td>11.80</td>
<td>Good</td>
<td>Possible</td>
<td>Very Poor</td>
</tr>
<tr>
<td></td>
<td>GTG (2532)</td>
<td></td>
<td></td>
<td>93</td>
<td>9.94</td>
<td></td>
<td>No</td>
<td>Very poor</td>
</tr>
<tr>
<td></td>
<td>GTG (2472)</td>
<td></td>
<td></td>
<td>73</td>
<td>7.78</td>
<td></td>
<td>No</td>
<td>Very poor</td>
</tr>
<tr>
<td>5</td>
<td>TGA (3493)</td>
<td>ATG (3577)</td>
<td>TAA (3892)</td>
<td>105</td>
<td>11.63</td>
<td>High</td>
<td>Yes</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td>GTG (3715)</td>
<td></td>
<td></td>
<td>59</td>
<td>6.36</td>
<td></td>
<td>No</td>
<td>Poor</td>
</tr>
<tr>
<td></td>
<td>GTG (3754)</td>
<td></td>
<td></td>
<td>46</td>
<td>5.01</td>
<td></td>
<td>No</td>
<td>Poor</td>
</tr>
<tr>
<td>6</td>
<td>TAA (1098)</td>
<td>ATG (1107)</td>
<td>TAA (1392)</td>
<td>95</td>
<td>11.21</td>
<td>Good</td>
<td>Possible</td>
<td>Poor</td>
</tr>
<tr>
<td></td>
<td>GTG (1302)</td>
<td></td>
<td></td>
<td>30</td>
<td>3.43</td>
<td></td>
<td>No</td>
<td>Very poor</td>
</tr>
<tr>
<td></td>
<td>GTG (1338)</td>
<td></td>
<td></td>
<td>18</td>
<td>2.17</td>
<td></td>
<td>No</td>
<td>Very poor</td>
</tr>
<tr>
<td>7</td>
<td>TGA (826)</td>
<td>ATG (859)</td>
<td>TGA (1107)</td>
<td>83</td>
<td>9.18</td>
<td>High</td>
<td>Yes</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td>ATG (919)</td>
<td></td>
<td></td>
<td>63</td>
<td>7.05</td>
<td></td>
<td>No</td>
<td>Good</td>
</tr>
<tr>
<td></td>
<td>ATG (1015)</td>
<td></td>
<td></td>
<td>31</td>
<td>3.60</td>
<td></td>
<td>No</td>
<td>Good</td>
</tr>
</tbody>
</table>

| a | Stop codon in-frame with the ORF-1, upstream of all potential start codons (coordinate position given in brackets). |
| b | The first three ATG or GTG codons are listed (coordinate positions given in brackets). |
| c | The stop codon at the end of the ORF which terminates translation (coordinate position given in brackets). |
| d | Number of amino acid residues in the protein. |
| e | Estimated molecular mass. |
| f | Qualitative score based on the statistical tests of the ORF (Codon Preference, Fickett's Testcode, Codon Improbability, Base Positional Preference and Shepherd's RNY Base Positional Preference). |
| g | Presence of Shine and dalgarno (SD) and SD-like sequences. |
| h | Qualitative score for the expression of the ORF based on the statistical test scores and on the presence of a potential SD or SD-like sequence. The presence of a promoter is not considered. |
| i | See Table A3:2 for other ORF-1 sizes. |
Statistical analysis of the coding region of repA indicates that the gene has a good probability of expression (Table A3:1). Examination of the secondary structure of the gene product (RepA) using Peplplot (Gribskov et al., 1986) suggests that the protein consists of a number of alpha helices and few beta-pleated sheets. Examination of RepA using Peptidestructure indicates that the protein can be cleaved by cyanogen bromide to produce five polypeptides of 2.7, 4.5, 5.8, 6.5 and 19.6kDa (cleavage is after methionine residues at positions 39, 91, 260 and 315), and by exposure to pH2.5 to produce three polypeptides of 5, 12.6 and 21.2kDa (after aspartic acid residues at positions 188 and 230). A map showing the cleavage products is given in Appendix Four, Figure A4:2).

<table>
<thead>
<tr>
<th>TABLE A3:2</th>
<th>Potential ORF-1 polypeptides</th>
</tr>
</thead>
<tbody>
<tr>
<td>Start codon</td>
<td>dbp^a</td>
</tr>
<tr>
<td>CTG</td>
<td>2279</td>
</tr>
<tr>
<td>GTG</td>
<td>2321</td>
</tr>
<tr>
<td>GTG</td>
<td>2345</td>
</tr>
<tr>
<td>GTG</td>
<td>2384</td>
</tr>
<tr>
<td>ATG</td>
<td>2393</td>
</tr>
</tbody>
</table>

a Coordinate position of the start codon.  
b Number of amino acid residues in the polypeptide.  
c Estimated molecular mass.  
* Initiation codon determined in this Thesis.

The amino acid sequence of RepA was compared against a prokaryotic DNA-binding protein consensus sequence (Nakata and Maizel, 1989) to determine whether RepA contained any domains which may enable the protein to bind to DNA (Ohendorf et al., 1983; and Pabo and Sauer, 1984). This comparison identified three regions of the protein which showed a degree of homology with Nakata and Maizel's consensus sequence (Table A3:3).

<table>
<thead>
<tr>
<th>TABLE A3:3</th>
<th>Amino acid sequence variations in RepA</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MENENSDIKK</td>
</tr>
<tr>
<td>51</td>
<td>LKNSKKNTLS</td>
</tr>
<tr>
<td>101</td>
<td>IIHSFARHNV</td>
</tr>
<tr>
<td>151</td>
<td>TVISFSRTDE</td>
</tr>
<tr>
<td>201</td>
<td>VLLQLKAINA</td>
</tr>
<tr>
<td>251</td>
<td>SQNQTVRAM</td>
</tr>
<tr>
<td>301</td>
<td>NPAPPSPAEEK</td>
</tr>
</tbody>
</table>

1 A in P307, D in pMA4322 (Ala → Asp).  
2 R in P307, S in R386 (Arg → Ser).  
3 F in P307, L in pH507 (Phe → Leu).  
4 L in P307, Q in F, R124, CoV3-K30 and CoV (Phe → Gin).

Residues in BLUE show pH2.5 cleavage positions; GREEN show cyanogen bromide cleavage positions. The start codon shown here was determined in this Thesis.

Sequence analysis of the E11 fragment from P307.
It is not obvious which of the three sites is more likely to act as a DNA-binding domain, although circumstantial evidence suggests that domain I is the more likely candidate. This evidence comes from the sequence obtained from a RepFIB copy mutant (pMA4322). This mutant differs from the wild type RepA sequence at only one residue position in the repA gene; the Ala→Asp (residue 74) change occurs at a conserved position within domain I (residues 70-89). Whilst both domains II and III differ at one of the three highly conserved residue sites within the consensus, both have a greater overall homology with the consensus sequence than domain I.

Although there is no biochemical evidence available, Perez-Casal and Crosa (1989) have also suggested that domain I represents a DNA-binding domain using a less well developed consensus sequence based on the alpha helix-turn-alpha helix motif found in the lambda Cro and C1 proteins. Domain I does not map to an obvious secondary structure feature on the Chou-Fasman (CF) and Garnier-Osguthorpe-Robson (GOR) predictions of RepA (Chou and Fasman, 1978; Garnier et al., 1978; and Gribskov et al., 1986), nor do the mutant RepA and wild type RepA predictions differ significantly.

The repA coding regions of a number of RepFIB replicons have been sequenced by Gibbs (MSc. Thesis, 1991). The gene is very highly conserved amongst the ten examples sequenced, with the majority of DNA sequence differences occurring at the 3rd 'wobble' position of the codons. Most of the base changes which have resulted in changes to the protein sequence have involved changes to similar amino acids.

A3.2.2 Other open reading frames

Maxicell analysis of the translational products of pMA4322 (which contains the E11 fragment) identified two small proteins of 10.8kDa and 9.8kDa (Saul et al., 1989). The relative molecular weights of these two proteins agree well with the predicted masses of proteins expressed from ORF-5 and ORF-7 (Table A3:1). These two ORFs have the highest probability of expression after ORF-1 (repA).

The statistical examination of the ORF-2 coding region suggests strongly that the ORF would not be expressed despite the presence of potential Shine and Dalgarno (SD) sequences upstream of three of the start codons. No suitable promoter sequences are located upstream of ORF-2. ORF-6 is also unlikely to be expressed as the coding region lacks a SD sequence, upstream promoter and has a low probability of expression. ORF-4 has a good probability of expression and includes a potential SD sequence. However, the ORF does not have a promoter sequence located within 100-200bp of the start codon.

ORF-3 has a higher probability of expression than either ORF-2 or ORF-6; statistically the region has a higher probability of expression, and one of the start codons has a suitable upstream SD sequence. No information on promoters is available, as the region 5' to the start of ORF-3 was not examined.
for potential promoter sequences. There is a possibility that the sequence of the E11 RepFIB fragment reported in Saul et al. (1989) contains a single base deletion in the region between 4020-4040bp. If this is the case, then ORF-3 would begin in the adjacent EcoRI fragment. This would result in a potential ORF encoding a protein of ≥ 29.6kDa showing considerable sequence homology with the mini-F resolvase protein (Section A3.7.2).

A3.3 DNA repeat elements

A3.3.1 The A-K repeat elements

The E11 fragment was searched for repeat elements using the programmes Comparison and Repeat. Repeat identified nine elements which shared considerable sequence homology with one another over a region of = 20bp. A consensus sequence derived from these nine elements (BCDEFGHIJ) was used to identify two further elements (A and K) sharing a lesser degree of homology (Spiers, MSc. Thesis, 1988; and Saul et al., 1989). The eleven repeat elements are clustered together to form a direct repeat group upstream of repA (BCD) and a complex group downstream of repA (EFGHIJ) which includes an inverted repeat (EF) and a direct repeat group (HIJ) (the grouping of the repeats is shown in Figure A3:1; the repeat sequences are listed in Table A3:4).

<table>
<thead>
<tr>
<th>Table A3:4</th>
<th>The A-K repeat elements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Element</td>
<td>Sequence</td>
</tr>
<tr>
<td>A</td>
<td>ATATGTGTGTTTCAATGCCAG</td>
</tr>
<tr>
<td>B</td>
<td>ACATAAGCTGTTCCATTTAC</td>
</tr>
<tr>
<td>C</td>
<td>GGATATAACTGACATATGTAAT</td>
</tr>
<tr>
<td>D</td>
<td>TGATATAGCTGTTATCCATATT</td>
</tr>
<tr>
<td>D'</td>
<td>CCATATATGACTAGTATAG</td>
</tr>
<tr>
<td>D'</td>
<td>ACATAAATAGCTATAGTATAG</td>
</tr>
<tr>
<td>D''</td>
<td>GACTCAGTTGTTAATACAG</td>
</tr>
<tr>
<td>E</td>
<td>ACATAAAACTGCAGTCGAGGCAGA</td>
</tr>
<tr>
<td>F</td>
<td>ATATAGCTGTTGATCGAGGCA</td>
</tr>
<tr>
<td>G</td>
<td>ATATAGCTGTTGATCGAGGCA</td>
</tr>
<tr>
<td>G'</td>
<td>GACTCAGTTGTTAATACAG</td>
</tr>
<tr>
<td>G*</td>
<td>GACGTAAGGGGACAGCAGGCA</td>
</tr>
<tr>
<td>H</td>
<td>ATATAGCATTGAGCGATGTTAAC</td>
</tr>
<tr>
<td>I</td>
<td>ATATAGCATTGAGCGATGTTAAC</td>
</tr>
<tr>
<td>J</td>
<td>ATATAGCATTGAGCGATGTTAAC</td>
</tr>
<tr>
<td>K</td>
<td>ATATAGCATTGAGCGATGTTAAC</td>
</tr>
<tr>
<td>Consensus:</td>
<td>ATATAGCATTGAGCGATGTTAAC</td>
</tr>
</tbody>
</table>

a, b Coordinate positions of the first and last base of the element.
c Number of identical bases to the consensus sequence.
The relative orientation of the repeat elements are indicated with arrows.
The consensus sequence was constructed using all repeats except D'', G' and G*.
The original analysis of the repeat elements considered that only the initial nine elements were significant (BCD, EFG, HIJ) in terms of sequence similarity. Repeat elements A and K shared lesser homology with the other elements, and in particular, included different nucleotides at otherwise invariant positions in the consensus sequence. Subsequent analysis suggests that three additional repeat elements are located adjacent to repeat D (repeats D', D'' and D'''). Both D' and D'' share a greater degree of similarity with the consensus sequence than repeat element A. D', D'' and D''' are spaced in a similar manner as are repeats BCD and form a single direct repeat structure (BCDDD'D''). G' and G'' are located between the G and H repeat elements. D'', G' and G'' have homology scores equal or less than repeat A to the A-K consensus sequence. For this reason, these three repeat elements have been omitted from RepFIB diagrams in this Thesis.

An analysis of the individual repeat elements has not revealed significant sequence relationships between individual repeat elements and the groupings of the repeats into the BCDD'D''EFGG'G'' and HIJ repeat groups (Figures A3:2 and A3:3).

![Cladogram of the ABCDD'D''EFGHIJK repeats](image)

**Figure A3:2** Cladogram of the ABCDD'D''EFGHIJK repeats

The UGWCG programme *Pileup* has been used to determine the relationships between the ABCDD'D''EFGHIJK repeat elements. Cladistic distances between repeats are shown along the x-axis. The A-K repeat consensus groups with the F repeat element. A schematic representation of the relative positioning and orientation of each repeat is shown above the cladogram.
The inclusion of the A-K repeat element sequence data from the plasmids sequenced by Gibbs (MSc. Thesis, 1991) does not significantly alter the consensus sequence for the repeat elements derived from the P307 E11 fragment alone. The A-K consensus sequence does not include the hexanucleotide sequence 5'-TGAGG/A-3' found in the major origin-repeats of several replicons (Filutowicz et al., 1985).

The EFGHIJ repeat elements appear to be associated with the incE incompatibility characteristic of the RepFIB replicon. Gardner et al. (1985) demonstrated that the cis acting elements of incE expression were located within a ≈ 450bp region of the F plasmid RepFIB replicon which includes the EFGHIJ repeats. However, Perez-Casal and Crosa (1989) have suggested that incE incompatibility is restricted to the EFG repeats (in the ColV3-K30 RepFIB replicon), and that the HIJ repeat elements form a second and distinct inc group (incF).

**FIGURE A3:3** Cladogram of the ABCDD'D''EFGG'G''HIJK repeats

The UWGCG programme *Pileup* has been used to determine the relationships between the ABCDD'D''EFGG'G''HIJK repeat elements. Cladistic distances between repeats are shown along the x-axis. Compare the relationships shown here with those in Figure A3:2 where the D'', G' and G'' repeats have not been included in the analysis. A schematic representation of the relative positioning and orientation of each repeat is shown above the cladogram. The D'', G' and G'' repeats are shown in grey.

Sequence analysis of the E11 fragment from P307
A3.3.2 Other DNA repeat elements

The E11 fragment contains two additional sets of repeat elements which do not share homology with the A-K repeat elements. The 123 repeats are located upstream of ORF-7, and each of the three repeats may in fact consist of two direct sub-repeat elements of 6-7bp (consensus sequence of 5'GGGCGGCGTGGCGG-3'; 1: 510-523, 2: 548-561 and 3: 591-604cbp). The only other significant set of repeats in the E11 fragment is associated with Dam methylation sites in the origin region of RepFIB (Section A3.6.3, Table A3:7).

A3.4 Sequences with homology to the \textit{E. coli} consensus promoter sequence

The E11 fragment sequence was searched for regions showing homology with the \textit{E. coli} consensus promoter sequence using \textit{TargSearch} (Hawley and McClure, 1983; and Mulligan \textit{et al}., 1984). Only portions of the E11 fragment have been searched and analysed by \textit{TargSearch}. The four search regions correspond to the region upstream of ORF-7 (section I), the region upstream of ORF-2 (II), the region encompassing the BCD repeats and upstream portion of repA (III) and the upstream region of ORF-5 including the EFGHIJ repeats (IV). \textit{TargSearch} assigned each potential promoter a percentage score indicating the likelihood of the identified sequence acting as a promoter element (Mulligan \textit{et al}., 1984). A value of 30\% or more was considered significant, and all potential promoter sequences identified by the programme with scores \(\geq 30\%\) are listed in Table A3:5.

The analysis of potential promoter sequences has been extended from that of Spiers (MSc. Thesis, 1989) particularly in the 2000-2200cbp region (contained within the \textit{BamHI-PstI} 200bp fragment) thought to contain the \textit{repA} promoter. The \textit{BamHI-PstI} fragment contains five promoter elements which are embedded within the BCDD'D" repeat elements (see Chapter Three, Figure 3:1).

A3.5 The RepFIB minimal replicon

A number of pNZ945 deletion-derivatives originally constructed to allow the sequencing of the E11 fragment carrying RepFIB were used to determine the left and right hand boundaries of the minimal region required for replication (the minimal RepFIB replicon). The left hand boundary was defined at 1727cbp using Xho36, whilst the right hand boundary was defined using XS61 at 3392cbp (Spiers, MSc. Thesis, 1989; and Saul \textit{et al}., 1989).

Unfortunately, XS61 has since been lost. However, a SnaB1-BalI deletion of pNZ945 (SnaB1-BalI \(\Delta\)) can be used in its place and has a deletion boundary at 3442cbp. The minimal RepFIB replicon is...
presumed to have an origin of replication and includes the BCDD'D' repeats, all of repA and the EFG repeat elements.

### Table A3:5

<table>
<thead>
<tr>
<th>code</th>
<th>ebp</th>
<th>-35 Sequence</th>
<th>-10 Sequence</th>
<th>ebp</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>I 1</td>
<td>788</td>
<td>ATAAAAGACTTGTTAA</td>
<td>AGGAGTATCTTTGA</td>
<td>828</td>
<td>62.0</td>
</tr>
<tr>
<td></td>
<td>836</td>
<td>TAAACAGTCTTGGAAG</td>
<td>CTTACATTTATCTT</td>
<td>875</td>
<td>55.0</td>
</tr>
<tr>
<td>II 3</td>
<td>1471</td>
<td>AATTCCGCCATGGAAA</td>
<td>AGCCTTAAAAGTGA</td>
<td>1511</td>
<td>39.0</td>
</tr>
<tr>
<td>III 4a</td>
<td>2026</td>
<td>CTGCAAAACCTTGTCAC</td>
<td>ACCACGATAACATA</td>
<td>2067</td>
<td>38.5</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>AATAAGGTCGTGGCAT</td>
<td>ATGCTAAATCATCC</td>
<td>2069</td>
<td>38.5</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>AGTTTATCCGTAACAT</td>
<td>AGCCTTAAAAGTGA</td>
<td>2055</td>
<td>38.5</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>AACATGCTAATGAACT</td>
<td>ATCCCATATCTGGA</td>
<td>2136</td>
<td>48.5</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>AATACAATATGCTGA</td>
<td>ATCTTTTACCTGTC</td>
<td>2191</td>
<td>46.7</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>ACACGAGTTCTCTGAC</td>
<td>ATCTTTTACCTGTC</td>
<td>2261</td>
<td>39.1</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>TAGGGGAGGCTGGATA</td>
<td>CTGGTGACCTGAC</td>
<td>2355</td>
<td>35.5</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>CTGCTGACACCTGAC</td>
<td>CACCGTACACCTG</td>
<td>2384</td>
<td>30.8</td>
</tr>
<tr>
<td></td>
<td>12a</td>
<td>TCGATAAGGTTGAGAT</td>
<td>GCTGAGATATGGAT</td>
<td>2554</td>
<td>38.5</td>
</tr>
<tr>
<td></td>
<td>12b</td>
<td>CTGGATATGGATAA</td>
<td>CTGGATATGGATAA</td>
<td>2556</td>
<td>37.9</td>
</tr>
<tr>
<td></td>
<td>13a</td>
<td>TTAATCCTTTCTGCC</td>
<td>TTGGTGACAAATCT</td>
<td>2623</td>
<td>36.1</td>
</tr>
<tr>
<td></td>
<td>13b</td>
<td>AGTCTTAAGATGGA</td>
<td>ATGGTGACAAATCT</td>
<td>2623</td>
<td>35.5</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>ACCTCGAACTATTGGA</td>
<td>TTGGTGACAAATCT</td>
<td>2645</td>
<td>32.0</td>
</tr>
<tr>
<td></td>
<td>15a</td>
<td>GGGTACAACTATTGGA</td>
<td>TTGGTGACAAATCT</td>
<td>2623</td>
<td>32.0</td>
</tr>
<tr>
<td></td>
<td>15b</td>
<td>GGGTACAACTATTGGA</td>
<td>TTGGTGACAAATCT</td>
<td>2654</td>
<td>37.9</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>GGGTACAACTATTGGA</td>
<td>TTGGTGACAAATCT</td>
<td>2667</td>
<td>53.3</td>
</tr>
<tr>
<td>IV 17</td>
<td>3326</td>
<td>AGTGCTCTTTCACTGAC</td>
<td>AGCGGGATTTGAAG</td>
<td>3283</td>
<td>44.0</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>TCATAAAACATGACTT</td>
<td>GCTTATATACCTG</td>
<td>3522</td>
<td>45.0</td>
</tr>
</tbody>
</table>

a Section: I - IV (regions of the EcoRI fragment searched); Promoter number: 1-18.
b Coordinate position of the first base of the -35 sequence.
c Coordinate position of the last base of the -10 sequence.
d Percentage score determined by TargSearch (Muligan et al., 1984).

### A3.6 The RepFIB origin of replication

Plasmid origins of replication typically contain three distinctive sequence elements: i) a set of repeats which will bind the plasmid-specific initiation protein; ii) several sequences which bind a number of host encoded replication factors; and iii) stretches of AT or GC-rich sequence involved in DNA melting or stability.

The origin region of RepFIB (1727-20194cbp) has been defined by the use of Xho36 (Spiers, MSc. Thesis, 1989; and Saul et al., 1989) and work in this Thesis. The origin region contains the BCDD'D' repeats as well as a single-strand initiation sequence, a dnaA box, a number of Dam methylation marks, and a number of sequence motifs that are involved in the binding of the host-encoded replication factors.
sites, one or more 13mer-like sequences and numerous potential primosome assembly sites (a detailed map of the origin region is given in Chapter Seven, Figure 7:2). In addition, the origin region contains a stretch of sequence which has a high AT content and a GC-rich sequence which may be important in maintaining the correct spacing between several host and plasmid-encoded protein binding sites.

**A3.6.1 Functional single-strand initiation sequence**

Nomura et al. (1991) have isolated a single, functionally active single-strand initiation sequence (ssi) responsible for priming DNA synthesis in the RepFIB replicon isolated from the F plasmid. The ssi sequence (ssIF) is located within a region which shares 96% sequence homology with the E11 fragment carrying the P307 RepFIB. The corresponding region in the E11 fragment is adjacent to the dnaA box, lying between 1812-1865bp.

The RepFIB ssIF sequence appears to be grouped into a set of ssi sequences which require = 100bp for activity and include ssi sequences from R6K (ssiA), R100 (ssiB), ColE2, mini-F (f5) and two sequences from the f2 EcoRI fragment of the F plasmid. In contrast, the second group of ssi sequences require 200-350bp for full activity (R100 ssiA, R6K ssiB, and ssi sequences from f3 and f6 fragments from F).

**Table A3:6**

<table>
<thead>
<tr>
<th>dbp</th>
<th>Sequence</th>
<th>c</th>
<th>d</th>
<th>dbp</th>
<th>Sequence</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>1944</td>
<td>TTATCCACA</td>
<td>9</td>
<td>1</td>
<td>156</td>
<td>CTGTACACA</td>
<td>7</td>
<td>0.2</td>
</tr>
<tr>
<td>2328</td>
<td>TTATCCAGA</td>
<td>1092</td>
<td>8</td>
<td>637</td>
<td>TTATCCAGT</td>
<td>629</td>
<td>7</td>
</tr>
<tr>
<td>2556</td>
<td>TTATCCAGG</td>
<td>2320</td>
<td>8</td>
<td>1472</td>
<td>TTATCCAGG</td>
<td>1464</td>
<td>7</td>
</tr>
<tr>
<td>1100</td>
<td>TTATCCATA</td>
<td>2548</td>
<td>8</td>
<td>1576</td>
<td>TTATCCACT</td>
<td>1568</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1690</td>
<td>CAATACACA</td>
<td>1698</td>
<td>7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3511</td>
<td>TTATATACA</td>
<td>3519</td>
<td>7</td>
</tr>
</tbody>
</table>

**Table A3:6**

The consensus sequence 5'-TTATCCACA-3' was used to locate potential dnaA box-like sequences in the E11 RepFIB fragment. A number of sequences share homology with the consensus sequence, although only one of these is a perfect match with the consensus (Table A3:6). The dnaA box-like sequences have been scored according to the quality matrix of Messer et al. (1991). This matrix is based on the relative binding affinity between DnaA and a variety of oligonucleotides containing dnaA box-like sequences. A matrix based on the ability of bound DnaA to terminate transcription (Schaefer and

**A3.6.2 Sequences with homology to the dnaA box**

The consensus sequence 5'-TTATCCACA-3' for the binding of DnaA protein (Fuller and Kornberg, 1983; Fuller et al., 1984) was used to locate potential dnaA box-like sequences in the E11 RepFIB fragment. A number of sequences share homology with the consensus sequence, although only one of these is a perfect match with the consensus (Table A3:6). The dnaA box-like sequences have been scored according to the quality matrix of Messer et al. (1991). This matrix is based on the relative binding affinity between DnaA and a variety of oligonucleotides containing dnaA box-like sequences. A matrix based on the ability of bound DnaA to terminate transcription (Schaefer and
Messer, 1991) gives similar scores to those in Table A3:6. The scores suggest that the only significant sequence is the perfect-match sequence located within the origin region of RepFIB.

A3.6.3 Dam methylation sites
The 5'-GATC-3' sequence recognised by Dam methylase occurs throughout the E11 fragment. Five of these sequences are clustered within the origin region and are arranged in an ordered array (Table A3:7). These methylation sites are associated with a repeat sequence (5'-AAnGATCCAnnAA-3'; where n is any nucleotide) which is not found with any of the other GATC sequences outside of the origin region. The repeat sequence is associated with two other repeats which are not Dam methylation sites due to a change in one of the four bases required for recognition.

<table>
<thead>
<tr>
<th>dbp</th>
<th>GATC-repeat sequence</th>
<th>b</th>
<th>dbp</th>
<th>13mer-like sequence</th>
<th>c</th>
</tr>
</thead>
<tbody>
<tr>
<td>1998-2012</td>
<td>AGAGATCCATTAA</td>
<td>11</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2010-2022</td>
<td>AAAGATCCCCATA</td>
<td>10</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Consensus:</td>
<td>AA-GATCCA--AA</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A3.6.4 Sequences with homology to the oriC 13mer sequence
The origin region appears to contain several sequences showing a high degree of sequence similarity with the consensus sequence of the oriC 13mer binding site (5'-GATCTxtTtyt/T/A/T/A/T/G-3'; where x is A, G or C; and y is A, T, or C). The oriC 'AT-rich' 13mer sites appear to serve two functions: DNA melting and protein binding (Kowalski and Eddy, 1989; and Hwang and Kornberg, 1990).

The E11 fragment 13mer-like sequences are located between the single perfect dnaA box (Table A3:7) and the BCDD' repeat elements. Each 13mer-like sequence also includes part of a Dam methylation repeat sequence (Table A3:7), although each of the Dam methylation repeats is not associated with an 13mer-like sequence. The spacing between each of the 13mer-like elements is not similar to the arrangement of 13mer elements in oriC where each element is separated by two or three bases. For this reason and because of the close proximity of the perfect dnaA box, only the first 13mer-like sequence listed in Table A3:7 is likely to be of significance.
A3.6.5 Sequences with homology to the primasome assembly site

The E11 fragment has twenty-three sequences showing homology with the primasome assembly site consensus sequences, 5'-GTGAGCG-3' and 5'-GnGGAGC-3' (n is any nucleotide; Imber et al., 1983). Five of these sites are clustered within = 150bp between the dnaA box and the B repeat element in the origin region of RepFlB (1740-1746, 1755-1749, 1829-1835, and 2215-2121bp).

A3.6.6 Origin repeat elements

The origin region of the E11 RepFlB fragment contains two sets of repeat elements. These include the BCDD' repeats (Section A3.3.1) and the repeat elements associated with the Dam methylation sites (Section A3.6.3). The RepFlB origin region does not appear to include the 5'-ACAGG-3' origin motif, nor do the BCDD' repeats contain the 5'-TGAGGfuG-3' hexanucleotide motif commonly found in the origins of several replicons (λdv, R6K, RK2, mini-F and mini-P1; Filutowicz et al., 1985b).

<table>
<thead>
<tr>
<th>Table A3:8</th>
<th>AT-richness of the origin region</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Region</strong></td>
<td><strong>Size (bp)</strong></td>
</tr>
<tr>
<td>General</td>
<td></td>
</tr>
<tr>
<td>1-4298</td>
<td>4298</td>
</tr>
<tr>
<td>1-1727</td>
<td>1727</td>
</tr>
<tr>
<td>1727-3392</td>
<td>1666</td>
</tr>
<tr>
<td>3392-4298</td>
<td>907</td>
</tr>
<tr>
<td><strong>Regions within the Origin</strong></td>
<td></td>
</tr>
<tr>
<td>1727-2194</td>
<td>468</td>
</tr>
<tr>
<td>1727-2002</td>
<td>276</td>
</tr>
<tr>
<td>1727-1944</td>
<td>218</td>
</tr>
<tr>
<td>1944-2022</td>
<td>79</td>
</tr>
<tr>
<td>2022-2063</td>
<td>42</td>
</tr>
</tbody>
</table>

a Coordinate positions from the first to the last base.

A3.6.7 Origin AT and GC-rich sequences

The overall AT-content of the E11 RepFlB fragment is fairly similar to that of E. coli chromosomal DNA. The RepFlB origin appears to have a high AT-rich region between the dnaA box and the right-hand GATC repeat element (60% AT-rich; Table A3:8). The analysis of the origin sequence composition also indicates that a high GC-rich region is located between the right-hand GATC repeat element and the B repeat element (45% AT-rich, 55% GC-rich).

A GC-rich 37bp region has been identified in the mini-P1 origin (Brendler et al., 1991b). The primary function of this sequence apparently is to act as a spacer between the last GATC repeat element...
and the first P1 RepA binding site. The RepFlB origin appears to have an analogous GC-rich element located in a similar position with respect to GATC repeat elements and RepA binding sites. However, it is not clear which of two possible alignments shares greater homology with the P1 37bp spacer sequence (Table A3:9).

Although the distance between the last GATC repeat element and the B repeat is approximately the same distance between the analogous P1 structures, both RepFIB alignments start = 13-20bp away from the GATC repeat, and alignment #2 includes most of the B repeat. In alignment #2, P1 and RepFlB share a well conserved 9bp element within the P1 37bp spacer region, 5'-GGCGC/GGCG/A-C-3' (superscripts show the P1 sequence).

**Table A3:9** Comparison of the P1 37bp GC-rich spacer and the analogous region in RepFIB

<table>
<thead>
<tr>
<th>dabp</th>
<th>Sequence</th>
<th>dabp</th>
<th>c</th>
<th>d</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1 37bp GC-rich spacer</td>
<td>TTGCAGGCCCGCAGGCTGTGGGTAGAATGTCG</td>
<td>502</td>
<td>12</td>
<td>22</td>
</tr>
<tr>
<td>RepFIB alignment #1</td>
<td>TTTG ... CG A A A A A C</td>
<td>2072</td>
<td>12</td>
<td>22</td>
</tr>
<tr>
<td>RepFIB alignment #2</td>
<td>TGGCGG G C T C A TA</td>
<td>2079</td>
<td>12</td>
<td>15</td>
</tr>
</tbody>
</table>

* Conserved AT or GC pair. Gaps are positioned where no match between the two sequences exists.
  a, b Coordinate positions of the first and last base of the element. P1 sequence coordinates are those used by Brendler et al. (1991).
  c Number of identical bases shared between P1 and RepFIB (out of 37).
  d Number of identical bases plus positions which maintain the same GC-content between P1 and RepFIB (out of 37).

### A3.7 RepFIB and E11 fragment homology with other DNA

Searches of the Genbank database for DNA sequences or proteins sharing homology with the sequence of the E11 fragment have been made regularly. To date (August, 1992) five sequences sharing significant homology with E11 RepFIB fragment have been found. One of these sequences is that of the mini-P1 RepFIA replicon (Abeles et al., 1984); three others are of mini-P1 replicon homologues; P7 (Froehlich and Scott, 1988), Rts1 (Kamio and Terawaki, 1983; Kamio et al., 1984; and Nozue et al., 1988), and R401 (Tabuchi, 1985). The fifth sequence is that of a resolvase gene located on the same EcoRI fragment which carries the F plasmid RepFIA replicon (mini-F).

#### A3.7.1 Homology between RepFIB and mini-P1 RepA proteins

The repA gene of RepFIB shares a significant degree of protein sequence similarity with the repA genes of mini-P1 (67% similar amino acids) and Rts1 (63%, Saul et al., 1989) (Table A3:10). Similar homology has been noted between the ColV3-K30 RepFIB repA gene and the genes of P1 and Rts1 (Perez-Casal and Crosa, 1989). Despite the similarity between the repA protein sequences, the
DNA sequence of the RepFlB gene shares vary little homology with the DNA sequences from P1 or Rts1.

In our initial analysis, repA of RepFlB appeared to start 12 codons after the start position of mini-P1 repA, and to extend some 44 residues beyond the position at which translation of the P1 repA gene is terminated (Saul et al., 1989). However, with the determination of the N-terminal sequence of the RepFlB RepA protein, it is now clear that the RepFlB repA translation start codon is four codons before the P1 repA start codon (Table A3:10).

Due to the poor DNA homology between the two genes, it is not possible to align the RepFlB and P1 (or Rts1) DNA sequences to determine whether a base change, insertion or deletion was responsible for the different translational start and termination codons used by the two proteins.

### Table A3:10

Amino acid sequence homology between RepA of mini-P1, mini-Rts1 and RepFlB

<table>
<thead>
<tr>
<th>RepFlB</th>
<th>1</th>
<th>MENENSDIKK LFGEVDKSSG ELVTLTPNN NTVPQVAMR LGVFVPTLKS</th>
</tr>
</thead>
<tbody>
<tr>
<td>mini-P1</td>
<td>1</td>
<td>MNQSF1 SDILYADIES KALETVSN NTVPQVAMR LGFVPKPSK</td>
</tr>
<tr>
<td>mini-Rts1</td>
<td>1</td>
<td>METOLVI SDVLFGNTEE KQKPLTVNEL NTIQPVAMR LGFVPKPSR</td>
</tr>
<tr>
<td>RepFlB</td>
<td>51</td>
<td>LKNSKKNTLS RTDATEELTR LSL1RAEGFD KVEITGPRLD MDNDFKTVWG</td>
</tr>
<tr>
<td>mini-P1</td>
<td>47</td>
<td>SKGESKE IDATKAFSQ LEIAKAEYD DIKITGPRLD MDTFKTVWG</td>
</tr>
<tr>
<td>mini-Rts1</td>
<td>48</td>
<td>SSDYSPM IDVSELSST FEFARLEGFT DIKITGERLD MDTFKTVWG</td>
</tr>
</tbody>
</table>

| RepFlB | 101 | IIFSARHNV IGDKVELPVF EFAKLCGIPS SQSRR2LRER ISPSLKRIAG |
| mini-P1 | 93 | VIYAFSKYGL SSNTIQLSFO EFAKACGFPS KRLDAKLRIT IHESLGLRLN |
| mini-Rts1 | 94 | IVKAFSKYGI SSNRILKLFSS EFAKDCGFPK KLKDKLRAH IDESLKLRIG |

| RepFlB | 151 |TVISFSRTDE KHRTREYIHTL VQSATTDER DIVQLQPDR LFELYQFDK |
| mini-P1 | 143 | KIAFKRGKD AKG GYQTL GLKVRFDADL DLIELEAASK LWELFQDYL |
| mini-Rts1 | 144 | KISIFKRGKD SQS AYHTGL IKIAYFNADT DWVLEADER LWELLYFDY |

| RepFlB | 201 | VLLLQKAINA LKRESAQAL YTIESLPRD PAPISLARLR ARLNKLKPSF |
| mini-P1 | 192 | VLLQHARLA LPKKEAQAQI YTIESLSPQN PLPSPEARIR ERLALQSAVG |
| mini-Rts1 | 193 | VVLQHLAIKA LPRLEVAQL YTFLASLPSN PAPISERL ERLSLISQV |

| RepFlB | 251 | SQNQTVRAM EQLREIGYLD YTEIQGRRTK 3FCIHYRPR LKAPNDES |
| mini-P1 | 242 | EQNRIKKAIE EQLKTIGYLD CSIEKKGRES FIVH5SNPK KKLPE |
| mini-Rts1 | 243 | EQNRIKKAIE TKLIDIGNLD ASMVKKKQEN YLIHCRSPK LSVINE |

| RepFlB | 301 | NP4PPSPAEEK VSPENAELA LLEKLGITLD DLEKLFKSR |

1 A in P307, D in pMA4322. 
2 R in P307, S in R386. 
3 F in P307, L in pH507. 
4 L in P307, Q in F, R124, ColV3-K30 and ColIV.

Gaps have been inserted in the sequences to obtain maximum homology. Underlined sequences mark potential DNA binding domains (see Table A3:3). Note that the RepA start codon used in this table was determined in this Thesis.

---

**APPENDIX THREE**
ORF-3 homology to the resolvase gene of mini-F

Several regions from mini-F (RepFlA) were identified as sharing some degree of homology with the E11 RepFlB fragment (Spiers, MSc. Thesis, 1988; and Saul et al., 1989). Of these, homology to the mini-F resolvase gene (D gene, Lane et al., 1986) was considered to be significant with 57% overall conservation between the two coding regions. This similarity is increased to 63% (over all) when the protein sequences are compared. If all three reading frames are compared with the D protein sequence upstream of the first start codon of ORF-3 (in frame g, GTG at 3981bp), homology between the g frame and the D gene extends up to 4020-4040bp. From this point, homology exists between the predicted protein sequence in the d frame and the D protein up to the EcoRl site at 42986bp. A single base deletion may have been responsible for the frameshift in the gene homologous to the mini-F resolvase gene. This gene may have been translated from a codon = 20bp beyond the EcoRl site to produce a protein of the same molecular mass of the D gene protein (29.6kDa).

Phylogenetic analysis of the relationships between different RepFlB replicons

The RepFlB replicon is present in a variety of different plasmids, most of which are in the IncFl group. RepFlB has been confirmed in twenty-five plasmids by southern hybridisation or PCR analysis (Lane and Gardner, 1979; Bergquist et al., 1986; and Gibbs, MSc. Thesis, 1991); in sixteen plasmids of the IncFl group, four IncFlII plasmids, two IncFlIII plasmids, one IncFlV plasmid, one IncI plasmid and in one IncP plasmid (Gibbs, MSc. Thesis, 1991). Of these twenty-five examples, the sequence of the RepFlB replicon has been determined from P307, F (Saul et al., 1989); R386, R124, pH502, pH507, ColVBtrp (Gibbs, MSc. Thesis, 1991); pColV (Kim et al., 1988; and Gibbs, MSc. Thesis, 1991) and pColV3-K30 (Perez-Casal and Crosa, 1989; Perez-Casal et al., 1991; and Gibbs, MSc. Thesis, 1991). The sequence data from these examples corresponding to the 1525bp 2046-3570bp region of the E11 RepFlB fragment (P307) has been analysed by using maximum parsimony techniques to describe the possible evolutionary relationships between the different examples of RepFlB (Figure A3:4).

As no outgroup sequence was available to root the tree, all examples have an equal chance of representing the most ancestral form of the replicon. Although the DNA sequence data from P1 and Rts1 can be aligned to the RepFlB sequences, the high level of divergence does not allow parsimony analysis to reliably predict the correct tree and determine the ancestral RepFlB replicon. This problem cannot be solved using the protein sequences to determine phylogenetic relationships. Use of the protein sequence of RepA from P1 or Rts1 as an outgroup to root the RepFlB RepA tree results in a simplistic dichotomeric grouping of P1 (or Rts1) and RepFlB RepA. As the difference between P1 and Rts1 RepA proteins to RepFlB RepA is so great, the inclusion of P1 or Rts1 RepA in the analysis does not effectively root the tree (Gibbs, MSc. Thesis, 1991).
Phylogram taken from Gibbs (MSc. Thesis, 1991) showing the single most parsimonious solution for the relationships between the nine examples of RepFIB for which sequence data is available. The analysis has grouped P386, ColV Btrp, pHH502 and pHH507 together, R124, ColV3-K30 and ColV form a second group, whilst P307 and the copy mutant pMA4322 (P307*) form the third grouping. Branch lengths are scaled horizontally.
A4.1 Introduction

This appendix briefly describes two Western-DNA experiments in which attempts were made to inhibit RepA DNA-binding. RepA antiserum was used in an attempt to block the DNA-binding site or sites of RepA before the Western-DNA membrane was incubated with probe DNA. In the second experiment, RepA-cyanogen bromide and low pH-cleavage products were tested in order to determine which of three potential DNA-binding domains were functional in the protein. Although neither experiment was successful, both are briefly described below.

A4.2 Antiserum inhibition of RepA DNA-binding

An attempt was made to inhibit RepA DNA-binding on Western-DNA membranes by pre-incubating the membranes after electrotransfer in blocking buffer containing various concentrations of RepA rabbit antiserum. However, the presence of RepA antiserum had no significant effect on RepA DNA-binding, regardless of whether DNA-binding was determined by measuring the peak intensity across the RepA-'band' on the autoradiograph (Figure A4:1), or by measuring the mean intensity of the 'band' (data not shown).

RepA DNA-binding insensitivity to the antiserum might be due to the fact that the antiserum was generated against denatured RepA and may not recognise renatured RepA. However, this explanation is unlikely since the antiserum can recognise RepA on Western-DNA membranes after autoradiography (data not shown). An alternative explanation is that the insensitivity may be due to a relatively poor immunogenicity of the DNA-binding domain of RepA. If this were the case, a relatively minor portion of the antibodies in the antiserum would be targeted against the domain and DNA-binding would not be affected except at very high antiserum concentrations. It is also possible that the DNA-binding domain is sufficiently well protected so that the access of antibodies directed against the DNA-binding domain is prevented.
RepA antiserum does not affect RepA DNA-binding

RepA DNA-binding inhibition by RepA antiserum was measured by incubating Western-DNA membranes with RepA antiserum before incubation with the probe DNA. After autoradiography, the peak intensity of the signal was measured as a function of film opacity (relative units, vertical axis). **Left to right:** no antiserum; 1/30,000, 1/25,000, 1/20,000, 1/15,000, 1/10,000, 1/5,000 and 1/2,000 dilutions of the antiserum.

**Method:** 8M urea extract protein samples containing RepA were separated by 12.5% SDS-PAGE in this Western analysis. The membrane strips were incubated in TBS, 0.5% Tween-20, 3% NFMP, 0.01 mg/ml sonicated calf thymus DNA (CTD) for two hours with various dilutions of antiserum before incubation with BamH1-cut, 32P-end labelled PCR-amplified DNA (primers #8 and #2) for two hours. The densitometric scans were made directly from the autoradiograph.

**A4:3 Identification of the RepA DNA-binding domain**

RepA has three potential DNA-binding domains which have been identified using a consensus sequence derived from the binding domains of known prokaryotic DNA-binding proteins (Nakata and Maizel, 1989). The three domains (I, II and III) can be separated by cleavage of RepA using either cyanogen bromide (CnBr) or incubation at pH 2.5. Cleavage using CnBr will result in a 19.6kDa polypeptide containing domains II and III, whilst cleavage at pH 2.5 will produce a 21.2kDa containing domains I and II (Figure A4:2). These two fragments are sufficiently large to be identified by SDS-PAGE using 12.5% gels. However, the smaller fragments resulting from cleavage with CnBr or at pH 2.5 would migrate with the dye front and remain unresolved by the gel. Therefore, it seemed possible that if either the 19.6kDa or 21.2kDa RepA cleavage products contained functional binding domains, DNA-binding should be observable using the Western-DNA technique. A comparison of the binding abilities of the 19.6kDa and 21.2kDa cleavage products could be used to identify which of the three domains are responsible for RepA DNA-binding.
RepA is cleaved by cyanogen bromide (CnBr) to produce five polypeptides whilst incubation at low pH (pH 2.5) will produce three polypeptides. Two of the three potential DNA-binding domains in RepA can be isolated on the largest CnBr-fragment (domains II and III) while domains I and II can be isolated on the largest pH 2.5-fragment.

8M urea extracts containing RepA were cleaved with CnBr or cleaved at pH 2.5 to provide the samples necessary for Western-DNA analysis. However, neither the 19.6kDa nor the 21.2kDa RepA cleavage products bound RepFlB probe DNA. In order to test whether the cleavage conditions resulted in a peptide unable to refold during the Western-DNA procedure, freeze-dried samples of 8M urea extracts containing RepA after cleavage were resuspended in urea buffer before SDS-PAGE. Presumably, this resuspension would denature the cleavage products to such an extent that they could then mimic the refolding of un-cleaved RepA during the course of the Western-DNA procedure. However, incubation for twenty minutes in 1M or 2M urea buffer prior SDS-PAGE did not result in binding of the probe DNA by the cleavage products. As a result, the identification of the binding domain or domains remains unresolved.
Evidence for a ccd-like function

A5.1 A brief description of the ccd system

Controlled cell death (ccd, formerly coupled cell division) functions are maintenance systems often associated with plasmid replicons or oriT (origin of transfer) regions of conjugative plasmids. In a population of plasmid-plus and plasmid-minus bacterium where there is no advantage in maintaining a plasmid, plasmid-plus cells are disadvantaged by the increased metabolic load required to maintain the resident plasmid. Under such conditions, plasmid-plus strains are readily cured of their plasmids, generally through the faster growth of plasmid-minus segregants resulting from the uneven partition of daughter plasmids during cell division. In order to prevent the generation of plasmid-minus segregants, plasmids can either integrate into the host chromosome or carry a ccd system which will destroy plasmid-minus cells before they can develop into a significant portion of the population.

Controlled cell death was first documented by Ogura and Hiraga (1983) in mini-F. The ccd function involves two proteins (Bex et al., 1983), one of which causes cell death (CcdB) and the other (CcdA) which inhibits the action of the first (Karouri et al., 1983). A further analysis of mini-F ccd by Jaffé et al. (1985) provided the Non-viable Segregant model of ccd action. In this model, CcdB reduces the viability of plasmid-minus segregants by interfering with the cellular division process. It is now known that CcdB destroys chromosomal DNA by interacting with Topoisomerase II and causing the accumulation of linearised DNA (Bernard and Couturier, 1992).

A second ccd-like system has been found amongst plasmids of the IncP group. One well studied representative from this group is the kil/kor system from R2K (for example, see Figurski et al., 1982; and Young et al., 1985). Although R2K appears to have at least three pairs of killer/saviour proteins (whose genes are not linked and are randomly positioned throughout R2K), the control of each killer protein (Kil) by the saviour (Kor)1 seems to occur in a manner similar to the control of CcdB by CcdA. However, there is no sequence similarity between mini-F ccd and R2K kil/kor genes.

There exists a third group of ccd-like functions which form a highly conserved class which is quite distinct from both mini-F ccd and R2K kil/kor (reviewed by Gerdes et al., 1990). The class relies on the expression of a short counter-transcript (or antisense) RNA to prevent the translation of a killer

1 Kil - killing, Kor - repressor of killing.
protein from a second more stable RNA molecule. In the case of the R1 plasmid hok/sok system\(^2\), Sok ctRNA prevents the expression of the Hok protein (Gerdes et al., 1986). Sok ctRNA prevents the accumulation of Hok mRNA by mediating RNase III-degradation of the mRNA (Gerdes et al., 1992). However, if Hok mRNA is translated, Hok will damage the cell membrane, which in turn leads to gross cellular changes and cell death.

The sequence of the E11 fragment has been compared with the sequences of the mini-F ccd, R2K kil/kor and the R1 Hok/sok system in an attempt to identify regions of homology. Although no such regions have been found, a ccd system similar to mini-F or R2K would only require two 10-12kDa genes, and a system similar to R1 would require a single = 10kDa gene.

### A5.2 Experimental evidence for the presence of a ccd-like system in the E11 fragment

In an experiment designed to determine whether RepFIB replication is dependent of host DnaA, evidence was obtained which suggested that the E11 fragment used to produce RepFIB mini-plasmids contained a ccd-like system (Chapter Seven, 7.2.2). A schematic diagram and explanation of the experiment is given in Figure A5.1. The comparison of a PB1849 strain carrying a RepFIB mini-plasmid which was dependent on host DnaA for replication (pSS3928) and a PB1849 strain carrying a DnaA-independent RepFIB hybrid plasmid (pNZ945) indicated that the E11 fragment used to produce both plasmids carried a ccd-like system (Figure A5.2, A and B). This conclusion was based on the fact that the PB1849 + pSS3928 secondary cultures failed to grow at 32°C regardless of selection, whereas the PB1849 + pNZ945 secondary cultures grew independently of whether the selective antibiotic was added to the cultures.

### A5.3 Probable location of the ccd-like region

The ccd-like function encoded by the E11 fragment is almost certainly located between 0-1900cbp. The remaining portion of the E11 fragment (1901-4298cbp) is not likely to contain the ccd-like system, since a RepFIB mini-plasmid constructed using the EcoRI 7.5kb f7 fragment is unstable without selection and plasmid-minus segregants are viable (observations from this group). The E11 and f7 fragments sequences are almost identical in the RepFIB replicon portions of the fragments (1900-3600cbp in E11). Southern analysis has demonstrated that the region to the right of the replicon (= 3600-4298cbp of E11) is highly conserved between the two fragments and that the sequence to the left of the replicon (0-1900cbp of E11) is the only region in which the two fragments are not homologous (observations from this group). If f7 contained a ccd-like function, then a f7 mini-

\(^2\) Hok - Host killing, Sok - suppressor of killing.
plasmid would not be unstable in the absence of selection and plasmid-minus segregants could not be grown. If the f7 fragment contains the same ccd-like function encoded by E11, then the ccd-like function must be located in a region of DNA that is common to both fragments. Therefore, since E11 contains a ccd-like function and f7 does not, the ccd-like function must be located in the 0-1900cbp region of E11.

A temperature-sensitive DnaA strain (PB1849) was used to determine whether RepFlB mini-plasmids required DnaA for replication (boxed portion). DnaA-dependent replication was recognised if the strain carrying the plasmid failed to grow at 42°C when a selective antibiotic for the plasmid was added to the culture. As a control, an identical culture was grown at 42°C (without the antibiotic) to demonstrate that incubation at 42°C did not prevent the growth of the culture. The presence of a ccd-like function was examined by using samples taken from the 42°C control culture at various intervals to determine the cell viability. The samples were used to inoculate fresh cultures which were then grown at 30°C. At this temperature, DnaA is active, and if the cells still contain the DnaA-dependent plasmid, the plasmid should now be able to replicate. If the plasmid is still present, the culture should grow in the presence of a selective antibiotic. If the plasmid has been lost, the culture will not grow. If the plasmid contained a ccd-like function, then cultures inoculated with cells which have lost the plasmid will not develop even in the absence of the antibiotic. It is important to note that the presence of a ccd-like function is only obvious by comparing PB1849 containing different plasmids. For this reason, a DnaA-dependent plasmid carrying the suspected RepFlB ccd-function is compared with a second, DnaA-independent plasmid which carries the identical region of RepFlB.

**Figure A5:1** An extension of the test for DnaA-dependent replication to demonstrate the presence of a ccd-like system.
Evidence for a ccd-like function.
The 0-1900cbp region active promoter the coding sequence. ORF-7 could express these. See Chapter Two, 2.3.5. See Appendix Three, A3.2, and Table A3.1.

**Analysis of the ccd-like function**

The E11 fragment ccd-like function can be investigated using the DnaA-temperature sensitive replication assay schematically described in Figure A5:1. PB1849 strains carrying various plasmids are first grown at 42°C with or without antibiotic selection (top graphs). At various time intervals, samples of the non-selective cultures (○) are used to inoculate new cultures grown at 30°C with or without selection (bottom graphs). The presence of the ccd-like function is seen by comparing the different growth of the 30°C cultures for pSS3928 (A) and pNZ945 (B). The ccd-function is not complemented or neutralised by the presence of a second plasmid carrying ORF-7 (pAS10, C) or the EcoRI-PstI 2.2kb portions of the E11 fragment (#288, D).

**A** PB1849 containing pSS3928 without selection (○) and with spectinomycin selection (●) to maintain the plasmid. Samples were taken from the ○-culture and were used to inoculate secondary cultures which were grown overnight at 30°C (lower graph; no selection - ○; with spectinomycin selection - ●).

**B** PB1849 containing pNZ945 without selection (○) and with ampicillin selection (●) to maintain the plasmid. Samples were taken from the ○-culture and were used to inoculate secondary cultures which were grown overnight at 30°C (lower graph; no selection - ○; with ampicillin selection - ●).

**C** PB1849 containing pSS3928 and pAS10 without selection for pSS3928 (○) and with spectinomycin selection (●) to maintain pSS3928 (pAS10 was maintained with ampicillin). Samples were taken from the ○-culture and were used to inoculate secondary cultures which were grown overnight at 30°C (lower graph; no selection for pSS3928 - ○; with spectinomycin selection for pSS3928 - ●).

**D** PB1849 containing pSS3928 and #288 without selection for pSS3928 (○) and with spectinomycin selection (●) to maintain pSS3928 (#288 was maintained with chloramphenicol). Samples were taken from the ○-culture and were used to inoculate secondary cultures which were grown overnight at 30°C (lower graph; no selection for pSS3928- ○; with spectinomycin selection for pSS3928 - ●).

**Method:** (Top graphs) Overnight cultures of PB1849 strains grown at 30°C were used to inoculate L-broth cultures (1/100 dilution) which were then grown at 42°C with or without selection for the test plasmid (pSS3928: 50µg/ml spectinomycin; pNZ945: 50µg/ml ampicillin). At intervals the growth of the culture was determined by measuring the culture density (OD600). (Bottom graphs) 10µl samples taken from the non-selective cultures grown at 42°C were used to inoculate 2.5ml L-broth cultures (with and without antibiotics). These secondary cultures were incubated overnight at 30°C before the growth of the culture was determined by measuring the culture density (OD600). pAS10 was maintained with 50µg/ml ampicillin and #288 with 25µg/ml chloramphenicol at all times.

The 0-1900cbp region of E11 contains three large open reading frames (ORF-2, ORF-6 and ORF-7)³. Of these three, ORF-7 has the highest probability of expression in terms of a statistical analysis of the coding sequence. ORF-7 could express a protein of 9.2kDa and the coding region is preceded by an active promoter (ORF-7p)⁴. ORF-6 is the next most likely open reading frame to be expressed, and

³ See Appendix Three, A3.2, and Table A3.1.
⁴ See Chapter Two, 2.3.5.
finally ORF-2. ORF-7<sub>p</sub> transcripts might be translated to express ORF-6, but this possibility has not been investigated. Although ORF-2 is in the correct orientation to be translated from ORF-7<sub>p</sub> mRNA, it is unlikely to occur since the open reading frame is ~700bp downstream of the promoter sequence.

ORF-6 and ORF7 appeared to be good candidates for the ccd-like system since the two open reading frames may be linked in so far that both may be translated from transcripts originating from ORF-7<sub>p</sub>. If these two ORFs represented the ccd-like function, then ccd-induced death caused by the loss of pSS3928 should be overcome if ORF-6 and ORF-7 are present in trans on a second plasmid. If ORF-7<sub>p</sub> is part of the ccd system, then ORF-7<sub>p</sub> might be autoregulated by either of the ORF-6 or ORF-7 proteins.<sup>5</sup>

ORF-7<sub>p</sub> regulation has been demonstrated using the ORF-7:β-galactosidase fusion promoter-probe plasmid pAS44<sup>6</sup>. The expression of the fusion protein from ORF-7<sub>p</sub> is significantly reduced when a RepFlB mini-plasmid is present in trans (Table A5:1). The E11 fragment only contains three open reading frames with a high probability of expression (repA, ORF-6 and ORF-7)<sup>7</sup>; if ORF-7<sub>p</sub> is repressed by a protein, then it is most likely to be a protein expressed by either ORF-6 or ORF-7 since the region around ORF-7<sub>p</sub> does not contain any sequences sharing homology to the RepA binding sites (the A-K repeat elements)<sup>8</sup>.

<table>
<thead>
<tr>
<th>Plasmids</th>
<th>β-Gal.</th>
<th>Repression</th>
</tr>
</thead>
<tbody>
<tr>
<td>pAS44</td>
<td>1160 ± 26</td>
<td>0%</td>
</tr>
<tr>
<td>pAS44 + pMA4322</td>
<td>380 ± 7</td>
<td>77%</td>
</tr>
<tr>
<td>pMA4322</td>
<td>0.0 ± 1.3</td>
<td></td>
</tr>
</tbody>
</table>

DH5α cells were transformed with pAS44, pAS44 + pMA4322, or pMA4322 CsCl DNA. β-galactosidase activities were determined from mid-log phase cells and mean Miller units with standard errors are shown. The plasmids were maintained with the appropriate antibiotics (pAS44: 100μg/ml ampicillin; pMA4322: 50μg/ml spectinomycin).

There are three repeat elements located upstream of ORF-7<sub>p</sub> (the 123 repeats)<sup>9</sup> which may act as binding sites for the ORF-7<sub>p</sub> repressor. However, as the repeats are located 200-300bp from the

---

5 Such an arrangement is found in the mini-F ccd system where CcdA autoregulates the expression of ccdA and ccdB (Tam and Kline, 1989; and Feyter et al., 1989). A second possibility is that ORF-7<sub>p</sub> is regulated by a repressor expressed by a second promoter located near ORF-7<sub>p</sub>. This possibility is less likely since an examination of the sequence around ORF-7 has not identified a potential promoter sequence (Appendix Three, A3.4).

6 See Chapter Two, 2.3.5. The details of all plasmids constructed for this Thesis are supplied in Appendix Two.

7 ORF-3 has a high probability of expression, but lacks a promoter and the first ~20 residues of the coding region (Appendix Three, A3.2.2, A3.7.2).

8 See Chapters Three, Four and Six for RepA DNA-binding.

9 See Appendix Three, A3.3.2; Figure A3:1.
promoter sequence, regulation of ORF-7\textsubscript{p} from these elements may be unlikely. An examination of the sequence around ORF-7\textsubscript{p} has not identified any other obvious sets of repeat elements.

### A5.4 Attempts to complement the ccd-like function

Attempts were made to complement or neutralise the E11 ccd-like function carried by pSS3928 with a second plasmid carrying DNA thought most likely to contain the ccd-like function genes. Attempts to complement pSS3928 were made with pAS10 and #288. The *Bgl* II 0.67kb fragment which contains ORF-7\textsubscript{p} and the majority of ORF-7 was inserted into pBS\textsuperscript{+} to produce pAS10. #288 contains the EcoRI-PstI 2.2kb (0-2194cbp) region of the E11 fragment which includes ORF-2, ORF-6, ORF-7, ORF-7\textsubscript{p}, and the origin of replication. Since both plasmids contain DnaA-independent replicons, they are able to replicate in PB1849 at 42\textdegree{}C and can be maintained with ampicillin or chloramphenicol respectively.

PB1849 strains containing pSS3928 and pAS10 or #288 grew very poorly compared to PB1849 + pSS3928 (Figure A5:2, C and D) and neither pAS10 nor #288 was able to complement or neutralise the ccd-like function carried by pSS3928.

### A5.5 Summary of the analysis of the ccd-like function in the E11 fragment

An experiment used to determine whether RepFIB replication was dependent on host DnaA was extended and used to demonstrate that the EcoRI 4.3kb E11 fragment used to produce RepFIB mini-plasmids also contained a ccd-like function. The ccd-like function serves to prevent the formation of a substantial sub-population of plasmid-minus segregants in a culture where the plasmid does not confer a selective advantage to the host. In the case of RepFIB, the ccd-like function appears to be located in the 0-1900cbp region of the E11 fragment, although attempts to confirm this location by trans complementation have not been successful. A comparison of the E11 sequence with the *GenBank* database (August, 1992) has not identified any regions of E11 which show homology with known ccd functions.

Further attempts to locate the ccd-function are currently underway (A. Spiers and N. Bhana).
A6.1 Replication control of ColE1

A6.1.1 Overview

The control of ColE1 plasmid replication is well understood and serves as an example for a class of related plasmids which include RSF1030, CloD13, p15A and pMB1 (most plasmid vectors available today utilise p15A or pMB1-derived replicons). ColE1 replication has been well reviewed by Rowbury (1978), Backman (1979), Davison (1984), Scott (1984), Cesareni and Banner (1985), Thomas (1988) and by Køes and Stahl (1989).

The control of replication involves three elements: a repressor RNA molecule (RNA-I), an initiator RNA (primer-RNA or RNA-II) and a small protein (Rom or Rop). The genetic organisation of these elements in ColE1 is shown in Figure A6:1. Copy control relies on the variation in concentration of RNA-I and the Rom-mediated repressive interaction by RNA-I with the primer-RNA. If the plasmid copy number is low, RNA-I concentrations will be low and as a result primer-RNA will form a RNA/DNA hybrid in the origin region which is then cleaved by ribonuclease H (RNase H1) to initiate DNA synthesis. However, if RNA-I levels are high, Rom-mediated hybridisation between RNA-I and the 5' region of primer-RNA will occur. This hybridisation prevents primer-RNA RNA/DNA hybridisation at the origin and DNA synthesis is not initiated.

A6.1.2 Primer-RNA

The primer-RNA (RNA-II) is an untranslated RNA molecule required for the initiation of replication. It is expressed from a constitutively active promoter containing three Dam methylation sites which can regulate promoter activity (Patnaik et al., 1990). Primer-RNA transcripts of various lengths have been isolated, ranging from ≥135n, 185-360n, 360-555n to >555n. Only the 360-555n and >555n forms can be modified to give a mature primer capable of initiating DNA synthesis.

A6.1.3 RNA-I

RNA-I is a 108 nucleotide, untranslated RNA molecule entirely complementary to the 5' terminal region of primer-RNA. RNA-I has three functional elements: three single-stranded loop domains; three double-stranded stems (arranged to form three stem-loop domains) and a single-stranded 5' tail. These elements have been confirmed by T1 ribonuclease, S1 nuclease, cobra venom nuclease and diethyl pyrocarbonate modification analyses (Tamm and Polisky, 1983). A comparison of the wild
type RNA-I with a wide variety of RNA-I mutants has indicated that the integrity of all three elements are necessary for RNA-I repressor activity (Dooley et al., 1985). In particular, removal of 5 bases from the 5' tail will greatly reduce RNA-I repressor activity (Tomizawa 1984). Lin-Chao and Cohen (1991) have demonstrated that the five 5' bases of RNA-I are not required for activity per se, but that the deletion of the bases produces a particularly unstable inactive 105n ΔRNA-I molecule. RNA-I is cleaved specifically by ribonuclease E (RNase E) to give the 105n form, and mutations altering the RNase E recognition site alter RNA-I half-life and plasmid copy number (Lin-Chao and Cohen, 1991).

RNA-I is transcribed from an efficient promoter but its half life is so short that the concentration of RNA-I closely follows a rise or fall in plasmid copy number. RNA-I concentration appears to be the controlling factor in copy number control of CoIE1 and RNA-I is degraded primarily by RNase E activity. However, RNA-I molecules lacking the RNase E recognition site can be eliminated by growth-rate dependent degradation resulting in growth-responsive control of plasmid copy number (Lin-Chao and Cohen, 1991).

A6.1.4 RNA-I/Primer-RNA interaction and the involvement of Rom (Rop)

In an investigation of the 600bp region of DNA containing the promoters for both RNA-I and primer-RNA, Cesareni et al. (1982) found that the RNA-1 promoter was insensitive to CoIE1 in trans, whilst the primer-RNA promoter was repressed. A small 63-amino acid polypeptide (6.5kDa) was responsible for this repression, and was named Rop (repressor of primer). Cesareni et al. suggested that Rop modulated the initiation of primer-RNA transcription by repressing promoter activity. Rop was isolated and identified by Som and Tomizawa (1980) in an analysis of the expression of galK from primer-RNA promoter fusions with the galK structural gene. Som and Tomizawa refer to the

**Additional reading: replicon genetics**
protein as Rom (regulator of modulation). Rom (Rop) lowered galactokinase activities when 135bp of primer-RNA sequence was included in the fusion gene, but had no effect when only 52bp was used. In further work, they demonstrated that Rom acted not by the repression of promoter activity, but by enhancing the binding of RNA-I to primer-RNA (Tomizawa and Som, 1984). For this reason, the name Rom is preferred for the protein.

An analysis of RNA-I/primer hybrids using RNase T1 revealed that pairing is initiated at or near the 5' end of RNA-I and progresses towards the 3' end of the molecule (Tomizawa, 1984). Prior to pairing, the two molecules interact in the regions of their single-stranded loops. This interaction is transient, but facilitates the complete pairing ("zipping") between molecules along the entire length of RNA-I (Figure A6:2). Binding of the two RNA molecules requires that the loop sequences are complementary, whereas the stem sequences can vary. All bases in the loop regions are involved in pairing and Rom binds and stabilises any complex formed by pairs containing fully complementary loop sequences (Eguchi and Tomizawa, 1991). The formation of the hybrid is particularly sensitive to mutations in the 5' region of RNA-I, suggesting that this region of RNA-I is also involved in inter-molecular pairing (Tamm and Polisky, 1985).

![Figure A6:2 'Zipping' of RNA-I and the 5' region of primer-RNA](image)

Schematic diagram representing the interaction of RNA-I with primer-RNA (A). From top to bottom (B), the two molecules first interact between their single-stranded loop regions. The 5' single-stranded 'tail' of RNA-I begins to hybridise to the complementary sequence in primer-RNA, and base-pairing 'zips' from 5'→3' along the length of RNA-I. The interaction of RNA-I and primer-RNA is mediated by the Rom protein which appears to bind to the 'stem' structures present in both molecules. Rom binding aligns the molecules and allows 'zipping' to proceed. The ribonuclease HI cleavage site is marked by '▼'.

![APPENDIX SIX](image)
Rom will bind to either RNA-I or primer-RNA, or to the hybrid molecule. The stem and loop structures of RNA-I alone are sufficient to bind Rom. The bound protein protects the stems from nuclease degradation, but leaves the loop portions sensitive to ribonuclease activity (Helmer-Citterich et al., 1988). Rom acts as an adaptor between the stem structures to position the two RNA molecules correctly to allow loop interaction and inter-molecular pairing. Mutagenesis analysis of Rom has shown that the dimeric protein is particularly insensitive to residue alterations and to small insertions or deletions. The monomer consists of two $\alpha$-helices connected by a sharp bend (Benner et al., 1987). Most changes which affect Rom activity are clustered together at the extremities of each of the four $\alpha$-helices of the dimer, presumably altering the RNA-binding domain (Castagnoli et al., 1989). Only a single domain is required to recognise both RNA-I and primer-RNA, as both molecules are synthesised from the same template containing an inverted repeat. As a result of this, RNA-I and primer-RNA have identical stem-structure sequences which are recognised by Rom.

Primer RNA is not always sensitive to repression by RNA-I. Early in the synthesis of primer-RNA, the molecule adopts an "anti-RNA-I" conformation which is then replaced by an alternative conformation (a stem-loop) which is maintained until the completion of transcription (Wong and Polisky, 1985). RNA-I can interact with primer-RNA whilst it is in the 'anti-RNA-I' state (the 'window of susceptibility' during which time RNA-I is 135-240n long), but cannot form a hybrid when primer-RNA is in the alternative conformation (changed by the time RNA-I is 241n long). This change of conformation during a discreet period in transcription affects the region of primer-RNA which is the complement of the RNA-I 5' single-stranded tail. If RNA-I binds to primer-RNA during the window of susceptibility, primer-RNA cannot adopt the necessary configuration for RNase H processing required for the initiation of replication. If RNA-I does not bind to primer-RNA, approximately half of the nascent primer-RNA molecules will form persistent hybrids with their DNA template near the origin region.

The pause pattern of RNA Polymerase during primer-RNA transcription of the 'anti-RNA-I' state differs between wild type and a particular class of mutant replicons which permit RNA-I to hybridise with primer-RNA yet still allow RNase H processing (Polisky et al., 1990). The difference in patterns suggests that the primer-RNA secondary structure differs between wild type and mutant primer-RNA. The altered secondary structure of the mutant primer-RNA molecule allows pairing with RNA-I but stops the subsequent alteration of primer-RNA structure which would normally prevent RNase H cleavage downstream.

A6.1.5 DNA synthesis

ColE1 replication is unidirectional and is initiated separately on the leading and lagging-strands. The components required for leading-strand initiation were fractionated in a high concentration ammonium sulphate precipitation of cell lysates, and included RNA Polymerase, DnaB (gyrase) and ribonuclease H (Hilenbrand and Stauderbauer, 1982). Itoh and Tomizawa (1980) demonstrated that primer-RNA was cleaved by RNase H in the origin region to initiate replication. RNase H cleaves only a portion of
primer-RNA molecules, as primer transcripts passing the origin can be isolated (the processed form of primer-RNA is 550n). Of those molecules which are cleaved at the origin to leave a 3'-OH terminus, no small RNA fragments have been isolated corresponding to the section of the primer which extended beyond the origin. This finding suggests that RNase H cuts the RNA/DNA hybrid at many sites downstream of the cleavage position used to initiate replication.

RNase H cleaves within a sequence of five adenosines in the origin to generate active primers. Any precursor primer-RNA still hybridised to the origin DNA is cleaved at secondary sites by RNase H and is then digested by the 5'→3' exonuclease activity of DNA Pol I during DNA elongation from the mature primer. Finally, further cleavage by RNase H will remove any remaining primer from the nascent DNA strand (Selzer and Tomizawa, 1982).

In an in vitro assay with DNA Polymerase (DNA Pol I), RNA Polymerase and RNase H, Itoh and Tomizawa (1980) found that DNA Pol I will add either dAMP or dCMP to the 3'-OH end of primer-RNA. This addition represents the first step in the synthesis of new leading strand DNA. DNA Pol I addition to the 3'-OH group is apparently very inefficient, as Inoue and Uchida (1991) have demonstrated that most of the cleaved primer-RNA is unable to prime the initiation of DNA synthesis effectively.

Nomura and Ray (1980) isolated the lagging-strand initiation site (rri-1) which was located 150bp downstream of the origin. Once the leading-strand synthesis has begun, complementary DNA synthesis of the lagging strand begins from rri-1 on the nascent leading strand DNA. Initiation from rri-1 requires the host dnaB and dnaG gene products (Nomura and Ray, 1980). Thereafter, DNA synthesis of the ColEl plasmid is similar to that of the E. coli chromosome.

Replication can also occur in a more complex manner in hosts lacking either or both of RNase H and DNA Pol I (reviewed by Kües and Stahl, 1989). Inoue and Uchida (1991) found that 30-40% of primer-RNA transcripts ended at or near the origin in an rnh mutant strain lacking ribonuclease H, suggesting that E. coli has residual RNase H activity sufficient to allow ColE1 replication. This residual activity is encoded by the rnhB gene which expresses a second ribonuclease H, RNase HII (Itaya, 1990).

A6.1.6 Summary of ColE1 replication control

The copy control mechanism used by ColE1 (and related replicons) is now well understood. Copy control relies on two RNA molecules, one of which is the initiator of replication (primer-RNA) and the other a repressor of replication (RNA-1). The two RNA molecules interact during a temporally restricted period in the transcription of the initiator. One of the more complex aspects of copy control in this replicon is the result of the annealing of the two RNA molecules which may be aided by the presence of a small protein. If the two molecules anneal, then the formation of a suitable DNA/RNA
hybrid recognised by RNase H1 at the origin does not occur. However, if annealing does not take place, the hybrid forms and the initiation of replication by RNase H1 cleavage of primer-RNA occurs.

Both the primer-RNA and RNA-1 are expressed at the same rates. However, RNA-1 is a very unstable molecule and has a shorter half-life than does primer-RNA. Random fluctuations from the normal copy number are detected as a result of RNA-1 degradation leading to a situation where not every primer-RNA is repressed by a RNA-1 molecule. This imbalance allows replication to occur and re-adjusts the copy number to the correct value. In situations where the copy number may vary dramatically from the normal (such as after conjugation or transformation), copy number re-adjustment requires a slightly different explanation. Immediately after conjugal transfer or transformation, primer-RNA and RNA-1 are expressed from the replicon in equal numbers. However, RNA-1 rapidly diffuses into the cell whereas primer-RNA remains linked to the replicon DNA. As a result of this (and the relative half-lives of the RNAs, plus the temporal requirement for RNA-1 repression), primer-RNA 'sees' a lower concentration of RNA-1 in the vicinity of the replicon, and is able to initiate replication and establish the correct copy number in the cell.

ColE1 copy control utilises changes in RNA or RNA/DNA secondary structures both to regulate and to initiate DNA replication. Future research needs to investigate the physical mechanism defining the 'window of susceptibility' in which RNA hybridisation can occur leading to repression, and how secondary structure changes in an incomplete primer-RNA molecule can affect subsequent secondary structures formed with the origin DNA. Both of these questions can only be answered by an investigation of the physical mechanisms, rather than genetic mechanisms, of RNA and RNA/DNA hybridisation.

A6.2 Replication control of mini-R1

A6.2.1 Overview
The control of mini-R1 plasmid replication is fairly well understood and serves as an example for a class of related plasmids which include R100 (NR1) and R6-5. Mini-R1 is also an example of an IncFII group replicon, often referred to as 'RepFIC'. Mini-R1 replication has been reviewed by Scott (1984), Nordström et al. (1984), Novick (1987) and Thomas (1988).

The copy number control system of mini-R1 exists on two levels and involves three elements: a repressor RNA molecule (CopA), a small repressor protein (CopB) and an initiator protein (RepA). The genetic organisation of these elements in mini-R1 is shown in Figure A6:3. At the first level of control, CopA will hybridise with the 5' region of the RepA mRNA preventing RepA translation and inhibiting the initiation of replication. At a low copy number, CopA/mRNA hybridisation does not occur, resulting in the expression of RepA and subsequent initiation of replication at the origin. This control mechanism appears to be active during times of minor fluctuation in plasmid copy number.

Additional reading: replicon genetics
During periods where copy number fluctuations are more severe, a second level of control appears to be dominant. In this second level of control, CopB represses the transcription of the RepA mRNA from the relatively strong repA promoter. This repression will lead to a decrease in the expression of RepA and hence lower the copy number of the plasmid.

A6.2.2 Identification of the repressors
Gustafsson and Nordström (1980) predicted that mini-R1 replication was under the control of a repressor molecule. In actual fact, mini-R1 control relies on the action of two repressor molecules referred to as CopA and CopB. The initial investigations of these repressors were confused as both CopA and CopB are coded by similar-sized DNA fragments resulting from digestion with PstI, and the activities of CopA and CopB are quite different.

CopA was mapped to the PstI F1 fragment and was considered to be either a 7kDa protein, a small RNA molecule or a small cis-acting DNA sequence. The repressor was identified as an untranslatable 80 nucleotide RNA molecule by Stougaard et al. (1981) who also demonstrated that CopA had a high degree of secondary structure and a half life of only a few minutes. Subsequently, Stougaard et al. (1982) identified a 200n form of CopA. This larger molecule did not share the repressor activity of the 80n CopA, and the switch between the two forms was shown to be the result of convergent transcription from copBp and copAp. Convergent transcription from these two promoters interferes with, or abolishes normal CopA transcription termination, resulting in a larger, inactive transcript. The secondary structure of CopA was investigated using both single-strand and double-strand
specific ribonucleases and revealed that CopA formed two stem-loop structures separated by a long single-stranded spacer and had a long single-stranded 3' tail (Wagner et al., 1986).

Molin et al. (1981) isolated a series of copy mutants which appeared to affect a repressor element with characteristics different to those of CopA. A lacZ gene fusion was used to demonstrate that this second repressor was a small protein of 11kDa (CopB) with the gene orientation the same as that of RepA. CopB is located on a small PstI (PstI F2) fragment with the same electrophoretic mobility as the fragment containing CopA (Burger et al., 1981). Light and Molin (1982a) showed that CopA and CopB act independently of one another; the target of CopB repression is a small 60bp region containing the repA promoter sequence (repA<sub>p</sub>), whilst CopA acted on a sequence in the 5' region of the repA mRNA.

CopB has been partially purified and used for footprint analysis to demonstrate in vitro moderately strong binding to its target site (Riise and Molin, 1986). The expression of CopB is independent of any plasmid-coded functions, is gene-dosage dependent and expressed from an efficient promoter, copB<sub>p</sub> (Light and Molin, 1982). Both CopB and RepA are translated from the same transcript originating from copB<sub>p</sub>.

A6.2.3 Level one: CopA/RepA mRNA interaction
If the copy number of mini-R1 is high, CopA will bind to CopT and as a result, the expression of RepA will be inhibited. CopA recognises and interacts with the 5' region (CopT) of the large RepA mRNA transcript originating from copB<sub>p</sub>. Persson et al. (1990a) has demonstrated that CopA/CopT interaction forms an RNA/RNA hybrid, and that the reaction involves two distinct steps. In the first step, a transient complex forms between the second stem-loop structure of CopA and its complement in CopT (the 'kissing' form). In the second step, the formation of the hybrid requires a stretch of single-stranded RNA 5' to the second stem-loop structure. The effects of salt concentration and temperature indicate that the CopA/CopT hybrid formation occurs through a mechanism of gradual intra-strand breaking and inter-strand formation of hydrogen bonds. The hybridisation constant for CopA/CopT is 1000-fold greater than for the kissing complex between the second stem-loop structure and CopT (Persson et al., 1990b). Hence, the formation of the kissing complex is rate-limiting for the interaction between CopA and CopT. RepA expression is normally inhibited by the formation of the CopA/CopT hybrid despite the fact that the formation of the 'kissing complex' alone appears sufficient to inhibit RepA translation (Wagner et al., 1992).

The CopA/CopT complex appears to affect the expression of RepA in two ways. In the first, the complex is cleaved by ribonuclease III (RNase III) to prevent translation (Blomberg et al., 1990). This has been demonstrated by the similarity of CopA/CopT cleavage patterns in vivo with the cleavage patterns observed in vitro after the addition of RNase III. The ribonuclease cleavage site is upstream of the repA initiation site, suggesting that cleavage changes the secondary structure of the mRNA in such a manner that ribosomes are no longer able to recognise the translation start codon of repA.

Additional reading: replicon genetics
This may involve the formation of a stem-loop structure downstream of the CopT region in the mRNA. In such a structure, the repA ribosomal binding site would be located on the stem where perhaps it is not recognised by the ribosomes (Rownd et al., 1985).

CopA/CopT complex formation also affects RepA expression in a second manner. Wagner et al. (1987) demonstrated that a small 7kDa open reading frame upstream of repA could be used to express a lacZ fusion protein. A translational start mutant resulting in a 10-fold decrease in the expression of RepA could not be reversed by supplying the 7kDa protein in trans. This observation (and others) suggests that the 7kDa protein itself is not required or involved in the regulation of mini-R1 replication. Wagner et al. (1987) suggest however, that the translation of the 7kDa open reading frame transiently alters the secondary structure of the leader mRNA sequence (including CopT) and thus alters the CopA/CopT interaction.

A6.2.4 Level two : CopB rescue by de-repression of repA<sub>p</sub>
CopB concentration is the primary copy control factor influencing mini-R1 replication at near-normal plasmid copy numbers (level-two control). Plasmids such as R1 and other IncFII plasmids have a special need for the ability to rapidly regain copy number because of their inherent low copy number makes down-fluctuations disastrous. This ability is provided by the transcription of RepA from two promoters, one of which is repressed by CopB at normal plasmid copy numbers. This promoter (repA<sub>p</sub>) is only de-repressed when the copy number falls below the normal level and allows the rapid expression of RepA required to regain a normal plasmid copy number (Light and Molin, 1982a).

A6.2.5 RepA

<i>De novo</i> protein synthesis is required for both <i>in vivo</i> and <i>in vitro</i> mini-R1 replication (Goebel, 1979; and Uhlin and Nordström, 1978). Mini-R1 replication depends upon an intact repA gene (Kolleck et al., 1978) which was shown to express a 33kDa protein detected by Brauner and Jaskunas (1982), and confirmed by Masai et al. (1983). The expression of RepA was also demonstrated by use of lacZ fusions (Light and Molin, 1981). RepA is expressed from one of two transcripts: the larger transcript codes for both the CopB repressor and RepA, whilst the shorter transcript initiates from an internal transcription site (repA<sub>p</sub>) and only codes for RepA (Light and Molin, 1981; and Light et al., 1985).

The intergenic region between copA and repA in the larger transcript has the potential to form a secondary structure which may be responsible for the low translation rate of RepA (Öhman and Wagner, 1991). Mutations in this region destabilising the secondary structure have higher copy numbers corresponding to greater levels of RepA expression.

A6.2.6 Origin of replication
The origin of replication for mini-R1 is located downstream of repA and contains many symmetrical DNA sequences allowing the formation of quite complex secondary structures. Burger (1983) has shown that a deletion of 218bp from this region does not affect replication, and that a 14kDa open
reading frame located in the origin region was not required for replication. Masai et al. (1983) defined the origin more closely using an in vitro replication system, and demonstrated that the minimal origin lay between 1424-1611 bp coordinates.

A6.2.7 Summary of mini-R1 replication control

The copy control mechanism used by mini-R1 is only partially understood and there appear to be two levels of copy control. The normal system relies on the interaction of two RNA molecules in a manner very similar to that of ColE1. However, in the case of mini-R1, the initiator RNA (RepA mRNA) encodes a protein whose interaction with the origin DNA is responsible for the initiation of replication. During periods in which there is a large fluctuation of plasmid copy number within the cell, the second level of control operates. Control at this level is an example of autoregulation of a polycistronic mRNA by a small repressor protein (CopB).

Although the physical alterations of RepA mRNA secondary structure caused by the interaction of CopA are now known, there is some evidence to suggest that the interaction between the two RNA molecules may be affected by the translation of a small ORF encoded by RepA mRNA. This point should be investigated in any future mini-R1 research. In addition, the interaction of RepA with the origin should also be investigated, as the physical mechanism of initiation of DNA synthesis is unknown.

A6.3 Replication control of pT181

A6.3.1 Overview

Many of the plasmid replicons which have been examined extensively have been isolated from Escherichia coli. A notable exception is pT181 which is a 4.4 kb tetracycline-resistant plasmid isolated from Staphylococcus aureus. The full sequence of pT181 is known, and the control of replication is well understood and is an example of the single-stranded or rolling circle replication (ssDNA or RCR) system found in a number of other plasmids. pT181 replication has been reviewed by Scott (1984), Novick (1987), Thomas (1988) and by Gruss and Ehrlich (1989).

The control of replication involves two elements: a repressor RNA molecule (CopA) and an initiator protein referred to as RepC. The genetic organisation of these elements in pT181 is shown in Figure A6:4. Copy control relies on the variation in concentration of Cop and interaction of Cop with the complementary sequence in the 5’ region of the RepC mRNA. High Cop concentrations results in the hybridisation of CopA and RepC mRNA. This hybridisation will prevent the translation of the RepC mRNA and subsequently, prevent the initiation of replication. If the plasmid copy number is low, the two RNA molecules will not hybridise, RepC will be expressed and replication will be initiated at the origin.
A spontaneous deletion mutant has been isolated resulting in a 50-fold increase in plasmid copy number. An analysis of the mutant has identified a 180bp deletion directly upstream of the RepC coding region, and has shown that this deletion renders the mutant compatible with the wild-type pT181 (Novick et al., 1984). This effect is probably due to the over-expression of the initiator gene, repC, and suggests that the 180bp region codes for a repressor (CopA) which might act on the expression of RepC. On the basis of this result, Novick et al. (1984) have suggested that RepC synthesis is negatively controlled by a repressor that is encoded upstream of RepC, and that this repressor acts on a target located in the same area as the repressor's coding region. Kumar and Novick (1985) have demonstrated that the repressor is, in fact, an untranslated RNA which is present in two forms; CopA-I (RNA-I), 80n; and CopA-II (RNA-II), 150n. Both CopA RNAs are transcribed from a single constitutive promoter (P_{I,II}; Kumar and Novick, 1985) and both RNA molecules are involved in the control of pT181 copy number.

![Figure A6:4](https://example.com/figure.png)

**FIGURE A6:4** Organisation of the control elements in pT181

Copy number control in pT181 involves a repressor RNA molecule (CopA) and an initiator protein (RepC). Both CopA and the 5' region of the RepC mRNAs have complex secondary structures which are pivotal in regulation. The origin of replication is located between the Cop and RepC coding regions. Key: P_{I,II} - CopA promoter (○). RepC mRNA promoters, P_{III} and P_{IV} (●). The coding region of repC extends to the right of the diagram. The primary region of interaction between the Cop RNAs and the RepC mRNAs is shown (▼). The direction of DNA synthesis at the origin is indicated by the arrow. Base pair coordinates are marked according to Novick (1987).

In *Staphylococcus aureus* strains carrying the plaC1 mutation, transcription from the P_{I,II} promoters is significantly reduced and leads to an elevated pT181 copy number. The plaC gene encodes a minor species of RNA polymerase σ (sigma) factor and appears to direct RNA polymerase transcription in a restricted class of promoter sequences (which do not include the P_{III} and P_{IV} promoters; Basheer and Iordanescu, 1991).
A6.3.3 RepC

Replication of pT181 in cell free extracts required a plasmid-encoded protein (Khan et al., 1981). This protein was mapped using a number of non-replicating pT181 mutants each of which were found to have mutations in the same cistron (repC). A number of deletions formed in vitro which were unable to replicate were used to further define the repC coding region which expressed the RepC 37.5kDa protein (Novick et al., 1982; and Khan and Novick, 1983). RepC is translated from either of two RepC mRNA transcripts (RNA-III and IV) originating from two weak promoters, PrIII and PrIV, respectively (Kumar and Novick, 1985).

RepC has both sequence-specific endonuclease and topoisomerase activities (Koepsel et al., 1985b). RepC introduces a single-strand nick within the origin region between nucleotides 70 and 71 on the bottom strand and RepC remains attached to the 5' end of the nicked strand (Koepsel et al., 1985). This attachment has been shown to involve a link between the 5' phosphate of the cut strand and a tyrosine residue in RepD, a homologue of RepC from the closely related plasmid pC221 (Thomas et al., 1990). Six residues in the C-terminal region of RepC 80 residues away from the active site tyrosine are responsible for the origin recognition specificity (Wang et al., 1992). In addition to nicking, RepC also has a ligase activity and is able to close nicks, although it is likely that the nicking activity is more efficient that the closing activity. This activity is pT181-specific and requires the presence of an intact pT181 origin of replication. The nucleotide sequence around the nick site can potentially form a hairpin structure with the nick site occurring within the single-stranded loop region. It is possible that such a transient hairpin structure may exist as a recognition signal for RepC nicking.

A6.3.4 CopA interaction with RepC

Carleton et al. (1984) have isolated dominant CopA mutants which have defects in the CopA target. Most of the mutations occurred within a very short GC-rich segment upstream of the RepC coding region. This region is single stranded in the computer-generated secondary structures of CopA and of incompletely transcribed RepC mRNA. Once the 200-180 region of repC has been transcribed, the optimal folding pattern changes and the target region (on the mRNA) is now partially double-stranded (Kumar and Novick, 1985). All of the dominant mutants examined by Carleton et al. (1984) contained base changes which would have prevented the formation of the partial double-stranded form of the target region.

Subsequent analysis has shown that the presence of CopA promotes the formation of a termination-causing hairpin 5' to the RepC translation initiation codon (sequence elements III and IV, forming the III-IV stem). In the absence of CopA, the target region pairs with the proximal arm of the terminator stem preventing the formation of the hairpin and allowing the translation of RepC (Novick et al., 1989). The hairpin formed as a result of CopA/mRNA hybridisation sequesters the repC Shine and Dalgarno (SD) sequence in such a manner as to block translation of the gene. In addition, the hairpin is followed by a thymidine stretch which may act as a transcription terminator preventing RNA polymerase from producing a full-length mRNA.

Additional reading: replicon genetics
When CopA is not present, the preemptor (sequence element I) which includes the SD sequence basepairs with element III allowing a functional mRNA to form (Novick et al., 1985). Novick et al. (1989) have shown that the III-IV hairpin is a very efficient terminator of transcription and suggest that the pT181 replication control system is functionally similar to the classical attenuators. However, in the case of pT181, attenuation is induced by CopA hybridisation rather than by tRNA-induced stalling of ribosomes.

A6.3.5 DNA synthesis

The 170bp origin of pT181 can form complex secondary structures. The origin contains a set of inverted repeats capable of forming a strong cruciform (hairpin) structure which has been detected in vivo and in vitro (Noirot et al., 1990). The formation of the cruciform appears to unwind the GC-rich origin region, and RepC binding greatly enhances the formation of the cruciform. Noirot et al. have suggested that RepC cruciform extrusion is involved in the initiation of pT181 replication.

Cruciform extrusion is dependant upon the superhelix density of the plasmid and at normal superhelical densities, the cruciform rarely forms at all in the absence of RepA. However, localised changes to the superhelical density induced by transcriptional activity may increase the formation of DNA cruciforms. Projan and Novick (1992) have suggested that RepC binds to double-stranded and cruciform origin DNA. When a cruciform formation is available, RepC binding leads to the initiation of replication. However, if the cruciform is absent, RepC binding will lead to a non-productive nick-closing event which does not lead to the initiation of DNA synthesis.

RepC binds to a 32bp sequence corresponding to the cruciform repeats (Koepsel et al., 1986), and Khan et al. (1982) have demonstrated that RepC nicks the bottom strand of the plasmid DNA within the origin region to initiate replication. Koepsel et al. (1985b) have suggested that free 3'-OH generated by RepC nicking is used as a primer for DNA synthesis via the rolling circle replication mechanism (RCR). This is supported by the observation that replication is unidirectional and proceeds in a clockwise manner (with respect to Figure A6:4; Kahn et al., 1982; and Koepsel et al., 1985a). RepC is attached to the 5' end and may ligate the newly synthesised strand in a fashion similar to the oX174 CisA protein (gene A, Eisenberg and Kornberg, 1979). The lagging strand synthesis would begin at specific sites on the nascent leading strand in a manner similar to the replication system of RF DNA of some bacteriophages (such as oX174, fd and f1).

A6.3.6 Summary of pT181 replication control

The copy control mechanism of pT181 is well understood. Like ColE1 and mini-R1, copy control relies on the interaction of two RNA molecules. Ultimately, pT181 copy number is determined by the relative numbers of RepC mRNA and CopA RNA molecules present near the replicon. Like mini-R1, the initiator RNA is an mRNA molecule which encodes a protein which acts at the origin to initiate replication. Unlike mini-R1, the action of the initiator protein (RepC) is well understood biochemically. RepC is able to nick DNA at the origin to initiate DNA synthesis by providing a 3'-OH required for
strand elongation. The origin region appears to have the ability to form complex secondary structures which may be required in order for RepC to recognise and bind to the very specific nicking site within the origin.

The replication of pT181 is also interesting as it is a very good example of rolling circle replication (RCR). This method of replication is used by many bacteriophages and other plasmids, but differs substantially from the theta-mode replication of bacterial chromosomes and of large plasmids containing replicons such as mini-F, mini-P1 or mini-R6K.

A6.4 Replication control of mini-F (RepFIA)

A6.4.1 Overview

The control of replication from the primary replicon of the F plasmid has been studied for some time. This replicon has been isolated on an EcoRI fragment and is referred to as mini-F. The experiments used to investigate the control of mini-F replication have been reviewed by Lane (1981), Scott (1984) and Kline (1985).

The control of mini-F replication involves four elements: an initiator protein referred to as the E protein, the E gene promoter (pE) and two sets of repeat elements located on either side of the E gene. The genetic organisation of these elements in mini-F is shown in Figure A6:5. The E protein binds to a set of inverted repeats near the E gene promoter to autoregulate the expression of the gene. The E protein also binds to the set of repeat elements located downstream of the E gene (the titration repeat elements, incC) and by doing so, lowers the frequency of replication. Finally, the E protein binds to the set of repeats upstream of the E gene to initiate replication at the origin of mini-F (ori-2).

Although the interaction of the E protein with the E gene promoter and with both sets of repeat elements have been investigated, it is not yet obvious how these interactions combine to control mini-F replication. The observation that the E protein is autoregulated, acts as a titrated repressor of replication as well as acting as the initiator of replication has posed a serious problem for a number of models proposed to explain mini-F replication.

A6.4.2 Mini-F

The F plasmid has two functional replicons, referred to as RepFIA and RepFIB, as well as a third replication region which has apparently been rendered non-functional by the presence of a Tn1000 transposon (Saadi et al., 1987). The RepFIA replicon is referred to extensively in the literature as 'mini-F', and the 9kbp EcoRI f5 fragment carrying the replicon was isolated as an autonomously replicating plasmid by Timmis et al. (1975) and Lovett and Helinski (1976). Mini-F contains two origins of replication referred to as ori-1 and ori-2 (oriS) although ori-1 can be deleted without interfering with mini-F replication (Eichenlaub et al., 1977; and Manis and Kline, 1978). Restriction
enzyme digestion of mini-F and the use of transposon mutagenesis demonstrated that mini-F replication requires only ≈ 2.7 kbp of DNA for replication, and that this region contained ori-2 and two different incompatibility loci (Figurski et al., 1978; Kahn et al., 1979; Bergquist et al., 1981; and Seelke et al., 1982). Kline and Lane (1980) have named the two inc loci incB and incC.

**Figure A6:5**

*Organisation of the control elements in mini-F*

Copy control in mini-F involves an autoregulated gene (E) and two sets of repeat elements, incB and incC. The E protein is known to bind to both incB and incC, as well as to a small inverted repeat located near the E gene promoter (PE). The incB repeat elements are part of the *ori*-2 (oriS) origin of replication (boxed). F kilobase pair coordinates are indicated.

The central region required for mini-F replication was sequenced by Murotsu et al. (1981). The sequence revealed the presence of nine 19 bp repeat elements arranged in a group of four elements all in one orientation, and a second group of five repeats all arranged in the opposite orientation. The spacings between elements in each group suggest that conserved sequences in the elements would all face the same side of the duplex DNA molecule. The two sets of repeat elements correspond to the two incompatibility loci mapped in mini-F: the four repeat element maps to incB, and the five repeat element maps to incC.

IncB and incC are separated by an open reading frame which can express a 29kDa protein, of which small amounts could be detected in vivo. Finally, incB is very close to the map position of the *ori*-2 origin of replication (Murotsu et al., 1981).

### A6.4.3 The E gene

Tolun and Helinski (1982) separated mini-F into two complementary fragments; an origin-containing fragment and a segment which functions *in trans* to support the replication of the origin fragment. The trans-acting fragment contains the coding region for the 29kDa gene referred to as the *E* gene (which expresses the E protein). The *E* gene is transcription is right-ward (with respect to Figure A6:5) and has been identified as a 29kDa protein in minicells (Wehlmann and Eichenlaub, 1980; Komai et al.,
1982; and Watson et al., 1982). The E protein is essential for mini-F replication, as amber mutations or Tn5 insertions into the E gene abolish the ability of mini-F to replicate (Watson et al., 1982; Maki et al., 1983, and 1984). In addition, Rokeach et al. (1985) have shown that an excess of E protein in trans increases the replication frequency of F plasmid.

The E protein has been over-expressed, purified and used in in vivo demonstrations to show that E protein binds to three unique sequences in mini-F (Tokino et al., 1986; Masson and Ray, 1986, and 1988). These regions correspond to the incB and incC repeat elements and to a small inverted repeat located between incB and the beginning of the E gene. All three sites have a 8bp sequence in common (which is part of the 19bp consensus sequence), which appears sufficient to bind E protein.

A6.4.4 Autoregulation of the E gene
Sogaard-Andersen et al. (1984) used an E gene:β-galactosidase fusion to locate the E gene promoter (pe). The promoter lies just downstream of the inverted repeat and it appears that when E protein is bound to the inverted repeat, pe is repressed (Sogaard-Andersen et al., 1984; and Rokeach et al., 1985). Initial attempts to demonstrate that the E protein was solely responsible for the autoregulation of pe proved difficult as the same fragment of DNA which contains the E gene also includes three other open reading frames, one of which might express a protein involved in repression. The three open reading frames code for proteins of approximately 9kDa, and have been called 9kDa-1, 9kDa-2 and 9kDa-3. Of these, 9kDa-3 has a Shine and Dalgarno sequence and active promoter. Bex et al. (1986) demonstrated that 9kDa-3 expresses a protein using a β-galactosidase fusion, but that the activity of the promoter was 1% of the unrepressed pe. An analysis of E gene mutants where the mutations affect only the E protein and not potential 9kDa proteins has shown that pe repression requires only the E protein (Bex et al., 1986) and that the E gene is autoregulated.

A6.4.5 RNA polymerase-σ32 transcription of pe
Mini-F cannot replicate in rpoH mutant strains and pe has a sequence more similar to the heat shock gene promoters than to the normal E. coli promoter consensus sequence (Cowing et al., 1985; Wada et al., 1986; and Kawasaki et al., 1990). These two observations have lead to the suggestion that transcription from pe is mediated by RNA polymerase containing σ32, a relatively minor species of RNA polymerase σ (sigma) factor previously only implicated in the transcription of heat shock genes.

However, Kawasaki et al. (1991) have isolated a number of mini-F mutants from rpoH strains, each of which overexpress a mutant E protein with a single residue alteration from the wild-type sequence. The mutant proteins have reduced repressor activities and an enhanced ability to initiate replication. These results suggest that in the absence of σ32, pe is recognised by RNA Pol-σ70 (the normal RNA polymerase), and the reduction in E gene transcription is overcome by the altered E protein activity. Finally, Kawasaki et al. have noted the presence of a weak promoter transcribed by RNA Pol-σ70 upstream of pe. This promoter is also repressed by E protein, and transcription from this promoter becomes more pronounced in the absence of σ32.

Additional reading: replicon genetics
A6.4.6  E protein activities
Tokino et al. (1986) suggested that the interaction between the E protein and \( p_E \), \( incB \) and \( incC \) regulates both the expression of the \( E \) gene as well as regulating plasmid copy number and the initiation of replication. Tolun and Helinski (1981) have shown that E protein will bind to as few as two of the 19bp repeats found in \( incC \), and that two repeats are sufficient to cause an incompatibility reaction with mini-F. Tsutsui et al. (1983) demonstrated that both cloned \( incB \) and \( incC \) fragments cause incompatibility reactions with \( F \), and that the strength of the reaction is proportional to the dosage of the repeats in the cell. In addition, an excess of E protein will cause an increase in the frequency of \( F \) plasmid replication which can be lowered by an increase in the dosage of either \( incB \) or \( incC \) (Rokeach et al., 1985).

The E protein appears to bind at all three sites in a similar manner. However, when considered individually, E protein binding has different consequences with regards to the control of copy number. Since the E protein is the initiator of replication from \( ori-2 \), autoregulation at \( p_E \) will maintain the level of E protein at a steady state. Initiation of replication is achieved by the binding of E protein at \( incB \) which is located on the right-hand border of \( ori-2 \). Disquile-Kochem et al. (1986) demonstrated that the \( incB \) repeats are required for replication, and as a result, that \( incB \) should be considered as an integral part of \( ori-2 \).

The combination of autoregulation and origin activation by E protein would appear to be sufficient to maintain a steady plasmid copy number. However, the \( incC \) repeat elements also bind E protein, and binding at \( incC \) can be viewed as a second means of limiting the amount of E protein free to bind at \( incB \) and initiate replication. This second mechanism of control, often referred to as titration, has lead to a major paradox concerning mini-F regulation: how can autoregulation and titration occur simultaneously?

A6.4.7  A brief summary of the mini-F replication control models
A number of models have been proposed in an attempt to solve the paradox caused by the apparent concurrent autoregulation and titration of the E protein (Figure A6:6). Any model for the control system of mini-F must include the following elements: the E protein as the rate-limiting initiation factor for replication; autoregulation of the \( E \) gene; and E protein binding to \( incB \) and \( incC \). It is interesting to note that although mini-F and replicons such as mini-P1 and R6K do not share sequence homology (except for limited regions in the origin of replication), these replicons appear to face the same paradox expressed by mini-F. A number of groups have been investigating these replicons, and the models presented below can apply to one replicon as well as to another.

A6.4.8  Hierarchical binding
It is possible that the paradox could be resolved by suggesting that the E protein recognises and binds to various regions of mini-F with different affinities (Chattoraj et al., 1984). In this model, the E protein would first bind to the inverted repeat in order to autoregulate the \( E \) gene (Figure A6:6, A).
However, binding at this site would not completely repress the synthesis of E protein and additional E protein would bind to the incC repeats. Once the incC sites were filled, E protein would then bind to the incB repeats and initiate a round of replication. The cellular concentration of E protein would slowly fall due to natural decay during growth and be further diluted by cell division. After division, the concentration of E protein will be low enough to de-repress pE and the cycle would begin again (this system is described by the Clepsydra analogy proposed by Chattoraj et al., 1985a; see Chapter Eight, 8.5). No firm evidence of hierarchical binding has been found, despite initial experiments which have shown that E protein binding to the inverted repeat occurs at a lower protein concentration than for the incB repeats (Masson and Ray, 1986).

A6.4.9  Post-translational modification (the two-stage model)
Trawick and Kline (1985) have suggested that the E protein exists in two forms: i) a transitory repressor form (Ei); and ii) a more stable initiator form (Ei). In this model the nascent E protein is capable of repression of pE (Figure A6.6, B). Some of the Ef protein not bound to the inverted repeats would be modified by host enzymes into a more stable form with initiator activity. The modified E protein would first bind to the incC repeats and then to the incB repeats to initiate replication. Ei protein would not be able to bind to the inverted repeats located upstream of pE; hence Ei protein would not be capable of autoregulation. After cell division, the concentration of Ef protein would be reduced and the process would begin again.

Although Trawick and Kline (1985) did not specifically proposed a mechanism of post-translation modification, they suggested that dimerisation of the E protein may be a sufficient modification to alter the E protein activity. Since the majority of DNA binding proteins are either dimers or monomers, Rokeach et al. (1985) have modified Trawick and Kline's proposal from a monomer/dimeric modification to a dimer/tetrameric modification.

At low E protein concentrations (low F copy number) a dimeric form of the E protein would be predominant which is capable of initiating replication; at higher concentrations (higher F copy numbers) a tetramer would be predominant which is incapable of initiation, but which would autoregulate the E gene leading to lower protein concentrations and ultimately, lower F copy numbers. The shift between these two forms would take place between a very narrow range of concentrations and would lead to the very strict control of copy number observed by Tsutsui and Matsubara (1981). However, a dimeric/tetrameric modification may not occur, as it is now known that the E protein is present as a dimer in solution, and that higher order polymers have not been observed (Masson and Ray, 1988).

A number of other modifications, including phosphorylation, endopeptidase cleavage or limited exopeptidase digestion, or the association of other cellular components could be included in this model. Womble and Rownd (1987) have tested the Two-stage model in a quantitative manner by the use of computer simulations. By proposing that the transition from Ei to Ef is irreversible, and by making a number of assumptions with regards to binding kinetics, RNA polymerase affinity for pE and the
availability of other host factors, Womble and Rownd have shown, theoretically at least, that the Two-stage model can account for the observed mini-F copy number control mechanism.

Recent work has demonstrated that the E protein can be isolated in two forms: Δ1-E, corresponding the to full-length protein (with fMet removed); and Δ17-E, where the next sixteen residues have been removed by some specific post-translation proteolysis mechanism (Kline et al., 1992). In vivo and in vitro experiments have shown that although Δ1-E and Δ17-E are both able to act as incompatibility factors, autorepress pE and bind mini-F DNA, Δ17-E has lost the ability to initiate replication from ori-2. Proteolysis of Δ1-E might therefore alter an initiator form of the E protein into a form engaged solely in repression. However, it is not yet known whether a specific Δ1-E → Δ17-E transition occurs or whether both forms of the E protein appear immediately after translation, nor is it known whether the two different E protein forms are responsible for mini-F copy number control.

A6.4.10 9kDa repressor action

It has been suggested that the E protein is not in fact autoregulatory, and that one of the 9kDa open reading frames located on the same segment of mini-F which carries the E gene is responsible for the repression of pE (Figure A6:6, C). However, although 9kDa-3 expresses a protein, Bex et al. (1986) have demonstrated that the autoregulation of pE is effected by the E protein, and that the autoregulatory region is located in the carboxyl terminus of the E protein. There is currently no evidence to suggest that either 9kDa-1 or 9kDa-2 express proteins, or that any 9kDa gene product is involved in mini-F copy control mechanism.

A6.4.11 Steric hindrance

As a result of a lack of evidence to support either of the three models suggested by Trawick and Kline, two additional models proposed for other replicons have been adapted to mini-F in order to solve the mini-F control paradox (Figure A6:6, D and E; see Chapter One, 1.2.7). In vitro binding studies using purified E protein and mini-F DNA have shown that as a result of E protein binding to both incB and incC, mini-F DNA will bend to form a loop joined at incB and incC by the bound E protein.

The first of the adapted models suggests that in the looped state, origin function is repressed by the physical presence of incC and bound E protein (Steric Hindrance; proposed by McEachern et al., 1989, for R6K). In addition, pE will be repressed and E protein expression will be reduced. The looped mini-F form is in equilibrium with a linear form which is capable of replication. As in the original models, pE repression is not complete. The gradual increase in E protein concentration will favour the formation of linear forms and will ultimately result in a round of replication. However, if the copy number of mini-F is too high, the majority of mini-F copies will be maintained in stable inter-molecular paired forms rather than in linear or looped forms. Like the looped forms, inter-molecular mini-F pairs are unable to replicate.
Additional reading: replicon genetics
Solutions to the mini-F paradox

Opposite page A number of models have been suggested to account for the apparent paradox generated by the observation of E gene autoregulation and E protein titration at ori-2 and incC (see Figure A6:5 for a description of the control elements of mini-F):

A Hierarchical Binding. In this model, the E protein has different affinities for pE, incB (ori-2) and incC: The E protein will first bind to pE (I), then to incC (II) and finally to incB (III).

B Post-translational Modification ('Two-Stage' model). In this model the nascent E protein represses the expression of the E gene at pE. The E protein (E-repressor) is then modified in some unspecified manner to give the initiator form (Ei) which then interacts with incB and incC.

C 9kDa Repressor. The segment containing the E gene also contains three other open reading frames, each of which could express a 9kDa protein (9kDa-1, 2 and 3). In this model, one of these open reading frames is expressed and the 9kDa protein (9kDa-2 in this figure) is responsible for repression of the E gene at pE, whilst the E protein interacts with incB and incC.

D Steric Hindrance. In this model, the replicon is in dynamic equilibrium between linear and loop structures. The loop is formed by E protein 'bridging' between the incB and incC repeat elements. The bridge may result from E protein-dimer interactions with the repeats, or by dimer-dimer interactions where each dimer is bound to one set of repeat elements. The replicon is able to replicate when it is in the linear state. However, with increasing concentrations of E protein, the looped form is favoured over the linear form and replication is prevented.

E At higher copy numbers, inter-replicon pairing (II) occurs rather than intra-replicon circularisation or looping (I).

The second model, based on the Anti-parallel Pairing model proposed by Abeles and Austin (1991) for mini-P1, maintains that inter-molecular mini-F pairs form immediately after a round of replication. Inter-molecular pairing prevents further replication and the pairs are very stable. Replication only occurs after cell partition physically separates the mini-F copies. These two new models are as yet untested, both for the original replicons and, in their adapted forms, for mini-F1.

A6.4.12 Summary of mini-F replication control

Although the copy control mechanism of mini-F has been studied nearly as long as the control mechanism of ColE1 and mini-R1, the mini-F control mechanism remains unknown. Two decades of research has only managed to dissect the elements of the control system that involve autoregulation and titration of an initiator protein. Early control models proposed for mini-F were unable to avoid the paradox resulting from concurrent autoregulation and titration, and current models attempting to solve the problem remain, as yet, untested.

1 These two models are discussed in Chapter One, 1.2.7-1.2.10.
Replication control of mini-P1

A6.5.1 Overview
Bacteriophage P1 is a temperate phage frequently used in transduction experiments with *E. coli*. P1 does not integrate into the host chromosome as does λ phage and is able to replicate autonomously within the bacterium as a plasmid. The replicon responsible for plasmid replication (vegetative replication) is referred to as mini-P1; this replicon has been well studied and has been reviewed by Scott (1984). The control of mini-P1 replication is similar to that of mini-F; it involves four elements: an initiator protein referred to as RepA, the repA gene promoter (repA_P) and two sets of repeat elements located on either side of the repA gene. Unlike the E gene promoter of mini-F, the repA promoter in mini-P1 is located within the *incC* set of repeat elements upstream of the repA gene.

The genetic organisation of these elements in mini-P1 is shown in Figure A6:7. RepA protein binds to the *incC* repeat elements to autoregulate the expression of repA. RepA also binds to the *incC* and *incA* repeats respectively to initiate replication and to regulate copy number. Although the interaction of RepA with both *incC* and *incA* repeats are well characterised it is not yet obvious how these interactions combine to control mini-P1 replication. As in the case of the E protein activity in mini-F, RepA activity in mini-P1 replication has posed a serious problem for a number of control models proposed to explain mini-P1 replication.

A6.5.2 Isolation of mini-P1
The mini-P1 replicon is located on two adjacent EcoRI fragments, EcoRI5 and part of EcoRI8 located to the right of EcoRI5 (Prentki *et al.*, 1977; Austin *et al.*, 1982; Sternberg and Austin, 1983). Early examination of the replicon was based on a plasmid recombinant referred to as λ-P1:5R (mini-P1). λ-P1:5R contains approximately 11kb of P1 DNA and the plasmid appeared to replicate with the same efficiency and low copy number as P1 (Sternberg and Austin, 1983). In addition, λ-P1:5R and P1 are incompatible. The incompatibility elements were located by deletion analysis of a λ-P1:5R derivative and Austin *et al.* (1982) determined that the derivative contained two separate incompatibility loci, *incA* and *incB*. Either of these elements are capable of causing an incompatibility reaction with P1 or mini-P1 (Austin and Abeles, 1983; and Abeles *et al.*, 1984), but only one (*incA*) is associated with mini-P1 replication.

A6.5.3 Sequence analysis of mini-P1
Deletion analysis of λ-P1:5R had shown that only ~ 3kb of the 11kb of P1 DNA present in the construction was required for replication (Austin *et al.*, 1982). The minimal region required for replication was shortened by Abeles *et al.* (1984) who analysed a further set of λ-P1:5R deletions which had been cloned into pBR322. From these, Abeles *et al.* were able to isolated several deletions which demonstrated that mini-P1 replication required just over 2kb of DNA. The DNA sequence of this region of mini-P1 was determined and extensively examined (Abeles *et al.*, 1984).
The minimal region required for mini-P1 replication includes a single gene encoding a 32kDa protein (RepA) and a series of 19bp repeat elements (A). The initiator gene (repA) is flanked on either side by two incompatibility loci, incA and incC and are comprised of a series of 19bp repeat elements (triangles). The boxed region encloses the mini-P1 origin of replication. The origin includes the incC repeat elements and recognition sites for host encoded DnaA and Dam methylase (B). DnaA binds to two dnaA boxes located at the left boundary of the origin. Between the dnaA boxes and incC are five short 7bp repeat elements and five 'GATC' sequences recognised by Dam methylase. The 7bp repeats and the incC repeats are separated by a short GC-rich sequence. The position of the repA promoter (repAp) is shown, along with the Shine and Dalgarno (SD) sequence and the start of the repA coding region (thick arrow). The scale is shown in base pairs.

The 2081bp minimal replicon contains a single large open reading frame (repA) flanked on either side by a series of repeated 19bp DNA elements (Figure A6:7). The repA gene encodes a protein with an expected mass of 32kDa. This estimate agreed with the experimental observation of a 32kDa protein by maxicell analysis (Abeles et al., 1984).

The set of repeats located downstream of repA includes five elements, each of which is in the same orientation and separated from one another by spacers of 2/3bp. The second set of repeat elements located downstream of repA contains nine elements of which three are in one orientation, and the other six in the opposite orientation (Figure A6:7). Each element is separated from the other by spacers of 11/12 or 21/22bp. Spacer lengths of 2/3, 11/12 or 21/22bp between elements result in each sequence facing the same side of the DNA helix. Abeles et al. (1984) were able to demonstrate in vivo that the downstream repeat elements corresponded to the incA incompatibility loci identified.
genetically. The upstream elements form a new loci, incC, which is capable of preventing P1 replication when present at high copy number, although incC induces a less strong incompatibility reaction than incA when present at a low copy number.

Abeles et al. also identified two other features associated closely with the incC repeat elements; i) two potential dnaA boxes near the incC repeat elements; and ii) a set of five 7bp repeat elements located between the dnaA boxes and incC. Four of these 7bp repeat elements include the ‘GATC’ sequence recognised by the host Dam methylase. There is a fifth ‘GATC’ sequence located to the right of the set of 7bp elements which is not contained within the 7bp repeat consensus sequence.

A6.5.4 RepA

The repA gene is essential for mini-P1 replication and an analysis of repA amber mutants indicated that a protein was expressed by the open reading frame (Austin et al., 1985). The RepA protein was identified by SDS-PAGE as a labelled band present in maxicell extracts of bacteria harbouring mini-P1, which was absent in extracts from a replication-deficient mini-P1 point mutation (Abeles et al., 1984). Gel filtration of RepA has demonstrated that the protein exists as a dimer in solution (Swack et al., 1987). DNase I footprinting and binding experiments have demonstrated that RepA binds to both incC and incA DNA (Chattoraj et al., 1984; Pal et al., 1986; and Abeles et al., 1989).

The incC repeats where found by primer extension to include the repA gene promoter (repAp, Chattoraj et al., 1985b). RepA is capable of repressing repAp, and appears to do so by displacing RNA polymerase which binds to incC in the region of repAp (Abeles et al., 1989). Although only \( \approx 20 \) RepA dimers are found per plasmid copy in vivo, this is sufficient to repress the repA gene promoter to 0.1% of its unrepressed activity (Swack et al., 1987).

A6.5.5 The incA repeat elements

If the incA repeats in mini-P1 are deleted, the copy number of the mini-plasmid will increase. The incA element can be reintroduced in a number of sites to return the copy number to normal levels (Pal et al., 1986). A single repeat element is sufficient to alter the copy number of a mini-P1A incA plasmid, and a single repeat is able to bind RepA in vitro. However, deletion of the 19bp repeat reduces its affect on plasmid copy number and ability to bind RepA (Pal et al., 1986). The incA repeat elements are also able to destabilise mini-P1 replication in trans, and Chattoraj et al. (1984) have shown that only three of the incA repeats are required for destabilisation. The ability of incA to bind RepA in vitro, coupled with the ability of incA repeats in trans to destabilise mini-P1 replication lead Chattoraj et al. (1984) to suggest that mini-P1 copy control might be via the titration or sequestration of RepA by the incA repeats. While the repeats were not saturated by RepA, there would be no free RepA to bind the ori repeats (incC) to initiate replication.

However, Pal and Chattoraj (1988) demonstrated that titration can not adequately explain the copy control mechanism of mini-P1. By increasing the concentration of RepA in trans, the copy number of a
plasmid containing the mini-P1 ori sequences increases. If the plasmid also contains a copy of the incA repeat elements, the plasmid copy number remains stable in direct contradiction of the titration model. This observation lead Pal and Chattoraj to suggest that the current RepA sequestration or titration model by the incA repeats was unsatisfactory in explaining mini-P1 copy number control (the current mini-P1 control models are discussed in Chapter 1).

A6.5.6 Involvement of host factors in mini-P1 replication
The mini-P1 origin contains a number of sequence elements methylated by the host Dam methylase (Figure A6.7, B). Mini-P1 DNA cannot be used to transform dam- strains, and both in vivo and in vitro assays have demonstrated that Dam methylation is essential for replication (Abeles and Austin, 1987 and 1988). Mini-P1 replication also requires the DnaA for both in vivo and in vitro replication (Hansen and Yarmolinsky, 1986; and Wickner and Chattoraj, 1987). Mini-P1 replication also requires RNA polymerase, DNA gyrase and DnaB, and will replicate in vitro in a cell free system capable of supporting oriC replication (Wickner and Chattoraj, 1987).

Mini-P1 replication requires three E. coli heat shock proteins in addition to the normal complement of host factors in order to replicate. DnaJ, DnaK and GrpE are directly involved in mini-P1 replication and are essential in in vitro replication assays (Wickner, 1990). The dnaJ, dnaK and grpE genes of E. coli were first identified by the isolation of mutants that were unable to replicate bacteriophage λ DNA (Georgopoulos, 1977; Saito and Uchida, 1977; and Sunshine et al., 1977). While the roles of these three genes in E. coli cell growth are largely unknown, they are involved in growth at normal temperatures and are essential for growth at high temperatures. λ replication in vitro requires DnaJ and DnaK under all conditions, and GrpE under some (Zylic et al., 1989). oriC replication also requires DnaJ and DnaK in in vitro replication assays (Sakakibara, 1988; and Zylic et al., 1985). However, the involvement of these proteins in plasmid replication was unsuspected until Tilly and Yarmolinsky (1989) and Bukau and Walker (1989) found that mini-P1 was not stably maintained in dnaJ, dnaK or grpE mutant strains.

Wickner et al. (1991b) have demonstrated that DnaK and DnaJ render RepA 100-fold more active in binding to the origin of replication. The activation of RepA by these two proteins is through the conversion of RepA dimers into monomers in an ATP-dependant reaction. RepA dimers bind rapidly with DnaJ dimers to form complexes which are then recognised by DnaK-ATP. DnaK then dissociates the tetrameric complex, releasing two RepA monomers and the DnaJ dimer in an ATP-dependent reaction. Once converted to a monomer, RepA is stable at low RepA concentrations and binds with high affinity to the origin DNA. At higher concentrations, dimerisation of RepA occurs in vitro.

Dimerisation of RepA also occurs in vivo and the dimer form is likely to be involved in the regulation of mini-P1 replication. Wickner et al. have suggested that dimerisation may be a mechanism for sequestrering RepA and thus limiting replication since RepA dimers bind poorly to origin DNA (Wickner et al., 1991a).
A6.5.7 Origin of replication

Mini-P1 can be divided into two regions, a 245bp origin of replication and a 959bp fragment carrying the repA gene (Chattoraj et al., 1985). The origin lies upstream of repA (in Figure A6:7) and is the only region of mini-P1 needed for replication if RepA is provided in trans (Austin et al., 1985). Replication from the origin proceeds unidirectionally (to the left in Figure A6:7; Wickner and Chattoraj, 1987). The origin of replication contains two distinct types of sequence elements: i) elements of specific DNA sequences which bind or are recognised by a variety of proteins; and ii) elements of non-specific sequence but which have defined lengths or AT composition, which serve to separate binding domains or to alter the stability of the DNA helix.

The right hand border of the origin region is defined by the presence of the incC repeat elements which bind RepA. Between these repeats and the left hand border are a set of 5'-GATC-3' sequences which are recognised by Dam methylase. In vitro experiments have demonstrated that the methylation status of these sites affects origin replication (Abeles and Austin, 1987 and 1988). Although it is not known if hemimethylated mini-P1 DNA will replicate, mini-P1 replication is blocked when the DNA is unmethylated (Abeles and Austin, 1988). In contrast, the only blocked state for oriC is when the DNA is hemimethylated, and both methylated and non-methylated oriC DNA can replicate (Russel and Zinder, 1987). If the sites are fully methylated or if the sites are unmethylated, replication of the origin can occur. The GATC sequences are also associated with a short 7bp consensus sequence repeated five times whose length and spacer size appears to place each repeat on the same side of the DNA helix. Alteration of any of the first six bases of these repeats with a single exception have a deleterious affect on mini-P1 replication (Brendler et al., 1991b).

Two dnaA boxes which bind DnaA are located to the left of the GATC sequences (Abeles et al. 1984; DnaA activity is reviewed in Sections A6.7 and A6.8). Subsequent analysis of the origin region has shown that these two boxes form the left boundary of the origin, and that there are three more boxes sharing less homology with the dnaA box consensus sequence = 200bp to the right in the leader region of the repA gene. The host DnaA protein is required for both in vivo and in vitro replication (Hansen and Yarmolinsky, 1986; and Wickner and Chattoraj, 1987). However, of the five dnaA boxes in or near the origin region, one of the perfect left hand dnaA boxes is inactive, and only one perfect box located to the left or right of the origin core sequence is required for replication (Abeles et al., 1990; and Brendler et al., 1991a).

The origin contains a 'non-specific sequence' element which is located between the right-most GATC sequence and the left-most incC repeat element. This region is a 37bp sequence with a high GC-content and appears to be important in the physical separation of the incC repeats and other origin elements (presumably the 7bp repeats and the dnaA boxes, Brendler et al. 1991b). Mutations involving base changes in this region are silent, including some C/G → A changes which show that not all the G and C bases are critical. In contrast, single base deletions completely block origin function (Brendler et al., 1991b).
A6.5.8  The mini-P1 control paradox

A variety of experiments have proven that mini-P1 control relies on the autoregulation of RepA as well as on the titration of RepA at the incA repeat elements to regulate replication. As for mini-F, these control mechanisms prove to be mutually exclusive and have posed a paradox for all mini-P1 control models. A number of solutions to this paradox have been proposed for both mini-F and for mini-P1. In the case of mini-P1 most attention has been directed at examining models based on DNA looping and inter-molecular pairing (see Section A6.4.11), rather than the investigation of suggestions similar to those posed for mini-F (see Sections A6.4.7-A6.4.10). The latest control model for mini-P1 replication is reviewed in Chapter 1 (Section 1.2.7-1.2.10).

The activation of RepA by DnaJ and DnaK (Wickner et al., 1991a and 1991b) through the monomerisation of RepA dimers (see Section A6.2.2) appears to be similar to the post-translational modification model suggested by Trawick and Kline (Section A6.4.9). However, Wickner et al. show that whilst dimerisation may sequester RepA in a form which will not allow replication, it is the monomeric form which binds mini-P1 DNA and presumably functions in both repressor and initiator roles.

A6.5.9  The mini-Rts1 replicon

The mini-P1 and mini-Rts1 replicons are close homologues sharing considerable organisational and sequence homologies. The homologies cover the general structure of repeat elements and initiator gene, as well as the organisation of the origin of replication. The sequence of the mini-P1 and mini-Rts1 repeat elements are similar, and the RepA proteins expressed by the two plasmids are 57% identical. Because of this homologies between the two replicons, the examination of mini-Rts1 replication control and comparison with that of mini-P1 may be useful in the elucidation of mini-P1 replication control.

Rts1 is the prototype of IncT plasmids (Coetzee et al., 1972) and is known as a plasmid with a number of temperature sensitive phenotypes (Terawaki and Rownd, 1972; Ishaq and Kajii, 1980; Okawa et al., 1987; and Terawaki et al., 1968). The minimal replicon was originally isolated on a = 1.8kb EcoRI-HindIII fragment (Terawaki et al., 1976) and is also temperature sensitive in replication (Itoh et al., 1982). The sequence of mini-Rts1 was obtained (Kamio and Terawaki, 1983; and Kamio et al., 1984), including a region beyond the EcoRI site not present in the original mini-Rts1 clone (Nozue et al., 1988). The organisation of mini-Rts1 is very similar to that of mini-P1 and other step function replicons. A single large open reading frame encoding the initiator protein (RepA) is flanked by two sets of repeat elements, incl and incll. Both the nucleotide sequence of the repeats and repA are similar to the repeats and repA gene of mini-P (Abeles et al., 1984).

---

2  Mini-P1 and Rts1 RepA proteins are compared in Appendix Three, Table A3:10.
Both incl and incll express incompatibility against mini-Rts1 (Kamio and Terawaki, 1983; and Terawaki and Itoh, 1985) and the downstream repeats (incl) are involved in copy number control (Nozue et al., 1988). The incll repeats are located just upstream of the repA promoter and the expression of repA is autoregulated (Kamio et al., 1988). DNase I protection studies and binding experiments have shown that RepA binds to both sets of repeat elements (Kamio et al., 1988; Nozue et al., 1988).

Terawaki and Itoh (1985) isolated a repA mutant with a single residue change from the wild type sequence (repAcop1). RepAcop1 shows lower incompatibility to cloned incl or incll repeat elements than does RepA suggesting that the mutation lowers the binding affinity of the protein. Two additional RepA mutants have been isolated which demonstrate that the autorepressor function and incompatibility phenotype associated with RepA is not linked to the initiator activities (Terawaki et al., 1990).

Ito and Terawaki (1989) have demonstrated that mini-Rts1 replication requires the host DnaA. Mini-Rts1 has two dnaA boxes located within the origin of replication, but both boxes can be removed without destroying the replicative ability of the mini-plasmid. The deletion of the boxes reduces the copy number of the mini-plasmid. However, if the incl repeats are also deleted, the copy number remains stable in the wild type mini-Rts1 plasmid and increases in a repAcp1 mutant. Even though DnaA is required for replication, these results imply that RepA can partially replace the DnaA function and that DnaA activity does not require specific ori binding sites (Itoh and Terawaki, 1989).

A6.5.10 Summary of mini-P1 and Rts1 replicon control
Like mini-F, the copy control mechanism of mini-P1 remains unknown. Although the elements of the control system involving autoregulation and titration of an initiator protein have been demonstrated, early control models faced the same paradox as mini-F and current models attempting to solve the problem remain untested. However, mini-P1 research is more advanced than mini-F research and much more is known about the involvement of other host factors with the initiator protein (RepA), and the interaction of the origin DNA with both RepA and other factors.

The related mini-Rts1 replicon has also been investigated and it appears that the mini-Rts1 RepA protein displays the same behaviour as does RepA from mini-P1. Replicons such as mini-P1, Rts1 and mini-F share a similar arrangement of origin binding sites, DNA repeat elements and initiator genes. The copy control response to abnormally high copy numbers of these replicons is also similar, placing the replicons in the Step function class3.

However, despite these similarities in organisation and in behaviour, a comparison of the differences (in initiator protein sequence, DNA binding-site sequences, and in the relative placement of promoters,
binding sites and other origin features) between these replicons has not been useful in solving the copy control paradox faced by each replicon.

A6.6 Other plasmid replicons

A number of other plasmid replicons not described here have been studied in detail. Among them are the replicons from R6K, RK2, pSC101 and λdv. For reviews on R6K see Scott (1984) and Thomas (1988); RK2 see Kües and Stahl (1979) and Thomas (1988); pSC101 see Manen and Carol (1991); and λdv see Matsubara (1981). The R6K replicon is very similar in genetic organisation and function to the mini-F and mini-P1 replicons, despite the fact that the three replicons share very little sequence homology.

A6.7 The DnaA protein and oriC replication

A6.7.1 DnaA involvement in the initiation of oriC replication

The DnaA protein is a host factor involved in the replication of a wide variety of plasmid replicons as well as in the replication of the E. coli chromosome. The initiation of E. coli chromosomal replication begins at the origin of replication, oriC. DnaA is implicated in the early stages of the initiation of DNA synthesis at oriC and there is strong evidence to suggest that DnaA controls the timing of initiation from oriC. DnaA involvement in oriC and in a variety of plasmid replicons is discussed in the following sections.

A6.7.2 DnaA in the cell cycle

The dnaA gene product was initially identified by conditional lethal mutations that were defective at an elevated temperature in the initiation of a cycle of chromosomal replication (Hirota et al., 1970). The gene product (DnaA) was purified from a strain over-expressing the protein, and identified as a 52kDa protein (Fuller and Kornberg, 1983). The purified protein binds strongly to the E. coli origin of chromosomal replication (oriC) where it is involved in the early stages of initiation, preceding DNA synthesis (Fuller and Kornberg, 1983; see Messer, 1987, for a review of oriC replication). In vitro replication of oriC-containing plasmids has demonstrated that DnaA has a preference for supercoiled DNA, binds at specific sites within oriC (at several 'dnaA boxes'; 5'-TTATC/λACA-3'), and appears to bind oriC in a cooperative manner (Fuller and Kornberg, 1983; and Matsui et al., 1985).

Fuller et al. (1984) have shown that when DnaA is added to oriC DNA in saturating conditions, DnaA protects a 250bp region of the origin from DNase I digestion. This region contains four dnaA boxes that could bind DnaA (labelled from left to right in oriC as R1, R2, R3 and R4), but it appears that only three sites (R1, R2 and R4) bind the protein according to in vivo footprinting analysis (Samitt et
al., 1989; and Matsui et al., 1985). Fuller et al. (1984) have identified a number of other dnaA boxes located in the regulatory regions of the dnaA gene, in the Tn5 transposon and in replication origins of a number of plasmids as shown by binding studies and electron microscopy. DNase I footprinting showed that DnaA binding is highly cooperative with complexes of 20-30 DnaA monomers binding at oriC and covering 250bp. A single dnaA box can protect 100bp when bound by DnaA (Fuller et al., 1984).

DnaA binding at oriC appears to be the trigger for the early stages of the initiation of chromosomal replication. Xu and Bremer (1988) demonstrated that over-expression of DnaA leads to extra initiation of chromosomal replication, and Løbner-Olesen et al. (1989) demonstrated that high levels of DnaA resulted in the initiation of chromosomal replication earlier in the cell cycle. Using a plasmid expressing DnaA under the control of an inducible promoter, Løbner-Olesen et al. showed that the concentration of DnaA determines the time of initiation and therefore the initiation mass required for replication. In addition, they were able to demonstrate that no cyclic variation in DnaA levels were required to control initiation in E. coli. This last fact suggests that the autoregulation of dnaA may be sufficient to regulate chromosomal replication in E. coli, and has been supported by the finding that DnaA concentration in E. coli is invariant over a wide range of growth rates (Hansen et al., 1991).

The dnaA gene is expressed from two promoters which are both autoregulated by DnaA. DNase I protection experiments have shown that a single dnaA box between the two promoters binds DnaA to regulate the transcriptional activity of both promoters (Braun et al., 1985). In addition, promoter 2 (the downstream promoter) is also regulated by the state of methylation of a 5'-GATC-3' sequence within the -35 element of the promoter. Hemimethylation of the GATC sequence by Dam methylase blocks transcription from both promoters (Campbell and Kleckner, 1990). Since hemimethylation occurs for a brief period immediately after a round of replication, DnaA expression is repressed momentarily after the passage of a replication fork through the dnaA gene. Finally, dnaA expression may also be linked to the SOS regulon, as the dnaA gene is inducible by DNA damage (Quinones et al., 1991).

A6.7.3 Physical role of DnaA at oriC
DnaA clearly has some physical role in the initiation of replication at oriC. Messer et al. (1991) have demonstrated that base changes to the dnaA boxes in oriC have little effect on initiation, but that mutations which change the distance between boxes can inactivate the origin. If deletions or insertions of one helical turn are made however, near-normal origin activity are obtained. Finally, DnaA-dependent unwinding of the left hand side of oriC has been observed in vivo and in vitro (Messer et al., 1991; and Gille and Messer, 1991). This unwinding of the origin may require the involvement of host-encoded IHF, HU and FIS proteins (Polaczek, 1990; Skarstad et al., 1990; Gille et al., 1991; Kano et al., 1991; Messer et al., 1991; and Masters, 1991) as well as DnaB, DnaC, gyrase (DnaG) and SSB (Baker et al., 1986). Such unwinding may lead to the formation of a nucleoprotein structure (replisome) which then initiates replication.
The initial binding of DnaA to the dnaA boxes requires that ATP is bound to DnaA, ATP is slowly hydrolysed in the presence of DNA and it is only the ATP-DnaA form which is active in replication (Sekimizu et al., 1987). The entire formation of the replisome in oriC may take place on the inner membrane of E. coli. DnaA has a high affinity for acidic phospholipids which may re activate ADP-DnaA to the ATP-DnaA form (Sekimizu and Kornberg, 1988; and Yung and Kornberg, 1988) and antibodies to DnaA have located DnaA on the inner membrane of E. coli (Kostyal et al., 1989). oriC DNA is also capable of binding to E. coli membranes via specific oriC sequences and a number of unidentified proteins (Yongqing, 1987; Wolf-Watz and Masters, 1979; Hendrickson et al., 1982; Kusano et al., 1984; see also Schaechter et al, 1991; and Kataoka et al., 1991)

Bramhill and Kornberg (1988a; reviewed in 1988b) have suggested that the binding of DnaA to oriC proceeds in three stages (Figure A6.8). In the first stage, DnaA binds to the dnaA boxes to form an initial complex. In the second, the bound protein successively melts three AT-rich 13mer repeats located in oriC to form an open complex. Finally, in the last stage, the bound protein guides the DnaB-DnaC complex to the melted region to form the prepriming complex which it turn mark the future forks of bidirectional replication from oriC. Georgopoulos (1989) has extended this model and suggests that the function of DnaA binding is to trigger a cascade of events which ultimately leads to the loading of DnaB helicase (in association with DnaC) onto oriC. Once DnaB has bound, primase (DnaG) recognises the DnaB-ssDNA complex and synthesises RNA primers which are then extended by DNA polymerase III to initiate replication.

Of the three AT-rich 13mer elements located in oriC, the left-most 13mer is thermodynamically unstable and is capable of melting in the absence of DnaA. Kowalski and Eddy (1989) and Asai et al. (1990) have demonstrated that the helical instability of this element, not the sequence, is required for origin function. Kowalski and Eddy have suggested that the leftmost 13mer be referred to as the DNA unwinding element (DUE) in recognition of its unique activity in oriC activity.

The rightmost 13mer is also stable but differs from the leftmost 13mer in its strict sequence conservation among related bacterial origins. The differences between the two 13mer elements suggests that the 13mer region serves two functions: i) helical instability (DUE) and ii) (right-most 13mer) protein recognition.

The 13mer region involvement in protein recognition has been demonstrated by the isolation of a 33kDa protein which specifically binds to the 13mer elements (Hwang and Kornberg, 1990). The gene encoding the 33kDa protein (IciA) has been identified and IciA shown to have sequence homology to the LysR family of regulatory proteins (Thöny et al., 1991). The transition from initial complex to open complex can be blocked by the presence of IciA. Between 10-20 IciA monomers bind per copy of oriC, and the binding blocks the opening of the 13mer region by DnaA. However, once the opening has occurred, IciA is no longer able to repress the initiation of replication from oriC (Hwang and Kornberg,
194

1990). It is not yet known whether the expression of *iciA* is regulated or restricted to certain periods of the cell cycle.

![Diagram of DNA replication and protein complexes](image)
A6.7.4 A second function of DnaA DNA-binding

Origin melting may not be the only function of DnaA binding, both in terms of initiation and the control of the expression of a number of other genes. When DnaA binds to a dnaA box, the complex is able to act as a terminator for transcription and terminate about 25% of the transcripts crossing the site (Schaefer and Messer, 1988; and Schaefer and Messer, 1991). DnaA transcription termination occurs irrespective of whether the dnaA box region is translated, but only if the dnaA box orientation is 5'-TTTTCCACA-3' on the non-coding strand (Schaefer and Messer, 1989), suggesting that DnaA binds to only one of the two strands in the DNA duplex.

Transcriptional regulation by DnaA influences oriC replication in two ways: i) through the autoregulation of dnaA expression; and ii) by repressing transcriptional activity of the 16kDa (mioC) promoter across oriC which is involved in the regulation of initiation of replication at oriC (Lother et al., 1985; see Messer, 1987).

A6.7.5 DnaA-independent initiation of chromosomal replication

It is now recognised that E. coli chromosomal replication is not entirely dependent on DnaA. This has been demonstrated by the isolation of a number of dnaA temperature-sensitive pseudo-revertants where the revertants still carry the dnaAts mutation (Lindahl and Lindahl, 1984). Each of these revertants carry mutations between the metD and proA genes of E. coli suggesting that the mutations all affect the same gene (sin). Although sin suppressors do not confer any particular phenotype in E. coli, they all appear to be deficient in RNase H. On the basis of this, Lindahl and Lindahl have suggested that dnaA mutations can be suppressed by decreasing the amount of cellular RNase H activity. RNase H may degrade RNA-DNA hybrids which have been shown to be involved in the activation of oriC replication (Skarstad et al., 1990).
Kogoma et al. (1981) have postulated that E. coli possess an alternative DNA-independent mechanism for the initiation of replication. This mechanism is used in sdrA or sdrT (stable DNA replication) mutants (Kogoma 1978; and Lark et al., 1981) and requires some unknown function of the rec+ gene product (Torrey and Kogoma, 1982; and Lark and Lark, 1979). The sdrA mutations map to the same region as da sin mutants (Kogoma et al., 1981) and like sin, are able to suppress dnaA mutations (Kogoma and von Meyenberg, 1983). Torrey et al. (1984) have shown that a third set of dnaA suppressor mutants (dasF) and sdrA mutants are allelic.

A6.7.6 Summary of DNA involvement in chromosomal replication

The DNA protein is perhaps the most important element involved in the initiation of replication from oriC. The expression DNA influences topological changes at the origin which then allow cellular replication machinery to begin DNA synthesis. Despite the importance of DNA in the recognition and preparation of the origin for replication, current research indicates that a number of other cellular elements are important for origin function. By combining all of the current knowledge of oriC activation, it is apparent that the origin DNA forms a very complex structure with a number of proteins, some of which are only transiently associated with oriC.

A6.8 DNA involvement in plasmid replication

A6.8.1 DNA-dependent and independent replicons

A number of plasmids and bacteriophage appear to phenotypically suppress dnaA temperature-sensitive (dnaA46ts) mutations by integrating into the host chromosome (Nishimura et al., 1971; and Chesney and Scott, 1978). At the permissive temperature (30°C) chromosomal replication is from oriC and involves DNA. However, at the non-permissive temperature (42°C) DNA is inactivated and the replication of the chromosome results from replication initiated from the integrated plasmid or bacteriophage replicon.

The in vivo replication of the plasmid R1 is DNA-independent and the plasmid is able to integrate into the host chromosome; hence, R1 can integratively suppress dnaA46ts mutants. At 42°C the rate of chromosomal replication and cell division of an integrated R1 dnaA46ts strain is lower that at 30°C. However, both replication and division rates can be restored by increasing the expression of RepA protein essential for R1 replication (Bernander et al., 1991). RepA appears to compensate for the absence of DNA in the initiation and assembly of the replisomes required for replication at oriR (Bernander et al., 1991). However, DNA association with oriR is initiated by interactions with a RepA-oriR complex which is capable of supporting DNA-dependent replication when the oriR dnaA box has been deleted (Ortega-Jiménez et al., 1992). It appears likely that the DNA-induced melting of oriC DNA is not required for R1 replication, and that RepA is able to melt oriR DNA and promote replisome formation instead.
Although R1 and the closely related plasmid R100 (NR1) can replicate independently of DnaA in vivo (Chandler et al., 1977; Tang et al., 1989; and Bernander et al., 1991), in vitro experiments suggest that mini-R1 replication is DnaA-dependent (Masai and Arai, 1987; and Ortega et al., 1986). It is likely that although mini-R1 can replicate in DnaA-dependent and DnaA-independent manners, R1 replication is usually in the DnaA-dependent mode and that replication in the absence of DnaA is very inefficient.

The F plasmid has also been used to suppress dnaA46<sup>ts</sup> mutants (Nishimura et al., 1971; and Goebel, 1974b). Nishimura et al. (1971) demonstrated that the F plasmid was unable to complement dnaA46<sup>ts</sup> mutants in trans, and that most dnaA46<sup>ts</sup> revertants were recombination donors (Hfr) where F had integrated into the chromosome. At 42°C these revertants were sensitive to both ethidium bromide and acridine orange, indicating that chromosomal replication was under the control of the integrated F plasmid. However, subsequent analysis of a number of the Hfr revertants by Tresguerres et al. (1975) showed that the strains could grow at either 40°C or 42°C on enriched medium but only at 40°C on minimal medium. Tresguerres et al. concluded from these observations that in some cases at least, components of the enriched medium allowed suppression of the dnaA46<sup>ts</sup> mutation and that growth of these strains did not require an integrated F plasmid.

This conclusion has since been supported by experiments which have shown that mini-F (RepFIA), the primary F plasmid replicon, and mini-P1, the lysogenic replicon of bacteriophage P1, both require DnaA for replication. Kline et al. (1986) demonstrated that a number of dnaA mutants used to test F plasmid or Hfr chromosomal replication were sufficiently 'leaky' to allow DnaA-dependant replication. Neither integrated F nor F plasmids were able to replicate in a strain in which Tn10 had disabled the dnaA gene. Mini-P1 replication in vitro has been shown to be dependent on DnaA (Wickner and Chattoraj, 1987), and that only one of the five dnaA boxes near the mini-P1 origin is required for replication (Abeles et al., 1990).

Some DnaA-independent plasmids are unable to suppress dnaA46<sup>ts</sup>, presumably due to an inability to integrate into the host chromosome (for example, ColE1 and ColE2; Goebel, 1974b). At non-permissive temperatures in a dnaA46<sup>ts</sup> host, the replication rate of ColE1 decreases and stabilises at a lower rate (Frey et al., 1979; and Polaczek and Ciesla, 1983). A more intensive examination of the replication of pBR322 (a ColE1-type plasmid) has shown that the accumulation of plasmid DNA ceases after 1-2 hours after elevation to 42°C in a dnaA46<sup>ts</sup> host (Chiang et al., 1991). Although pBR322 has a dnaA box present in the origin region, deletion of the box has no effect on plasmid replication. Chiang et al. have therefore concluded that DnaA plays no specific role in pBR322 replication in vivo, and that the gradual cessation of replication results from indirect effects on chromosomal replication resulting from the dnaA46<sup>ts</sup> mutation.

Despite the fact that dnaA boxes were initially identified by their ability to bind DnaA, a number of DnaA-dependent plasmids appear to be able to replicate when some or all of the dnaA boxes have
been deleted from the origin region. Deletion of either of the two dnaA boxes in mini-F does not prevent mini-F replication (Murakami et al., 1987), whilst the deletion of all of the dnaA boxes from mini-Rts1 results in replication in a dnaA+ host at half the normal copy number (Itoh and Terawaki, 1989). Mini-P1 can still replicate with only one of the five dnaA boxes present, suggesting that the boxes play a different role in plasmid replication than do the boxes in the oriC origin, where the number and position of the boxes are critical (Abeles et al., 1990).

A6.8.2 Summary of DnaA involvement in plasmid replication

DnaA action differs in oriC replication and in the replication of plasmid origins. At oriC, DnaA prepares the DNA in such a manner which will allow the recognition and binding by a variety of proteins required to start DNA synthesis. DnaA involvement in the replication of DnaA-dependent plasmid origins appears to be more limited. Although DnaA binding has been demonstrated at a number of origins, it is not known whether DnaA causes the same cascade of events there as it does at oriC. The possibility of a different DnaA activity has been suggested by the fact that the number and positions of plasmid origin dnaA boxes appear to be less strictly regulated than those at oriC, and by the fact that an increase in plasmid initiator protein can sometimes overcome a deficiency of DnaA or dnaA boxes.

Early examination of DnaA-dependent replication in a number of plasmids was made confusing through the use of leaky dnaA mutant strains. The development of tight mutants has demonstrated that plasmid replicons can be divided into three classes according to their reliance on DnaA for the initiation of replication. These classes are i) strictly dependent, represented only by oriC plasmids; ii) directly dependent, such as mini-F, mini-P1; mini-Rts1; and iii) independent, such as ColE1. However, in this last class DnaA may still be involved indirectly in initiation and therefore effects plasmid replication.

These observations have lead to the suggestion that DnaA-dependent plasmid origins have evolved from oriC-like origins where initiation results from DnaA action. In the case of DnaA-dependent origins DnaA is still involve directly in initiation, but some of the DnaA functions are now carried out by the plasmid initiator protein. DnaA-independent replicons which have dnaA boxes (such as ColE1) are seen as examples of further evolution, where DnaA is no longer required for replication, yet still indirectly influences initiation.

The development of tight dnaA mutant strains has demonstrated that E. coli can also replicate without DnaA. This DnaA-independent replication is not directed from oriC and does not appear to be associated with any particular chromosomal location. It is not known whether any plasmids are able to utilise this secondary system for their own replication (other than when integrated into the host chromosome).
During the time spent undertaking the research presented in this Thesis, I have been involved in the publication or submission of several papers. The abstracts of each are provided below and the extent of my involvement in each is indicated.

**Nucleotide sequence and replication characteristics of RepFIB, a basic replicon of IncF plasmids**


(Abstract) A second autonomous replicon of P307, RepFIB, has been isolated that has significant homology with other replicons in IncF group plasmids. Eleven homologous repeats of 21 base pairs are present on the sequence and flank an open reading frame capable of coding for a protein of about $M_t = 40,000$. This protein was identified by maxicell analysis of cloned RepFIB. A series of deletion mutations of RepFIB were inserted into a DNA polymerase I-dependent vector and examined for their replication proficiency in a polA1 strain. These experiments defined a minimal replication region of 1.6 kilobases which includes the three repeats immediately upstream and downstream of the open reading frame. Deletion of a second set of repeats further downstream doubled the copy number of a chimeric plasmid replicating under RepFIB control. It was concluded that these repeats control the copy number of the replicon. Incompatibility tests showed that all three sets of repeats could express incompatibility with a resident RepFIB plasmid.

Sequence analysis, some experimental work, manuscript preparation.

**Expression and Regulation of the RepA Protein of the RepFIB Replicon from Plasmid P307**


(Abstract) The control of RepFIB replication appears to rely on the interaction between an initiator protein (RepA) and two sets of DNA repeat elements located on either side of the repA gene. Limited
N-terminal sequence information obtained from a RepA:β-galactosidase fusion protein indicates that although the first residue of RepA is methionine, the initiation of translation of RepA occurs from a 'CTG' codon rather than from the predicted 'GTG' codon located further downstream. Over-expressed RepA in trans is capable of repressing a repA: lacZ fusion plasmid in which the expression of the fusion protein is under the control of the repA promoter. The repA promoter has been located functionally by testing a series of repA: lacZ fusion plasmids. Both in vivo genetic tests and in vitro DNA-binding studies indicate that repA autoregulation can be achieved by RepA binding to one or more repeat elements which overlap the repA promoter sequence.

Accepted for publication by the Journal of Bacteriology.

RepFIB: A basic replicon of large plasmids

M. D. Gibbs, A. J. Spiers and P. L. Bergquist

(Abstract) The distribution of the RepFIB replicon amongst a total of 20 plasmid incompatibility groups was determined using Southern blot and polymerase chain reaction analysis. The presence of the replicon was confirmed in 25 out of 55 plasmids tested. The majority of plasmids carrying RepFIB are from the IncF incompatibility groups, but one plasmid from IncI and one plasmid from IncP also possess the replicon. Seven different examples of RepFIB were sequenced in the minimal replicon region to obtain 1525bp of sequence information covering the repA gene and flanking repeat regions for comparison. An analysis of these sequences plus three sequences previously reported showed almost perfect conservation of the predicted protein sequence of RepA and of the flanking DNA repeats. DNA sequence data was analysed using maximum parsimony techniques to describe the possible evolutionary relationships of the ten examples of RepFIB.

Accepted for publication by Plasmid.

Regulatory interactions between RepA, an essential replication protein, and the DNA repeats of RepFIB from the plasmid P307

A.J. Spiers, N.Bhana and P.L. Bergquist

(Draft abstract) The control of RepFIB replication appears to rely on the interaction between an initiator protein (RepA) and two sets of DNA repeat elements located on either side of the repA gene (BCDD'D'' and EFGHIJ). In vivo genetic tests demonstrate that the BCDD'D'' repeats form part of the origin of replication, whilst some of the downstream repeat elements (HIJ) are involved in the sensing and setting of plasmid copy number. RepA DNA-binding to these groups of repeats has been investigated in vivo utilising the fact that the replicon contains three active promoters (oriP, repAp
and EFp), one of which has previously been shown to control the expression of repA (repAp). All three promoters are closely associated with the repeat elements flanking repA, and an investigation using lacZ or cmI gene fusions has demonstrated that RepA expressed in trans is able to repress each promoter. However, these assays suggest that the transcriptional response of ori and repAp to RepA repression is significantly different, despite the fact that both promoters are embedded within the BCDD'D repeat elements. Extra copies of the BCDD'D or EFG repeats in trans has no effect on RepA repression of repAp embedded in a second copy of the BCDD'D repeats, but copies of the HIJ or EFGHIJ repeats are able to de-repress repAp suggesting that there is a fundamental difference between RepA-BCDD'D or -HIJ complexes and RepA-EFG or -EFGHIJ complexes.

To be submitted to the Journal of Bacteriology.  

This paper reports some of the work presented in this Thesis.


References


References


References 213


References


References


References


References

221


---

References

---

225


