

**A Thesis Submitted for the Degree of PhD at the University of Warwick**

**Permanent WRAP URL:**

<http://wrap.warwick.ac.uk/108622>

**Copyright and reuse:**

This thesis is made available online and is protected by original copyright.

Please scroll down to view the document itself.

Please refer to the repository record for this item for information to help you to cite it.

Our policy information is available from the repository home page.

For more information, please contact the WRAP Team at: [wrap@warwick.ac.uk](mailto:wrap@warwick.ac.uk)

**A study of plasmid biology in**  
***Bacillus thuringiensis***  
**subspecies *kurstaki* HD1-Dipel**

**by**

**David Gordon McDowell**  
**B.Sc. (Hons.) University College London**

**A thesis presented for the degree of**  
**Doctor of Philosophy**

**University of Warwick**

**October 1990**

THE BRITISH LIBRARY DOCUMENT SUPPLY CENTRE

## BRITISH THESES N O T I C E

The quality of this reproduction is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print, especially if the original pages were poorly produced or if the university sent us an inferior copy.

Previously copyrighted materials (journal articles, published texts, etc.) are not filmed.

Reproduction of this thesis, other than as permitted under the United Kingdom Copyright Designs and Patents Act 1988, or under specific agreement with the copyright holder, is prohibited.

THIS THESIS HAS BEEN MICROFILMED EXACTLY AS RECEIVED

THE BRITISH LIBRARY  
DOCUMENT SUPPLY CENTRE  
Boston Spa, Wetherby  
West Yorkshire, LS23 7BQ  
United Kingdom

## Contents

List of figures	VIII
Acknowledgments	XII
Declaration	XIII
Dedication	XIV
Abstract	XV
Abbreviations	XVI

## Chapter 1

## Introduction

1.1 Foreword	1
1.2 The history of <i>Bacillus thuringiensis</i>	2
1.3 Introduction to plasmids	4
1.4 General properties of plasmids	5
1.5 <i>Bacillus thuringiensis</i> plasmids	6
1.6 Classification of <i>Bacillus thuringiensis</i>	11
1.7 Physiology and Development of <i>B. thuringiensis</i>	13
1.8 <i>Bacillus thuringiensis</i> as a biological pesticide	14
1.9 Crystal Proteins of <i>Bacillus thuringiensis</i>	18
1.10 Plasmid host range	18
1.11 Plasmid replication in ssDNA plasmids	21
1.12 Plasmid stability and replication control	24
1.13 Plasmid partitioning	28
1.14 Incompatibility	32
1.15 Recombination events in ssDNA plasmids	36

## 1.16 Outlook

38

## Chapter 2

## Materials and Methods

2.1 Bacterial strains	40
2.2 Plasmids used in this study	40
2.3 Commonly used biochemical buffers	41
2.4 Culture media and growth of bacterial strains	42
2.5 Preparation and storage of antibiotics	44
2.6 Maintenance and storage of bacterial strains	45
2.7 Preparation of plasmid DNA	45
2.7.1 <i>Escherichia coli</i> :- Small scale	45
2.7.2 <i>Escherichia coli</i> :- Large scale	47
2.7.3 <i>Bacillus subtilis</i> and <i>Bacillus thuringiensis</i> plasmid preparations	50
2.7.4 <i>B. thuringiensis</i> protoplast lysis plasmid preparation	50
2.8 Quantitation of DNA	51
2.9 Restriction endonuclease digestion of DNA	52
2.10 Treatment of DNA with Calf Intestinal Phosphatase	53
2.11 Ligation of DNA	54
2.12 Transformation of <i>Escherichia coli</i>	54
2.13 Transformation of <i>B. subtilis</i> competent cells	55
2.14 Electroporation of <i>Bacillus thuringiensis</i>	56
2.15 Purification of DNA by phenol	58
2.15.1 Preparation of phenol	58

2.15.2	Extraction with phenol	59
2.16	Precipitation of DNA	59
2.17	Agarose gel electrophoresis	59
2.18	Visualization of DNA in agarose gels	61
2.19	Purification of DNA from agarose gels	61
2.20	Southern blotting procedure	62
2.20.1	Preparation of the gel	62
2.20.2	Preparation of the filter	62
2.20.3	Southern blotting procedure	63
2.21	Preparation of radio-labelled DNA probes	63
2.21.1	Nick translation	63
2.21.2	Purification of Nick translated DNA	64
2.22	Hybridisation of Southern blots	65
2.23	Radiolabelling of plasmid-encoded polypeptides using the Maxicell procedure	67
2.24	Polyacrylamide Gel Electrophoresis of proteins	68
2.24.1	Gel preparation	69
2.24.2	Sample preparation and running of gel	71
2.24.3	Visualization of polypeptide bands	71
2.24.4	Detection of <sup>35</sup> S methionine labelled polypeptides by autoradiography	72
2.24.5	Determination of polypeptide molecular weights	73
2.25	Preparation of dialysis tubing	73
2.26	M13 cloning and sequencing techniques	73
2.26.1	Preparation of replicative form M13 DNA	74

2.26.2	Generation of random clones based on a sonication procedure	74
2.26.3	Transfection of M13 DNA into <i>E. coli</i> TG2	75
2.26.4	Selection of M13 recombinants	75
2.26.5	Storage of M13 isolates	76
2.26.6	Preparation of template DNA	76
2.26.7	Sequencing reactions	77
2.26.8	Gel electrophoresis	77
2.26.9	Autoradiography	78
2.26.10	Computer assisted sequence compilation	78
2.27	Preparation of cell lysates for ssDNA detection	79

### Chapter 3

#### Selection and cloning of a *Bacillus thuringiensis* plasmid

3.1	Introduction	80
3.2	Plasmid profiles	81
3.3	Selection of the 2 kb plasmid	83
3.4	Cloning of the 2 kb plasmid	86
3.5	Restriction analysis of the 2 kb plasmid	92
3.6	Alternative cloning of the 2 kb plasmid	98
3.7	Southern blotting analysis	102
3.8	Summary and discussion	106

## Chapter 4

## Sequencing studies

4.1	Introduction	109
4.2.1	Random cloning in M13 vectors	112
4.2.2	Directed cloning in M13 vectors	117
4.3	Computerised analysis of the sequence	123
4.3.1	The new restriction map	123
4.3.2	%A+T and %G+C analysis	125
4.3.3	Two possible open reading frames	131
4.3.4	-10 and -35 promoter sequences	133
4.3.5	Ribosome binding sites	137
4.4	Maxicell analysis	139
4.5	Summary and Conclusions	143

## Chapter 5

## A model for the replication of pHD2

5.1	Introduction	145
5.2	Preliminary data bank searches	147
5.3	Protein sequence alignments using the Clustal sequence alignment package	148
5.4	Further protein sequence alignments and significances with ORF-A	158
5.5	Further protein sequence alignments and significances with ORF-B	162



5.6 A proposed nick site and ORF-A as a Rep protein	165
5.7 A possible minus origin	171
5.8 Summary and discussion	174

## Chapter 6

### Replication of pHD2 chimeras in Gram-positive hosts

6.1 Introduction	175
6.2 Cloning of a marker functional in <i>Bacillus</i>	176
6.3 Construction of shuttle vectors	180
6.4 Transformation into <i>B. subtilis</i> and <i>B. thuringiensis</i> subsp. <i>israeliensis</i> 1PS-78/11	187
6.5 Deletional events in ssDNA plasmids	192
6.6 Repeated sequences	193
6.7 Plasmid stability studies	196
6.8 Summary and Discussion	197

## Chapter 7

Overall conclusions and outlook	200
---------------------------------	-----

Appendix 1	207
------------	-----

Bibliography	212
--------------	-----

## List of figures

Figure 3.1	<i>Bacillus thuringiensis</i> subspecies <i>kurstaki</i> HD1-Dipel plasmid profiles	82
Figure 3.2	Comparison of DNA obtained from <i>Bacillus thuringiensis</i> subspecies <i>kurstaki</i> HD1-Dipel using small scale and large scale plasmids preparation procedures	84
Figure 3.3	Preliminary restriction analysis of the <i>Bacillus thuringiensis</i> subspecies <i>kurstaki</i> HD1-Dipel plasmid pHD2	87
Figure 3.4	Restriction map of pLG338	89
Figure 3.5	Restriction map of pBR322	91
Figure 3.6	Restriction map of pHD2	95
Figure 3.7	Restriction analysis of pDM200	96
Figure 3.8	Derivation of pDM200	97
Figure 3.9	Restriction analysis of pDM100	100
Figure 3.10	Derivation of pDM100	101
Figure 3.11	A Southern blot of total <i>Bacillus thuringiensis</i> subspecies <i>kurstaki</i> HD1-Dipel plasmid DNA probed with nick translated pDM200 DNA	104
Figure 3.12	A Southern blot of the constructs pDM100 and pDM200 probed with nick translated pHD2 plasmid DNA	105
Figure 4.1	Synthetic oligonucleotide primers and M13 clones used in the determination of the DNA sequence of pHD2	116

Figure 4.2	Synthetic oligonucleotides used as sequencing primers	120
Figure 4.3	Improved restriction map of pRD2	124
Figure 4.4	The DNA sequence and preliminary analysis of pRD2	126
Figure 4.5	The genetic code	127
Figure 4.6	Codon usage in ORF-A	129
Figure 4.7	Codon usage in ORF-B	130
Figure 4.8	Possible open reading frames within pRD2	132
Figure 4.9	Promoter sequences in <i>B. subtilis</i> and <i>E. coli</i>	135
Figure 4.10	Putative promoter and ribosome binding site sequences upstream of the proposed open reading frames	136
Figure 4.11	Expression of polypeptides from pRD2 based constructs using the Maxicell system	142
Figure 5.1	An alignment of the ORF-A and RepM (pC223) amino acid sequences using the Clustal4 programme	150
Figure 5.2	An alignment of the ORF-A and RepD (pC221) amino acid sequences using the Clustal4 programme	151
Figure 5.3	An alignment of the ORF-A and RepC (pT181) amino acid sequences using the Clustal4 programme	152
Figure 5.4	An alignment of the ORF-A and RepE (pS194) amino acid sequences using the Clustal4 programme	153
Figure 5.5	A multiple sequence alignment of the ORF-A and <i>S. aureus</i> Rep protein amino acid sequences using the Clustal4 programme	154

Figure 5.6	An alignment of the amino acid sequences of four <i>S. aureus</i> Rep proteins using the Clustal4 programme	155
Figure 5.7	An alignment of the ORF-B and RepA amino acid sequences using the Clustal4 programme	157
Figure 5.8	An alignment of the RepA (pLS1) and 46 codon ORF (pT181) using the Clustal4 programme	157
Figure 5.9	An alignment of the pT181 46 codon ORF and ORF-B amino acid sequence using the Clustal4 programme	157
Figure 5.10	An alignment of the ORF-A and four <i>S. aureus</i> Rep protein amino acid sequences using the Gap programme	159
Figure 5.11	An alignment of the ORF-A and four <i>S. aureus</i> Rep protein amino acid sequences using the Gap programme	160
Figure 5.12	An analysis of the significance of the alignment of the ORF-A and RepC amino acid sequences using the SGap programme	163
Figure 5.13	An analysis of the alignment of the ORB and RepA amino acid sequences using the Gap and SGap programmes	164
Figure 5.14	Conserved regions of the Rep proteins of pC194-like origins	167
Figure 5.15	Consensus sequences within the plus origins of sDNA plasmids	168
Figure 5.16	A possible plus-origin sequence within pHD2	169

Figure 5.17	A possible hairpin structure within the proposed origin of pBD2	170
Figure 5.18	Minus origin sequences and palindromes	172
Figure 5.19	A long inverted complimentary repeat sequence present within pBD2. A possible minus origin	173
Figure 6.1	Derivation of pCC100	178
Figure 6.2	Restriction endonuclease analysis of pCC100 and pCC150	179
Figure 6.3	Restriction endonuclease analysis of pCC200	182
Figure 6.4	Derivation of pCC200	183
Figure 6.5	Restriction endonuclease analysis of pCC300	184
Figure 6.6	Derivation of pCC300	185
Figure 6.7	Restriction endonuclease analysis of pCC300 derivatives following electroporation of <i>Bacillus thuringiensis</i> subspecies <i>israelensis</i> IPS-78/11	191

### Acknowledgments

I would like to thank Dr Nick Mann for his guidance and supervision during the course of this project and without whose encouragement and enthusiasm I would not have got this far.

Many people in the Department of Biological Sciences at the University of Warwick, both past and present, have been of invaluable assistance during the course of this project and I would like to thank them all. In particular: Andy Stainthorpe for his guidance through the delights of DNA sequencing; Andy Morby and Malcolm McCray for their help with the DNA sequence analysis; and Greg Watson with whom sharing a lab was quite an experience. Last, but far from least, my thanks to Sue for her patience despite spiral staircases.

Financial assistance for this project was provided by the Science and Engineering Research Council and by Novo Nordisk.

Declaration

I hereby declare that the work described in this thesis was conducted by myself, under the supervision of Dr N.H. Mann.

None of the information contained herein has been used in any previous application for a degree.

To my parents for their patience and understanding during a particularly chaotic period of my life.



### Abstract

The work presented in this thesis involves the characterization of a small plasmid, pH<sub>D2</sub>, of approximately 2 kb in size from *Bacillus thuringiensis* subspecies *kurstaki* HD1-Dipel. The plasmid was cloned and sequenced and compared to other plasmids from Gram-positive bacteria for which sequence information was available and for which replication functions had been assigned. Homology between the predicted amino acid sequence of an open reading frame within pH<sub>D2</sub> and the *rep* gene products of the pT181 group of staphylococcal plasmids suggested a common method of plasmid replication. The further identification of possible plus and minus origins indicated that pH<sub>D2</sub> was a member of a family of plasmids replicating via a single-stranded DNA intermediate. Plasmid replication control in the staphylococcal plasmids pT181, pC221 and pS194 involves a negative control circuit using countertranscripts. Replication control in the case of pH<sub>D2</sub> may utilize an alternative system involving the gene product of a second small open reading frame with homology to RepA of pLS1 in which replication control is achieved by the binding of the *repA* gene product to the promoter region of the replication protein *repB*.

pH<sub>D2</sub> is the first plasmid from a *Bacillus thuringiensis* isolate to which replication functions have been assigned. The work presented here suggests pH<sub>D2</sub> to be a member of the ssDNA family of plasmids and extends the range of such plasmids which have been characterized with the suggestion that this group contains, in addition to the highly related staphylococcal plasmids, a number of more distantly related members. Additionally, chimeric plasmids containing pH<sub>D2</sub> and pBR322 have been demonstrated to show structural instability, although not segregational instability, in an alternative *Bacillus thuringiensis* host. Consequently, the use of such constructs in the cloning of heterologous genes in *Bacillus thuringiensis* in such a system may prove impractical at this stage with further work being required in order to overcome these problems and extend the exploitation of this industrially important family of entomopathogenic bacteria.

Abbreviations

ccc	covalently closed circular
l	linear
oc	open circular
s	sensitive
r	resistant
Kcal	kilo calories
hrs	hours
UV	ultra violet
mRNA	messenger RNA
rRNA	ribosomal RNA
Tc	tetracycline
Ap	ampicillin
Ca	chloramphenicol
PBG	polyethylene glycol
rpm	revolutions per minute
DTT	DL-dithiothreitol
BCIG	5-bromo-4-chloro-3-indolyl- $\beta$ -D-thiogalactopyranoside
IPTG	isopropyl $\beta$ -D-thiogalactopyranoside
SDS	sodium dodecyl sulphate
EDTA	ethylene diamine tetraacetic acid
U	uracil
A	adenine
G	guanine
T	thymine
C	cytosine
CAT	chloramphenicol acetyl transferase
ss	single-stranded
ds	double-stranded
ori	origin of replication
→	denotes orientation of gene

## Chapter 1

### Introduction

#### 1.1 Foreword

*Bacillus thuringiensis* is an industrially important organism comprising a large number of subspecies and has been used commercially as a biological pesticide for nearly forty years. *B. thuringiensis* has a complex array of plasmids, however, very little is yet known of the function or mode of replication of the many plasmids which occur within the many subspecies that have so far been categorized. It was the basic aim of this project to characterize one or more of the low molecular weight plasmids from *B. thuringiensis* subspecies *kurstaki* HD1-Dipel in terms of genetic organization and mode of replication.

During the course of this introduction I will outline the historical landmarks which have lead to our present understanding of *B. thuringiensis* as well as its industrial and commercial significance. A review of the many features and functions of plasmids will be presented although this will concentrate predominantly on those found within Gram-positive species of bacteria and will include plasmid replication and copy number control, incompatibility, and plasmid partitioning as well as structural stability and recombinational events. A comprehensive classification of this species is beyond the scope of this introduction but the principles which have been used in an attempt to classify the many isolates of *B. thuringiensis* will be outlined.

### 1.2 The history of *Bacillus thuringiensis*

The first isolation of a *Bacillus thuringiensis* strain was described by Ishiwata (1902, as cited by Bulla *et al.*, 1980) in Japan, who isolated the organism from the diseased larvae of the silkworm *Bombyx mori*. This isolate is classified as var. *sotto* within the current classification system. At this point in time the implicit threat of this pathogen to the important silk industry must have caused considerable worry in marked contrast to the anticipated benefits of using *B. thuringiensis* as a biological pesticide which were to follow in future years.

Berliner isolated a *B. thuringiensis* strain from the larvae of the Mediterranean flour moth (*Anagasta kuehniella*) in 1915 (as cited by Bulla *et al.*, 1980). The larvae originated from a mill in the district of Thuringen and hence he named the pathogen *B. thuringiensis*. The presence of a parasporal inclusion was noted, but not associated with the insecticidal properties of the organism at this stage. At about the same time it was noted by Aoki and Cgigasaki (as cited by Lüthy *et al.*, 1982) that only sporulated cultures were toxic to the larvae of the silkworm (*Bombyx mori*) and they suggested that the organism might be added to infested flour, and so the possibility of using the pathogen in the control of insect pests was appreciated even as early as 1915.

Agostino Bassi (1838) was the first to suggest that microorganisms might be of use in the control of insects (as cited by Angus 1965).

Numerous attempts to use *Bacillus thuringiensis* in the control of insect pests between 1920 and 1950 met with only limited success. It became apparent that only the larvae of certain members of the order Lepidoptera were susceptible to the pathogen although the mechanism of pathogenicity was not to be elucidated until much later.

In 1953 the presence of a parasporal inclusion or crystal was again noted by Hannay. His observation that a number of pathogenic isolates of *B. cereus*, as *B. thuringiensis* was then known, produced a crystal whereas other acrySTALLIFEROUS strains were not toxic led him to speculate that the crystal and pathogenicity were in some way related. Angus (1954), working with old cultures of the *sotto* variety which he had treated with clarified silk worm gut extract, showed that this was only capable of paralysing silkworm larvae when administered *per os* and not when injected. It was also noted that the toxin was heat labile and non-dialysable. In association with Fitz-James (1955), Hannay showed that this inclusion was in fact protein. However, it was not until 1956 when Angus separated the inclusion from the spores that the parasporal protein was identified as being the pathogenic component in *B. thuringiensis*. The exploitation of the entomopathogenic properties of *B. thuringiensis* and its use as a biological pesticide is discussed in Section 1.8.

### 1.3 Introduction to plasmids

A bacterial plasmid is a type of nonessential, extrachromosomal deoxyribonucleic acid that replicates autonomously as a stable component of the cell's genome (Novick, 1987). The existence of plasmids in bacteria has been known for more than three decades and although first identified in the Enterobacteriaceae have since been found in almost all bacterial groups which have been screened (see Hardy, 1981). Watanabe (1963) reviewed research on *Shigella* (the causative agent of dysentery) which showed isolates were simultaneously resistant to two or more drugs and that this resistance could be transferred to other bacteria. An analogy was drawn between this and a report by Lederberg *et al.* (1952), postulating the involvement of an extra-chromosomal element in the transfer of genetic information between strains of *Escherichia coli*. The realization of the medical significance of the transfer of drug resistance markers between pathogenic bacteria provided the impetus for the first considerable expansion of interest in the study of plasmids. Research into plasmids carrying drug resistance was facilitated by virtue that host cells could be easily identified and enriched for by their drug resistance. Bacterial plasmids have since been found to confer a wide variety of other phenotypes including plant tumor formation, biological nitrogen fixation and the metabolism of organic compounds (see Hardy, 1981).

#### 1.4 General properties of plasmids

Plasmids occur in the size range of between one and several hundred kilobases and a copy number of between one and several hundred per cell. Whilst some large plasmids such as an F' may be 400 kb long and carry up to 600 genes (Low, 1972), the essential feature common to all plasmids is a replication region capable of propagating itself and any genes linked to it (Cohen, 1976).

Most plasmids found in bacteria occur as covalently closed circles (ccc) and exist mainly in the negative supercoiled form. Such a form is generated when one of the strands of the DNA molecule is broken, the double helix structure partially unwound and the molecule religated. This process involves DNA gyrases. Several plasmids including ColE1 (Clevell, 1969) exist *in vivo* as supercoiled molecules associated with proteins which can act as specific topoisomerases. These proteins may become activated upon treatment with proteases or alkali giving rise to an open circular (oc) or relaxed form of the plasmid and are hence termed relaxation complexes.

The presence of linear plasmids has also been noted in *Bacillus thuringiensis* subspecies (Gonzalez *et al.*, 1981; Gonzalez and Carlton, 1984) by the increased mobility of a plasmid band relative to ccc and oc plasmid forms during electrophoresis in increasing concentrations of agarose, and in *Streptomyces rimosus* by electron microscopy and exonuclease analysis (Chardon-Loriaux *et al.*, 1986). Single-stranded forms of plasmids have also been noted in *Bacillus subtilis* and

*Staphylococcus aureus* (te Riele *et al.*, 1986) although these are intermediates in plasmid replication and are discussed later (Section 1.11).

### 1.5 *Bacillus thuringiensis* plasmids

In general, species of *Bacillus thuringiensis* have been shown to contain a substantial proportion of their genetic information in the form of plasmid DNA. The plasmid content of a number of *B. thuringiensis* isolates has been demonstrated by a number of workers and the differences between the plasmid profiles observed in different laboratories suggested to reflect differences in isolation technique particularly with regard to the high molecular weight plasmids (Aronson *et al.*, 1986).

Little information has been published assigning functions to the majority of the *B. thuringiensis* plasmids although most interest has so far been towards the location and function of the plasmid encoded toxin genes. The toxin gene has been localized to one of the higher molecular weight plasmids in certain species (see Aronson 1986) with plasmid curing experiments shown to give rise to acrySTALLIFEROUS derivatives.

The exchange of plasmids among strains of *B. thuringiensis* and *B. cereus* by conjugation has been shown to occur at high frequency (Gonzalez *et al.*, 1982). Further work by Battisti *et al.* (1985) showed that two plasmids, pX011 and pX012 found in *B. thuringiensis* subsp.



*thuringiensis* were fertility plasmids responsible for plasmid mobilization. Recipient strains containing pX012 were found to be more efficient donors although in strains containing both plasmids pX011 appeared to be the more fertile suggesting that the two plasmids might be competing in some way in the transfer process. The modulation of plasmid transfer by other plasmids in a number of *B. thuringiensis* isolates has also been observed (Gonzalez *et al.*, 1982; Gonzales and Carlton, 1984).

Work carried out by Minnich and Aaronson (1984) showed that the presence of the toxin gene was not in itself sufficient for normal toxin production in *Bacillus thuringiensis* subsp. *kurstaki* HD1 and demonstrated, by plasmid curing and cell mating experiments, that the plasmids of 4.9, 29 and 110 MDa were implicated in the regulation of synthesis of the major protoxin from the 44 MDa plasmid. Loss of the 110 MDa plasmid conferred a conditional phenotype in that toxin synthesis occurred at 25°C but not at 32°C. In mating experiments, involving the transfer of plasmids to a *B. cereus* recipient, none of the high molecular weight plasmids were not found to be transferred and a similar conditional phenotype was noted. The presence of the 29 and 44 MDa plasmids appeared to result in toxin synthesis at 25°C with the presence of the the 4.9 MDa plasmid significantly increasing the level of expression of the major protoxin from the 44 Mda plasmid.

Similar experiments using the subspecies *kurstaki* HD73, which is closely related to HD1, implied no involvement of the other plasmids in the regulation of toxin synthesis from the 50 MDa plasmid although the

level of expression in *B. cereus* was not as high as in the original HD73 isolate. The presence of the 29 MDa plasmid from HD1 in a *B. cereus* isolate also containing the 50 MDa plasmid from HD73 was also found to have the conditional phenotype implying the 29 MDa plasmid was responsible in some way for the conditional phenotype. HD73 has a simpler plasmid profile than HD1 and it may be that in this isolate regulatory genes involved in protoxin synthesis have been incorporated onto this higher molecular weight plasmid.

In the isolate *B. thuringiensis* subsp. *finitimus* there are only two plasmids with the 98 MDa plasmid encoding the toxin gene. Transfer of this plasmid to *B. cereus* is sufficient to allow protoxin synthesis although at a lower level than in the parental strain (Debro *et al.*, 1986). Regulatory elements may be present on the 77 MDa or on the chromosome in this strain but this has not been confirmed.

The occurrence of conserved sequences amongst the plasmids of different *B. thuringiensis* serotypes has also been demonstrated (Lereclus *et al.*, 1982). A molecular relationship was observed between plasmids, using Southern hybridization techniques, which placed them into two groups, one of greater than 15 MDa and one of less than 15 MDa, between which no cross-homology was found. The subdivision of these plasmids by size may reflect a common origin or function for these plasmids (Aronson *et al.*, 1986). The presence of related sequences between the plasmids and chromosome was also investigated (Lereclus *et al.*, 1982). Whilst the presence of homologous sequences between some of the larger group plasmids and the chromosome was demonstrated, no such

sequences could be detected between the chromosome and the smaller group of plasmids. The presence of insertion sequences on some of the *B. thuringiensis* plasmids was postulated to account for these findings (Kronstadt and Whiteley, 1984). The presence of such sequences had previously been indicated by Debabov *et al.* (1980) who demonstrated the presence of an inverted repeat structure on a 16.4 kb plasmid from *B. thuringiensis* subsp. *galleriae*; and by Lereclus *et al.* (1983) following the recovery of a piece of DNA originating from a *B. thuringiensis* plasmid from the *Streptococcus faecalis* plasmid pAMB1 after it had been transferred to a *B. thuringiensis* isolate and subsequently reisolated. The presence of two sets of inverted repeat structure flanking the crystal toxin gene of *B. thuringiensis* subsp. *kurstaki* HD73 designated IR2150 and IR1750 was demonstrated by Kronstadt and Whiteley (1984). The presence of homologous inverted repeat structures in a number of other *B. thuringiensis* subspecies was investigated using IR2150 and IR1750 as probes. Such sequences were shown to be present on the plasmid containing the crystal toxin gene in subspecies *sotto*, *darmstadtensis*, *tolworthi*, *thuringiensis*, *morrisoni* and *galleriae* although not in subspecies *alesti*. The presence of homologous sequences was also found on a number of plasmids not containing a crystal toxin gene as well as the chromosome (Kronstadt and Whiteley 1984).

A DNA segment of 3 MDa termed the Th-sequence, which had been cloned *in vivo* into the *S. faecalis* plasmid pAMB1 (Lereclus *et al.*, 1983), was later shown to be flanked by inverted repeat sequences and suggested to have a transposon-like structure (Lereclus *et al.*, 1984).

It was generally found to be located in close proximity to the crystal gene in a variety of *B. thuringiensis* isolates as well as having a number of additional locations within the same subspecies. This genetic element was later determined to be a transposon which was also functional in *E. coli* and termed Tn4430 (Lereclus *et al.*, 1986).

Few attempts have been made to determine anything of the function or structure of any of the many plasmids found within the *B. thuringiensis* subspecies. Three small, cryptic plasmids from *B. thuringiensis* subsp. *thuringiensis*, pG11 pG12 and pG13 with molecular sizes of 8.2, 9.2 and 10.6 kb respectively were partially characterized by Mahillon *et al.* (1988). The entire pG12 and pG13 plasmid sequences were successfully cloned in the *E. coli* pBR322 vector, although the pG11 plasmid could only be cloned in pieces, and comprehensive restriction maps were generated for all three plasmids. Cross-hybridization experiments revealed the presence of sequences homologous to pG12 in one of the higher molecular weight plasmids in this isolate as well as plasmids from *B. thuringiensis* subsp. *berliner* 1715. The results also suggested homology to sequences within the chromosome although this could have resulted from contamination with sheared DNA from large plasmids. Homology of pG11 and pG13 was also shown to *B. thuringiensis* subspecies *berliner* 1715 and *B. thuringiensis* subspecies *kurstaki* HD1-Dipel. The complete nucleotide sequence of pG12 has also been determined (Mahillon and Seurinck, 1988) and shown to contain the transposon Tn4430.

*Bacillus thuringiensis* plasmids are therefore non-essential with few functions being assigned to them except bacteriocin production (Aronson, 1986), and  $\delta$ -endotoxin production and regulation as discussed above. The reason for the maintenance of such coding potential in the form of plasmids in the *B. thuringiensis* subspecies and the significances of differences and similarities between the plasmid profiles still remains to be explained.

#### 1.6 Classification of *Bacillus thuringiensis*

The classification of a group of closely related entomopathogenic Bacilli as varieties of *B. cereus* and *B. thuringiensis* initially led to confusion and hampered the understanding of the biology of this group of Bacilli. A number of attempts have been made over the years to classify the many isolates using a number of criteria. The first important attempt to resolve this problem was made by Heimpel and Angus (1958, 1960) who used morphological and biochemical characteristics in an attempt to differentiate and classify the isolates known at that time. The retention of the name *Bacillus thuringiensis* was already widely accepted, as the presence of the crystal at sporulation clearly differentiated this group from *B. cereus*. Later attempts to differentiate and classify *B. thuringiensis* isolates were based upon flagellar agglutination tests (de Barga and Bonnefoi, 1962; as cited by Luthy *et al.*, 1982) and a further classification system was suggested by Norris (1964) based on the electrophoretic patterns of esterase enzymes present in extracts of vegetative cells. Classification systems based

on flagellar antigens in association with other biochemical properties has proved to be the most useful system for the subdivision of the serotypes into varieties or subspecies. This system, however, gives little information about the pathogenicity of the isolate.

Differences in pathogenicity between isolates reflect variations in the toxicity of the crystal protein for insect larvae which may be reflected in antigenic variations in the crystal protein. An attempt to classify the subspecies on the basis of crystal antigens has been made by Pendleton and Morrison (1966). Whilst good correlation has been observed between the various classification systems each additional system results in the re-classification of a number of isolates to other or to new groupings.

Further attempts to resolve the classification problem have been made in a study carried out by Krywienzyk (1977, as cited by Luthy *et al.*, 1982) in which the antigenic composition of more than 300 *B. thuringiensis* isolates was determined. This gave a correlation between crystal antigens and the serotype of about 85 % of the strains investigated with the remainder not containing the variety-specific antigens or containing mixed antigens preventing definite assignment to a variety. Significantly, strains previously assigned to the variety *kurstaki* were demonstrated to consist of two serologically distinct groups designated K-1 and K-73 (Krywienzyk *et al.*, 1978). Recent attempts to classify the many *B. thuringiensis* isolates have been based on a classification of their toxin genes (for review see Hofte and Whiteley, 1989).

### 1.7 Physiology and development of *B. thuringiensis*

*B. thuringiensis* is a Gram-positive, highly motile, soil bacterium which is characterized by its ability to produce crystalline inclusions during sporulation. These inclusions consist of proteins exhibiting a highly specific insecticidal activity (see section 1.9). The vegetative cell is approximately 2-5  $\mu\text{m}$  long and 1.0  $\mu\text{m}$  wide and undergoes a process of sporulation in order to survive periods of environmental stress. The dormant bacterial spores are able to survive unfavourable conditions for long periods of time. Dehydration of the spore protoplast during sporulation is responsible for the ametabolic state and for the resistance to heat, chemicals, irradiation and desiccation.

During the process of sporulation, the nucleoid condenses and the forespore septum forms giving rise to the 'incipient forespore'. Spore wall development and subpolar engulfment then ensues at which stage the cells are committed to sporulation (Bechtel and Bulla, 1976). During the final stages of sporulation, the various layers of the spores start to form until after maturation it contains several distinct layers which, starting from the forespore cytoplasm are: the inner membrane, primordial cell wall, cortex, outer membrane, lamellar spore coat, outer fibrous spore coat and exosporium. During the final stages of spore development, the mother cell nucleoid fragments. The mature spore is released by lysis of the sporangium during the maturation process.

In *Bacillus thuringiensis*, sporulation is accompanied by the development of a bipyramidal, parasporal crystal known as the  $\delta$ -

endotoxin (for review see Aronson, 1986). The time and site of appearance of the parasporal crystal varies among the different strains, but in subspecies *kurstaki* often first starts to develop during engulphment. A number of other crystal proteins have been shown to occur in other subspecies and which also have different activity spectra (for review see Hofte and Whiteley, 1989). The crystal has been shown to be responsible for the insecticidal properties of this group of organisms. The haemolymph of the target insect larvae is an excellent nutritional environment and may be one of the rare environments in which soil bacteria can reach sufficiently high concentrations for genetic exchange via cell mating to occur (Carlton and Gonzalez, 1984; Gonzalez, *et al.*, 1980)

#### 1.8 *Bacillus thuringiensis* as a biological pesticide

A wide variety of microbes have been found to be pathogenic against insects including viruses, mycoplasmas, rickettsias, spirochaetes, protozoa, fungi, and bacteria (Bulla *et al.*, 1975) although it is the bacterial insect pathogens which have received the most attention. A considerable number of bacteria pathogenic to insects were tested under field conditions between 1920 and 1940 (see Bulla *et al.*, 1975). Only *B. thuringiensis* and *Bacillus popilliae* emerged as possible candidates for plant protection. The fact that *B. popilliae* can only be grown in the host, the Japanese beetle (*Popillia japonica*) eliminated this bacterium from large-scale industrial production leaving



*B. thuringiensis*, which is easily grown on a variety of media, as the only suitable candidate for large-scale production.

Sporeine, developed by a French company in 1938 was the first commercial preparation of *Bacillus thuringiensis*. More preparations came onto the market after the Second World War including Bakthane 1-69, Biotrol B7B, Parasporin, Thuricide, Agritol, Larvatrol, Bactospeine and Entobacterin-3 (see Luthy 1982). The early success promised by these formulations did not materialize primarily because of inconsistencies in activity combined with inadequate standardization and for the next 30 years *B. thuringiensis* stood in the shadow of readily available, cheaper and more efficient chemical alternatives. The situation has radically changed in the last two decades with the emergence of resistance to chemical insecticides and a greater awareness of the undesirable impact of these chemicals on the environment. Pesticides have been linked to mutation, cancer, human disease, damage to wildlife and toxicity to plants and this has recently pointed the way towards alternatives in pest control.

It had initially been thought that *B. thuringiensis* subspecies were only active against lepidopteran species, somewhat limiting their application. With the discovery of new strains with specificities towards different insect orders, such as *B. thuringiensis* subsp. *israelensis* which is highly active against some dipterans including mosquito and blackfly larvae (Goldberg and Margalit, 1977), and subspecies *tenebrionis* and *san diego*, which are active against various coleoptera, the applications of this group of bacteria continue to grow.

In 1985 there were 410 registered formulations of *Bacillus thuringiensis* approved for use against insect pests in the United States (Aronson, 1986). It was estimated at this time that the annual market for pesticides was  $\$10 \times 10^9$  with biological pesticides accounting for approximately 1 % of this, but expected to rise to 10 % or more by the end of the decade.

The overriding concern in the use of bacterial pathogens is their safety. *B. thuringiensis* and *B. popilliae* have both been demonstrated to be virtually harmless to vertebrates and other non-target organisms. The relatedness of *B. thuringiensis* to *B. anthracis* (Somerville and Jones, 1972) has lead to the suggestion that there is a danger of genetic change, either by transduction or transformation, which could give rise to human pathogens. Current practices prevent the introduction of animal pathogens by the testing of each preparation before release. Certain serotypes of *B. thuringiensis* produce a heat stable ATP analogue known as the  $\beta$ -exotoxin with a much wider toxic activity. Consequently only strains lacking this toxin have been used in commercial preparations.

The future development of biological forms of pest control looks very promising. The toxin gene has been cloned from a number of *B. thuringiensis* subspecies (for review see Hofte and Whiteley, 1989). In 1982, Gonzalez *et al.* demonstrated the operation of a high frequency plasmid transfer system in *B. thuringiensis*. Klier *et al.* (1983) took advantage of this system and produced strains active against lepidopteran and dipteran species. The portion of the polypeptide

responsible for toxicity in *B. thuringiensis* subsp. *kurstaki* HD-1 has been shown to be encoded by the amino terminal 55 % of the toxin gene (Schnepf and Whiteley, 1985) using deletions and fusions of the toxin gene expressed in *E. coli*. A similar result was obtained with subsp. *berliner* 1715 (Hofte *et al.*, 1986).

Modern technology has enabled a further development in the use of *Bacillus thuringiensis* to control insect pests with the transgenic plants in which modified crystal protein genes have been introduced into the plant genome. Vaeck *et al.* (1987) introduced a truncated toxin gene, encoding the N-terminal 610 amino acids from *Bacillus thuringiensis* subsp. *berliner*, into tobacco plants and obtained tobacco plants that were protected from attack by *Manduca sexta*. Introduction of the entire gene for an unknown reason was not found to give rise to resistant plants. No instances of insect/host resistance to *Bacillus thuringiensis* toxin in the environment have been reported, although this may reflect the low persistence of *Bacillus thuringiensis* in the environment and the situation may change with the arrival of transgenic plants. Resistance has been shown to arise with the Indian mealmoth (*Plodia interpunctella*) when fed a diet containing a dose of *Bacillus thuringiensis* subsp. *kurstaki* HD1-Dipel expected to give 70-90 % larval mortality (McGaughey, 1985). The many strains of *Bacillus thuringiensis* with their variety of activity spectra and toxicities provide a wealth of genetic information which may be exploited in a number of ways to yield new target species increased activities or modes of delivery which may ultimately revolutionize pest control.

### 1.9 Crystal proteins of *Bacillus thuringiensis*

Most strains of *Bacillus thuringiensis* show larvicidal activity towards certain members of the Lepidoptera with some isolates active against dipteran or coleopteran species (see review by Hofte and Whiteley, 1989). On sporulation, a parasporal inclusion is produced, often referred to as the crystal toxin or  $\delta$ -endotoxin. Upon ingestion, these crystalline inclusions become solubilized in the alkaline reducing environment of the larval midgut. The  $\delta$ -endotoxin is initially present in the form of a protoxin and it is here that it is proteolytically cleaved into the active form by proteases present in the larval gut. The active form of the toxin interacts with the gut wall in susceptible larvae. High affinity binding sites have been shown to exist on the midgut epithelium of such insects (Hofman *et al.*, 1988a; 1988b) and may account for the specificity of these toxins. Electrophysiological and biochemical data (Harvey, 1983; Knowles and Ellar, 1987) suggest that the toxins act by introducing pores into the cell membrane disturbing the osmotic balance causing the cells to swell and lyse. The crystal proteins and genes of *B. thuringiensis* has been extensively reviewed recently (see Hofte and Whiteley, 1989).

### 1.10 Plasmid host range

The control and mechanism of plasmid replication has been the subject of extensive research, for reviews see Scott (1984) and Thomas (1988). The functional sequence of a number of replication origins from

a number of prokaryotic replicons have been determined including ColE1 (Tomizawa, 1977; Oka, 1979), R1 and R1drd19 (Oertel, 1979; Rosen, 1979), R6K (Stalker, 1979), the single-stranded phages G4 (Simms and Dressler, 1978) and  $\phi$ X174 (Sanger, 1978), the lambdaoid bacteriophages (Grosshieldl and Hobom, 1979) and the chromosomal origins of *E. coli* (Meijer, 1979) and *S. typhimurium* (Zykind, 1980).

The host range of most plasmids is restricted to one or a few closely related species, although certain plasmids have also been shown to be able to replicate in a large number of species. RK2 is a self-transmissible plasmid of 56 kb expressing resistance to ampicillin, kanamycin and tetracycline in *E. coli* (Meyer *et al.*, 1977) where it replicates unidirectionally from a unique origin (Meyer and Helinaki, 1977). RK2 is capable of transfer between and maintenance in a wide variety of Gram-negative hosts (Beringer, 1974; Datta and Hedges, 1972) and if the identified origin is the only functional origin on this plasmid then it must be recognized and utilized by replication enzyme systems from a number of hosts. RK2 is a low copy number plasmid and therefore has a tightly regulated replication system (Meyer *et al.*, 1977). It differs from other such plasmids in that the replication origin present within a 393 bp *Hpa*II fragment, whilst not in itself a complete replicon, is physically separable from two other regions *tfrA* and *tfrB* which when present *in trans* in the same cell allow replication to occur (Thomas 1980).

Derivatives of the plasmids pLAB1000 and pLAB2000 isolated from *Lactobacillus hilgardii* were shown to replicate in *Enterococcus* as well

as *Lactobacillus* species. Replication was also shown to occur in *Bacillus subtilis* although the growth rate of the cells was reduced due to a reduction in copy number in this species (Josson *et al.*, 1989). RP4, a plasmid from *Pseudomonas aeruginosa*, can be transferred to other Gram-negative bacteria such as *Escherichia coli*, *Salmonella typhimurium*, *Klebsiella aerogenes*, *Rhizobium leguminosarum*, and *Agrobacterium tumefaciens* (see Chakrabarty, 1976).

Another instance of transfer of naturally occurring plasmids across the species barrier involves plasmids from *Staphylococcus aureus*. Ehrlich (1977) showed that five *S. aureus* plasmids; pC194, pC221, pC223, and pUB112 carrying chloramphenicol resistance markers, and pT127 carrying a tetracycline resistance marker previously described by Novick (1976), were capable of replication and expression in *Bacillus subtilis*. All of these *S. aureus* plasmids have since been shown to be segregationally unstable in *B. subtilis* (Novick *et al.*, 1986).

Attempts to cross the Gram barrier were later to be attempted by the construction of shuttle vectors carrying replicons specific for each target species. Most modern DNA technology is based around the Gram-negative bacterium *E. coli*. DNA sequences from all manner of sources have been cloned in this organism and whilst structural analysis of cloned genes can be done in a heterologous system such as this, functional analysis is best done in a homologous system. To this end a number of vectors capable of crossing the species barrier have been developed. These generally involve the generation of chimeric plasmids by the fusion of plasmids functional in both the Gram-positive and Gram-

negative hosts and therefore supplying the required sequences for the chimaeric plasmid to function in either host. Such a construct has been developed using a small cryptic plasmid from *B. thuringiensis* subspecies *israelensis* of 3.65 MDa and the *E. coli* replicon pJH101 for use in the investigation of the expression and regulation of toxin genes in the parental organism (Miteva and Grigorova, 1988). This construct was found to be genetically unstable in the *Bacillus* host although not in *E. coli*. Such instabilities are common in ssDNA plasmids carrying insertions of foreign DNA (see section 1.15).

#### 1.11 Plasmid replication in ssDNA plasmids

Numerous plasmids have been isolated from different species of Gram-positive bacteria and the functions and mode of replication of some of these determined. In particular, plasmids from *S. aureus* have been the subject of intense investigation with regard to their method of replication and provide a model system for the method of replication in this group of highly related plasmids. The similarity in structural organization and sequence homology between these plasmids and those from other bacteria has led the speculation that these plasmids may have a common ancestry (Iordanescu *et al.*, 1978; Projan and Novick, 1988). This group of highly interrelated plasmids appears to replicate using a rolling-circle method of replication (RCR) involving a single-stranded DNA intermediate (te Riele *et al.*, 1986).

Gros *et al.* (1987), Khan *et al.* (1988), and Koepsel *et al.* (1985) have demonstrated that the *S. aureus* plasmids pT181, encoding tetracycline resistance and pC194, encoding chloramphenicol resistance, replication occurs by RCR. Rolling circle replication requires three plasmid encoded elements: a plus origin, a replication protein (Rep) and a minus origin. Replication initiation in the case of pT181, requires the replication protein, RepC (Khan *et al.*, 1981). This protein has been shown to have a site-specific nicking-closing or topoisomerase activity and introduces a single-stranded nick in the origin of pT181 or heterologous plasmids containing the same origin but not in plasmids lacking the origin (Koepsel *et al.*, 1985).

The introduction of a single-stranded nick at a specific site in the origin sequence initiates plasmid replication. The plus strand becomes displaced and a new plus strand is synthesised using 3'-OH extension from the nick site. The termination sequence which overlaps the origin sequence is recognised by the Rep protein which then introduces a second nick yielding a fully replicated strand and the displaced strand as a single-stranded monomer. The origin and termination sequences overlap although are non-identical as demonstrated for pC194 (Gros *et al.*, 1987) and the analogous ssDNA *E. coli* phages (Baas and Janaz, 1988). The ends of this ssDNA monomer are finally ligated by the Rep protein giving a circle which is detectable as a free molecule (te Riele *et al.*, 1986). The nicks which initiate and terminate plus strand replication must by necessity occur at the same site (Gros *et al.*, 1987) although replication can terminate at a sequence homologous to the replication origin. This feature may be



responsible for the structural instability of many replicons based on members of the ssDNA group of plasmids (see Section 1.15).

Replication is completed by the conversion of the single-stranded intermediate into the double-stranded product. The minus origin (M-O), which is recognised by host factors including RNA polymerase since rifampicin blocks the conversion of the ssDNA intermediate to the dsDNA form, functions as an efficient initiation site for the initiation of minus strand synthesis. The minus origins of all plasmids are large and contain imperfect palindromic structures. Whilst many of the ssDNA plasmids are capable of replication in two or more hosts (del Solar *et al.*, 1987; Ehrlich, 1977; Goze and Ehrlich, 1980; Gryczan *et al.*, 1978) the M-Os of most of these plasmids only appear to function in the native host (del Solar *et al.*, 1987; Gruss *et al.*, 1987). Consequently, in hosts where the M-O is not functional, the ssDNA intermediate accumulates with conversion to the double-stranded form occurring only by non-specific initiation at a reduced frequency. (Boe *et al.*, 1989; Gruss *et al.*, 1987; Ray *et al.*, 1982). Only the M-O of pUB110 has so far been shown to be functional in more than one host (Boe *et al.*, 1989) and is functional in both *S. aureus* and *B. subtilis*.

Completion of minus strand synthesis ends the cycle of replication in which two plasmids are generated from one. If minus strand synthesis does not occur there is no net synthesis of dsDNA and the cycle results in the accumulation of ssDNA intermediate.

The high level of homology between the Rep proteins and the plus origins from other plasmids which have been sequenced strongly suggests that they may all replicate in a similar manner (Gros *et al.*, 1987; Projan and Novick, 1988). Conservation of sequences at the plus origin as well as an amino-acid motif around the active site of the  $\phi$ X174 Rep protein and the ssDNA plasmids has also been noted (Gros *et al.*, 1987) suggesting the method of nicking is analogous in these two systems. The plus origins which have been studied so far can be grouped into three types according to their homology with other plus origins with the regions of strongest homology in the regions surrounding the nick sites (see Chapter 5).

#### 1.12 Plasmid stability and replication control

A plasmid is maintained at a certain average copy number under a given set of conditions. An understanding of the physiological factors which determine copy number enables the development of models which may be used to predict the copy number and stability of hybrid plasmids prior to their industrial exploitation. Plasmid shedding is a potential problem in large-scale fermentations with organisms carrying recombinant plasmids. The 'metabolic burden' placed upon cells containing plasmids may reduce the growth rate such that they become outgrown by cells which have lost the plasmid if this is not stably maintained (Ataai and Shuler, 1987).

Interest in the mechanisms which determine plasmid copy number has been stimulated by the exploitation of plasmids as cloning vehicles. The ability to regulate the number copy such that it could be maintained at a low level and subsequently amplified can aid the segregational stability during the first phase of the a bacterial culture and help to optimize product yield in the expression phase (Thomas, 1988).

All plasmids so far studied appear to fit with the prediction of a negative control circuit in the control of plasmid replication as exemplified by the Inhibitor Dilution model (Pritchard *et al.*, 1969) and the Autoregulatory model (Sompayrac and Maalo, 1973). In the inhibitor dilution model, a plasmid-encoded repressor is synthesized in proportion to the copy number of the plasmid and eventually shuts off replication. As the cell volume increases the repressor becomes diluted eventually allowing replication to continue. The copy number of the plasmid therefore oscillates around a certain average number.

Studies with the plasmid R1 suggested that control of plasmid replication might not conform closely to this model as the probability of replication was not reduced to zero as the copy number was increased (Nordstrom *et al.*, 1984). Work carried out by Highlander and Novick (1987) however, showed that a temperature-sensitive mutant of the *S aureus* plasmid pT181 continued plasmid replication at the non-permissive temperature. Plasmid replication was halted on being returned to the permissive temperature until the copy number fell below a certain level at which point replication was resumed and thus showed the presence of a plasmid encoded repressor. The replication rate of a high copy number

mutant of pT181, lacking the repressor, was observed to decrease as it attained a new higher level rather than proceeding exponentially. This suggested that control of replication may be as much under the control of positive elements as negative elements. It maybe that in a natural, stressful environment, physiological constraints may mean that the copy number never attains a level subject to negative control.

The most intensely studied plasmid from a Gram-positive bacterium is the *S. aureus* plasmid pT181. pT181 replication is controlled at the level of RepC synthesis by a negative regulatory system that is functionally similar to the ColE1 and IncFII plasmids of *E. coli*. Since the *Bacillus thuringiensis* subspecies *kurstaki* HD1-Dipel 2 kb plasmid, pKD2, which was investigated during the course of this project, appears to be most closely related to the pT181 group of plasmids (see Chapter 5) it is this system which will be described as a model system here.

pT181 is a 4437 bp plasmid with a copy number of 20-25 per chromosome encoding resistance to tetracycline (Khan and Novick, 1983; Projan *et al.*, 1983). The replication of the ssDNA plasmids of which pT181 is a member is described later (Section 1.11). The copy control system is similar to that described to those of the ColE1 and IncFII groups of *E. coli* in that they utilize regulatory circuits of the inhibitor-target type in which a counter-transcript is the inhibitor. In the pT181 system, the plasmid replication control circuits are located in the region known as Cop just 5' to *repC* (Novick *et al.*, 1984; Carleton *et al.*, 1984). The primary replication control gene is known as *copA* and specifies two inhibitory counter-transcripts, RNAI and

RNAII. These RNA molecules are complementary to and can bind to the 5' untranslated *repC* leader so as to interfere with the translation of this mRNA. The RepC protein acts *in trans*, is rate limiting for plasmid replication (Novick *et al.*, 1982; Novick *et al.*, 1984) and has a site-specific nicking-closing or topoisomerase activity (Koepeel *et al.*, 1985). When pT181 is cloned into a vector that is maintained at a higher copy number than pT181, the chimeric replicon prevents the establishment of the autonomous pT181 due to the negative activity of the regulator of replication, referred to as CopA (Novick *et al.*, 1984).

The *repC* messenger RNA is initiated from two tandem promoters analogous to the R1 system in *E. coli* although there is no evidence for or against transcriptional control of either (See Novick, 1987) as found in this system (Light and Molin, 1982). Two counter-transcripts have been found for pT181 within the *copA* region, RNAI and RNAII, having common 5' start sites and different 3' termination sites. Work with copy number mutants suggested that the inhibitory activity expressed by *copA* is due to the counter-transcripts and that both may be required for normal copy control (Carlton *et al.*, 1984). Additionally, a striking parallel was observed between this pattern of mutations and that observed for ColEI and the IncFII plasmids, suggesting a similar control mechanism may be in operation (Carlton *et al.*, 1984).

An analysis of the possible secondary structures of the *repC* mRNA leader revealed two possible secondary structures, an upstream and a downstream loop. In the upstream loop structure, a sequence complementary to the Shine-Dalgarno (SD) sequence is paired with an

upstream precise repeat of the SD sequence. In this structure the SD sequence itself is in a single-stranded form and available for translation initiation. The loop in this structure was suggested to be the target site of RNAI/RNAPII with the formation of the RNA-RNA duplex leading to a secondary structure further downstream in which the SD sequence was fully base paired and unavailable for translation (Novick *et al.*, 1985). It was also proposed that the formation of this hairpin structure could serve as an early termination site or attenuator for *repC* mRNA transcripts. In this model, the critical inhibitory interaction occurs before transcription of the leader sequence has progressed beyond a certain point, at which time the structure is predicted to refold into the active form and be insensitive to the inhibitory action of RNAI/II (Novick *et al.*, 1985).

### 1.13 Plasmid partitioning

The segregational stability of plasmids depends upon the efficient partitioning of plasmid copies between daughter cells at cell division. The simplest way of achieving this partitioning would be to maintain the copy number of the plasmid at a sufficiently high level that the chance of generating a daughter cell with no copies of the plasmid would be very low. It is possible that some *E. coli* plasmids are maintained in this way, however, many plasmids show alternative ways of achieving efficient partitioning. Efficient site-specific recombination systems have been described which maximize the pool of plasmids available for partition by resolving multimeric forms of the plasmids (Austin *et al.*,

1981; Summers and Sherratt, 1984). Other mechanisms have been reported in which plasmids are responsible for the production of a both a toxin and an unstable blocking agent such that daughter cells which do not contain the plasmids are killed as the blocking agent decays away (Jaffe *et al.*, 1985; Gerdes *et al.*, 1986).

Plasmids which are maintained at a low copy number stand a much larger chance of being lost in a random partitioning and the killing of such plasmid-free cells would place an intolerable burden on cell propagation. Differences between the actual frequency of plasmid loss and that which would be expected if a random segregation system was in operation is often the result of an active plasmid partitioning system. For instance, the P1 plasmid is lost at a frequency of less than 1 in  $10^6$  cell divisions (Austin *et al.*, 1981) whereas the random distribution of plasmids would lead to a loss rate of 25 % per cell division.

A number of models have been proposed for plasmid partitioning. In the model proposed by Austin and Abeles, (1983a; 1983b), a specific plasmid-encoded protein (Par protein) binds to the plasmid partition site. This promotes a specific pairing of the plasmids from a free pool by a protein-protein interaction which may then bind to one of a number of similar host sites with which they are then co-partitioned. In the case of P1 there are two partition proteins, *parA* and *parB*. A minimal partition site of less than 29 bp is sufficient to promote active partition of a plasmid so long as the two Par proteins are supplied *in trans* (Martin *et al.*, 1987). In P1, the partition site consists of a 13 bp perfect palindrome with a 3 bp spacer although, in other cases such

as the F plasmid the situation may be more complex (for review see Austin, 1988).

Evidence for the noncovalent association of F plasmid DNA to the folded host chromosome structure has been presented (Kline and Miller, 1975). Additionally, evidence has also been presented suggesting a specific association of various replicons with the outer membrane of the cell envelope (Wolf-Watz and Masters, 1979). In the cases of R1 and pSC101, a specific cell fraction containing outer membrane material has also been shown to contain these plasmids. This association does not occur with a plasmid lacking the *par* region, suggesting that in this case at least, membrane attachment is a key event in plasmid partitioning (Gustafsson *et al.*, 1983).

The situation with regard to plasmids from Gram-positive bacteria does not seem to be as well resolved. A number of *S. aureus* plasmids, whilst capable of replication in *B. subtilis*, have been found to segregationally unstable in this host (Novick *et al.*, 1986) suggesting that the normal mechanism for partitioning plasmids between daughter cells is not functioning normally in this host.

Evidence for the membrane association of the replication origin and the initiation of DNA replication has been demonstrated in *B. subtilis* (Winston and Sueoka, 1980) for both the chromosome and a composite plasmid, pSL103, consisting of the *S. aureus* plasmid pUB110 and a *Bacillus pumilus trpC*<sup>+</sup> DNA fragment. Tanaka and Sueoka (1983)



later showed four specific regions in the pUB110 portion of pLS103 which were capable of membrane binding *in vitro* and postulated that this may have a role in plasmid partition in a similar way to the *par* locus of pSC101 (Meacock and Cohen, 1980).

Chang *et al.* (1987) observed the same partitioning deficient phenotype for a number of their constructs based upon the *S. aureus* plasmids pC194 and pUB110 in *B. subtilis* and identified a partition element from the *Bacillus* plasmid which complemented this *Par*<sup>-</sup> phenotype. The element was shown to complement *in cis* but not *in trans* and as it was therefore suggested that it did not specify a diffusible gene product. Its function was also shown to be strand-specific although this was not due to a dependence on transcriptional activation since this could not be overcome by placing a promoter adjacent to it. This 167 bp *par* fragment contained a number of inverted and direct repeats as shown by sequencing although no homology to other *par* sequences could be detected.

Devine *et al.* (1989) postulated that the segregational instability of many *S. aureus* plasmids in *B. subtilis* might be due to suboptimal host-plasmid relationship and decided to pursue an investigation of the replication and segregational stability of the plasmid pBAAL which was present within an industrial strain of *B. subtilis* and had been maintained under industrial fermentation conditions without any apparent selection pressure. The deletion or mutation of a specific sequence was observed to result in the production of large quantities of the single-

stranded DNA intermediate characteristic of the ssDNA group of plasmids. This region, identified as the minus origin, was required for the correct initiation of second strand synthesis. A 350 bp fragment was shown to be highly homologous to a region of pLS11 which was reported to contain the partition function (Chang *et al.* 1987) showing differences at only four positions. It therefore seems highly probable that the partitioning function reported for pUB110 was in fact the minus origin.

There does not appear to be any evidence of a membrane attachment mechanism or *par* function aiding plasmid segregation for the ssDNA plasmids. In the case of pT181, mutations which affect stability all appear to affect replication functions (Gennaro and Novick, 1986; Gennaro and Novick, 1988; Gruss *et al.*, 1987; Iordanescu, 1986). Stability in the case ssDNA plasmids appears to be coupled with replication, rather than with a discrete *par* like function (Gruss and Ehrlich, 1989).

#### 1.14 Incompatibility

The copy number of a plasmid under a given set of conditions is characteristic of that plasmid. The rate of plasmid replication initiation is controlled by the plasmid and because plasmid replication is potentially autocatalytic, replication also has to be autoregulated by an inhibition of the primary replication mechanism (Pritchard *et al.*, 1969). Plasmid incompatibility has been defined as 'the inability of two different plasmids to stably coexist in the same host cell in the

absence of continued selection pressure' (Novick *et al.*, 1976). Plasmid incompatibility is a sign of relatedness and such plasmids share common elements involved in replication or partitioning. It is because of this fact that plasmids compete with each other with the eventual result that one of the plasmids is displaced. Consequently plasmids which are incompatible with each other are assigned to the same incompatibility group in a system which provides a useful method for plasmid classification. The availability of DNA probes for specific incompatibility determinants may be of great assistance in the classification of plasmids (for review see Couturier *et al.*, 1988).

Plasmids of the IncFII and ColE1 groups have been shown to be selected at random apparently without regard for their previous replication history in the process of replication (Gustafsson *et al.*, 1978; Rownd, 1969; Bazaral and Helsinki, 1970), a feature which is consistent with that suggested for the Staphylococcal plasmids pT181 and pC221 (Projan and Novick, 1984). Similarly, plasmids are assumed to be drawn in a random fashion from the plasmid pool in the partitioning process. Consequently, plasmids which are unable to compete effectively against another plasmid may be lost during this randomization process and it is because of this process that certain plasmids appear to be incompatible with each other. Isologous plasmids, which are identical in all regions involved in replication maintenance, can generally be established in a heteroplasmid strain by growth in a selective medium. In the absence of selection, isolates containing only one plasmid are predicted to arise due to the random nature of the partitioning process even if one plasmid is not more competitive than the other whilst

plasmids which are not of the same incompatibility group do not appear to segregate. These predicted patterns were confirmed using the incompatible plasmids pT181 and pSA500 which are isologous plasmids, differing only in their antibiotic resistance markers, and the compatible plasmids pT181 and pC194 carrying tetracycline and chloramphenicol resistance markers (Projan and Novick, 1984).

Attempts to classify the staphylococcal plasmids which have been shown to replicated via a single-stranded DNA intermediate (for review see Gruss and Ehrlich, 1989) based on incompatibility revealed the presence of at least seven distinct incompatibility groups, designated Inc1-Inc7 (Novick *et al.*, 1976). The situation for a number of these plasmids, for which incompatibility against the other plasmids could not be tested since they all specified resistance to chloramphenicol, was later resolved by the use of recombinant plasmids carrying different resistance determinants. The result of this was that three additional incompatibility groups were determined for *S. aureus* plasmids (Iordanescu *et al.*, 1978). These were Inc8, Inc9 and Inc10 having as prototype plasmids pC194, pUB112 and pC223 respectively with pC221 having previously assigned to Inc4.

The staphylococcal plasmid pT181 belongs to the incompatibility group Inc3 and expresses two types of incompatibility termed Inc3A and Inc3B. As stated, plasmids which are incompatible, and therefore belong to the same incompatibility group compete for the same elements involved in plasmid replication or partitioning. As discussed (Section 1.12, pT181 replication is controlled at the level of RepC production and

regulated by two counter-transcript RNAs from the *copA* region which bind to the *repC* mRNA leader. Novick *et al.* (1984) suggested that *copA* was an incompatibility determinant. In a further study, Highlander and Novick (1990) cloned the *copA* region from eleven copy number mutants of pT181 into the compatible plasmid pE194. The ability of these constructs to inhibit the replication of pT181 and its copy number mutants and showed that mutants which altered the sequence or production of the two counter transcripts greatly reduced or eliminated Inc3 activity. The IncB incompatibility determinant was additionally suggested to be the plus origin, *ori*, which is the target of RepC (Highlander and Novick, 1990; Carlton *et al.* 1984).

As discussed, plasmid incompatibility is a sign of relatedness. Plasmids of the ssDNA family have many common features (see Gross and Ehrlich, 1989) and yet members of this family fall into a number of incompatibility groups. This is presumably a function of the relatedness or specificity of the elements in the replication process of this large family of interrelated plasmids. A study by Projan and Novick (1988) suggested, based on hybridization studies, that the *S. aureus* plasmids pC223, pUB112, pS194, pT181, and pC221 may have similar replication proteins and plus origins. An alignment of their sequences showed them to be highly conserved, with the replication proteins of all five plasmids being 63 % identical at the amino acid level and the least similar pair showing 75 % identity. However, despite this homology, the replication proteins and plus origins of these plasmids did not show complementation *in vivo*. Additionally, all five of these plasmids fall into different incompatibility groups, as would be expected in this

instance. Consequently, whilst incompatibility indicates a degree of relatedness, sequence homology suggesting functional similarity is not sufficient to predict incompatibility since relatively small differences are sufficient to alter functional specificity.

#### 1.15 Recombination events in ssDNA plasmids

The mechanism by which ssDNA plasmids replicate, that is to say the production of single-stranded DNA and the rolling circle mechanism, appears in itself to stimulate greatly the levels of homologous and illegitimate recombination compared with that observed in the chromosome. In every recombination process single-stranded DNA is a reactive intermediate and in nearly all steps in the replication of ssDNA plasmids, errors are known or thought to occur.

The initiation and termination steps in the synthesis of plus strand plasmid intermediate have both been suggested to be responsible for the generation of deletion mutants of hybrid plasmids between pC194, pBR322 and the bacteriophage  $\phi 1$  (Michel and Ehrlich, 1986a; 1986b). In such constructs, deletions are thought to occur when a nick is made after erroneous plus origin recognition followed by a round of DNA synthesis and termination at the correct origin sequence. The reverse is also thought to occur where correct initiation is followed by premature termination at a pseudo-origin.

As stated, single-stranded DNA is a reactive intermediate in DNA recombination. During the replication process in ssDNA plasmids, the elongation and displacement of the plus strand generates such a single-stranded piece of DNA and it has been suggested that homologous recombination with the plus strand may occur in this situation (Meselson and Radding, 1975). Indeed the frequency of deletions between long direct repeats in ssDNA plasmids is approximately 1,000 times greater than if a similar structure was present on the chromosome (Figac *et al.*, 1988).

Recombination can also occur during minus strand synthesis. Recombination between 9 bp direct repeats flanking an inactive transposon was found to be 150-1,500 times more frequent in ssDNA plasmids rather than non ssDNA plasmids or the chromosome (Janniere and Ehrlich, 1987). Deletion frequencies between such repeated sequences have also been shown to be proportional to the length of the repeated sequences (Peeters *et al.*, 1988). In *E. coli*, the frequency of recombination between short direct repeats is approximately the same as that occurring between long direct repeats (D. Brunier, as cited by Gruss and Ehrlich, 1989). However, the frequency of recombination between short direct repeats may be  $10^5$ - $10^6$  fold lower in *B. subtilis* suggesting that the copy choice mechanism of recombination, which is active in the deletion of short direct repeats (Brunier *et al.*, 1988), may be less frequent, or better monitored, in *B. subtilis* than in *E. coli* (Gruss and Ehrlich, 1989). In this method of recombination, the DNA replication mechanism is proposed to switch from one copy of one repeat to another on a single-stranded DNA template. The final

recombinant molecule, carrying only one copy of the repeat sequence, is then generated by another round of replication or by cleaving away the single-stranded DNA loop (Albertini *et al.*, 1982). Further details of recombinational events involving repeated sequences are given in Chapter 6.

#### 1.16 Outlook

A greater understanding of the genetics of replicons which are functional in Gram-positive bacteria may in the long-term lead to a better understanding and further exploitation of this group of organisms. If an organism is to be useful in the production of biologically important products, it must possess certain specific properties. It must be amenable to genetic manipulation, adaptable to various nutritional and physical conditions of growth, be genetically stable, be able to produce large quantities of the desired product, and not be pathogenic or toxic to humans (Doi *et al.*, 1986). Most attention has been focussed on *B. subtilis* which possesses most of these characteristics and has the capability of secreting specific foreign proteins relatively free of other cellular proteins although a greater understanding of the events causing both structural and segregational instability in *Bacillus subtilis* will be required if this organism is to be exploited to its full potential.

The use of *B. subtilis* in the production of biochemically or medically important products has no practical or psychological barrier



since it has been used historically in the production of foodstuffs (Debatov, 1982). The use of other *Bacillus* species such as *B. thuringiensis* and *B. cereus* may however pose further problems as they are more closely related to the pathogen *B. anthracis*. The exploitation of other species may have advantages. Recombinant DNA technology in the case of *B. thuringiensis* may yield protein toxins with new host specificities or by gene fusion with the toxin gene provide a mechanism by which recombinant proteins could be released into the medium. The experience gained in the use of *B. subtilis* should find application in the further exploitation of this interesting group of organisms.

Methodology for the cloning, analysis, manipulation, and propagation of DNA sequences has been well established in *E. coli* and so attempts to introduce recombinant sequences into *Bacillus* species have often been made using bifunctional or shuttle vectors capable of functioning in both of these organisms. Many instances in which deletions have occurred when chimeric plasmids have been introduced into *B. subtilis* and which are thought to be the consequence of the replication processes of this group of plasmids. A greater understanding of these processes is the primary aim of this thesis.

## Chapter 2

### Materials and Methods

#### 2.1 Bacterial strains

The following bacterial strains were used:

*E. coli* DH1 (Hanahan, 1983)  
F<sup>-</sup>, *recA1*, *endA1*, *gyrA96*, *thi-1*, *hsdR17(r<sub>K</sub>m<sub>K</sub>)*, *supE44*,  $\lambda^{-}$

*E. coli* CSH26<sup>-</sup>F6 (Jones and Holland, 1984)  
*ara*, *thi*,  $\lambda^{(recA-srl)}$ ,  $\lambda^{(lac-pro)}$ , *Str<sup>R</sup>*; *Sup0*

*E. coli* TG2 (See Sambrook et al., 1989)  
 $\lambda^{(lac-pro)}$ , *supE*, *thi*, *hsdD5*, F<sup>'</sup>*traD36*, *proA<sup>+</sup>D<sup>+</sup>*, *lacI<sup>q</sup>*,  
*lacZ(M15)*

*B. thuringiensis* subsp. *kurstaki* HD1-Dipel obtained from Microbial Resources Ltd.

*B. thuringiensis* subsp. *israelensis* strain IPS-78/11 (Ward and Ellar, 1983)

*B. subtilis* 168 obtained from the National Collection of Industrial Bacteria.

#### 2.2 Plasmids used in this study

Plasmid	Characteristics	Phenotype	Reference
pBR322	Intermediate copy no. cloning vector	Ap <sup>R</sup> , Tc <sup>R</sup>	Bolivar <i>et al.</i> (1977)
pUC8	High copy no. cloning vector	Ap <sup>R</sup>	Veira & Messing, (1982)
pUC9	High copy no. cloning vector	Ap <sup>R</sup>	Veira & Messing, (1982)

pLG338	Low copy no. cloning vector	Km <sup>R</sup> , Tc <sup>R</sup>	Stoker <i>et al.</i> , (1982)
pC194	Antibiotic resistance plasmid	Cm <sup>R</sup>	Horinouchi & Weisblum, (1982b)
pHD2	<i>B. thuringiensis</i> 2 kb plasmid		McDowell & Mann, Submitted
pDM100	pBR322/pHD2 fused at <i>Cla</i> I site	Ap <sup>R</sup>	This study
pDM200	pBR322/pHD2 fused at <i>Hind</i> III site	Ap <sup>R</sup>	This study
pCC100	pBR322 + pC194 Cm <sup>R</sup> gene	Cm <sup>R</sup> Ap <sup>R</sup>	This study
pCC200	pCC100/pHD2 fused at <i>Cla</i> I site	Cm <sup>R</sup> Ap <sup>R</sup>	This study
pCC300	pCC200/pHD2 fused at <i>Hind</i> III site	Cm <sup>R</sup> Ap <sup>R</sup>	This study

### 2.3 Commonly used biochemical buffers

#### TES buffer

Tris-HCl	pH 8.0	30 mM
EDTA	pH 8.0	5 mM
NaCl		50 mM

#### TE buffer

Tris-HCl	pH 8.0	10 mM
EDTA	pH 8.0	1 mM

#### 10 x TBE running buffer

Tris base	108 g
Boric acid	55 g
EDTA 0.5 M (pH 8.0)	40 ml

20 x SSC 1 litre

Na <sub>3</sub> Citrate	87.5 g
NaCl	175.6 g

The solution was adjusted to pH 7.0 using citric acid powder

SSPE 20 x

NaCl	174.0 g
NaH <sub>2</sub> PO <sub>4</sub> .2H <sub>2</sub> O	27.6 g
EDTA	7.4 g

Distilled water was added to a final volume of 1 litre with the pH being adjusted to 7.4 with NaOH

#### 2.4 Culture media and growth of bacterial strains

*Bacillus thuringiensis* and *Bacillus subtilis* were grown at 30°C in a broth described by Singer and Rogoff (1968) and referred to as SR broth.

SR Broth

Yeast extract	5.0 g
Tryptone	5.0 g
Glucose	1.0 g
K <sub>2</sub> HPO <sub>4</sub>	0.8 g

Distilled water was added to 1000 ml and the pH adjusted to 7.2 with 1 M (w/v) NaOH. Solid medium was prepared by the addition of 15 g l<sup>-1</sup> agar.

*Escherichia coli* was grown in Luria Bertani medium referred to as LB broth.

## LB Broth

Tryptone	10.0 g
Yeast extract	5.0 g
NaCl	10.0 g

Distilled water was added to 1000 ml and the pH adjusted to 7.5 with NaOH and the broth sterilized by autoclaving. Solid medium was prepared by the addition of 15 g l<sup>-1</sup> agar.

## Spizizen's minimal medium

Used in the preparation of *B. subtilis* competent cells

0.5 %	Glucose
1.4 %	K <sub>2</sub> HPO <sub>4</sub>
0.6 %	KH <sub>2</sub> PO <sub>4</sub>
0.072 %	MgSO <sub>4</sub>
0.2 %	(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>
0.19 %	Na <sub>2</sub> citrate.2H <sub>2</sub> O
50 µg ml <sup>-1</sup>	Tryptophan

The medium was prepared as a 10 fold concentration in the absence of magnesium and tryptophan filter sterilized and stored at -20°C until use. Additionally, 0.02 % casamino-acids (Difco) and 50 µg ml<sup>-1</sup> histidine was added prior to use.

## Hershey Salts

Used in the preparation of maxicells

NH <sub>4</sub> Cl	1.1 g
CaCl <sub>2</sub> .2H <sub>2</sub> O	15.0 mg
MgCl <sub>2</sub> .6H <sub>2</sub> O	0.2 g
FeCl <sub>3</sub> .6H <sub>2</sub> O	0.2 mg
Tris-HCl	12.1 g

Distilled water was added to a final volume of 1000 ml with the pH being adjusted to 7.4. The pH was adjusted to 7.4.

**M9 Salts 10x stock**

Used in the preparation of maxicells

$\text{Na}_2\text{HPO}_4$	60 g
$\text{KH}_2\text{PO}_4$	30 g
$\text{NaCl}$	5 g
$\text{NH}_4\text{Cl}$	10 g

Distilled water was added to a final volume of 1000 ml with the pH being adjusted to 7.4.

**Hershey Medium**

Used in the preparation of maxicells

Glucose	20.0 %	2.0 ml
Proline	2.0 %	1.0 ml
Thiamine-HCl	0.1 %	0.1 ml
Hershey salts		100.0 ml

**K-medium**

Used in the preparation of maxicells

Sterile distilled water		90 ml
M9 Salts	20 x	5 ml
Casamino acids	20.0 %	5 ml
Glucose	20.0 %	2 ml
$\text{CaCl}_2$	50.0 mM	200 $\mu\text{l}$
$\text{MgSO}_4$	1.0 M	200 $\mu\text{l}$
Thiamine-HCl	0.1 %	10 $\mu\text{l}$

**2.5 Preparation and storage of antibiotics**

Where appropriate, antibiotics were used at the following concentrations unless otherwise stated.

Antibiotic	Stock concentration	Final concentration
	mg ml <sup>-1</sup>	µg ml <sup>-1</sup>
Ampicillin	25	100
Kanamycin	25	50
Chloramphenicol	30	5
Tetracycline	10	10

## 2.6 Maintenance and storage of bacterial strains

Bacterial strains were grown overnight on appropriate agar plates and stored at 4°C until required. The plates were restreaked from a single colony every 4-5 weeks. For longterm storage, glycerol was added to broth cultures and aliquots stored at -20°C.

## 2.7 Preparation of plasmid DNA

### 2.7.1 *Escherichia coli*: Small scale

Small quantities of plasmid DNA for identification or screening of recombinant plasmids were prepared by a modification of the alkaline lysis method described by Birnboim and Doly (1979). The method used was as follows:

## Solution 1

Glucose	50 mM
EDTA	10 mM
Tris-Cl (pH8.0)	25 mM

## Solution 2

NaOH	0.2 M
SDS	1.0 %

## Solution 3

Potassium acetate 5 M	60.0 ml
Glacial acetic acid	11.5 ml
Distilled water	28.5 ml

The pH of this solution was approximately 4.8.

1. A 10 ml LB broth containing the appropriate selective antibiotics was inoculated with a single bacterial colony and incubated at 37°C with shaking overnight.

2. 1.5 ml of the overnight culture was transferred to a 1.5 ml Eppendorf tube and centrifuged for 1 minute at full speed in an Eppendorf centrifuge. The medium was removed from the bacterial pellet by aspiration.

3. The bacterial pellet was resuspended in 200  $\mu$ l of solution 1.

4. The suspension was incubated at 37°C for 5 minutes.

5. 400  $\mu$ l of freshly prepared solution 2 containing 0.2 M NaOH and 1 % SDS were added and the contents of the tube were mixed by gentle inversion before being placed on ice for 5 minutes.

6. 300  $\mu$ l of an ice-cold solution of potassium acetate was added and the tube inverted to mix the contents. The tube was placed on ice for a further 5 minutes.



7. The tube was centrifuged for 10 minutes and the supernatant transferred to a new tube.
8. 15  $\mu$ l of DNase-free RNase (20 mg ml<sup>-1</sup>) was added to the cleared lysate which was then incubated at 37°C for 30 minutes.
9. Further purification was by 2 phenol/chloroform extractions and a single chloroform extraction.
10. The plasmid DNA was precipitated by the addition of 600  $\mu$ l of propan-2-ol followed by incubation on ice for 5 minutes and centrifuged for 10 minutes in an Eppendorf centrifuge.
11. The DNA pellet was washed with 70 % ethanol followed by further centrifugation for 5 minutes.
12. The tubes were capped with parafilm in which small holes were made using a syringe needle and the DNA pellets dried under vacuum for 5 minutes.
13. 10 -20  $\mu$ l of TE was generally added to the DNA pellet which was then left to resuspend for 1 hour at room temperature.
14. Restriction endonuclease digestions were generally performed on 5-10  $\mu$ l of the DNA solution. The remainder of the preparation was stored at -20°C.

#### 2.7.2 *Escherichia coli*:- Large scale

1. A 500 ml LB broth in a 2 litre flask was inoculated with a 10 ml seed culture and incubated at 37°C with shaking until late log/stationary phase was attained. Where the plasmid being purified was pBR322-based 75 mg of chloramphenicol powder was added and incubation

continued for a further 14-16 hours. Under these conditions chromosomal DNA replication is halted and plasmid replication continues leading to an amplification of the amount of plasmid DNA which can be purified.

2. The bacterial cells were harvested by centrifugation of the cell suspension at 10,000 rpm for 10 minutes in a 6 x 250 ml rotor in an MSE HS18 centrifuge.

3. The supernatant was removed by aspiration and the cells resuspended in 6.5 ml of solution 1 containing 5 mg ml<sup>-1</sup> lysozyme as described previously and transferred to an Oakridge centrifuge tube. Incubation was at 37°C for 10 minutes.

4. 13 ml of solution 2 was added to the tube and the contents mixed by gentle inversion. The tube was stored on ice for 10 minutes.

5. 6.5 ml of ice cold solution 3 as described above was added to the tube and the contents mixed by gentle inversion. The tube was stored on ice for 15 minutes.

6. The lysate was cleared by centrifugation of the tube in an 8 x 50 ml rotor at 15,000 rpm in an MSE HS18 centrifuge.

7. 0.6 volumes of propan-2-ol was added to the cleared lysate. After mixing by inversion the tube was incubated on ice for 15 minutes.

8. The DNA was pelleted by centrifugation for a further 10 minutes at 15,000 rpm. The pellet was rinsed with 70 % ethanol; if this was done carefully the ethanol could be poured off without the need of a further centrifugation step.

9. The pellet was dried briefly under vacuum. 5 minutes was generally sufficient.

10. The DNA pellet was taken up in TE to a final volume of 4.2 ml.

11. 4.7 g of caesium chloride was added to the solution and mixed by gentle swirling until fully dissolved.

12. 0.5 ml of a 10 mg ml<sup>-1</sup> solution of ethidium bromide in distilled water was added and after mixing by gentle swirling left to stand at room temperature for 20 minutes to allow protein complexes to form.

13. The protein complexes were pelleted by a further centrifugation step at 10,000 rpm for 15 minutes. The cleared supernatant was loaded into a 5 ml Beckman quickseal tube using a 5 ml syringe and needle. A less efficient method was sometimes used in which the protein complexes were removed by loading the mixture directly into the quickseal tubes using a syringe in which a small cotton wool plug had been placed. This had the effect of filtering out the majority of the protein complexes

14. The tubes were sealed according to the manufacturers instructions and centrifuged at 55,000 rpm for 16 hours or for 4 hours at 65,000 rpm at room temperature in a Beckman VTi65 rotor.

15. 2 bands were generally visible after centrifugation. Irradiation with long wave ultra-violet light was not always necessary. The lower of the two bands containing covalently closed circular plasmid DNA. After piercing the top of the tube this band was removed using a syringe into a 1.5 ml Eppendorf tube.

16. If necessary, the sample was rebanded by making up the volume with 1 g ml<sup>-1</sup> caesium chloride in TE and repeating the centrifugation step at 55,000 rpm for 4 hours.

17. The ethidium bromide was removed by repeated extractions against isoamyl alcohol. Generally 3 extractions were sufficient to remove all traces of pink colouration.

18. The caesium chloride was removed by dialysis of the samples against TE. This was carried out at 4°C in the cold room overnight with at least one change of buffer.

19. The DNA was precipitated with ethanol and taken up in a suitable volume of TE giving a final concentration in the region of 1 µg µl<sup>-1</sup>. The DNA solution was stored at -20°C until required.

### 2.7.3 *Bacillus subtilis* and *Bacillus thuringiensis* plasmid preparations

The preparations of plasmids from *Bacillus* species were basically the same as the methods described for *E. coli* with the following modifications:

1. Cells were grown to an OD<sub>650</sub> of 1.0-1.5 in SR broth.
2. Lysozyme was used at 10 mg ml<sup>-1</sup>.
3. The concentration of SDS in solution 2 was increased from 1 % to 2 %.

### 2.7.4 *B. thuringiensis* protoplast lysis plasmid preparation

This was carried out as described by Eckhardt (1978) and modified by Gonzalez *et al.* (1981) and was designed to prevent the shearing of high molecular weight plasmids.

#### Lyszyme mixture in TES buffer

Lyszyme	2 mg ml <sup>-1</sup>
Sucrose	20 %
RNaseA	100 µg µl <sup>-1</sup>

#### SDS mixture in 1x TBE

SDS	2 %
Sucrose	5 %
Bromophenol blue	0.05 %

A loopful of cells was taken from an overnight plate and resuspended in 50 µl of lyszyme mixture and incubated at 37°C for 30 to 120 minutes. The generation of sphaeroplasts was monitored by phase-contrast microscopy. 20 µl of the SDS lysis mixture was transferred to each of the wells of the agarose gel and allowed to stand for 20 minutes. 10 µl of the sphaeroplast mixture was then pipetted under this layer and electrophoresis carried out at 15 volts for 90 minutes before being increased to 60 volts overnight when using the large gel format (400 ml agarose). The gel did not contain ethidium bromide, but was subsequently stained by soaking in a 1 µg ml<sup>-1</sup> solution to allow visualization of the DNA.

#### 2.8 Quantitation of DNA

A sample of the DNA was made up in TE. The concentration was then calculated from the absorbance of the solution according to the following equation.

$$OA_{260} \times \text{dilution factor}^{-1} \times 50 = \mu\text{g ml}^{-1} \text{ DNA.}$$

The OA<sub>280</sub> was also checked and the OA<sub>260</sub>/OA<sub>280</sub> ratio calculated. This value should be 1.8. Where the value was found to be significantly lower, the sample was assumed to be contaminated with protein or phenol and the DNA concentration could not be measured using this method.

Alternatively, a sample of the DNA and a sample of known concentration, normally lambda DNA restricted using *Hind*III were run on a mini-gel containing ethidium bromide. The different molecular weight bands in the lambda *Hind*III track are present in equimolar amounts and therefore represent different amounts of DNA. Comparison of the fluorescence of the test with the standards generally made a good estimation of DNA concentration to be made. Where the DNA to be quantified was a plasmid a dilution series of pBR322 of known concentration was used for comparison since different DNA forms have different capacities to bind ethidium bromide.

## 2.9 Restriction endonuclease digestion of DNA

Restriction endonucleases were obtained from Amersham International and Gibco-BRL. Restrictions were carried out according to the manufacturers recommendations using the supplied restriction buffers. A typical restriction digest was set up as follows:

DNA	2 $\mu$ l (1 $\mu$ g ml <sup>-1</sup> )
Restriction buffer (10x)	2 $\mu$ l
Sterile distilled water	15 $\mu$ l
Restriction enzyme	1 $\mu$ l (10 units)

All restriction digests were carried out at 37°C for 1 hour unless otherwise recommended by the manufacture eg. *TaqI* digests were performed at 65°C.

Lambda DNA restricted with *HindIII* was used as a molecular size marker and used to determine the molecular size of other DNA fragments

#### 2.10 Treatment of DNA with Calf Intestinal Phosphatase

Calf Intestinal Phosphatase (CIP) was obtained from BCL. This enzyme was used to remove the 5' phosphate from vector DNA molecules prior to ligation. This has the effect of preventing self-ligation of vector DNA and so dramatically increases the proportion of recombinant molecules obtained.

The restriction digest mixture, normally containing 2 µg of restricted vector DNA in a volume of 20 µl, was increased in volume to 100 µl using TE following incubation. 1 µl of CIP (BCL) was then added and incubation continued at 37°C for a further 30 minutes. The DNA was then purified using phenol/chloroform and precipitated with ethanol. The DNA was usually taken up in 20 µl of TE buffer and stored at -20°C until required.

## 2.11 Ligation of DNA

T4 DNA ligase was obtained from Amersham International and Gibco-BRL and was used according to the manufacturers recommendations. T4 DNA ligase from Amersham International was used in conjunction with the following ligation buffer whereas that of Gibco-BRL was used with the supplied buffer.

### Ligation buffer (5x)

Tris-HCl pH7.5	250 mM
MgCl <sub>2</sub>	50 mM
2-mercaptoethanol	50 mM
ATP	5 mM

All ligations were carried out overnight at 15°C. A typical ligation reaction contained the following:

Vector DNA	1 µl (1 µg ml <sup>-1</sup> )
Insert DNA	1 µl (40 ng ml <sup>-1</sup> )
Ligation buffer (5x)	2 µl
Sterile distilled water	5 µl
T4 DNA ligase	1 µl (2.5 units)

The amounts of vector and insert DNA were as advised by Maniatis *et al.* (1982).

## 2.12 Transformation of *Escherichia coli*

The method used for the transformation of *E. coli* was a modification of the calcium chloride method described by Mandel and Higa (1970).



1. Generally a 10 ml culture was inoculated with 200  $\mu$ l of an overnight culture of the strain to be transformed. This was incubated at 37°C with shaking.
2. When the OD<sub>550</sub> had reached 0.5 the cells were chilled briefly on ice and harvested by centrifugation at 5,000 rpm for 5 minutes in an MSE Microcentaur centrifuge.
3. The cells were resuspended in 5 ml of ice cold CaCl<sub>2</sub> and left on ice for 15 minutes.
4. The cells were pelleted again and resuspended in 1 ml of ice cold CaCl<sub>2</sub> (50mM) and left on ice until required. Leaving the cells for 24 hours gives maximum competence.
5. The DNA to be transformed was added to 100  $\mu$ l of competent cells mixed briefly by vortexing and incubated on ice for 15 minutes.
6. The cells were incubated at 42°C for 2 minutes.
7. 900  $\mu$ l of LB broth was added to the cells and the suspension incubated at 37°C for between 30 minutes and 2 hours to allow expression of antibiotic resistance.
8. Aliquots of the transformation mix were plated out onto LB agar plates containing the selective antibiotic at the required concentration. Generally 100-200  $\mu$ l were used per plate with the remainder being stored at 4°C overnight.

#### 2.13 Transformation of *B. subtilis* competent cells

*B. subtilis* competent cells were prepared by the method of Bott and Wilson (1967). Spizizen minimal medium with the addition of 50  $\mu$ g

ml<sup>-1</sup> histidine, 50 µg ml<sup>-1</sup> tryptophan and containing extra magnesium (0.072 % (w/v) anhydrous MgSO<sub>4</sub>) was inoculated from an overnight plate to an OD<sub>500</sub> of 0.1. The flask was incubated at 30°C with shaking and the growth monitored. Competence was attained three hours after the culture had left exponential phase.

Transformation of the competent cells was achieved by mixing 100 µl of the DNA mixture (10 ng µl<sup>-1</sup>) with 900 µl of the competent cells and incubating at 37°C in a shaking water bath for 45 minutes. 1 ml of double concentration Penassay broth (Difco) was then added and incubation continued for a further 60 to 90 minutes. Transformants were selected for by plating on antibiotic plates.

#### 2.14 Electroporation of *Bacillus thuringiensis*

Protocols for the transformation of *Bacillus thuringiensis* species have tended to be complex and not readily reproducible. Electroporation is however a relatively simple procedure in which a high voltage electric discharge through a cell suspension results in a transient permeabilization of the cell membrane allowing entry of DNA into the cells. The efficiency, however, appears to be highly strain dependent (Bone and Ellar, 1989).

Electroporation was carried out using a Bio-Rad Gene Pulser<sup>TM</sup>, using the method described by Bone and Ellar (1989). Briefly, 100 ml of LB broth were inoculated from an overnight plate and incubated at 30°C

with shaking at 200 rpm until an  $OD_{600}$  of approximately 0.5 was attained. The cells were harvested, washed in ice-cold sucrose electroporation buffer (SPB) and resuspended in 5 ml of SPB. Alternatively the cells from a plate incubated at room temperature for 14-16 hours were harvested in SPB and used. 0.8 ml of the cell suspension was chilled on ice in a 0.4 cm cuvette and 0.5  $\mu$ g of DNA in TE added. The cell suspension was then pulsed with the machine set at 25  $\mu$ F and 2 kV. This generally gave a discharge time of 2.5  $\mu$ sec depending on the cell density. The cell suspension was added to 5.6 ml of LB broth and incubated at 30°C to allow expression of antibiotic resistance. The cells were plated out on LB agar containing the appropriate selective agent and incubated overnight at 30°C.

In these experiments an isolate of *Bacillus thuringiensis* subsp. *israelensis* IPS-78/11 kindly supplied by D.J. Ellar was used. This strain has been cured of most of its low molecular weight plasmids (Ward and Ellar, 1982) and is transformed at high frequency by the electroporation technique (Bone and Ellar, 1989).

#### Sucrose Electroporation Buffer (SPB)

Sucrose	272mM
Sodium phosphate buffer pH 7.4	7 mM
MgCl <sub>2</sub>	1 mM

## 2.15 Treatment of DNA by phenol

The following method was used to remove protein contamination or restriction endonucleases from DNA solutions where required before subsequent manipulations were carried out.

### 2.15.1 Preparation of phenol

1. An approximately 50 % v/v mixture of phenol was prepared in 0.5 M Tris-HCl pH8.0. After being well mixed by shaking the solution was allowed to stand and the aqueous and organic phases to separate.
2. The aqueous phase was removed by aspiration and the organic phase extracted against a further volume of 0.5 M Tris-HCl pH 8.0.
3. The aqueous phase was again removed and the organic phase extracted twice with equal volumes of TE buffer. The TE saturated phenol was stored in the dark in the cold room.

### 2.15.2 Extraction with phenol

1. The solution to be purified was made up to a suitable volume such as 100  $\mu$ l or more with TE and an equal volume of phenol added.
2. The solution was mixed by flicking or, where the risk of shearing the DNA was less or unimportant, by vortexing.
3. The phases were separated by centrifugation in an Eppendorf centrifuge for 1 minute at full speed and the upper aqueous phase

transferred to a fresh tube, taking care not to disturb the protein precipitated at the interface.

4. The extraction was repeated using a 50 % (v/v) mixture of phenol and chloroform/isoamyl alcohol 24/1 (v/v) until the interface became clear. A final extraction was then carried out with chloroform/isoamyl alcohol (24/1 v/v) to remove traces of phenol, and the sample precipitated with ethanol.

## 2.16 Precipitation of DNA

DNA was precipitated by the addition of 1/10 volume of 3 M sodium acetate and 2 volumes of absolute ethanol (stored at  $-20^{\circ}\text{C}$ ). After mixing, the solution was stored at  $-20^{\circ}\text{C}$  for 2 hours or overnight. The DNA was then pelleted by centrifugation at 15,000 rpm for 10 minutes in an Eppendorf centrifuge or equivalent. The pellet was rinsed in 70 % (v/v) ethanol and the centrifugation step repeated for 5 minutes. After removal of the ethanol, the pellet was dried under vacuum and taken up in a suitable volume of TE buffer.

## 2.17 Agarose gel electrophoresis

DNA fragments were generally separated on 1 % (w/v) agarose (type 11 Sigma) / TBE gels. Plasmid profiles from *Bacillus thuringiensis* were resolved on 0.5 % (w/v) agarose / TBE gels. The gel apparatus used was

obtained from BRL Ltd. The gel bed measured 25 x 20 cm and was generally filled with 400 ml of agarose solution.

The gel was prepared by boiling 400 ml of 1 x TBE and an appropriate amount of agarose until a clear solution was obtained. The solution was allowed to cool to approximately 50°C and poured into the gel bed the ends of which had been sealed with autoclave tape. An appropriate 20 well comb supplied by Gibco-BRL was inserted at one end to form wells for the application of the samples. 1/10 volume of loading dye (0.25 % bromophenol blue in 50 % glycerol) was added to the samples to be run before loading. The gels were run at 70 volts overnight until the loading dye approached the end of the gel.

Alternatively, small samples were run on a mini-gel format holding 50 ml agarose solution. Electrophoresis was generally at 150 volts for 1 - 2 hours. Due to the speed of running the quality of the resolution was not as good as that obtained from the large gel format. However, this system was more appropriate to check results before proceeding with the experiment. The availability of combs with small teeth enabled smaller amounts of DNA to be visualized. Ethidium bromide was generally included in the gel allowing instant visualization of DNA bands on the UV-transilluminator.

## 2.18 Visualization of DNA in agarose gels

DNA was visualized by the fluorescence of ethidium bromide bound to the DNA upon irradiation with ultra-violet light. Ethidium bromide was either added to the gel before pouring to a final concentration of  $0.5 \mu\text{g ml}^{-1}$  or the gel soaked in distilled water to which ethidium bromide had been added to the same concentration. A stock solution of ethidium bromide was made at  $10 \text{ mg ml}^{-1}$  and stored in the dark. Gels were examined over a long wave ultra-violet light box or photographed on a short wave ultra-violet light box using Polaroid P665 positive/negative film.

## 2.19 Purification of DNA from agarose gels

Where a DNA sample was to be recovered from an agarose gel, the region containing the sample was excised in as small a volume of agarose as possible. The DNA sample was then recovered by elution into a well containing 7.5 M ammonium acetate using an electroeluter (IBI) according to the manufacturers recommendations. The DNA sample was recovered by ethanol precipitation and stored in TE buffer.

## 2.20 Southern blotting procedure

The method used for blotting DNA from agarose gels onto nitrocellulose/nylon membranes was essentially as described by Southern (1975) and is described as follows:

### 2.20.1 Preparation of the gel

1. Depurination to facilitate the transfer of large DNA fragments was performed by a 15 minute wash in 0.25 M HCl. The gel was then rinsed 3 x in distilled water. If the fragments to be transferred were less than 10 kb this step was omitted.
2. Denaturation of the DNA was achieved with a 30 minute incubation in 0.5 M NaOH / 1.5 M NaCl.
3. Neutralization of the gel involved 2 x 30 minute washes in 0.5 M Tris-HCl (pH 7.4) 1.5 M NaCl.

### 2.20.2 Preparation of the filter

A piece of nitrocellulose (Schleicher and Schuell) or nylon blotting membrane (Amersham Hybond-N) was cut to the same size as the gel. Using gloves and tweezers the membrane was floated on distilled water until completely wetted from beneath and then submerged. The membrane was then transferred to 3 x SSC.



### 2.20.3 Southern blotting procedure

A double layer of 3MM (Whatman) paper wetted with 20 x SSC was placed over a support in a tray of 20 x SSC to act as a wick. The gel was inverted and placed on this and surrounded with Cling film to ensure that the buffer passed through the gel. The blotting membrane was placed on the gel and 2 sheets of 3MM paper the same size as the membrane placed on this. Care was taken not to trap any air bubbles in this system. A pile of paper towels was placed on top of the sandwich and a glass plate and weight placed on top of the pile. The blot was left overnight to ensure complete transfer of DNA from the gel to the membrane. The membrane was washed for 15 minutes in 3 x SSC without shaking and allowed to dry on a sheet of 3MM paper. The DNA was fixed to the membrane by baking for 2 hours at 80°C under vacuum for nitrocellulose or by irradiating face down on a long wave ultra-violet transilluminator for 5 minutes with nylon membranes.

### 2.21 Preparation of radio-labelled DNA probes

#### 2.21.1 Nick translation

The following method, which is based on that of Rigby *et al.* (1977), was used to radio-label DNA for use in hybridization experiments.

1. A dilution of DNase I was prepared to give a final concentration of 0.1  $\mu\text{g ml}^{-1}$  and used immediately.

2. The following reaction mix was prepared on ice:

Sterile distilled water	64.5 $\mu$ l
Nick-translation buffer	10.0 $\mu$ l
dNTP stock	7.5 $\mu$ l
DNA	4.0 $\mu$ l (0.5 $\mu$ g)
DNaseI freshly diluted	3.0 $\mu$ l
[ $\alpha$ - <sup>32</sup> P] dCTP	

3. The mixture was incubated at 12°C for 10 minutes.
4. 1  $\mu$ l of DNA polymerase I was added and the incubation continued for a further 75 minutes.
5. The reaction was stopped by the addition of 150  $\mu$ l of 10 mM Tris-HCl pH8.0, 1 mM EDTA pH8.0, 0.1 % (w/v) SDS. The labelled DNA was separated immediately on a sephadex G50 (Pharmacia) column.

**dNTP stock**

dGTP 10 mM	1 $\mu$ l
dATP 10 mM	1 $\mu$ l
dTTP 10 mM	1 $\mu$ l
Sterile distilled water	27 $\mu$ l

**Nick-translation buffer**

Tris-HCl pH7.8	0.5 M	5 ml
2-mercaptoethanol		1 ml
MgCl <sub>2</sub>	0.5 M	1 ml
Bovine serum albumin	50 mg ml <sup>-1</sup>	100 $\mu$ l

**2.21.2 Purification of Nick translated DNA**

A column of G50 fine sephadex was prepared in a siliconized glass Pasteur pipette plugged with siliconized glass wool. The sephadex was allowed to swell in a large volume of distilled overnight and

equilibrated against TE before use. The column was washed through with TE + 0.1 % SDS before use.

The DNA sample was added to the top of the column and allowed to run in. 150  $\mu$ l aliquots of TE + 0.1 % SDS were added to the top of the column and the fractions collected in separate Eppendorf tubes. Two peaks were generally obtained, the first being the radiolabelled DNA and the second being the unincorporated nucleotides. The peaks were either monitored using a hand-held radiation monitor or scintillation counter. The fractions of the first peak were pooled and stored at -20°C until use.

## 2.22 Hybridization of Southern blots

The following protocol was used to hybridize nick-translated DNA to DNA to homologous DNA immobilized on filters.

### Prehybridization solution 100 ml

SSPE	20 x	25.0 ml
SDS	20 % (w/v)	0.5 ml
Denhardt's solution	100 x	5.0 ml
Salmon sperm DNA	10 mg ml <sup>-1</sup>	1.0 ml
Distilled water		68.5 ml

### Hybridization solution 100 ml

SSPE	20 x	25.0 ml
SDS	20 % (w/v)	0.5 ml
Denhardt's	100 x	2.0 ml
Salmon sperm DNA	10 mg ml <sup>-1</sup>	1.0 ml
Dextran sulfate	50 % (w/v)	20.0 ml
Distilled water		51.5 ml

1. The filter to be hybridized was wetted using 3 x SSC and sealed in a plastic bag with a suitable volume of prehybridization solution, preheated to 65°C, taking care not to trap any air bubbles. Generally, 100 µl of prehybridization solution was used for each cm<sup>2</sup> of filter.
2. The filter was left to shake at 65°C for at least 4 hours.
3. The prehybridization solution was removed and replaced with the hybridization solution preheated to 65°C. Generally 50 µl of hybridization solution was used cm<sup>-2</sup> of filter.
4. The labelled probe, which had been boiled for 10 minutes to denature the DNA, was added to the hybridization solution and the bag sealed again taking care not to trap any air bubbles.
5. Shaking was continued overnight at 65°C overnight.
6. The filter was removed from the bag and placed in 500 ml of washing solution prewarmed to 65°C and shaking continued until the unhybridized probe had been removed. This was judged using a hand-held radiation monitor. Several changes of 0.1 % SSC (w/v) + 0.1 % SDS (w/v) were generally used where the probe was highly homologous to the DNA on the filter. Where this was not the case, the stringency of the wash could be reduced by increasing the SSC concentration accordingly.
7. The filter was allowed to dry on 3MM paper (Whatman), wrapped in Cling film and autoradiographed overnight with intensifying screens at -80°C.

**Denhardt's solution**

Ficoll 400,000	2.0 g
Polyvinyl pyrrolidone	2.0 g
Bovine serum albumin	2.0 g

Distilled water was added to a final volume of 500 ml,  
the solution filter sterilized and stored at -20°C.

**2.23 Radiolabelling of plasmid-encoded polypeptides using  
the Maxicell procedure**

A method to analyse the polypeptides expressed from *E. coli* plasmids was described by Sancar *et al.* (1979). Briefly, the plasmid was transformed into an *E. coli* strain incapable of UV repair. The principal of the technique is that, following irradiation with short-wave ultraviolet light, the chromosome degrades and polypeptides synthesized after this point are plasmid encoded and identified by the addition of  $^{35}\text{S}$  methionine. These labelled polypeptides are then analyzed using SDS-PAGE and identified by autoradiography. The method used for this analysis was as follows.

1. The plasmid to be analyzed was transformed into the *E. coli* strain CSH26<sup>-</sup>F6.
2. 15 ml of k-medium with antibiotics was inoculated with 750  $\mu\text{l}$  of an overnight culture, grown to an optical density of 0.5 at 600 nm and placed immediately on ice.
3. 10 ml of the cells were irradiated with approximately 50  $\text{J m}^{-2}$  short wave UV in a Petri-dish.

4. The cell suspension was then transferred to a clean Universal tube and cycloserine, made up fresh in sterile distilled water, added to a concentration of  $200 \mu\text{g ml}^{-1}$ .
5. Incubation was then continued for a further 14-16 hours at  $37^{\circ}\text{C}$  with gentle shaking.
6. The cells were washed twice in 5 ml of Hershey salts (Morice and Burigi, 1974) and finally resuspended in 5 ml of Hershey medium containing  $200 \mu\text{g ml}^{-1}$  cycloserine.
7. Incubation was continued at  $37^{\circ}\text{C}$  with shaking for 1 hour.
8.  $30 \mu\text{Ci}$  of  $^{35}\text{S}$  methionine was added and incubation continued for a further hour.
9. The cells were harvested and washed twice in 10 mM Tris-HCl pH8.0. The cells were taken up in 100  $\mu\text{l}$  of sample buffer and placed in a boiling water bath for 5 minutes.
10. The samples were stored at  $-20^{\circ}\text{C}$  until analysis on 10 % linear polyacrylamide gels was carried out. Generally 12.5  $\mu\text{l}$  of the sample was used.

#### 2.24 Polyacrylamide Gel Electrophoresis of proteins.

10 % linear polyacrylamide gels were used to resolve polypeptides which had been radiolabelled using the maxicell procedure previously discussed. The apparatus used in this procedure has been previously described by Porter (1984).

### 2.24.1 Gel preparation

A 10 % gel mixture was prepared and degassed for 30 minutes prior to use. The polymerization reaction was started by the addition of 12  $\mu$ l TEMED (N,N,N',-tetramethylethylenediamine) and 120  $\mu$ l ammonium persulphate solution (10 % w/v) immediately before use. The solution was transferred into the assembled casting plates using a 25 ml pipette and overlaid with 1 ml of butan-2-ol until polymerization was complete. The gel surface was washed thoroughly with distilled water to remove all traces of Butan-2-ol and the internal plate space dried, using filter paper taking care not to disrupt the gel surface.

#### 10 % gel solution: (50.0 ml)

High bisacrylamide stock	8.30 ml
Distilled water	34.90 ml
Lower gel buffer	6.25 ml
10 % (w/v) SDS	0.50 ml

A stacking gel solution was then prepared and degassed for 30 minutes before use. The polymerization reaction was initiated by the addition of 5  $\mu$ l TEMED and 100  $\mu$ l ammonium persulphate (10 % w/v). The mixture was layered onto previously set gel and a Teflon comb inserted taking care not to trap any air bubbles. The gel was allowed to set for at least a further 30 minutes before use. The gel was assembled into a gel tank which was then filled with 1 x running buffer. The wells were rinsed thoroughly using a syringe before use.

## Stacking gel solution: (10 ml)

Stacking gel acrylamide	3.0 ml
Distilled water	4.4 ml
Stacking gel buffer	2.4 ml
10 % (w/v) SDS	0.1 ml

The following stock solutions were prepared and used during the course of this procedure.

## 60 % (w/v) high bisacrylamide acrylamide

Acrylamide	60.0 g
Bisacrylamide	1.6 g
Distilled water to final volume	100.0 ml

## 10 % stacking gel acrylamide

Acrylamide	10.0 g
Bisacrylamide	0.5 g
Distilled water to final volume	100.0 ml

## Lower gel buffer

Tris base	36.6 g
Distilled water to final volume	100.0 ml
pH adjusted to 8.8 with HCl	

## Stacking gel buffer

Tris base	5.98 g
Distilled water to final volume	100.0 ml
pH adjusted to 6.8 with HCl	

## 10x running buffer

Tris base (25 mM final conc.)	60.4 g
Glycine (194 mM final conc.)	288.0 g
SDS (0.1 % (w/v) final conc.)	20.0 g
Distilled water to final volume	2.0 l



#### 2.24.2 Sample preparation and running of gel

Final samples were generally 100  $\mu$ l final volume in 1 x sample buffer. This was obtained by the addition of 50  $\mu$ l of 2 x sample buffer and a suitable volume of sterile distilled water to the sample giving a final volume of 100  $\mu$ l and a final concentration of 1x sample buffer. The sample buffer used was a modification of that of Haider *et al.* (1986) lacking phenylmethanesulphonyl fluoride.

##### 2x sample buffer

Tris-HCl	1 M (pH7)	0.10 ml
SDS	10 % (w/v)	0.20 ml
EDTA	50 mM (pH8.0)	0.04 ml
Glycerol		0.20 ml
Bromophenol blue	0.05 % (w/v)	0.10 ml
Distilled water		0.32 ml
2-mercaptoethanol (just prior to use)		0.04 ml

The proteins were solubilized by boiling for 5 minutes in a water bath. Insoluble material was pelleted by centrifugation of the sample for 5 minutes at full speed in an Eppendorf micro centrifuge and a suitable volume loaded on the gel. Gels were generally run at 30 mA in the cold room until the marker was approaching the bottom of the gel.

#### 2.24.3 Visualization of polypeptide bands

After running, the gels were stained for between 15-60 minutes depending upon the intensity of staining required in the following solution.

Methanol	45 % (w/v)
Acetic acid	10 % (w/v)
PAGE blue 83	2 g l <sup>-1</sup>

After staining the gels were destained in several changes of the following mixture over a period of time depending upon the initial degree of staining and the clarity of the result required.

Propan-2-ol	10 % (w/v)
Acetic acid	10 % (w/v)

#### 2.24.4 Detection of <sup>35</sup>S methionine-labelled polypeptides by autoradiography

After staining to visualize the molecular weight markers the gels were dried at 80°C under vacuum on a BioRad gel drier and water pump vacuum line. This generally took 1 hour. The dried gel was taped into an autoradiography cassette and a sheet of  $\beta$ -max hyperfilm (Amersham International) taped on top of this. The film and gel were orientated by piercing through the film into the gel beneath in a number of places. Autoradiography was at room temperature overnight initially. Longer exposures were carried out for longer periods depending on the strength of the overnight signal obtained. Autoradiographs were developed in Dektol liquid developer (Kodak Ltd.) and fixed in FX-40 liquid fixer (Kodak Ltd.) according to the manufacturer's instructions.

#### 2.24.5 Estimation of polypeptide molecular weights

Estimations of  $M_r$  were made from a calibration curve obtained by plotting  $\log_{10} M_r$  of known molecular weight standards obtained from Pharmacia against the distance migrated on the gel concerned.

Phosphorylase b	$M_r$ 94,000
Bovine serum albumin	$M_r$ 67,000
Ovalbumin	$M_r$ 43,000
Carbonic anhydrase	$M_r$ 30,000
Soybean trypsin inhibitor	$M_r$ 20,100
$\alpha$ -lactalbumin	$M_r$ 14,400

#### 2.25 Preparation of dialysis tubing

Pieces of dialysis tubing cut to the desired length were boiled for 10 minutes in 2 % sodium bicarbonate and 1 mM EDTA, rinsed with distilled water and boiled for a further 10 minutes in distilled water. The dialysis tubing was stored in distilled water in the cold room until required. The tubing was always handled with gloves and rinsed thoroughly before use.

#### 2.26 M13 Cloning and sequencing techniques

The following techniques were used in the generation of M13 clones for use in sequencing reactions.

### 2.26.1 Preparation of replicative form M13 DNA

The *E. coli* TG2 cells from which the replicative form (RF) of the M13 vectors to be used was to be prepared was a scaled up version (500ml) of that described in the template preparation procedure. However, rather than purifying single-stranded DNA from the supernatant, the intra-cellular double-stranded RF was prepared using the large scale *E. coli* plasmid preparation described.

### 2.26.2 Generation of random clones based on a sonication procedure

The DNA to be cloned was self-ligated prior to sonication so as to yield a random array of fragments. A 1.5 ml Eppendorf centrifuge tube containing the ligation mixture to be treated was fixed 3 mm above the probe of a cup horn sonicator and the bath filled with water. The sample was sonicated for 3 bursts of 30 seconds with the water being replaced between bursts to prevent heating and the sample being collected at the bottom of the tube by centrifugation between bursts. The DNA was size fractionated by agarose gel electrophoresis. The fragments were purified by electroelution and rendered blunt-ended using T4 polymerase I according to the manufacturer's recommendations. The fragments were cloned into M13 vectors in the standard way.

### 2.26.3 Transfection of M13 DNA into *E. coli* TG2

Competent cells of *E. coli* TG2 were prepared using the  $\text{CaCl}_2$  method as previously described. 300  $\mu\text{l}$  of the competent cells were added to the M13 DNA to be transfected and mixed by gently tapping the tube. The cells were then heat-shocked by incubating for 2 minutes at  $42^\circ\text{C}$ . The heat shocked cells were stored on ice until they could be plated out using the X-gal/IPTG overlay agar method.

### 2.26.4 Selection of M13

The entire transfection mix obtained previously was added to overlay agar held at  $42^\circ\text{C}$  to which 50  $\mu\text{l}$  each of X-gal and IPTG had been added. The mixture was mixed briefly by vortexing and used to overlay LBA plates.

#### X-gal (5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactoside)

2 % solution (25  $\text{mg ml}^{-1}$ ) in dimethylformamide. Stored at  $-20^\circ\text{C}$  in the dark

#### IPTG (isopropyl- $\beta$ -D-thiogalactopyranoside)

100 mM (23.8  $\text{mg ml}^{-1}$ ) in water.  
Stored at  $4^\circ\text{C}$

#### Overlay agar

1 g NaCl  
1 g Bacto tryptone (Difco)  
100 ml distilled water

#### 2.26.5 Storage of M13 isolates

Individual M13 plaques were picked into 1.5 ml Eppendorf centrifuge tubes containing ml LB broth using sterile Pasteur pipettes. These stocks may be kept at 4°C for periods of up to 1 year.

#### 2.26.6 Preparation of template DNA

Single-stranded M13 DNA was purified from *E. coli* TG2 isolates essentially according to the method of Bankier *et al.*, (1986). 5 µl of the M13 stock cultures were spotted onto LB plates which had been overlaid with soft agar containing *E. coli* TG2 cells. 2 ml LB cultures, containing a 1/10 dilution of an overnight culture of *E. coli* TG2, in glass test-tubes were inoculated with strips of infected overlay agar from these plates and incubated at 37°C with vigorous shaking (300rpm) for 6 hours.

1.5 ml of the culture was transferred to a 1.5 ml Eppendorf centrifuge and centrifuged in an MSE Microcentaur centrifuge at high speed for 15 minutes. 1.2 ml of the supernatant was transferred to a clean Eppendorf tube and 300 µl 20 % PEG/2.5 M NaCl added. The contents of the tube were mixed by inversion and incubated for 20 minutes at room temperature. The DNA was recovered by centrifugation for a further 10 minutes at high speed. All traces of liquid were removed with a drawn out Pasteur pipette.

The viral pellet was taken up in 100  $\mu$ l TE buffer and extracted once with 50  $\mu$ l TE-saturated phenol with vigorous vortexing. The organic and aqueous phases were separated with a further 5 minutes centrifugation at 15,000 rpm and 80  $\mu$ l of the supernatant transferred to a clean Eppendorf tube. The DNA was precipitated by a standard ethanol precipitation and taken up in 20  $\mu$ l of TE.

#### 2.26.7 Sequencing reactions

The sequencing reactions were performed using a dideoxynucleotide sequencing kit (Gibco-BRL) according to the manufacturer's instructions. Additional nucleotides (Sigma) and Klenow fragment of *E. coli* DNA polymerase I (Gibco-BRL) were obtained as required. Additional sequence data was obtained using the extended sequencing protocol of Stambaugh and Blakesley (1983).

#### 2.26.8 Gel electrophoresis

Denaturation of the DNA in the sequencing reaction mix was achieved by the addition of 2  $\mu$ l of formamide dye mixture and heating at 95°C for 5 minutes. The sample was cooled rapidly on ice and loaded on the sequencing gel using a drawn out glass capillary. The gels used were 55 cm 6 % polyacrylamide buffer gradient gels which were prepared and run as described in the Gibco-BRL sequencing manual. 1000 cm gels were also used to obtain extended sequence data.

**Formamide dye mixture**

Xylene cyanol	0.1 %
Bromophenol blue	0.1 %
EDTA	10.0 mM
Deionized formamide	95.0 %

**2.26.9 Autoradiography**

The sequencing gels were transferred to a sheet of Whatman 3MM paper and dried in a Bio-Rad gel drier at 80°C for 1 hour. The gels were placed in X-ray cassettes in intimate contact with Fuji RX X-ray film overnight in the first instance. The X-ray films were developed using Kodak LX-24 developer and fixed using Kodak FX-40 fixer according to the manufacturer's instructions.

**2.26.10 Computer-assisted sequence compilation**

The DNA sequence data was read from the autoradiographs by eye. The data was entered into the Microgenie sequence analysis program (Queen and Korn, 1986) running on an IBM XT personal computer. The analysis performed using this program is described in chapter 4. Further computer analysis of the sequence using additional programmes is discussed in chapter 5.



## 2.27 Preparation of cell lysates for ssDNA detection

The method used for the detection of ssDNA plasmid intermediates was that described by Te Riele (1986).

1. A culture of the transformed strain was grown to an  $OD_{650}$  of approximately 1 and the cells from 3 ml of culture harvested and washed in EDTA (0.1 M, pH8.0) / NaCl (0.15 M).
2. The cells were resuspended in 0.5 ml EDTA (0.1 M, pH6.9) / NaCl (0.15 M) and lysostyme added to a concentration of  $10 \text{ mg ml}^{-1}$ . Incubation was at  $37^\circ\text{C}$  for 45 minutes.
3. Sarkosyl was added to a final concentration of 1 % and the lysates incubated at  $65^\circ\text{C}$  for twenty minutes.
4. The samples were extracted with phenol and chloroform and precipitated with ethanol.
5. The samples were taken up in 90  $\mu\text{l}$  of TE buffer and 10  $\mu\text{l}$  of RNase ( $10 \text{ mg ml}^{-1}$  w/v) added. Incubation was at  $37^\circ\text{C}$  for 20 minutes.
6. 50  $\mu\text{l}$  samples were digested with 50 units of S1 nuclease in 1 x S1 buffer (0.28 M NaCl, 0.05 M sodium acetate pH 4.6, 4.5 mM  $\text{ZnSO}_4$ ). at  $37^\circ\text{C}$  for 45 $^\circ\text{C}$ .
7. Samples were extracted with phenol and chloroform, precipitated with ethanol and stored at  $-20^\circ\text{C}$  in TE buffer until required.
8. Samples were analyzed on 1 % agarose gels and the disappearance of single-stranded DNA bands monitored using Southern blotting and hybridization techniques.

### Chapter 3

#### Selection and cloning of a *Bacillus thuringiensis* plasmid

##### 3.1 Introduction

*Bacillus thuringiensis* subspecies contain a substantial proportion of their genetic information in the form of plasmids and exhibit complex plasmid profiles. As well as there being a great deal of variation in plasmid size and number between the subspecies of *B. thuringiensis*, there also seem to be differences in the plasmid profiles obtained from the same subspecies as reported from different laboratories. Lereclus *et al.* (1982) report that *B. thuringiensis* subspecies *kurstaki* HD1 contains 8 plasmids with a size range of 1.5 - 54 MDal whilst Kronstadt *et al.* (1983) report the presence of 12 plasmids of between 1.9 and 120 MDal. Such variations probably reflect, in part, differences in plasmid isolation technique between laboratories and the recovery, in particular, of the higher molecular weight plasmids.

Two isolates of the same subspecies of *B. thuringiensis* subspecies *kurstaki*-HD1 obtained from Bulla and Dulmage have been shown not to have identical plasmid profiles (Kronstadt *et al.*, 1983). Distinction has been made between these with the one obtained from Bulla being named *kurstaki* HD1-Dipel.

It was initially hoped to identify a suitable low molecular weight plasmid for further study. The parameters for the selection of such a plasmid were size and ease of purification. Work on plasmids from other Gram-positive organisms has generally localized minimal replicons, containing all the functions for replication and maintenance, to regions with a size in the region of 1.5 kb. No work on minimal replicons in *B. thuringiensis* has been reported. The analysis of essential functions and sequences on a low molecular weight plasmids in the *B. thuringiensis* plasmid profile would therefore be uncomplicated by DNA sequences of unknown and unessential function. An investigation into the replication mechanisms of such a plasmid, and its possible use in the development of cloning vehicles, suggested that it would have to be available in isolation from the other plasmids of *B. thuringiensis*. Cloning of the whole plasmid in an *E. coli* system was seen as a good way of achieving this and of facilitating further genetic manipulations.

### 3.2 Plasmid profiles

The protoplast lysis procedure, as described by Eckhardt (1978) and modified by Gonzalez and Carlton (1981), has been reported to be the best suited to displaying the entire array of plasmids, including those of high molecular weight. A modification of the alkaline lysis mini-preparation procedure, as described by Birnboim and Doly (1970), was found to give better results in the purification of the lower molecular weight plasmids (Figure 3.1) although the isolation of those of higher molecular weight was found to be inconsistent using this system.

Figure 3.1

*Bacillus thuringiensis* subspecies *kurstaki* HD1-Dipel plasmid profiles.

Total plasmid DNA was prepared from ten identical cultures using the alkaline lysis mini preparation procedure and separated by electrophoresis on a 0.5 % agarose gel. The three duplex bands may, or may not, represent different forms of two plasmids.

In order to perform any analysis of the plasmids, such as their mode of replication and partitioning, the presence of homologous sequences or the assignment of functions to particular sequences, the plasmid or plasmids selected have to be available in a pure form. The isolation of individual plasmid bands from mini-preparation profiles resolved on agarose gels and purified by electroelution could only ever yield small quantities of DNA. By scaling up the size of the plasmid preparation to 500 ml of culture, and introducing a CsCl gradient purification step, highly concentrated plasmid DNA was obtained that was essentially free of RNA and protein contaminants. Using this purification system a large amount of total plasmid DNA could be loaded on agarose gels, giving much improved resolution of the small molecular weight plasmids. The plasmids of intermediate and high molecular weight were not well purified in this system. If future work was to be based on the study of one of the smaller plasmids in isolation, such a plasmid profile should provide a better source of raw material from which individual low molecular weight plasmids could be purified, for example by electroelution (Figure 3.2).

### 3.3 Selection of the 2 kb plasmid

The purification of specific *B. thuringiensis* plasmid DNA's from agarose gels was very time consuming and gave a low yield compared to an *E. coli* replicon propagated in an *E. coli* host. It was decided that the next step should be to clone a *B. thuringiensis* subsp. *kurstaki* HD1-Dipel plasmid into an *E. coli* vector. This would allow larger

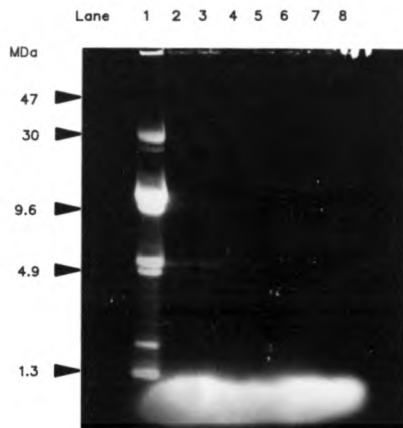


Figure 3.2

Comparison of DNA obtained from *Bacillus thuringiensis* subspecies *kurstaki* HD1-Dipel using small scale and large scale plasmid preparation procedures.

Lanes 2-8 show plasmid DNA from *Bacillus thuringiensis* prepared using the alkaline lysis mini preparation procedure. Lane 1 shows plasmid DNA prepared on a large scale with the inclusion of a caesium chloride/ethidium bromide buoyant density gradient purification procedure. All samples were resolved on 0.5 % agarose gels. Molecular weights were estimated from the profiles of Kronstadt *et al.*, (1983).

quantities of plasmid to be obtained in isolation from other *B. thuringiensis* plasmids. The cloned plasmid could be separated from the *E. coli* vector DNA by agarose gel electrophoresis after restriction endonuclease digestion and purified by electroelution in much greater quantities than would have been possible had it been isolated directly from *B. thuringiensis*.

Minimal replicons containing all the functions required for plasmid replication and partitioning, as previously discussed, have generally been found to be fairly small with a size of perhaps 1.5 kb. It was hoped that the selection of a low molecular weight plasmid would provide a model system for plasmid replication and partitioning in *B. thuringiensis* without the presence of large amounts of DNA of unknown and unessential function to complicate the picture. To this end, the lowest molecular weight plasmid was electroeluted from plasmid profiles resolved on agarose gels using plasmid DNA purified from caesium chloride gradients. This was then screened with a number of restriction endonucleases in the hope of finding a unique restriction site which could be used to clone the whole plasmid in a suitable *E. coli* vector. The selected plasmid, of unknown function, was subsequently determined to be approximately 2 kb in size and was designated pHD2.



### 3.4 Cloning of the 2 kb plasmid

Figure 3.3 shows gel-purified pH2 plasmid DNA which has been digested with a number of restriction endonucleases. Whilst none of the lanes show a single band, as would be expected if a unique site was present for that enzyme, the band pattern in lanes 2, 3, 5, and 6 can be explained as ccc and oc forms of the plasmid by comparison with uncut plasmid DNA on other gels. It therefore seemed likely that there were no restriction sites for any of these restriction enzymes. The presence of an additional band in lane 4, corresponding in size to the 2 kb band in the lambda *HindIII* markers, could be interpreted as the linearised form of the plasmid generated as a result of the presence of a unique *HindIII* site. The lack of complete digestion may result from the carry over of a contaminant from the agarose during electroelution. Alternatively, the lack of detection of restriction sites in the DNA, which had been purified from *B. thuringiensis*, may be the result of host modification rather than merely a lack of that restriction site.

The *E. coli* cloning vector, pUC8, has a number of unique sites available in a polylinker, including *HindIII*. As a cloning vehicle it has a number of features which make it attractive for use in the cloning of the *B. thuringiensis* 2 kb plasmid, pH2. In the host strain of *E. coli*, TG2 in this instance, a mutation in the 5'-region of the gene coding for  $\beta$ -galactosidase (*lacZ*) prevents assembly of four subunits into the functional  $\beta$ -galactosidase. The association of a short polypeptide, corresponding to the N-terminal portion of the enzyme, with the subunits of the enzyme, is able to restore  $\beta$ -galactosidase activity



Figure 3.3

Preliminary restriction analysis of the *Bacillus thuringiensis* subspecies *kurstaki* HD1-Dipel plasmid pHD2.

- Lane 1 Lambda DNA restricted with *HindIII*  
2 pHD2 DNA restricted with *EcoRI*  
3 pHD2 DNA restricted with *PstI*  
4 pHD2 DNA restricted with *HindIII*  
5 pHD2 DNA restricted with *BamHI*  
6 pHD2 DNA restricted with *SalI*

Lane 4 shows the appearance of an extra DNA band following restriction endonuclease digestion with *HindIII* corresponding to the linear form of pHD2. The digest is incomplete and cc and ccc forms of the plasmid may still be seen in this lane.

by  $\alpha$ -complementation allowing the conversion of the colourless substrate '5-bromo-4-chloro-3-indolyl- $\beta$ -D-galactoside' (X-gal) to a blue chromophore (Messing *et al.*, 1977). The vector, pUC8, contains the portion of the *lacZ* gene allowing  $\alpha$ -complementation to occur as described. The introduction of an insert into a site in the polylinker of pUC18 interrupts the *lacZ* gene resulting in a peptide incapable of complementation. Transformants of this type are incapable of producing the blue chromophore from X-gal and consequently appear white allowing positive identification of recombinant isolates. The high copy number of this plasmid (200-400 per cell) means that large quantities of plasmid DNA can readily be purified.

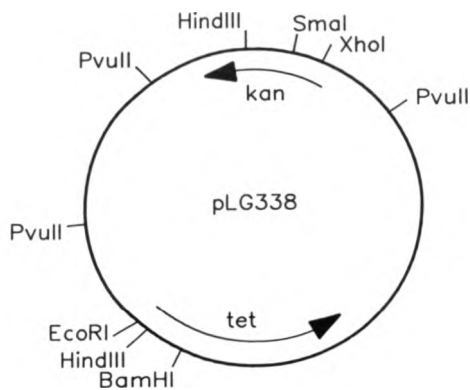
Two recombinant classes were obtained using the pUC8 system, one which grew normally and contained a plasmid not significantly different in size from native pUC8 and one producing microcolonies which could not be subcultured. The vector, pUC8, is a high copy number plasmid and it is possible that the 2 kb *B. thuringiensis* plasmid produces a protein which is toxic to *E. coli* when present at high copy number.

There are a number of genes which are unclonable at high copy number but which have been cloned using low copy number plasmid vectors, as outlined in the discussion. To overcome this problem, a versatile low copy number vector, pLG338, with a copy number of six to eight per chromosome was developed by Stoker *et al.* (1982). This vector is derived from pSC105 and carries genes conferring resistance to tetracycline and kanamycin within which there are a number of unique restriction endonuclease sites. The restriction map of pLG338 (Figure

Figure 3.4

Restriction map of pLG338

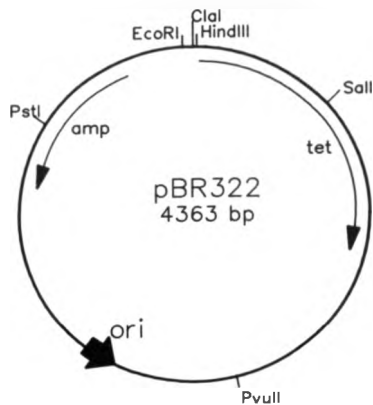
89



3.4] shows that there are two *Hind*III sites in this vector, one interrupting the kanamycin resistance gene and one interrupting the promoter of the tetracycline resistance gene. Deletion of the intervening fragment and its replacement with the target or the original sequence in the reversed orientation would destroy both selectable markers for this plasmid. To overcome this problem, the *B. thuringiensis* plasmid pHD2, which had been linearised using *Hind*III was ligated into the vector, pLG338, which had also been restricted with *Hind*III. Transformation of the products of this ligation into *E. coli* DH1 and the selection of transformants on kanamycin plates ensured that this region had been properly reconstructed. Subsequent screening of these transformants on plates containing tetracycline at  $10 \mu\text{g ml}^{-1}$  revealed which of these isolates were recombinant as those which had an insert in the *Hind*III site, located in the promoter of the tetracycline resistance gene, would no longer be resistant to tetracycline. A number of recombinant forms of this plasmid were identified in this manner. Digestion of pLG388 with *Hind*III gave two fragments of 2.6 and 4.7 kb. Digestion of these tetracycline sensitive pLG338 derivatives with *Hind*III, revealed the presence of an additional 2 kb band on agarose gel electrophoresis. It seemed highly likely that the cloning of the pHD2 plasmid had been successful in this low copy number system.

Having discovered that pHD2 was clonable at low copy number, but not apparently at high copy number, attempts were made to reclone this plasmid in the intermediate copy number vector pBR322. This vector has a copy number of 20 to 40 copies per chromosome and may be amplified using chloramphenicol by a factor of approximately ten. Unlike pLG338, pBR322

Figure 3.5  
Restriction map of pBR322



possesses a unique *HindIII* site making the cloning and purification of the pHD2 sequence more straightforward. The *HindIII* restriction site in pBR322 (Figure 3.5) is located in the promoter of the tetracycline resistance gene, as previously described for pLG338, and so the loss of tetracycline resistance could still be used to identify recombinants with an insertion in this site.

Attempts to clone the sequence in pBR322 yielded a number of kanamycin-resistant/tetracycline-sensitive clones when transformed into *E. coli* DH1. The plasmid DNA was purified from a number of these tetracycline-sensitive isolates by the mini-preparation procedure and analyzed by restriction mapping. Digestion with the restriction enzyme *HindIII* revealed the presence of an additional 2 kb band, when resolved using agarose gel electrophoresis, from a number of these isolates. It seemed highly likely that this was the *B. thuringiensis* pHD2 plasmid sequence which had successfully been cloned in pBR322 at intermediate copy number. One of these clones was selected for further investigation and termed pDM200.

### 3.5 Restriction analysis of the 2 kb plasmid

The availability of a good restriction map is a prerequisite if any genetic manipulation or analysis of a cloned DNA sequence is to be carried out. The determination of a restriction map for the cloned 2 kb plasmid sequence from *B. thuringiensis* subsp. *kurstaki* HD1-Dipel involved two strategies. If the plasmid was to be transformed back into



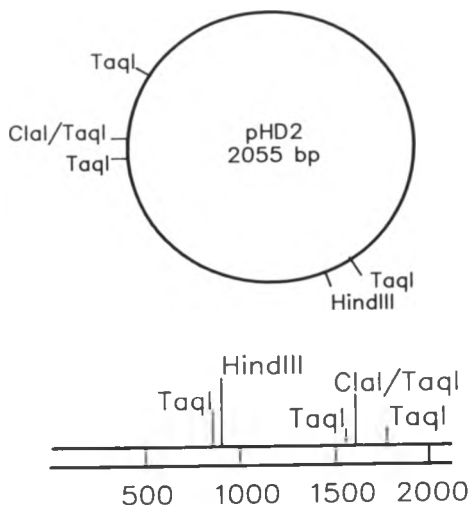
a *Bacillus* species for functional analysis, the presence of an antibiotic resistance marker which was functional in a Gram-positive host was essential. The introduction of such a marker required the availability of a suitable, unique restriction site into which this could be inserted using 'sticky-end' or 'blunt-end' ligation. A possible problem was envisaged at this stage. If there were in fact two restriction sites sufficiently close together that one had not been detected, cleavage and religation of the DNA could result in the inactivation of sequences vital for plasmid replication, either by their interruption or by the generation of a small deletion. The availability of additional unique sites would have allowed alternative strategies to be developed in the hope of circumventing these potential problems.

In the search for unique restriction sites, it was advantageous to perform the analysis directly on pDM200 rather than on purified insert. The detection of a restriction site in close proximity to the *HindIII* terminus of the linear fragment might go undetected if the individual 2 kb sequence was being screened. Performing the analysis on the whole clone would enhance the probability of detection as its presence would either linearise or release a piece of DNA from the clone, depending upon the presence and position of such a site within the pBR322 portion of the clone. This of course depends upon any such restriction site being sufficiently far away from the insert to avoid any additional confusion. The sequence of pBR322 is readily available and so any such ambiguity can readily be predicted.

In practice, the presence of a restriction site for any restriction enzyme could be screened for in this way, however, the fragment pattern becomes more complicated with the presence of multiple sites particularly within the pBR322 portion of the clone. Restriction enzymes which are likely to have multiple recognition sites within the 2 kb insert are those which have a 4 bp recognition sequence. These are expected to occur once every  $4^4$  or 256 bp on average and are expected to occur approximately 17 times in pBR322 which is 4.362 kb in size. In practice, the number of restriction sites is not that predictable. *TaqI*, which has a 4 bp recognition sequence, has only 7 sites in pBR322. In practice, purified 2 kb fragment which had been gel-purified after digestion of the clone with *HindIII* was used for any enzyme which had more than one restriction site within pBR322.

These analyses revealed the presence of a unique *ClaI* site and three *TaqI* sites although the *ClaI* site is also a *TaqI* site as it contains the *TaqI* recognition sequence. Partial digestion of the purified 2 kb sequence with *TaqI* allowed these sites to be mapped with respect to the terminal *HindIII* sites. Digestion of the 2 kb sequence with *ClaI* enabled this site to be mapped with respect to the *HindIII* site and thereby established which of the *TaqI* sites was also a *ClaI* site. The 2 kb plasmid was therefore determined to have three restriction sites for *TaqI*, and only one for each of *HindIII* and *ClaI*. The presence of a fourth *TaqI* site in close proximity to the *HindIII* site was not revealed until a further clone, pDM100 (Section 3.5), had been restriction mapped. The restriction map of pDM200, including the fourth *TaqI* site is shown (Figure 3.6). No sites were found for any of

Figure 3.6  
Restriction map of pHD2



No restriction sites could be detected within pHD2 for any of the following restriction endonucleases:

*AvaI*  
*BamHI*  
*BglII*  
*EcoRI*  
*HhaI*  
*HpaI*  
*PstI*

*PvuI*  
*SacI*  
*Sall*  
*ScaI*  
*SphI*  
*SetI*  
*XhoI*

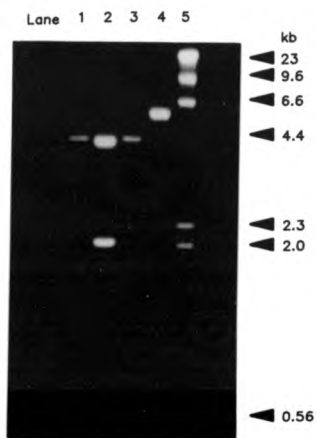


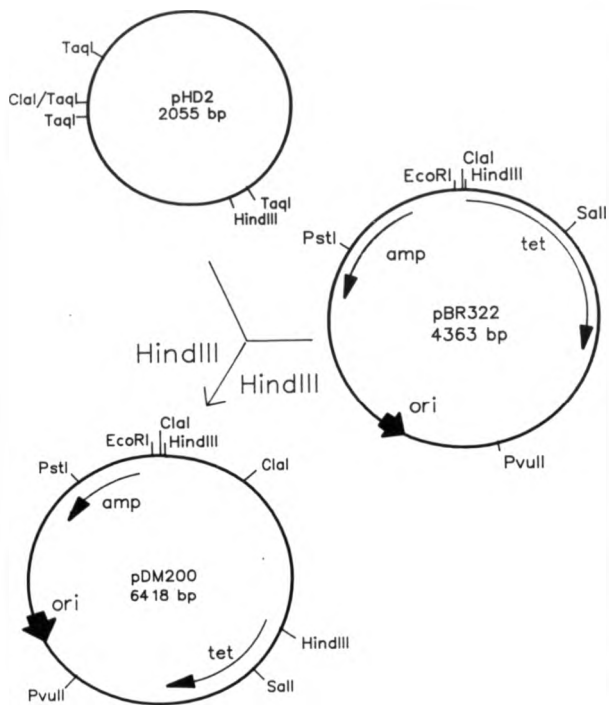
Figure 3.7

Restriction analysis of pDM200

- Lane 1 pBR322 restricted with *Hind*III
- Lane 2 pDM200 restricted with *Hind*III
- Lane 3 pBR322 restricted with *Cla*I
- Lane 4 pDM200 restricted with *Cla*I
- Lane 5 Lambda DNA restricted with *Hind*III

Restriction endonuclease digestion of pBR322 with *Cla*I and *Hind*III is shown to linearize the plasmid, giving a band of 4.3 kb. Digestion of pDM200 with *Hind*III reveals the presence of an additional band of 2 kb corresponding to pHD2 which is orientated by digestion with *Cla*I, see Figure 3.8.

Figure 3.8  
Derivation of pDM200



the restriction enzymes sites listed. The orientation of the 2 kb sequence within pBR322 was identified by determining the position of the internal *Cla*I site in relation to that within pBR322 (Figure 3.7). The restriction map of pDM200 is shown (Figure 3.8).

### 3.6 Alternative cloning of the 2 kb plasmid

As discussed in section 3.4, the cloning of the 2 kb plasmid after it had been linearised at the *Hind*III site may have disrupted a sequence vital for plasmid maintenance or replication either by insertional inactivation, or the generation a small deletion if there was a second *Hind*III site in close proximity to the first. Recloning of the plasmid, linearised at the subsequently detected unique *Cla*I site, would avoid interruption of the sequence at this site. This would allow verification of the fidelity of the previous clone, pDM200, in the region of the *Hind*III site in subsequent sequencing studies. The same purpose would be served by pDM200 with respect to the *Cla*I site.

The pHD2 plasmid, purified by gel electroelution from the *B. thuringiensis* plasmid profile, was cleaved with *Cla*I to give a 2 kb linear molecule. This was ligated into pBR322 which had also been cleaved with *Cla*I from which the 5' phosphate group had been removed using calf intestinal phosphatase (CIP), preventing self-ligation of the vector. The ligated molecules were transformed into *E. coli* DH1 and selected for using ampicillin resistance. Comparison with the control, vector treated with CIP and self-ligated in the absence of an insert

molecule, suggested that all of the ampicillin-resistant colonies obtained were likely to be recombinant. Screening of a number of these colonies using the plasmid mini-preparation method, followed by restriction analysis of the plasmids obtained, showed that all of the colonies tested contained a plasmid giving two fragments when restricted with *Cla*I. These corresponded in size to the 4.3 kb of pBR322 and the 2.0 kb of the pHD2 *B. thuringiensis* plasmid sequence. One of these, in future referred to as pDM100, was selected and purified on a large scale using caesium chloride gradients for further restriction analysis. The presence of an internal *Hind*III site the correct distance from the *Cla*I site as predicted was confirmed (Figure 3.9) and this was used to confirm the orientation of the 2 kb sequence in pDM100 (Figure 3.10). Surprisingly, the *Taq*I restriction pattern obtained upon restriction analysis of the purified 2 kb sequence from this source was similar to that obtained when the 2 kb sequence was purified from pDM100. Previous evidence, based on the restriction patterns obtained from the cloned 2 kb sequence from pDM200 when digested with *Taq*I and resolved on agarose gels, suggested the presence of two *Taq*I sites within this sequence. Restriction analysis of the 2 kb cloned sequence purified from pDM100 was consequently expected to reveal a single, internal *Taq*I site. When this restriction analysis was performed, three fragments were obtained with similar sizes to those previously observed with 2 kb insert purified from pDM200 when resolved on agarose gels. This suggested the presence of an additional *Taq*I site in close proximity to the *Hind*III site. This was later confirmed by the sequencing studies discussed in Chapter 4.



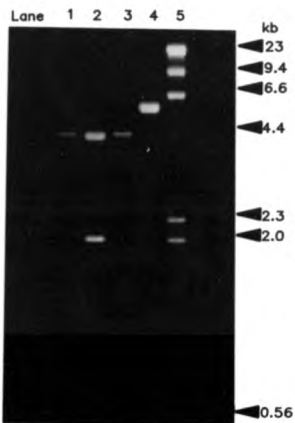


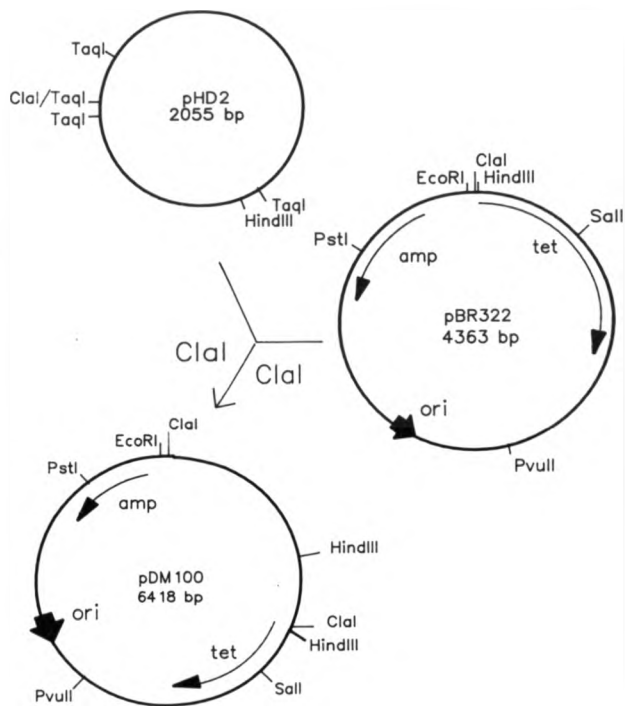
Figure 3.9

Restriction analysis of pDM100

- Lane 1 pBR322 restricted with *Cla*I
- Lane 2 pDM200 restricted with *Cla*I
- Lane 3 pBR322 restricted with *Hind*III
- Lane 4 pDM200 restricted with *Hind*III
- Lane 5 Lambda DNA restricted with *Hind*III

Restriction endonuclease digestion of pBR322 with *Cla*I and *Hind*III is shown to linearize the plasmid giving a band of 4.3 kb. Digestion of pDM200 with *Hind*III reveals the presence of an additional band of 2 kb corresponding to pHD2 which is orientated by digestion with *Hind*III, see Figure 3.9.

Figure 3.10  
Derivation of pDM100



Previous attempts to clone the pKD2 plasmid sequence, generated by restriction with *HindIII*, in the high copy number plasmid, pUC8, had been unsuccessful. The colour selection system for recombinant molecules based on X-gal and IPTG had been used in this instance. It was suggested (D. Whitcome personal communication) that some sequences which appeared to be unclonable in the pUC vectors using this identification procedure were clonable if IPTG was omitted and the *lac* operon repressed by including glucose in the medium. Attempts were made to clone the sequence in pUC8 under these conditions and the desired recombinants identified by restriction analysis of plasmids purified using the mini-preparation procedure after transformation into *E. coli* DH1. Vector which had been treated with CIP was used to reduce the number of transformants needed to be screened. The sequence did prove to be clonable in pUC8, however, as there appeared to be no advantage in using this system as the plasmid sequence did not contain any further restriction sites also present in the pUC8 polylinker, it was decided to proceed using the pBR322 based clones. Screening of the 2 kb plasmid did not reveal the presence of restriction sites for any of the restriction enzymes listed in Figure 3.6, suggesting that previous lack of digestion resulted from the absence of these restriction sites rather than a modification in the DNA from *B. thuringiensis*.

### 3.7 Southern blotting analysis

In order to confirm that the 2 kb sequence which had been cloned and restriction mapped was also present within the *B. thuringiensis*

plasmid profile, it was decided to hybridize the cloned 2 kb sequence against a Southern blot of the *B. thuringiensis* plasmid profile resolved on an agarose gel.

The plasmids were isolated from 10 identical cultures of *B. thuringiensis* subsp. *kurstaki* HD1-Dipel using the mini-preparation procedure, displayed on a 0.5 % agarose gel, and the plasmid profiles transferred to a nylon filter using the Southern blotting technique. Hybridization of the 2 kb sequence, which had been radiolabelled by 'nick translation', to DNA sequences immobilized on the nylon filter revealed that between three and five of the small molecular weight plasmids had sequences related to the cloned *B. thuringiensis* plasmid (Figure 3.11). The difference in the number of bands showing hybridization probably reflects difference in transfer or hybridization efficiency as all the lanes represent identical preparations of the same isolate. The bands showing hybridization on the autoradiograph are masked on the gel photograph by RNA contamination and it is difficult to be certain which of the hybridizing bands corresponds to the cloned plasmid.

In order to confirm the origin of the cloned sequence it was decided to hybridize pHD2, purified from *B. thuringiensis*, to the cloned sequences in pDM100 and pDM200. The clones, pDM100 and pDM200, were restricted with *Cla*I and *Hind*III respectively to release the 2 kb inserts, and these separated from the pBR322 portion of the clone using agarose gel electrophoresis. The DNA sequences were transferred to a nylon filter by Southern blotting and hybridized against the original *B.*

Lane 1 2 3 4 5 6 7 8 9

MDa

4.9



Figure 3.11

A Southern blot of total *Bacillus thuringiensis* subspecies *kurstaki* HD1-Dipel plasmid DNA probed with nick translated pDM200 DNA.

Plasmids were purified from nine identical cultures of *B. thuringiensis* using the alkaline lysis mini preparation procedure and separated on a 0.5 % agarose gel as shown in Figure 3.1.

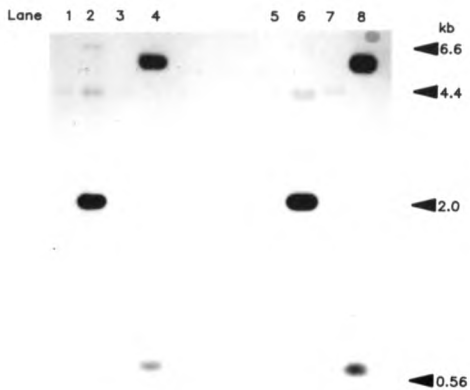




Figure 3.12

A Southern blot of the constructs, pDM100 and pDM200, probed with nick translated pHD2 plasmid DNA.

Lane 1 pBR322 restricted with *Hind*III  
Lane 2 pDM200 restricted with *Hind*III  
Lane 3 pBR322 restricted with *Cla*I  
Lane 4 pDM200 restricted with *Cla*I  
Lane 5 pBR322 restricted with *Cla*I  
Lane 6 pDM100 restricted with *Cla*I  
Lane 7 pBR322 restricted with *Hind*III  
Lane 8 pDM100 restricted with *Hind*III

The molecular size markers were fragments of Lambda DNA restricted with *Hind*III, visualized by ethidium bromide staining and visualization under U.V light prior to Southern blotting.

*thuringiensis* pKD2 plasmid. The 2 kb sequence released from these clones hybridized strongly to the purified *B. thuringiensis* plasmid (Figure 3.12) suggesting that these two sequences were highly related and consistent with their being identical. It appeared from this evidence that the two clones, pDM100 and pDM200, were indeed clones of the *B. thuringiensis* pKD2 plasmid.

### 3.8 Summary and discussion

A low molecular weight plasmid of unknown function was chosen, primarily because of its small size, from the *B. thuringiensis* plasmid profile for further study. The larger plasmids, having more presumed unessential DNA of largely unknown function, were likely to be unnecessarily complicated for investigation in the first instance. The minimal replication regions of a number of Gram-positive replicons have been localized to regions of a similar size to this plasmid for example; pLAB1000 and pLAB200 (Jansson, *et al.*, 1989), pSL1 (Shindoh *et al.*, 1987), pFTB14 (Murai *et al.*, 1987), pE194 (Villafane *et al.*, 1987), pC194 (Alonso and Taylor, 1987), and pUB110 (Maciag *et al.*, 1988).

The plasmid was demonstrated to be 2 kb in size and a preliminary restriction map generated. Only two unique restriction sites were found amongst the commonly used restriction enzymes, *Hind*III and *Cla*I and the plasmid was successfully cloned into these sites in the vector pBR322 for further analysis. The plasmid was also cloned at low copy number in the vector pLG338 but not at high copy number in the plasmid pUC8 unless

the *lacZ* gene used in the colour selection procedure was repressed. Additionally, the presence of two *TaqI* sites were demonstrated and the presence of at least one more deduced by restriction mapping. The cloned sequence was shown to hybridize strongly to the plasmid from which it was thought to have been derived and to have sequences homologous to four other low molecular weight plasmids. It has previously been shown that the low molecular weight plasmids are inter-related as are those of higher molecular weight (Aronson, 1986).

A number of genes have proved to be unclonable at high copy number. These include *polA* (Murray and Kelly, 1979) *ompA* (Beck and Bremner, 1980), and *dnaA* (Hansen and von Meyerburg, 1979). It seems likely that multiple copies of many genes which code for regulatory or membrane proteins are deleterious to the cell.

Assuming no sequences vital to plasmid replication or maintenance in the pHD2 plasmid have been destroyed in the cloning procedure, the two recombinant plasmids, pDM100 and pDM200, provide the basis of shuttle vectors in that they contain elements allowing replication and maintenance in both *E. coli* and *B. thuringiensis*. Ampicillin resistance is available as a selective marker in *E. coli* although this gene is not functional in Gram-positive organisms such as *B. thuringiensis*. The absence of a functional antibiotic resistance marker poses a problem if these clones are to be detected when transformed into a *Bacillus* or other Gram-positive host. Further studies regarding the functional integrity of the cloned *B. thuringiensis* plasmid made it essential to clone a marker such as an antibiotic resistance gene functional in Gram-

positive species into the constructs pDM100 and pDM200. The lack of alternative, unique restriction sites in the cloned *B. thuringiensis* plasmid reduced the number of strategies available in overcoming any problems of insertional inactivation as previously discussed.

## Chapter 4

### Sequencing studies

#### 4.1 Introduction

The sequencing of a piece of DNA can yield valuable information with regard to the presence of restriction endonuclease sites, open reading frames or structural features within that sequence. The comparison of such features with others of known function may give valuable evidence as to the possible role of these sequences. No data was available regarding the 2 kb plasmid, pH<sub>D2</sub>, from *Bacillus thuringiensis* subspecies HD1-Dipel although plasmids from other Gram-positive organisms have been extensively studied. Sequencing of the plasmid pH<sub>D2</sub> was hoped to yield much information about this plasmid and allow comparisons to be made with other plasmids from Gram-positive organisms.

The DNA sequencing protocols used to determine the sequence of the *B. thuringiensis* plasmid, pH<sub>D2</sub>, were based upon the dideoxy chain termination method described by Sanger (1977). In this procedure a DNA strand complementary to that being used as a template is synthesized by the extension of a synthetic oligonucleotide primer annealed in this case just upstream of a multiple cloning site. Termination of the polymerization reaction at specific bases is achieved by the incorporation of the dideoxy-form of the nucleotide into the reaction mixture. An analysis of the products of such reactions after separation

on denaturing polyacrylamide gels and visualization by autoradiography allows the sequence downstream of the polylinker to be determined.

Most sequencing protocols require the use of single-stranded template DNA for use in the reactions. The vector of choice in the cloning of DNA fragments for sequencing is the *E. coli* bacteriophage M13. M13 is a single-stranded DNA phage and, following infection, the M13 (+) strand DNA serves as a template for the synthesis of the complementary (-) strand. The phage is easily purified from the supernatants of infected *E. coli* cultures and is an ideal source of single-stranded DNA which may then be used in the sequencing reactions. M13 DNA has been modified (Messing *et al.*, 1977) to generate a series of derivatives known as the M13mp vectors which allow a number of sequencing strategies to be easily followed. Using this system, vectors carrying an insert are identifiable using a blue/white colour screen as previously described for the pUC series of vectors (Section 3.4). More recent developments have enabled sequencing reactions to be carried out directly from clones in double-stranded plasmids (Chen and Seeburg, 1985) although M13 remains the vector of choice and pieces of DNA are often subcloned into this vector for sequencing. In this instance, the use of a plasmid-based sequencing strategy using the pUC series of vectors was not thought to be a viable proposition as difficulties had previously been encountered possibly resulting from the high copy number of these vectors or high levels of polypeptide expression from them.

The aim of most sequencing strategies is to sequence both strands of the piece of DNA in question using a series of overlapping clones. A

number of strategies exist for generating such a series of clones. Random fragments from the target DNA may be generated by sonication (Deininger, 1983) or DNase treatment (Anderson, 1981) and are ligated into an M13mp vector after being size fractionated and converted to blunt end molecules. Non-random cloning strategies may also be used. A procedure based on a double-stranded exonuclease, *Bal-31*, has been described (Poncz *et al.*, 1982) as has a similar strategy based on exonuclease III (Henikoff, 1984) for use in the generation of a series of consecutive overlapping clones. In a targeted sequencing strategy, individual restriction fragments may be cloned selectively into an M13mp vector.

There are advantages and disadvantages to both random and directed sequencing strategies. A random strategy is likely to generate sequence from all regions of a clone with equal frequency and rapidly gives a large amount of sequence data. As the sequence data continues to be accumulated, it is often noted that whilst some sequences are often repeatedly sequenced, other regions are apparently absent from the available clones. Such a distribution is the result of the random nature of the cloning strategy and this method is therefore unsuitable for determining specific sequences within a larger sequence. A targeted strategy is often considered at this stage to fill in specific missing gaps.

Targeted strategies tend to suffer a reverse set of problems. Repeated cloning procedures of DNA cut with the same restriction enzymes result in the generation of the same clones. Regions proximal to these

restriction sites tend to be extensively sequenced whereas regions distal to these sites are not sequenced. Such strategies are therefore good in determining specific sequence data between selected restriction sites but not regions which are by chance distant from suitable restriction sites.

In deciding which strategy to employ in the determination of the pHD2 sequence, the following points were considered. A *HindIII*, *ClaI*, and two additional *TaqI* sites were available for use in a targeted strategy although the distribution of these sites meant that some areas would be unsequenceable and others difficult to orientate with respect to neighbouring fragments. A random sequencing strategy should rapidly supply sequence data from throughout the sequence possibly accessing regions remote from known restriction sites. Furthermore, the presence of additional restriction sites might be determined which could be employed in a later, directed sequencing strategy. It was therefore decided to pursue a random sequencing strategy in the first instance to generate most of the sequence and to use the improved restriction map data to access any regions which were found, by chance, not to be occurring at a suitable frequency within the available clones.

#### 4.2.1 Random cloning in M13 vectors

The pHD2 plasmid sequence was released from the construct pDM100 using the restriction enzyme *HindIII* and separated from the pBR322 portion of the clone by agarose gel electrophoresis. The 2 kb fragment



was then obtained by electroelution and purified by phenol extraction and ethanol precipitation. The generation of a random series of clones based on sonication relies upon the release of fragments, the ends of which have also generated in a random fashion. A piece of DNA which is linear and therefore has two defined termini does not satisfy this criterion. To avoid the terminal sequences of this piece of DNA being over represented in the population of clones generated, the DNA was self ligated to give a series of circular and multimeric molecules. The DNA was then subjected to sonication resulting in random shearing of the DNA. Small DNA fragments generated during this procedure yield little information on sequence analysis and consequently dramatically increase the number of clones required to be sequenced. To overcome this problem, a size fractionation step was included in which the fragments were resolved on a 1.2 % agarose gel. A slot was cut in the gel just below the position at which DNA fragments of approximately 300 bp were expected to have reached by comparison with a marker lane containing *Hae*III restricted  $\phi$ X174 DNA. The slot was filled with running buffer and electrophoresis continued with the buffer being removed and replaced periodically until all the DNA fragments up to a size of approximately 600 bp had been eluted. The DNA solution removed from the slot was extracted using phenol to remove any contaminants from the agarose and precipitated with ethanol. The ends of the purified DNA fragments were rendered blunt using the 3' and 5' exonuclease activity of T4 DNA polymerase and ligated into *Sma*I-restricted M13mp18 and mp19 which had been treated with calf intestinal phosphatase to prevent self ligation of the vector without the presence of an insert molecule.

The ligated molecules were transfected into *E. coli* TG2 and, under appropriate conditions, recombinant molecules were identified by the production of white plaques as opposed to blue plaques from non-recombinant phage based on the X-gal/IPTG protocol. 62 white plaques were obtained and these were transferred into 1.5 ml LB broths in bijoux using sterile Pasteur pipettes and stored at 4°C until required.

The number of clones required to give a certain probability of covering the whole sequence varies according to number of bases which may be read from each clone, but assuming 250 bases could be read, 50 clones would be expected to yield 95 % of the 2 kb sequence. When most of the sequence has been determined, a targeted strategy is generally considered to be the most efficient method of obtaining the remaining sequence.

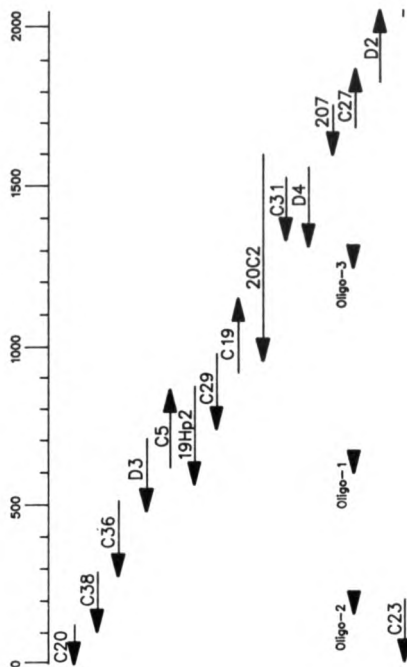
Small scale phage preparations were performed on 48 of the putative clones and single-stranded DNA purified. Sequencing reactions were performed on these DNA samples (Section 2.26.7) and the ladder of DNA fragments produced separated on 6 % polyacrylamide/7.8 M urea TBE gradient gels. [ $\alpha$ -<sup>35</sup>S]-dCTP was used as the labelling nucleotide and the DNA fragments visualized by autoradiography. The sequence data was read by eye and entered into the DNA sequence analysis program of Queen and Korn (1984). M13 sequences at the beginning or end of the sequences entered were detected and deleted before further analysis was carried out. Difficulties were encountered in obtaining readable sequence of more than 80-100 bases. Additionally more than 50 % of the clones sequenced showed only very small inserts. Attempts were made to

increase the amount of sequence obtained from clones which appeared to contain larger inserts by reducing the proportion of ddNTP to dNTP in order to extend the period over which termination was occurring. This appeared to have no significant effect on the amount of sequence obtained. The quality of the Klenow fragment of DNA polymerase 1 being used in the reaction can have an effect on the efficiency of the polymerization reaction. The use of a different batch of Klenow enzyme failed to solve the problem. Attempts to join the fragments of sequence obtained into longer contiguous sequences gave little encouragement and consequently it was decided to proceed with the directed sequencing strategy earlier than had previously been envisaged.

Further work on directional clones as described (Section 4.2.2) employed an extended sequencing protocol (Stambaugh and Blakealey, 1983). Following success using this protocol, further attempts were made to sequence the randomly generated clones. Using this procedure it was possible to increase the amount of sequence data available from these clones to 2.595 kb. This was eventually determined to correspond to 926 bases on one strand and 742 bases on the other and represented 1588/4110, approximately 40 %, of the final sequence. The role of these sequences in the determination of the final sequence is shown diagrammatically (Figure 4.1)

Figure 4.1

Synthetic oligonucleotide primers and M13 clones  
used in the determination of the DNA sequence of pHID2



#### 4.2.2 Directed cloning in M13 vectors

Discrete fragments of DNA may be cloned directly into compatible sites within the M13 polylinker for sequencing. The selection of such fragments based on the restriction map may give a more directed strategy in the sequencing of a given piece of DNA. In order to give a greater sense of direction it was decided to pursue such a directed sequencing strategy based upon the limited number of restriction sites known to be present within the pDH2 plasmid sequence which were compatible with those sites present within the M13 polylinker. A number of M13 vectors are available which contain different restriction sites in their polylinkers or which contain the same polylinker in the opposite orientation. Where a selected fragment is excised using two restriction enzymes producing non-compatible ends these are clonable only in a given orientation in the selected polylinker. By selecting such a pair of M13 vectors the DNA may be sequenced in both orientations. By sequencing both strands of defined fragments it was hoped to speed up progress since it would be known which pieces of sequence should overlap and approximately how much sequence data would be required to achieve this.

The cloned 2 kb *B. thuringiensis* plasmid sequences were gel-purified after excision from pDM100 using *Hind*III and pDM200 using *Cla*I. DNA fragments were produced using the restriction enzymes *Taq*I and *Cla*I and *Hind*III in single and double digests as appropriate. The rationale behind using both of these cloned sequences was that both were obtained initially from pDH2 using either *Cla*I or *Hind*III and so could be used to confirm the sequence across the other restriction site used in the

original cloning process and confirming that there had not initially been two sites present which would have resulted in the generation of a small deletion in cloning process. The ends generated using *TaqI* are compatible with those obtained using *AccI*. DNA fragments were cloned into M13 mp18 and mp19 which had been cleaved with the appropriate restriction enzyme or pairs of enzymes. M13 vectors cleaved with single restriction enzymes were treated with calf intestinal phosphatase to prevent self ligation of the vector in the absence of an insert. This was not carried out where a double digestion was being used to force directional cloning since the ends generated in this process were incompatible. The number of clones possible in these ligations varied between 2 and 6. Portions of the ligation mixtures were transfected into *E. coli* TG2 as previously described (section 2.26.3) with the remainder being stored at  $-20^{\circ}\text{C}$ . Between 10 and 30 representatives of these clones were picked into LB broth in the first instance and stored at  $4^{\circ}\text{C}$ . Single-stranded template DNA was prepared from these and sequencing reactions performed. Where certain sequences appeared not to be represented amongst the clones, the transfection was repeated and further clones screened. When specific fragments or orientations were still observed to be absent, the appropriate fragments were gel purified and attempts made to clone these selectively. In order to maximize the amount of sequence obtained from these clones and find overlapping regions, use was made of the extended sequencing protocol as described (Stambaugh and Blakesley, 1988). Using this system, 943 nucleotides of sequence information was obtained.

Despite repeated attempts to obtain the 0.6 kb and 1.2 kb *TaqI* fragments in both orientations in M13 mp18 and mp19 only one of the orientations was obtained. In order to overcome this problem, an attempt was made to obtain the 0.6 kb region as a *Clai/HindIII* fragment and the 1.2 kb region as a 1.4 kb *HindIII/Clai* fragment in both orientations using M13mp18 and mp19. Whilst it was possible to clone the 0.6 kb *Clai/HindIII* fragment in M13 mp18, it was not possible to clone it in mp19. In order to maximize the amount of sequence data obtained from the available 0.6 kb and 1.2 kb clones it was decided to synthesize 3 oligonucleotide primers (Figure 4.2) to anneal downstream from the usual sequencing primer towards the terminus of the previously determined sequence.

The use of the synthetic primer Oligo-3 enabled the sequence of the whole of this fragment to be determined in one direction. Attempts to gain the other strand by forced directional cloning in mp19 or by the use of random clones generated using sonication were unsuccessful and it would appear that the cloning of the sequence in this orientation was for some reason problematical. This fragment would contain all of ORF-B (See Chapter 5) and whilst the start of the gene is 300 bp downstream from the *HindIII* site it may be that a high level of transcription is being achieved from the *lacZ* promoter with subsequent translation causing a toxic effect. Deletion events often occur in fusion molecules containing DNA sequences replicating via a rolling circle mechanism such as ssDNA plasmids and M13 phage. This is thought to result from the aberrant recognition of origin sequences. An alternative explanation for the apparent unclonability of this DNA sequence in this orientation

Figure 4.2

Synthetic oligonucleotides used as sequencing primers

		Position	Strand
Oligo-1			
5'	AATATCCATACCATAAA	3' 612-596	Lower strand
Oligo-2			
5'	CATTTGTACACGATT	3' 197-181	Lower strand
Oligo-3			
5'	TATAACATCAGTTTAAA	3' 1234-1250	Lower strand



is that a deletion may be occurring during replication either removing the inserted sequence or rendering the phage inviable.

In a similar manner, the 1.4 kb *ClaI/HindIII* fragment could be cloned in M13 mp18 but not mp19. Attempts to gain the whole of the sequence from this clone involved the use of two oligonucleotide primers Oligo-1 and Oligo-2. Oligo-1 yielded good sequence data downstream from the initial sequencing start site whilst Oligo-2 which should have given sequence data even further downstream yielded no further information. This region was eventually sequenced using a clone generated from a blunt end ligation (C20) which confirmed the sequence of the synthetic oligonucleotide to be correct. In the opposite orientation, a number of short random clones enabled the beginning of the fragment to be sequenced whilst the cloning of the larger fragment in the same orientation did not appear to be possible. The reasons for the apparent unclonability of certain DNA fragments still remains unknown.

As previously discussed, a problem of sequencing clones obtained by the use of restriction endonucleases is the difficulty in accessing regions distant from such restriction sites. The 0.6 kb and 1.2 kb sequences released using *TaqI* gave an extra problem since they appeared to be clonable in only one orientation. To overcome this problem, two oligonucleotide primers were synthesized corresponding to regions downstream from the start of these clones and used to extend the amount of sequence data obtainable. Using these primers, the amount of sequence information obtained from the directional clones was increased from 943 bases to 1458.

Using the protocol outlined it should have been possible to obtain overlapping clones covering the whole sequence. It was hoped that sequences sufficiently distant from a suitable restriction site and which could be easily obtained from these clones could be determined using the blunt end clones previously described. Where appropriate, primers specific to previously determined sequences downstream from the M13 primer, were synthesized chemically and used to determine the DNA sequence further downstream from this position and which was otherwise unobtainable. The cloning and sequencing strategy which was used to determine the sequence of pH2 is shown (Figure 4.1). By combining the sequence data obtained from these directional clones and the extended sequences obtained using the additional primers, with further data from the random clones using the extended sequencing protocol, it was possible to merge the sequences into one contiguous sequence. From this, the size of the plasmid was determined as being 2055 bp of which approximately 60 % had been sequenced in both orientations. Attempts to extend the sequence from the available clones into regions unsequenced in both orientations using the extended sequencing protocol and additional oligonucleotide primers suggested that there may have been some reason why these regions were unsequenceable in a particular direction such as secondary structure leading to sequence pile-ups. The 2055 bp circular plasmid sequence was subjected to a number of forms of analysis as described in the following sections and chapter 5.

#### 4.3 Computerized analysis of the sequence

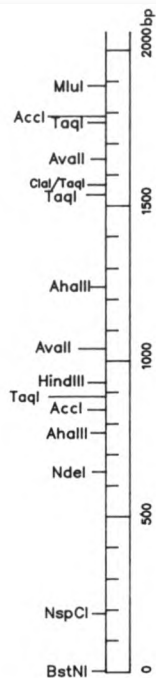
The final contiguous nucleotide sequence of the *B. thuringiensis* plasmid, pHD2, with an actual size of 2055 bp was compiled into a double-stranded circular molecule using the DNA sequence analysis program of Queen and Korn (1984). This program was then used to determine some of the structural features of the sequence. The DNA sequence of pHD2 and some of its features is shown in Figure 4.4.

##### 4.3.1 The new restriction map

The Microgenie sequence analysis programme is capable of carrying out a search for specified sequences. By using this facility in conjunction with a file specifying restriction endonuclease recognition sites, a detailed restriction map was generated (Figure 4.3). The presence of unique *Hind*III and *Cla*I sites was confirmed as was the presence of 4 *Taq*I sites. The absence of a number of other restriction sites such as *Bam*HI, *Eco*RI, *Pst*I and *Sal*I within the sequence was confirmed. Preliminary restriction analysis had previously suggested that this was the case but did not rule out the possibility of a host modification system preventing digestion of the DNA which had been purified directly from *B. thuringiensis*. The sequence data therefore confirmed the previous restriction analysis data.

Figure 4.3

Improved restriction map of pHD2



#### 4.3.2 %A+T and %G+C analysis

It has been proposed that the nucleotide frequency found within a bacterial genome forms an integral part of its classificational attributes (Williams *et al.*, 1989) and in a similar way is often reflected in the plasmids found in a given host. The A+T content of the plasmid was determined as being 65.4 % (Figure 4.4). Because of this feature, it would be expected that restriction enzymes with a similar A+T content in their recognition sequence would cut more frequently within this sequence than others which had a high G+C content. Another feature of the A+T/G+C ratio results from the degeneracy of the genetic code. Using a triplet codon, 4<sup>3</sup> or 64 combinations are possible using the four nucleotide bases. 61 of these represent amino acids codons with the remaining 3 being stop codons. There are in fact only 20 amino acids which occur in protein and as a result of degeneracy of the genetic code, options are available for all of these except methionine and tryptophan (Charnay *et al.*, 1979), (Figure 4.5). It has been noted that different codons coding for the same amino acid are not used with equal or random frequencies. It is possible that the use of degenerate codons may affect, or may be used to control, mRNA expressivity and a strong correlation has been made in *E. coli* between the relative abundance of tRNAs and codon choice (Grantham *et al.*, 1980). Codon selection may also be influenced by differences in codon-anticodon binding strengths and tRNA availability within the cell (Bennetzen and Hall, 1982; Blake and Hinds, 1984; Sharp *et al.*, 1986). In *Streptomyces coelicolor* the anticodon Leu<sup>TTA</sup> is only found to be expressed in secondary metabolism. Organisms with a high GC content in



Figure 4.4**The DNA sequence and preliminary analysis of pHD2**

The sequence contains 2055 nucleotides with the following distribution

	No.	(%)
A	707	(24.4)
C	224	(10.9)
G	488	(23.7)
T	636	(30.9)
A+T	1343	(65.4)
C+G	712	(34.6)

No. (%)	No. (%)	No (%)	No (%)
AA 254 (12.4)	CA 73 (3.6)	GA 190 (9.2)	TA 190 (9.2)
AC 85 (4.1)	CC 30 (1.5)	GC 49 (2.4)	TC 60 (2.9)
AG 162 (7.9)	CG 57 (2.8)	GG 93 (4.5)	TG 176 (8.6)
AT 206 (10.0)	CT 64 (3.1)	GT 156 (7.6)	TT 210 (10.2)

Figure 4.5

## The genetic code

Amino acid	Code	Codons						
Alanine	Ala	A	GCA	GCC	GCG	GCU		
Arginine	Arg	R	AGA	AGG	CGA	CGC	CGG	CGU
Aspartic acid	Asp	D	GAC	GAU				
Asparagine	Asn	N	AAC	AAU				
Cystine	Cys	C	UGC	UGU				
Glutamic acid	Glu	E	GAA	GAG				
Glutamine	Gln	Q	CAA	CAG				
Glycine	Gly	G	GGA	GGC	GGG	GGU		
Histidine	His	H	CAC	CAU				
Isoleucine	Ile	I	AUA	AUC	AUU			
Leucine	Leu	L	UUA	UUG	CUA	CUC	CUG	CUU
Lysine	Lys	K	AAA	AAG				
Methionine	Met	M	AUG					
Phenylalanine	Pha	F	UUC	UUU				
Proline	Pro	P	CGA	CCC	CCG	CCU		
Serine	Ser	S	AGC	AGU	UCA	UCC	UCG	UCU
Threonine	Thr	T	ACA	ACC	ACG	ACU		
Tryptophan	Trp	W	UGG					
Tyrosine	Tyr	Y	UAC	UAU				
Valine	Val	V	GUA	GUC	GUG	GUU		
	Stop	U	UAA	UAG	UGA			



their genome have a high percentage of G and C in the third-codon position whilst AT rich sequences have a high percentage of A and T in this position, for example the *E. coli* bacteriophage  $\phi$ X174, favours the use of codons with thymine in the third position of the triplet (Sanger *et al.*, 1978). The other two positions which are much more constrained by the amino acid sequence, differ much less in GC content between GC-rich and GC-poor organisms. Additionally, a strong link has been observed between codon usage and tRNA abundance in many organisms (Ikemura, 1981a,b, 1985, Bulmer 1987).

Two open reading frames are present within the 2055 bp sequence (Section 4.2.3). The distribution of A+T at the first second and third positions within the codons in ORF-A is 60.5 %, 72.2 %, 75.8 % (Figure 4.6). The maximum frequency of A+T, occurs at the third position in the codons on ORF-A. This is as would be expected in an A+T rich sequence since the degeneracy of the genetic code allows more variation at this base which is therefore more likely to be an A or T in an A+T rich sequence. In ORF-B, the distribution is 60.0 %, 78.7 %, 61.2 % with the highest frequency occurring at position 2 rather than position 3. However, the distribution of stop codons within the alternate reading frames in this region suggests that this is not the result of a mistake in the sequence resulting in a frame shift. The open reading frame is small coding for only 80 amino acids and the apparent difference in base distribution may result from differences in amino acid composition rather than any difference in codon bias (Figure 4.7). Alternatively, differences in codon usage may provide a means of regulating gene expression based on tRNA availability.

Figure 4.6

## Codon usage in ORF-A

TTT Phe 9 (4.0)	TCT Ser 6 (2.7)	TAT Tyr 16(7.1)	TGT Cys 0 (0.0)
TTC Phe 1 (0.4)	TCC Ser 0 (0.0)	TAC Tyr 0 (0.0)	TCT Cys 0 (0.0)
TTA Leu 5 (2.2)	TCA Ser 2 (0.9)	TAA End 0 (0.0)	TGA End 1 (0.4)
TTG Leu 5 (2.2)	TCG Ser 3 (1.3)	TAG End 0 (0.0)	TGG Trp 2 (0.9)
CTT Leu 0 (0.0)	CCT Pro 3 (1.3)	CAT His 8 (3.6)	CGT Arg 2 (0.9)
CTC Leu 0 (0.0)	CCC Pro 0 (0.0)	CAC His 0 (0.0)	CGC Arg 0 (0.0)
CTA Leu 1 (0.4)	CCA Pro 0 (0.0)	CAA Gln 4 (1.8)	CGA Arg 1 (0.4)
CTG Leu 0 (0.0)	CCG Pro 2 (0.9)	CAG Gln 2 (0.9)	CGG Arg 0 (0.0)
ATT Ile 10(4.5)	ACT Thr 3 (1.3)	AAT Asn 9 (4.0)	AGT Ser 7 (3.1)
ATC Ile 0 (0.0)	ACC Thr 1 (0.4)	AAC Asn 2 (0.9)	AGC Ser 1 (0.4)
ATA Ile 7 (3.1)	ACA Thr 4 (1.8)	AAA Lys 14(6.3)	AGA Arg 6 (2.7)
ATG Met 6 (2.7)	ACG Thr 2 (0.9)	AAG Lys 10(4.5)	AGG Arg 3 (1.3)
GTT Val 10(4.5)	GCT Ala 1 (0.4)	GAT Asp 15 (6.7)	GGT Gly 4 (1.8)
GTC Val 0 (0.0)	GCC Ala 0 (0.0)	GAC Asp 3 (1.3)	GGC Gly 0 (0.0)
GTA Val 3 (1.3)	GCA Ala 2 (0.9)	GAA Glu 11(4.9)	GGA Gly 5 (2.2)
GTG Val 4 (1.8)	GCG Ala 0 (0.0)	GAG Glu 6 (2.7)	GGG Gly 2 (0.9)

The codons represent 224 amino acids

Ala 2 (1.3).	Leu 11 (4.9)
Arg 12 (5.4)	Lys 24 (10.7)
Asn 11 (4.9)	Met 6 (2.7)
Asp 18 (8.0)	Phe 10 (4.5)
Cys 0 (0.0)	Pro 5 (2.2)
Gln 6 (2.7)	Ser 19 (8.5)
Glu 17 (7.6)	Thr 10 (4.5)
Gly 11 (4.9)	Trp 2 (0.9)
His 8 (3.6)	Tyr 16 (7.1)
Ile 17 (7.6)	Val 17 (7.6)
End 1 (0.4)	

Acidic (Asp + Glu)	35 (15.6)
Basic (Arg + Lys)	36 (16.1)
Aromatic (Phe + Trp + Tyr)	28 (12.5)
Hydrophobic (Ala + Ile + Leu + Met + Val)	79 (35.3)

The distribution of A+T at the first second and third positions within the codons is 60.5 %, 72.2 %, 75.8 %.

Molecular weight 26.447 kDa

**Figure 4.7****Codon usage in ORF-B**

TTT Phe 2 (2.6)	TCT Ser 0 (0.0)	TAT Tyr 2 (2.6)	TGT Cys 0 (0.0)
TTC Phe 0 (0.0)	TCC Ser 0 (0.0)	TAC Tyr 0 (0.0)	TGT Cys 0 (0.0)
TTA Leu 4 (5.2)	TCA Ser 0 (0.0)	TAA End 1 (1.3)	TGA End 0 (0.0)
TTG Leu 5 (6.5)	TCG Ser 0 (0.0)	TAG End 0 (0.0)	TGG Trp 1 (1.3)
CTT Leu 0 (0.0)	CCT Pro 0 (0.0)	CAT His 1 (1.3)	CGT Arg 2 (2.6)
CTC Leu 0 (0.0)	CCC Pro 0 (0.0)	CAC His 0 (0.0)	CGC Arg 0 (0.0)
CTA Leu 0 (0.0)	CCA Pro 1 (1.3)	CAA Gln 7 (9.1)	CGA Arg 0 (0.0)
CTG Leu 0 (0.0)	CCG Pro 1 (1.3)	CAG Gln 1 (1.3)	CGG Arg 0 (0.0)
ATT Ile 0 (0.0)	ACT Thr 1 (1.3)	AAT Asn 7 (9.1)	AGT Ser 4 (5.2)
ATC Ile 1 (1.3)	ACC Thr 0 (0.0)	AAC Asn 0 (0.0)	AGC Ser 2 (2.6)
ATA Ile 0 (0.0)	ACA Thr 1 (1.3)	AAA Lys 4 (5.2)	AGA Arg 0 (0.0)
ATG Met 9 (11.7)	ACG Thr 1 (1.3)	AAG Lys 4 (5.2)	AGG Arg 0 (0.0)
GTT Val 4 (5.2)	GCT Ala 1 (1.3)	GAT Asp 1 (1.3)	GCT Gly 1 (1.3)
GTC Val 0 (0.0)	CCC Ala 0 (0.0)	GAC Asp 1 (1.3)	GCC Gly 0 (0.0)
GTA Val 2 (2.6)	GCA Ala 0 (0.0)	GAA Glu 4 (5.2)	GGA Gly 0 (0.0)
GTG Val 1 (1.3)	GCG Ala 0 (0.0)	GAG Glu 3 (3.9)	GGG Gly 1 (1.3)

The codons represent 80 amino acids

Ala 1 (1.3)	Leu 9 (11.7)
Arg 2 (2.6)	Lys 8 (10.4)
Asn 7 (9.1)	Met 9 (11.7)
Asp 2 (2.6)	Phe 2 (2.6)
Cys 0 (0.0)	Pro 2 (2.6)
Gln 8 (10.4)	Ser 6 (7.8)
Glu 7 (9.1)	Thr 3 (3.9)
Gly 2 (2.6)	Trp 1 (1.3)
His 1 (1.3)	Tyr 2 (2.6)
Ile 1 (1.3)	Val 7 (9.1)
End 1 (0.4)	

Acidic (Asp + Glu)	9 (11.7)
Basic (Arg + Lys)	10 (13.0)
Aromatic (Phe + Trp + Tyr)	5 (6.5)
Hydrophobic (Aromatic + Ile + Leu + Met + Val)	31 (40.3)

The distribution of A+T at the first second and third positions within the codons is 60.0 %, 78.7 %, 61.2 %.

Molecular weight 9.470 kDa

#### 4.3.3 Two possible open reading frames

The initiation of protein synthesis usually starts at the codon AUG although ribosome binding experiments (Clark and Marcker, 1966) showed that GUG and UUG were also found to stabilize the binding of fMet-tRNA to *E. coli* ribosomes. The A+T content of the plasmid is manifested in an open reading frame (ORF) by the non-random distribution of bases within that ORF. If this region is read out of frame however, the AT rich nature of the sequence results in a high frequency of stop codons. The non-random occurrence of codons within a gene can be exploited in the prediction of coding regions (Staden, 1984). Using the sequence analysis programme of Queen and Korn (1984), the sequence was translated in all six reading frames and the occurrence of all start and stop codons noted to identify any possible open reading frames. The high frequency of stop codons in the different reading frames meant that the distance between start and stop codons was generally quite small. The occurrence of a number of comparatively much longer sequences lacking the apparent random distribution of stop codons noted elsewhere suggested that these were in fact open reading frames. Based on this analysis, 2 possible open reading frames were identified. These were termed ORF-A and ORF-B, and could have encoded polypeptides with predicted molecular weights of Mr 26,447 and 9,470 (Figure 4.8).

The presence of a start and stop codon is not sufficient to determine if an ORF is in fact a gene. In order to be expressed, a sequence must also carry the correct signals to be capable of interacting with the relevant enzymes allowing transcription and

Figure 4.8

Possible open reading frames within pHD2

Predicted amino acid sequence of ORF-A

MQVYLDRLMIKYKDVTEKQFSDVLTKISSQIFLPNTPIRSEHGTSVRDYHRVHIGYG  
EGAVYIGWKHNSEKEKDSYDMKVDFNPSKFENNELQKDSYEKVFETVFHTLNAVLSKNR  
VVYGMEDIAFDIERHMSDIVSYSTGKQQDRHKGTVYYGNRNKDGYLKIYDKKKELYNHFK  
RMIEEENLTRIEYSWRDSDGVVDEIRKSPFIDESYTFSLIU

ORF-A is located between nucleotides 235-906 as shown in Figure 4.4 and encodes a polypeptide with a molecular weight of 26.447 kDa.

Predicted amino acid sequence of ORF-B

MSEMVRVNTRISKKLNDWLDEYSKESGVPKSTLVHLALENVVNQKVMLEQMPKMQQLSM  
MFENVTOOQLNQKGNMFELK

ORF-B is located between nucleotides 1281-1522 as shown in Figure 4.4 and encodes a polypeptide with a molecular weight of 9.470 kDa.

translation to occur. These features are discussed in later sections (Sections 4.3.4 and 4.3.5). Attempts were also to correlate the polypeptides which were shown to be expressed from the various constructs, using the maxicell system (Section 4.4), with those expected to be expressed from the putative open reading frames.

#### 4.3.4 -10 and -35 promoter sequences

In order for a gene to be expressed it must first be transcribed by RNA polymerization. Transcriptional specificity involves the interaction between the RNA polymerization and the promoter site of genes or operons. In bacteria as distantly related as *E. coli*, *B. subtilis*, *S. coelicolor* and the cyanobacterium *Anacystis nidulans*, the RNA polymerization consists of a core polymerization of four subunits ( $\alpha_2\beta\beta'$ ) and an additional subunit, the sigma factor. The sigma factor binds reversibly to the core RNA polymerization and whilst all prokaryotic RNA polymerases appear to share a conserved promoter specificity (Wiggs *et al.*, 1979), imparted by the predominant sigma factor, recognition of alternative promoter sites often requires one of a number of alternative sigma factors (Doi and Wang, 1986). Hence, these sigma factors play a determining role in the specificity of transcription initiation. In *B. subtilis* and another low G+C Gram-positive species, *Lactobacillus curvatus*, an additional component of the holoenzyme, delta, appears to compensate for the absence from their principal sigma factors of an amino terminal region present in the *E. coli* protein and helps to confer specificity of interaction of the

polymerase with promoter sequences (Hyde *et al.*, 1986; Hilton and Whiteley, 1985; Schinkel *et al.*, 1987).

The first evidence of multiple sigma factors came from phage infected and uninfected *B. subtilis* cells (Losick and Pero, 1981) based on the observation of a number of temporally regulated biochemical events during phage replication or sporulation. The majority of cellular transcription requires the predominant, or primary sigma factor (Harris *et al.*, 1978; Osawa and Yura, 1981). In broad terms, the majority of genes that are expressed during normal vegetative growth of *E. coli* and *B. subtilis*, and many of those of *S. coelicolor*, have -10 and -35 sequences that resemble the consensus sequence for the major holoenzyme of *E. coli*. Sigma factors corresponding to the major sigma factor ( $\sigma^{70}$ ) of *E. coli* have been biochemically demonstrated for each species. In *E. coli* the primary sigma factor is referred to as  $\sigma^{70}$  and in *B. subtilis* as  $\sigma^{34}$ . In many bacterial species there are also alternative sigma factors that control the transcription of coordinately regulated sets of genes from promoter sequences that are quite distinct from those recognized by the primary sigma factor for example phage development or sporulation in *B. subtilis* (Doi and Wang, 1986; Reznikoff *et al.*, 1985; Losick *et al.*, 1986; Losick and Pero, 1981).

Detailed sequence information is not available for *B. thuringiensis* promoters. However, since plasmids must be able to maintain themselves through vegetative cycles of growth and the -10 and -35 sequences recognized by the appropriate sigma factors in both *E. coli* and *B. subtilis* share the same consensus sequence (Figure 4.9), it

Figure 4.9

Promoter sequences in *B. subtilis* and *E. coli*.

Conserved regions in promoters			
Holoenzyme	-35 region	Spacing (bp)	-10 region
<i>B. subtilis</i>			
Eo <sup>43</sup> (11)	TTGACA	17-19	TATAAT
Eo <sup>37</sup> (5)	AGGATTTNA	11-15	GGAATTNTTT
Eo <sup>32</sup> (2)	AAATC	14-15	TANTGNTTNTA
Eo <sup>29</sup> (4)	TTNAAA	14-17	CATATT
Eo <sup>28</sup> (2)	CTAAA	16	CCGATAT
SPO1Eo9p28(5)	TNAGGAGANNA	15-16	TTTNTTT
SPO1Eo9p33-34(5)	CGTTAGA	17-19	GATATT
<i>E. coli</i>			
Eo <sup>70</sup> (168)	TTGACA	16-18	TATAAT
Eo <sup>32</sup> (6)	TNTCNCCTTGAA	13-15	CCCCATTTA
T4Eo9p55(4)			TATAAATA

The total number of sequences analyzed for each class of promoters is shown in parentheses

For T4Eo9p35, there is no apparent consensus sequence in the -35 region.

Taken from Doi and Wang 1986



**Figure 4.10**

Putative promoter and ribosome binding site sequences  
upstream of the proposed open reading frames

Open reading frame A

184

TCG TGT ACA AAA TGA CAT GTT TAA TAA AAA ATA AGG AGC GGG ATA  
Ttg ACA Ta TAA T GG AGg  
-35 -10 SD

GAT TTT ATG CAA GTT TAT TTG 249  
Met Gln Val Tyr Leu

**Open reading frame B**

1211

GTG TAT CTT ATG TTG ATA TTG TGT TTA AAC TGA TGT TAT ATT TAT GTA GTA CGA  
TTG Aca TAT ATT  
-35 -10

TAT ACA AGA GGT GAT TAG ATG AGT GAA ATG GTT CGT GTT AAT 1136  
gGA GG Met Ser Glu Met Val Arg Val Asn  
SD

seemed likely that the -10 and -35 sequences exhibited within pH2 would at least be related to these sequences.

The presence of possible -10 and -15 sequences upstream of ORF-A and ORF-B are shown (Figure 4.10). Comparison of these sequences with the consensus sequences for *B. subtilis* Eo<sup>43</sup> promoter reveals good homology with the -10 sequence showing only one mismatch out of six bases and the -35 sequence showing two mismatches out of six bases. The spacing between the two sequences, however, is rather short at only 10 bases whereas the expected spacing would be 17-19 bases.

The second open reading frame shows a -10 sequence with 5 out of the 6 bases matching the *B. subtilis* Eo<sup>43</sup> consensus sequence. 18 bases upstream from this sequence there is a possible -35 sequence also showing 5 matches out of 6 to the consensus sequence with the spacing of these two sequences being within the expected 17-19 bases.

#### 4.2.5 Ribosome binding sites

Following transcription, the mRNA is translated. A purine rich region known as the Shine-Dalgarno sequence (Shine and Dalgarno, 1974) of 5-10 residues centered about 10 bases upstream from the start codon has been shown to be required for translation by virtue of its complementarity to a highly conserved region near the 3' end of the 16S rRNA (Steitz and Jakes, 1975). Extensive statistical studies have been carried out on a sample of 124 known *E. coli* ribosome-binding sites

(Stormo *et al.*, 1982). The Shine-Dalgarno sequences complimentary to the 16S rRNA are centered around the conserved 5'-CUCCU-3' sequence near the 3' end of the 16S rRNA (Steitz and Jakes, 1975) some 6-13 bases upstream from the start codon AUG.

The sequences of 16S rRNA are highly conserved between *E. coli*, *B. subtilis*, and *Streptomyces* and it is therefore not surprising that the ribosome-binding sites are also similar and that cross generic recognition can take place. There are some limitations to this; more extensive ribosome-binding sites are generally found in low (G+C) Gram-positive bacteria (McLaughlin *et al.*, 1981) than in enteric bacteria (Gold and Stormo, 1987) or *Streptomyces* (Hopwood *et al.*, 1986) and *B. subtilis* is more limited than the other two groups in its ability to initiate translation of transcripts of foreign bacterial genes (Hopwood *et al.*, 1986; McLaughlin *et al.*, 1981). The upstream region is generally rich in A residues and lacking in G residues and a similar trend in the early parts of the coding region may result from a biased third-letter codon choice (Rodier *et al.*, 1982). A number of mutations in or affecting this region have been shown to affect the efficiency of translation (Kozak, 1983; Tessier *et al.*, 1984).

As discussed, the sequences of 16S rRNA from organisms as diverse as *E. coli*, *B. subtilis* and *Streptomyces* are highly conserved as are the ribosome binding sites. It does not seem unreasonable therefore to assume that the differences in RBS sequences between *B. subtilis* and *B. thuringiensis* would therefore be minimal. No comprehensive sequence data is available from *B. thuringiensis*. Sequence data is however

available for some of the plasmid encoded crystal toxin genes from this group of organisms. The Shine-Dalgarno sequence from *B. thuringiensis* subspecies *kurstaki* HD-73 crystal-toxin gene given as GATGGAGGTAA 3 bases upstream of the translational start site. The 130-kd  $\delta$ -endotoxin gene of *B. thuringiensis* subsp. *israelensis* is given as GGAGG within the SD region TATGGGAGGAATA (Angathanasombat *et al.*, 1987,) which was also identical to that found in the 72-kDa *B. thuringiensis* subspecies *israelensis* gene (Thorne *et al.*, 1986). The five base pair motif, GGAGG, was taken to be the most important region to be conserved due to its complementarity to the specific binding sequence on the 16S rRNA. Such a region, rich in G residues, should be fairly easy to spot in a sequence which overall is 65.4 % A+T.

The sequence of the 3'-OH end of the *B. subtilis* 16 S rRNA is known to be UCUUCCUCCACUAG (Moran *et al.*, 1982). The free energies of interaction of Shine-Dalgarno sequences with the 16 S rRNA sequence of *B. subtilis* has been calculated in a number of instances according to Tinoco *et al.* (1973), with an observed range of -11.6 and -21 kcal/mol (McLaughlin *et al.*, 1981). The free energies of interaction of the proposed Shine-Dalgarno sequences with the *B. subtilis* 3'-OH terminus 16 S rRNA were calculated as being 12.8 kcal/mol for ORF-A and 17.8 kcal/mol for ORF-B which is consistent with this observation. This data in conjunction with the availability of -10 and -35 promoter sequences lends weight to the argument that these two open reading frames are in fact expressed.

### 5.7 Maxicell analysis

Maxicells are generated from certain strains of *E. coli* following UV irradiation and are used to determine the polypeptides which are expressed from plasmids in *E. coli* under the given conditions. The maxicell procedure was first described by Sancar *et al.* (1979) and was based upon the observation that when irradiated with UV light (254 nm), *E. coli* *recA uvrA* cells stop DNA synthesis and chromosomal DNA is extensively degraded such that only a small amount remains several hours after irradiation.

Plasmids contained within a UV irradiated maxicell strain and which do not receive a UV hit do not degrade and subsequently continue to replicate. Plasmid levels increase about 10 fold by about six hours after irradiation. Maxicells continue to synthesize plasmid encoded polypeptides for up to 24 hours following irradiation in the absence of host protein synthesis. These plasmid-encoded polypeptides may be radiolabelled by the addition of  $^{35}\text{S}$ -methionine to the medium. Whole cell extracts are then prepared and the polypeptides separated on polyacrylamide gels following denaturation. The polypeptides synthesized after irradiation and which are therefore encoded by the plasmid are detected by autoradiography.

The addition of cycloserine to the medium (Roberts *et al.*, 1979) after irradiation kills any cells which are still growing and which had therefore been sublethally irradiated. These cells would have been capable overgrowing the maxicells and would have given a high background

of chromosomally encoded polypeptides which could have obscured plasmid-specific products.

The presence of three polypeptide bands are reported to be produced in *maxicells* by the plasmid pBR322. Derivatives of pBR322 with deletions of the *amp* and *tet* genes revealed that the 28- and 31-kilodalton polypeptides were synthesized from the *amp* or *bla* gene and the 37-kilodalton polypeptide from the *tet* gene (Sancar *et al.*, 1979).

The construction of the plasmids pDM100 and pDM200 has been previously described (Chapter 3). Three further constructs; pCC100, pCC200 and pCC300, are included in this section for convenience although their construction is not described until later (Chapter 6). Briefly, pCC100 is a derivative of pBR322 in which the chloramphenicol acetyl transferase gene of pC194 has been cloned in the *Cla*I site of pBR322. The constructs pCC200 and pCC300 are further derivatives of pCC100 in which the pHD2 sequences from pDM100 and pDM200 have been subcloned respectively.

The genetic manipulations described in which pHD2 and the CAT gene from pC194 were cloned in the plasmid pBR322 all involve insertions in either the *Hind*III or *Cla*I sites (See Chapters 3 and 6). The *Hind*III and *Cla*I sites in pBR322 are both in the promoter of the *tet* gene. The presence of an insert in either of these two sites may therefore be expected to affect the synthesis of the 37 kd polypeptide visualized in this system. The *amp* gene should not be affected by any of these genetic manipulations. The presence of any other polypeptide bands on

Lane 1 2 3 4 5 6

kDa

30

14.5

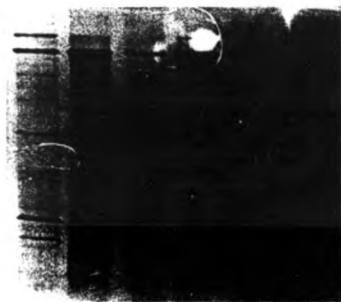


Figure 4.11

Expression of polypeptides from pHD2 based constructs using  
the Maxicell system

Lane 1	Expression of polypeptides from pBR322
Lane 2	Expression of polypeptides from pDM200
Lane 3	Expression of polypeptides from pDM100
Lane 4	Expression of polypeptides from pCC100
Lane 5	Expression of polypeptides from pCC200
Lane 6	Expression of polypeptides from pCC300



the autoradiograph from the maxicell analysis should theoretically have been synthesized from coding regions of the inserts present in the pBR322 based constructions. In reality, comparison of these bands with the control, pBR322, is required to check for background synthesis from any undegraded chromosomal host DNA.

The plasmids pBR322, pDM100, pDM200, pCC100, pCC200 and pCC300 were transformed into the *E. coli* host strain CSH26<sup>-</sup>F6 and the expressed polypeptides labelled with <sup>35</sup>S methionine and separated on 10 % polyacrylamide SDS gels. The radio-labelled polypeptides were visualized by autoradiography (Figure 4.11). A small polypeptide of less than 14.5 kd was clearly visible from the constructs pDM100, pDM200, pCC200 and pCC300 and was presumed to correspond to the ORF-B gene product which had a predicted molecular weight of 9,470 kd (See Chapter 5). The ORF-A gene product could not be distinguished using this system and may only have been poorly labelled if expressed at all. The CAT gene product with a molecular weight of 26,200 kd (See Chapter 6) was not observed to be expressed in this system, however, this is not surprising since this gene is known to be inducible (Morinouchi and Weisblum, 1982). The Maxicells were prepared without the use of chloramphenicol.

#### 4.5 Summary and Conclusions

The plasmid, pHD2, which was isolated from *Bacillus thuringiensis* subspecies *kurstaki* HD1-Dipel was shown to be 2055 bp in size with a G+C

content of 34.6 %. The presence of two possible open reading frames was noted which could encode polypeptides with a calculated molecular mass of 26447 (ORF-A) and 9470 (ORF-B). The codon usage within these open reading frames showed a greater frequency of A+T usage at positions 2 and 3 than at position 1 within the codons which was consistent with the expected pattern based on the high A+T content of the whole sequence and the constraints and redundancy in the genetic code. The presence of sequences upstream of the proposed ORFs required for both transcription and translation were detected suggesting that these putative ORFs might well be expressed. The possible role of these open reading frames is discussed in detail (Chapter 5). A small highly labelled polypeptide with a molecular weight of less than 14.5 kd, and thought to be the ORF-B gene product, was found to be expressed using the Maxicell system although expression of the ORF-A gene product could not be confirmed. A comprehensive restriction map was also generated confirming the presence and absence of a number of restriction sites. It had been thought that the unexpected lack of a number of restriction sites resulted from a host modification system. Sequencing of the plasmid revealed that these restriction sites were not present rather than being protected by such a mechanism. It was essential to resolve any such ambiguities if further cloning or manipulation work was to be undertaken such as in the construction of shuttle vectors (Chapter 6).

## Chapter 5

### A model for the replication of pKM2

#### 5.1 Introduction

The existence of plasmids has been demonstrated in many species of Gram-positive bacteria. Whilst the range of Gram-positive bacteria from which plasmids have been isolated is diverse, the properties of most of these plasmids with regard to their replication functions are very similar. This may indicate that many of these replicons have a common ancestry (Iordanesco *et al.*, 1978; Projan and Novick, 1988). The properties of these plasmids are discussed fully in Chapter 1 and will only be summarized here.

The majority of plasmids which have been characterized from Gram-positive species replicate via a rolling circle mechanism involving a single-stranded DNA intermediate. These replicons have consequently been termed ssDNA plasmids. At the onset of replication, the Rep protein recognizes a plus origin sequence and introduces a nick to initiate the process. Displacement of the old plus strand and the synthesis of a new plus strand then follows. A second nick is introduced by Rep upon the recognition of a termination sequence, which overlaps the origin, and ligates the ends of the displaced strand to form the single-stranded plasmid intermediate. Conversion of the intermediate to the double-stranded end product involves the recognition

of a minus origin by host factors for the initiation of minus strand synthesis.

The analysis of sequence data obtained from some of the ssDNA plasmids has allowed a comparison of some of the structures and functions of these replicons to be made. A homology comparison between the sequences at the plus origins of a number of these plasmids has enabled them to be divided into three groups. The position of the nick sites within these plus origin sequences has also been determined and sequence data is also available for the minus origins and Rep proteins of these plasmids allowing similarities in structure and in function to be compared.

To date, sequence data is only available for one other plasmid from a *B. thuringiensis* isolate. The plasmid, pGI2, contains the transposon Tn44330 and was isolated from *B. thuringiensis* subsp. *thuringiensis* strain H1.1. This isolate contains three plasmids; pGI1, pGI2 and pGI3 which have been cloned and partially characterized in *E. coli* (Mehillon *et al.*, 1988). Cross hybridization studies of two of these plasmids, pGI1 and pGI3, with the plasmids from other *B. thuringiensis* strains suggested that these contained sequences related to plasmids from *B. thuringiensis* subsp. *kurstaki* HD1-Dipel and HD-2 although not to strain HD-14. No such cross hybridization was observed using the plasmid pGI2. Sequence data for this plasmid did not reveal any significant homologies to Rep proteins or nick sites found in other plasmids from Gram-positive bacteria replicating via a ssDNA intermediate although one of the open reading frames could have coded

for a site specific recombinase (Mahillon and Seurinck, 1988). No significant homology could be detected between pGI2 and pHD2.

At this time, the method of replication of the many plasmids from *B. thuringiensis* subspecies remains unclear and is not certain whether these plasmids fit into the ssDNA category of replicons or replicate by some novel mechanism.

Sequencing studies and subsequent computer analysis (Chapter 4) suggested that the plasmid pHD2 contained two open reading frames. The function of these open reading frames was unknown. As discussed, the majority of plasmids which have been studied from Gram-positive organisms replicate via a rolling circle method involving a single stranded DNA intermediate (For review see Gruss and Ehrlich, 1989). It seemed possible that pHD2 would replicate in a similar manner, in which case a comparison of the sequence with published sequences for other plasmids from Gram-positive species would be expected to show similarities. The following sections describe the computer-aided analyses and comparisons which were made and the conclusions which were made from these with regard to the structures and function of pHD2.

## 5.2 Preliminary data bank searches

Vast amounts of DNA and protein sequence data are now accessible using computers and the functional role of many of these sequences has also been determined. It was therefore decided to compare the

polypeptide sequences predicted by the two open reading frames, ORF-A and ORF-B, against other published sequences held within a computer data bank. It was decided that such a comparison would allow the presence of conservative substitutions at the DNA level without affecting the protein sequence and so would give the best chance of finding alignments with proteins of related function. The alignment was performed using the computer facility at Daresbury. The output from this analysis gave the 100 best alignments against the protein sequences held within the data base. Interestingly, ORF-A showed the greatest homology to a group of Rep proteins from a number of *Staphylococcus aureus* plasmids and ORF-B to the RepA gene product of *Streptococcus pneumoniae* and *Streptococcus agalactiae*. The four Rep proteins found to be homologous were all members of the same pT181 group of *S. aureus* plasmids. These five protein sequences were obtained from the database for further analysis (Appendix 1).

### 5.3 Protein sequence alignments using the Clustal sequence alignment package

The Clustal sequence alignment package contains 4 programmes of which three are run consecutively to produce an alignment (Higgins and Sharp, 1988; Higgins and Sharp, 1989). The three programmes correspond to logically distinct stages in the multiple alignment process: 1. calculation of a similarity matrix containing values for each possible pairwise comparison of sequences (Clustal1); 2. construction of a dendrogram by cluster analysis (Clustal2) and 3. alignment according to

the branching order in the dendrogram (Clustal3 or Clustal4). The alignment programs (clustal3 and clustal4) can be used interchangeably. Clustal3 is an older program and is "quick and dirty" where as Clustal4 is a later, improved version which is much more sensitive but also much slower.

The protein sequences obtained from the data base were in single letter amino acid code in text-only computer files. These were loaded into the wordprocessing package Microsoft Word5 where they were formatted according to the requirements of the Clustal sequence alignment programme and stored as unformatted text-only files which could be handled directly by the Clustal programme. The sequences of ORF-A and ORF-B were exported from the DNA sequence analysis programme of Queen and Korn (1984) as text only files and manipulated using Microsoft Word5 in the same way.

The Clustal package is capable of aligning many sequences simultaneously but in order to obtain a meaningful alignment, the sequences need to be similar. Consequently the initial alignments were made in pairs and the output imported into Word5. The alignments of ORF-A with the *S. aureus* Rep proteins (Figures 5.1-5.4) were far from being identical, however, shorter regions were present in which the amino acid alignment showed a higher degree of identity combined with conservative amino acid substitutions. Moreover, these conserved regions appeared to be consistent between all of the alignments. It was therefore decided to perform a multiple sequence alignment between all of the Rep proteins and ORF-A. This alignment confirmed that the

**Figure 5.1**

An alignment of the ORF-A and RepM (pC223) amino acid sequences using the Clustal4 programme.

ORF-A  
RepM pC223

M--QVYLDRLMIKYKDVTEKQFSV-----LTKISSKKQIFL  
MSKNRYTNW-SNHLWEDSDMSFKTGYSGNSRLDAMFVCTSNPKLSFD  
.....\*

PNTPI--RSEGTGSVDYRHH-----VINIGYG-----EGAVYIGWKNHSEKEKDSY  
AHTIVGNLGNLKNAAKLSDFMSHLDQIRLDLQTFKFAKALQEKYIIEYDKVADTDWR  
.....\*

DMKVDVSPKSFENNE-----LQKDSYEKVFETVFHTLNAVLSKNSKRV--YCMDIAPDIERN  
NHRVFPENPKLTHDEMLWKHNIIDYMEDGGFTRLDLADFPEDLSDYALS-----EKA  
.....\*

MSDIVSYSKTGKQDRHKGTVYVGNRWKDGKYLKYDKKELVWFKRHIIEENLRIE-  
LKRTVFGTTGAETK-----YFGSRDSNRFIRIYNKKKERKENADVDSAHLRWVEIE  
.....\*

-----Y-----SW  
LKRMDVYVWKKFNDNLILPKAMATLESLEKQAMVYLLHSEESNGELRHRNRKYQII  
.....\*

RDSQGVVVDEIRKSPFSDISEY-----TFSILI  
QEISSIDLTLMKSTLTDNEENLQKQINFWQRKFPAW

- \* = Identity
- . = Conservative substitution





**Figure 5.3**

An alignment of the ORF-A and RepC (pT181) amino acid sequences using the Clustal4 programme.

```

ORF-A      MQVYLDRLMIKYKDVTEKQFSDVL-----TKISSKQIFLPNTPIRS
RepC pT181 M--YIGNNHANHSNHNLDNFNFKTGYNSRLDAHTVCISDPKLSF
          * * . . . . .

EHGTSVRDYHR-----VINIGYG---EGAVYIGWGNSEKEKDS
DAMTIVGNLNRDNAAQLSKPHEVEFQIRLMDILOTKFKAKALQEKVYIEYDKVKADSWDR
. * * . . * . . . . .

YDMKVDFNPSKFENNE---LQKDSYEKVFETVFNHTLNAVLKSNKRVV--YGMDFAFDIER
RNHRIEFNPNKLTREMIWLKQNIISYMEDDGFTRLDLAFDFEEDLSDYIAMS-----DK
. . . . . * . . . . .

HMSDIVSYSKTKGQDDRHKGTVYVYGNPNKDGYLKIYDKKKE-----LYNHFKRMIE
AVKKTIFYGRNGKPKETK----YFGVNRDSNRFIRIYNKKQBERUDNADAEVMSHLWRVEI
. . . . . * . . . . .

EENLTRIEY-----SW---RDSGQVVV-----DE
ELKRDMMVDYVWDCFSDLHILQPDWKTIQRTADRAIVFMLSDEEEMGKLRNSRTKYKNL
* . . . . * . . . . .

IRKSPFFSIDE-----SYTFSILI
IKEISPDVLDLHKSTLKANEKQLQKQIDPMQHEFKFWK
* . . * . . . . .

```

\* = Identity

. = Conservative substitutions

Figure 5.4

An alignment of the ORF-A and RepE (pS194) amino acid sequences using the Clustal4 programme.

```

ORF-A      MQVYLDRLHIKYK-DVTEKQFSDV---LTKISSKQIFLPNTPIRSEH
RepE pS194 HSKKAEETIQAKQSLKENSFNSTGYNSRLNRHIMYTEPKLHFDA
          *   .   .   .   .   .   .   .   .   .   .   .   .   .   .   .

GTSVRDYHR-----VIHIGYG---EGAVYIGWQHNSEKEDSYD
MTIVGNLKKQNAHKLSEFMSIAFQIRLWDILQTKFKAKALQEKVYIEYDKVKADAWDRRN
* * . . . . .   . . . . .   . . . . .   . . . . .   . . . . .

MKVDVFNPSKFENNE---LQKDSYEKVFETVFHTLNAVLSKNKRVV--YQMDIAFDIERHM
MRVEFNPKNKLTHEEMLWKQNIIDYMEDDGFTRLDLAFDFEDDLSDYVAMT----DKSV
*. . . . . * . . . . .   . . . . .   . . . . .   . . . . .   . . . . .

SDIVSYSKTKQQDRHKGTVYVYGNRNKDGYLKIYDKKKE-----LYNHFKRMIEEE
KCTIFYGRNGKPETK-----YFGVRDSDFIRIYNKKQERKNADIEVMSCHLWVEIEL
. . . . * . . . . .   . . . . .   . . . . .   . . . . .   . . . . .

NLTRIEYSWRDS-----DGVVVDIERKSPFFS-----IDESY----
KRDMDVY--WNCDFNDLHILKPDWSSLEKVKQAMHYMLIHEESTWGKLERRTKMKYREML
. . . . * . . . .   . . . .   . . . .   . . . .   . . . .

-----TFSILI
KSISEIDLTLMKLTLKENEKQLQKQIEFWOREFRWE
          . . . .

```

\* = Identity

. = Conservative substitution

Figure 5.5

A multiple sequence alignment of the ORF-A and *S. aureus* Rep protein amino acid sequences using the Clustal4 programme.

```

ORF-A      MQVYLDRLMIKYKDVTEKQ-FS-----DVLTKISSKQIFLP
RepM pC223 HSKQNYTHSNHLENHDSQNFSTGYNSRLDAHFVCTSNPKLSFDA
RepD pC221 HST---EMHSNYLQNMIDNFSTGYNSRLSGNFFTTPOPELSFDA
RepC pT181 MYKNNKAMHSHNLENMIDNFSTGYNSRLDAHTVCTSDPKLSFDA
RepE pS194 HSKKAEETQAKQSLKENSNSFSTGYNSRLNRHIMYTPPEKLHFDA

```

```

NTPI---RSEHGTSVRDYHR-----VIHIGYG---EGAVYIGWKHNSEKDSYD
MTIVGMLNONSAKKLSDFMSLDPQIRLMDILQTKFKAKALQEKVYI EYDKVKADTWDRRN
MTIVGMLNKTNAKKLSDFMSTEPQIRLMDILQTKFKAKALQEKVYI EYDKVKADSWDRRN
MTIVGMLNRDQAQLSKFMSEVPQIRLMDILQTKFKAKALQEKVYI EYDKVKADSWDRRN
MTIVGMLNKNNAHKLSEFMSIAPQIRLMDILQTKFKAKALQEKVYI EYDKVKADAWDRRN

```

```

MKVDFNPSKFENNE---LQKDSYEKVFETVPFHTLNAVLKSNKRUV--YGMIDAFDIERHH
MRVEFNPWKLTDEMLWKQNIIDYMEDDGFTRDLAFAFDLSDYYALS-----EKAL
MRVEFNPWKLTDEMLWKQNIIDYMEDDGFTRDLAFAFDLSDYYAMT-----DKAV
MRLEFNPWKLTDEMLWKQNIISYMEDDGFTRDLAFAFDLSDYYAMS-----DKAV
MRVEFNPWKLTDEMLWKQNIIDYMEDDGFTRDLAFAFDLSDYYAMT-----DKSV

```

```

SDIVSYSKTGKQQRHKGTVYGYGHRNKDGYLKIDYDKKKE-----LYNHFKRMIEE
KRTVFFGTTGKAETK----YFGSRDSNRFIRIYNKKKERKENADVDVSAEHLWRVEIEL
KKTIFYGRNGKPETK----YFGVRDSNRFIRIYNKKQERKKNADAEVMSHHLWRVEIEL
KKTIFYGRNGKPETK----YFGVRDSNRFIRIYNKKQERKKNADAEVMSHHLWRVEIEL
KKTIFYGRNGKPETK----YFGVRDSNRFIRIYNKKQERKKNADAEVMSHHLWRVEIEL

```

```

NLTRIEY-----SWRDSGV-----VVDEIRKSPFISIDESYTSILI-
KRDHVDVYNNCFNDLHLIKPAAWATLESLSKEQAMVYLLHEESKWGLHRNSRRYKQIIQ
KRDHVDVYNNCFDLDLHLIKPDWTTPEKVKQAMVYLLNEEGTWGLERHAKYKQIIK
KRDHVDVYNNCFSDLHLIQPDWKTIQRTADRAIVFMLLSDEEWGKLHRNSRTYKQNLK
KRDHVDVYNNCFNDLHLIKPDWSSLEKVDQAMVYLLHEESTWGLERRTKNKYREHLK

```

```

EISSIDLTLMKSTLTDSNEENLQKQINFWRKKEFWK
EISPIDLTLMKSTLTKEKELQKQIDFWQREFRFWK
EISPVDTLTLMKSTLTAKKELQKQIDFWQHEFFFWK
SISEIDLTLMKSTLTKEKELQKQIEFWQREFRFEW

```

\* = Identity  
. = Conservative substitution

Figure 5.6

An alignment of the amino acid sequences of four *S. aureus* Rep proteins using the Clustal4 programme.

```

RepM pC223  MSKQNYTHNSHNLENHSDSNFSKTGYNSRLDAHFVCTSNPKLSFDA
RepD pC221  MST---ENKSNYLQNKDLDNFSKTGYNSRLSGNFFTPQPELSFDA
RepC pT181  MYQNWHANSHNLENHDLNFSKTGYNSRLDAHTVCISDPKLSFDA
RepE pS194  MSKKAEEIQAKQSEKENSNFSTGYNSRLNRHIMYTFEPKLHFDA
          * . . . . . ***** . . . . . *
MTIVGNLNKNSAKKLSDFMSLDPQIRLWDILQTKFKAKALQEKVYIEYDKVKADTWDRRN
MTIVGNLNKNTAKKLSDFMSTEPQIRLWDILQTKFKAKALQEKVYIEYDKVKADSWDRRN
MTIVGNLNKDNQAALSKFMSVEPQIRLWDILQTKFKAKALQEKVYIEYDKVKADSWDRRN
MTIVGNLNKQNAKLSDFMSIAPQIRLWDILQTKFKAKALQEKVYIEYDKVKADAWDRRN
***** . . . * . . . . . *****
MRVEFNPKNKLTDEMLWLKQNIIDYMEDDGFTRLDLAFDFEDDLSDDYALSEKALKRTVF
MRVEFNPKNKLTDEMLWLKQNIIDYMEDDGFTRLDLAFDFEDDLSDDYAMTKAVKKTIF
MRIEFNPKNKLTDEMLWLKQNIIDYMEDDGFTRLDLAFDFEDDLSDDYAMSDKAVKKTIF
MRVEFNPKNKLTDEMLWLKQNIIDYMEDDGFTRLDLAFDFEDDLSDDYAMTKSVKKTIF
** . . . . . * . . . . . * . . . . . * . . . . . * . . . . . *
FGTTGKAETKYFGVSDSNRFIRIYNKKERKENADVDVSAENLWRVEIELKRDMDVYVWN
YGRNGKPKETKYFGVSDSNRFIRIYNKKQERKDNADVEVMSEHLWRVEIELKRDMDVYVND
YGRNGKPKETKYFGVSDSNRFIRIYNKKQERKDNADVEVMSEHLWRVEIELKRDMDVYVND
YGRNGKPKETKYFGVSDSNRFIRIYNKKQERKDNADIEVMSEHLWRVEIELKRDMDVYVND
. * . . . . . * . . . . . * . . . . . * . . . . . * . . . . . *
CFNDLHILKPAWATLESLEQAMVYLLHEESKMGELHRNSRRKYKQIQEISSIDLTDL
CFDDLHILKPDWTFPEKYQAMVYLLNEEGTWGKLERNAKYKYKQIKEISPIDLTEL
CFSDLHILQPDWTKIQRTADRAIVFMLSDSEEWGKLRNSRTKYKQIKEISPIDLTEL
CFNDLHILKPDWSSLEKYQAMVYLLHEESTWGKLERNTKQYREMLKSEISIDLTDL
** . . . . . * . . . . . * . . . . . * . . . . . * . . . . . *
MKSTLTDNEENLQKQINFWRQKFEFWK
MKSTLKENEKQLQKQIDFWQREFFFWK
MKSTLKANEKQLQKQIDFWQREFFFWK
MKSTLKENEKQLQKQIDFWQREFFFWK
** . . . . . * . . . . . * . . . . . * . . . . . *

```

\* = Identity  
 . = Conservative substitution

homologous regions were indeed conserved across all of the Rep proteins (Figure 5.5). An alignment of the four *S. aureus* Rep proteins revealed them to be highly conserved showing 65 % identity across all four sequences when aligned using the Clustal programme (Figure 5.6). The alignment also showed some regions to be more highly conserved between the four sequences than others. An alignment of ORF-A against the Rep proteins was far less striking, however, the regions observed to be to be more highly conserved between ORF-A and the Rep proteins also corresponded to those regions which were highly conserved between the Rep proteins. This suggested that these regions might correspond to functional domains and that whilst the amino acid sequences were not highly conserved, ORF-A might be distantly related to other Rep proteins and could be assumed to function as one.

ORF-B was found, using the Clustal alignment programmes, to exhibit considerable homology to a small polypeptide, RepA, from the plasmid pLS1 (Lacks *et al.*, 1986). RepA is 45 amino acids in size and is transcribed from a polycistronic mRNA together with the replication protein RepB. It has been proposed that RepA may be involved in the control of replication (Stassi, *et al.*, 1981). Whilst ORF-B is significantly larger (80 amino acids) than RepA, the Clustal alignment suggests that the most significant portion of homology occurs within the first half of the sequence suggesting this to be the more important region (Figure 5.7). A more convincing alignment was obtained when using the Gap alignment programme and the significance of this is discussed later (Section 5.5). The presence of a small, 46 codon ORF just upstream of the translational start site of the pT181 repC gene has

ORF-B MV-RVNTRI SKKLNIDWLD EYSKESGVPKSTLVH LAL ENYVQKVHLE  
RepA pL81 MKRLTITLESVLNLE KMAREGLSKSAMISVALENY-----  
..... = . . . . . \*\*\*\*\*

CMPIHQQLSHMFENVTQQQLNQGNMFLK  
-----KKGE---K  
... \*

**Figure 5.8**

RepA MKKRLTITLSEVLE---NLEKMAREMGLSKSAMISVALENYKKGQEK  
pT181 ORF MTKVCERHHSKKTLSCFYNLVYLDIKRYLNIHEDIYLG-ERFLK--RN  
\* \*

**Figure 5.9**

ORF-B MVRVNRTRISKKLNDMLDEYSKESGVPKSTLVNLALEMNYNQKVMLEQ  
pT181 ORF MTKVCERHKSST--LSCFYN-----LVYLDIKRYLN-----  
          \*       \*                \*                    \*                    \*  
MPKMQQLSMHFENVTOGGLNQKNGMFEK  
-----IHEDIVLGERFLKRN  
                                    \*                    \*

\* = Identity  
 . = Conservative substitution

been postulated to be involved in the control of plasmid replication (Projan and Novick, 1989). Alignments were made between this polypeptide and those of RepA (Figure 5.8) and ORF-B (5.9) since these were all proposed to regulate plasmid replication. No significant homology was found between any of these sequences.

#### 5.4 Further protein sequence alignments and significances with ORF-A

The sequence alignments produced by the Clustal programmes suggested that ORF-A was related to the Rep proteins of the pT181 group of *S. aureus* plasmids. However, the significance of the alignment could not be determined. For this reason, it was decided to pursue a further series of protein sequence alignments using the programmes of the University of Wisconsin Genetics Computer Group (UWGCG) (Devereaux *et al.*, 1984) and accessory programmes of Roach and Parkhill (unpublished data).

The *S. aureus* Rep proteins were aligned with ORF-A from pHD2 using the Gap programme (UWGCG), based on the algorithms of Needleman and Wunsch (1970) and consensus sequences derived requiring absolute conservation of amino acids across all five sequences (plurality of 5) (Figure 5.10) and conservation of four out of the five amino acids (plurality of 4) (Figure 5.11). The alignment produced between the five sequences using the Gap programme was not identical to the alignment produced using the Clustal programmes, however, the major conserved



Figure 5.10

An alignment of the ORF-A and four *S. aureus* rep protein amino acid sequences using the Gap programme.

```

ORF-A.....MQVYLD.RLMIKYKDVTEKQFS...
RepM pC223...MSKNNYTNHSHNLEHDSDFSKTGYNSRLDAHFVCTSNFKLSFDA
RepD pC221...MSTENHSMYLNQKLDLDFSKTGYNSRLSGNFFTTPOPLSFDA
RepC pT181...MYKNHIAHSHNLEHDLDFSKTGYNSRLDAHTVCISDPKLSFDA
RepE pS194...MSKKAEEIQAKQSEKENSFSKTGYNSRLNRIHWYTPFKLHFDA
-----Y---RL-----

...DVLTKISSKQI...FL...PNTPIRSEHGTSVRDYHRVINIGYGEQAVYIGWQHSE
MTIVGNLNNKSAKKISDFMSLDPOIRLWDILQTKFKAKAL.....QEKVYIEYDKVKA
MTIVGNLKNTHAKKISDFMSVEPOIRLWDILQTKFKAKAL.....QEKVYIEYDKVKA
MTIVGNLNRDNAQALSKFMSVEPOIRLWDILQTKFKAKAL.....QEKVYIEYDKVKA
MTIVGNLNNKHAHKLSEFMSIAPQIRLWDILQTKFKAKAL.....QEKVYIEYDKVKA
-----L-----F---P---T---VYI-----

KEKDSYDMKQVDFNPSKFENNE...LQKDSYKVFETVFTLNNAVLSKNKRVV...YGMEDIA
DTWDRNRNKRVEFNPKNLTHEEMIWLKQNIIDYMEDDGFTRLDAFDFEDDLSDYAALS...
DSWDRNRNKRVEFNPKNLTHEEMIWLKQNIIDYMEDDGFTRLDAFDFEDDLSDYAH...
DSWDRNRNKRVEFNPKNLTHEEMIWLKQNIIDYMEDDGFTRLDAFDFEDDLSDYAMS...
DAWDRNRNKRVEFNPKNLTHEEMIWLKQNIIDYMEDDGFTRLDAFDFEDDLSDYAH...
---D---M---FNP-K---E---L---F---L---Y-----

FDIERHMSDIVSYSKTKQODRHKGTVVYCHRNKQDGLKIVDKKELYNHFKRMIEENL
...EKALKRTVFCTTKGAETK...YFGSRDSNRFIRIYNKKERKNADVDVSAEHL
...TDKAVKNTIFYGRNGKPKETK...YFGVRDSRFIRIYNKKQERKDNADVEVMSEHL
...DKAVKNTIFYGRNGKPKETK...YFGVRDSNRFIRIYNKKQERKDNADVEVMSEHL
...TDKSVKNTIFYGRNGKPKETK...YFGVRDSRFIRIYNKKQERKDNADVEVMSEHL
-----GK-----Y-G-R-----IY-KK-E-----EL

.....TRIEYS.....WRD...SDGVVDEIRKSPFFSIDESYTPSILI...
.....WR...VEIELKRDHVDYWNCFNDLHILKPAMATLESLEKQAMVYLLHEESK
W.....RVEIELKRDHVDYWNCFDLDLHILKPDWTTPEKVKEQAMVYLLNEEGT
W.....RVEIELKRDHVDYWNCFDLDLHILKPDWTTIQRADRAIVFHLSDSEE
W.....RVEIELKRDHVDYWNCFDLDLHILKPDWSSLEKVQDQAMVYMLIHEEST
-----E-----W---D-----D-----

WGLHLRLHNSRRKYKQIIQEISSIDLTLMKSTLTDNEELQKQINFWQRKFEPWK
WGLHLRLERHAKYKYKLIKEISPIDTLMLKSTLKENEQQLQKQIDFWQREFRPMK
WGLHLRLHNSRKYKYNLIKEISPIDTLMLKSTLKENEQQLQKQIDFWQREFRPMK
WGLHLRLERTKNNKYREMLKSISEIDLTLMLKSTLKENEQQLQKQIDFWQREFRPMK
-----

```

Consensus sequence derived using a plurality of 5.

Figure 5.11

An alignment of the ORF-A and four *S. aureus* rep protein amino acid sequences using the Gap programme.

```

ORF-A .....MQVYLD.RLMIKYKDVTEKQFS...
RepM pC223...MSKNNTYTHNSHLEHDSDFSKTGYNSRLDAHFCTSNPKLSFDA
RepD pC221.....MSTENHNSNYLQKDLNFSKTGYNSRLSGNFPTQPPELSFDA
RepC pT181...MYKNHNAHNSHLEHDLNFSKTGYNSRLDAHFTVCISDPKLSFDA
RepE pS194...MSKKAEEIQAKQSLKENSFSKTGYNSRLRHMITYPEPKLHFDA
-----NFSKTGYNSRL-----P-LSFDA

...DVLTKISSKQI...FL...PMTPIRSEKGTSVRDYHVIHIGYGEAVYIGWKHNS
MTIVGNLNIKNSAKKLSDFMSLDPOIRLWDILQTKFKAKAL.....QEKVYIEYDKVKA
MTIVGNLNIKNAKLSDFMSVEPOIRLWDILQ...KFKAKAL.....QEKVYIEYDKVKA
MTIVGNLNRDNAQALSKFMSVEPOIRLWDILQ...KFKAKAL.....QEKVYIEYDKVKA
MTIVGNLNIKNAKLSDFMSIAPQIRLWDILQ...KFKAKAL.....QEKVYIEYDKVKA
MTIVGNLNIK--A--LS-FMS--PQIRLWDILQ-----QEKVYIEYDKVKA

KEKDSYDMKVDVFNPSKFENNE...LQKDSYEKVFETVPHLMAVLKSMGRVV...YGMEDIA
DTWDRNRMRVEFNPKNLTDEMIWLKQNIIDYMEDDGFTRDLAFDFEDDLSDYAALS...
DSWDRNRMRVEFNPKNLTHEEMIWLKQNIIDYMEDDGFTRDLAFDFEDDLSDYAM...
DSWDRNRMRVEFNPKNLTDEMIWLKQNIIDYMEDDGFTRDLAFDFEDDLSDYAMS...
DAWDRNRMRVEFNPKNLTHEEMIWLKQNIIDYMEDDGFTRDLAFDFEDDLSDYAM...
D-WDRNRMRVEFNPKNLT--EM-WLK-NII-YMEDDGFTRDLAFDFE-DLSDYAM--

FDIERHMSDIVSYSKTGQKQDRKHGTYYVGNRHKDYLIVDKKELYNHFKRMIEENL
...EKALKRTVFCCTCKAETK.....YFGSRDSRFRIRIYNKKERKENADVDVSAEHL
TDKAVKKTIFYGRNGKPKETK.....YFGVRDSRFRIRIYNKKQERKDNADVEVMSHL
...DKAVKKTIFYGRNGKPKETK.....YFGVRDSRFRIRIYNKKQERKDNADVEVMSHL
TDKSVKKTIFYGRNGKPKETK.....YFGVRDSRFRIRIYNKKQERKDNADVEVMSHL
---K--K-T-FYG--GK-ETK-----YFG-RDS-RFRIRIYNKK-ERK-NAD--V--EHL

.....TRIEYS.....WRD...SDGVVVDIEIRKSPFSDSYTFSILI...
...WR...VEIELKRDHVDYWNDCFDLHLILKPAWATLESLEKQAMVYLLNEESK
W.....RVEIELKRDHVDYWNDCFDLHLILKPDWTTPEKVKEQAMVYLLNEEGT
W.....RVEIELKRDHVDYWNDCFDLHLILKPDWTTQRTADRAIVFHLSDSEE
W.....RVEIELKRDHVDYWNDCFDLHLILKPDWSSLEKVQDQAMVYHLEHEEST
-----RVEIELKRDHVDYWNDCF-DLHLIL-P-W-----A-----LL--E-

.....
WGELHRNSRRYKQIIQEISSIDLTLMKSTLDNEENLQKQINFQWRKFEPWK.....
WGKLERHAKYKYLKIEISPIDLTLMKSTLKENEKQLKQIDFWOREFRPWK.....
WGKLERHNSRTYKQNLKIEISPVOLDTLMKSTLKENEKQLKQIDFWOREFRPWK.....
WGKLERRTNKYREMLKISIEIDLTLMKSTLKENEKQLKQIDFWOREFRPWE.....
WG-L-R---KY-----IS--DLT-LMK-TL--NE--LQKQI-FWQ--F-FW-----

```

Consensus sequence derived using a plurality of 4.

motifs were the same in both alignments. In order to determine which of the alignments was best, a further analysis was carried out using the Filofax programme (Roach and Parkhill, unpublished data). This programme produces a ratio of the alignment score, based on the Dayhoff matrix (Dayhoff 1972), over the length of the alignment. An output from this programme was presented in the form of a dendrogram using the Root66 programme (Roach and Parkhill, unpublished data). This revealed that the four *S. aureus* Rep proteins were closely related to each other and that ORF-A was only distantly related to these. The significance of this similarity was at this point uncertain. The ORF-A sequence was found to be most closely related to the RepC protein from pT181 and so this sequence was selected for further comparison.

In order to determine the significance of this similarity, a statistical analysis of the alignment produced using the Gap programme was carried out. The pT181 RepC sequence was compared to ORF-A using an adaption of the Gap programme called SGap (Roach and Parkhill unpublished data). Using this programme, the optimal alignment of the two sequences was compared with the alignment of 100 randomizations of the sequence. The significance of the optimal alignment is then given as:

$$\text{Significance} = \frac{\text{Maximal Quality} - \text{Average Quality}}{\text{Standard Deviation of Average Quality}}$$

Where the significance has a value of 3 or greater the comparison is thought to be, although may not necessarily be, significant. This

value has been determined empirically by comparing sequences of known functions. Using this equation the significance of the optimal alignment between Orf-A and pT181 RepC (Figure 5.12) was determined as being approximately 6.4 suggesting there was a functional similarity between the Rep proteins of the pT181 group of *S. aureus* plasmids.

#### 5.5 Further protein sequence alignments and significances with ORF-B

Alignments using the Clustal programmes had suggested that ORF-B was related to the RepA protein from pLS1. Attempts to determine the significance of this similarity was based on the programmes of the UWGCG package and accessory programmes of Roach and Parkhill as previously discussed (Section 5.4)

The Gap programme was used for the alignment of the RepA (pLS1) and ORF-B sequences and a statistical analysis of the alignment based the alignment of 100 randomizations of the sequences using the SGap programme was performed (Figure 5.13). The significance was determined as being 7.2. As previously discussed a value of 3 or above is usually regarded as being significant consequently this suggests that ORF-B had as structural similarity to RepA (pLS1).

RepA is a 5.1 kDa polypeptide required for replication of the plasmid pLS1 (Lacks *et al.*, 1986, del Solar *et al.*, 1987). RepA is stated to have a strong affinity for DNA and a role in the regulation of

Figure 5.12

An analysis of the significance of the alignment of the ORF-A and RepC amino acid sequences using the SGap programme.

```

ORF-A... ..MQVYLDRLMIKYKQVT. 16
RepC  MYKWHNHNHNSHLEHMDLDNFSKTYGSMRLDAHTVCI SDPRLSFDAMTI 50

17 ...EKQFSVDLTKISSKQIFLPNTPIRSEHGTSVRDYNRVIHIGYGEGA 62
51 VGNLNDRDQAQLSKFMSE...PQIRLMDILQTKFKA.....KALQEK 90

63 VYIGWKHNSKEKDSYDMKVDNPNPSKFENNELQKDSYEKFETVFHTLNA 112
91 VYIEYDKVKADSMRRRMRRIEFPNPKLTRDEM.....IWLQN1 129

113 VLKSNKRVVYQMDIAPDIERHMSDIVSYSTGKQ.....QDRHKGTVVY 156
130 ISYMDGGPTRLDLAPDFEEDLSQYAHSDKAVKKTIFYGRNGKPKETKYF 179

157 GNRNKDGYLKIYDKKKELYNHFKRMIEENLTRYEYSWRD..... 196
180 GVRDSNRFIIRIYNKKQERKDNADAENVSMZHLMRVVEISLKRMDVYNDCF 229

197 SDGVVVDIEIRKSPFSDIESYTFSLI..... 223
230 SDLHILQPDWRTIQTADRAIVFMLLSDEEWGKLHRNSRTKYKNLKEI 279

```

## Optimal alignment

Quality	93.3
Percent similarity	49.505
Percent identity	22.277

## Randomized alignments (100)

Average quality	73.5
Standard deviation	3.1
Significance	6.4



replication (Puyet *et al.*, 1988), although the mechanism for this is uncertain. The evidence regarding the strong affinity of RepA for DNA (Puyet *et al.*, 1986) is not stated. The homology between RepA and ORF-B and proposed related function suggested that ORF-B might also have an affinity for DNA. A computer search for the presence of structural features and characteristics with the amino acid sequence associated with known functions was made using the Profilescan programme (UWGCG). No evidence was obtained using this programme that either RepA or ORF-B had a DNA-binding capacity. The molecular structure of the protein-DNA recognition complex has been determined in a number of instances and involves a helix-turn-helix motif with one  $\alpha$ -helix lying in the major groove and the other  $\alpha$ -helix lying across the major groove (Pabo and Sauer, 1984). A structure prediction was made using the Peptidestructure and Plotstructure programmes (UWGCG) based on the algorithms of Kyte and Doolittle (1982) for RepA and ORF-B. A striking structural similarity was observed centering around the middle of the RepA sequence and the terminal region of the ORF-B sequence which was reminiscent of such a DNA-binding motif. However, the region of amino acid sequence homology did not correspond to these regions of structural similarity. It was not possible to obtain a hard copy of this output.

#### 5.6 A proposed nick site and ORF-A as a Rep protein

Replication of the *E. coli* phage  $\phi$ X174 has been shown to involve a catalyzed cleavage and ligation of the DNA and it has been observed that ssDNA plasmids which have homologous plus origins also have homologies

in their Rep proteins. A region has been shown within the Rep protein to be homologous to an amino acid motif around the active site of the  $\phi$ X174 Rep protein (van Manafeld *et al.*, 1986). Such a similarity (Figure 5.14) has led to the suggestion that the mechanism of nicking at the active site between these plasmids and  $\phi$ X174 is homologous. Numerous plus origins from ssDNA plasmids have been analyzed and have been found to be localized just upstream of or within the Rep open reading frame. The sequences around the nick sites of a number of such plasmids are similar and can be grouped into three classes based on sequence homology (Gruss and Ehrlich, 1989) (Figure 5.15). A homology search was made to see if any similar sequences were present within pHD2. This analysis suggested a possible nick site localized around base 170 in the sequence. The sequence around this region shows homology to the pT181 group of origins (Figure 5.16) which is the same group of plasmids to which homology had been observed between ORF-A and the Rep proteins. Additionally, the proposed nick site is just upstream of ORF-A which is consistent with that observed in other ssDNA plasmids of the pC194 and pE194 types although the nick site occurs towards the 5' end of the Rep proteins of the pT181 group of plasmids. The plus origins of most ssDNA plasmids contain sequences having the potential for secondary hairpin structures. A similar hairpin structure is possible within the proposed origin of pHD2 (Figure 5.17). However, the role of such structures with regard to plasmid replication is unknown.



**Figure 5.14****Conserved regions of Rep proteins of pC194-like origins**

pIJ101	Ala	Glu	Tyr	Ile	Ala	Lys	Thr	Gln	Asp	( <i>S. lividans</i>	72% GC)
<u>φX174</u>	<u>Ala</u>	<u>Lys</u>	<u>Tyr</u>	<u>Val</u>	<u>Asp</u>	<u>Lys</u>	<u>Lys</u>	<u>Ser</u>	<u>Asp</u>	( <i>E. coli</i>	51% GC)
pBAAI	Ser	Lys	Tyr	Pro	Val	Lys	Asp	Thr	Asp	( <i>B. subtilis</i>	42% GC)
pC194	Ala	Lys	Tyr	Ser	Gly	Lys	Asp	Ser	Asp	( <i>S. aureus</i>	34% GC)
pUB110	Ala	Lys	Tyr	Pro	Val	Lys	Asp	Thr	Asp	( <i>S. aureus</i>	34% GC)
pCB101	Phe	Lys	Tyr	Met	Thr	Lys	Val	Thr	Gly	( <i>C. butyricum</i>	27% GC)
pLP1	Ala	Lys	Tyr	Glu	Val	Lys	Ser	Ala	Asp	( <i>L. plantarum</i>	nd)

The φX174 residues are underlined. Homologous residues in the other sequences are shown in bold print.

Figure taken from Gruss and Ehrlich (1989).

**Figure 5.15**

**Consensus sequences within the plus origins of ssDNA plasmids.**

pT181	105	AGACAATTTTCTAAACCGGCTACTCT	AATAGCCGGTTGG	145
pC221	1246	AGACAATTTTCTAAACCGGCTACTCT	AATAGCCGGTTAA	1286
pS194	3404	ctcaAATTTTCTAAACCGGCTACTCT	AATAGCCGGTTAA	3444
4E174	4280	aTg TgC TccCccAaCTTG	ATA tTA	4303
pC194	1428	CTT TTC TTCTTAATCTTG	ATA ATA	1451
pUB110	4313	CTTGCTC TTCTTAATCTTG	ATACATA	4280
pRAA1	811	tcgGgTctTTCTTAATCTTG	ATAC TA	836
pCB101		TT TTC TTCTTAATCTTG	ATA ATA	
pLP1	2045	gTT TTC TTCTTAATCTTG	ATAC TA	2067
pE194	870	AAAaAtaGGGGGG	TACTACGA	CctcCCCCCTAG GTGtCCattGT CCAT 918
pLS1	427	AAAAtATgGGGGGGC	TACTACGA	CC CCCCTATaGPG CC GAGTGCCAA 474
pADB201	57	tcacgAccaGGGGGGC	TACTACGAtagC	CCCCAaTGAgtGatt TGTGACAT 106
pSB71	7	AAAAaAT GGGGG &	TACTACGA	C CCCCATTAAGTG CC GAGTGCCAA

' = Nick sites

— = Localisation of nick site

& = an additional sequence of 23 bp aaaggaagcgaattttgcttccg absent in other origins.

Nucleotides in capital letters are homologous; those in lowercase are nonhomologous.

Taken from Gruss and Ehrlich (1989).

**Figure 5.16**

### A possible plus-origin sequence within pHD2

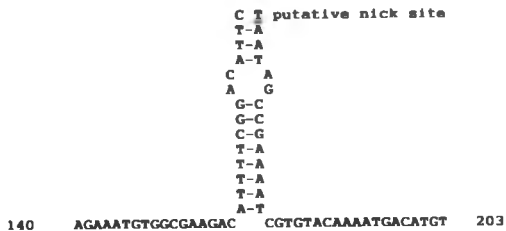
pC121	448	CTTAAATATGACGATGACATCTTTCTTAAACGGATCTCT	ATTAGCCGGTTCAGCGATCATCTGCGATCTTATCTTCCGCA	1350
pC121	1234	-----AAAGAGCTTTGACATCTTTCTTAAACGGATCTCT	ATTAGCCGGTTCAGCGATCATCTGCGATCTTATCTTCCGCA	1358
pH181	1126	-----AAAGAGCTTTGACATCTTTCTTAAACGGATCTCT	ATTAGCCGGTTCAGCGATCATCTGCGATCTTATCTTCCGCA	27
pH92	126	TGCTTGTAATATGACGATGACATCTTTCTTAAACGGATCTCT	ATTAGCCGAAATGCTTCGATGATGATCTTATCTTAAATAT	315
pH194	3406	ATGCTTGTAATATGACGATGACATCTTTCTTAAACGGATCTCT	ATTAGCCGGTTCAGCGATCATCTGCGATCTTATCTTCCGCA	2495
pH1212	1305	AGCTATGCGACGAAATATGAAATTTCTTAAACGGATCTCT	ATTAGCCGGTTCAGCGATCATCTGCGATCTTATCTTCCGCA	1394

Nucleotides in pHD2 which are conserved in one or more pT181 plasmid family sequences are indicated by |. Non-conserved nucleotides are indicated by -.

Sequence data derived from Projan and Novick (1988).

Figure 5.17

A possible hairpin structure within the proposed plus origin  
of pHD2.



### 5.7 A possible minus origin

The role of plus origins, nick sites and Rep proteins has been discussed (Chapter 1). There is also some evidence of inverted repeat structures being implicated in plasmid replication. Insertion of a pair of inverted complimentary repeat sequences from the plasmid pC194 along with an additional 200 nucleotides and the chloramphenicol resistance determinant into pBR322 was found to give a construct capable of autonomous replication in *B. subtilis* (Morinouchi and Weisblum, 1982). A similar structure was also found to be present in the plasmid pE194 (Morinouchi and Weisblum, 1982b). The structural similarity of these inverted repeat sequences, and evidence linking both of these regions with plasmid replication, suggested that they played a similar role in plasmid replication.

The conversion of the single-stranded plasmid intermediate to the final double-stranded form requires a minus origin. This involves host factors (Boe *et al* 1989; del Solar, 1987; Gruss *et al* 1987) and is initiated by RNA polymerase since rifampicin blocks conversion (Boe *et al.*, 1989). The minus origins of a number ssDNA plasmids have been analyzed (Boe *et al.*, 1989; del Solar *et al.*, 1987; Devine *et al.*, 1989; Gruss *et al.*, 1987). In all cases these minus origins contain imperfect palindromic structures covering a region 130 to 220 bp in size (Figure 5.18). The presence of a long inverted complimentary repeat sequence in pHD2 between bases 1565 and 1643 (Figure 5.19) may have a similar role in plasmid replication although no further evidence is available to support this.

**Figure 5.18**

### Minus origin sequences and palindromes

pC194 M-O (palA)

2209 1  
 >-----  
 GCATTTATGCCGAGAAAATTTATGTGCGTTGAGAAAGACCCCTTACTAACTTGCAGACGAATGTCGGCAT  
 -----1  
 AGCGTGAGCTATTAAAGCCACCATTCGACAAAGTTTGGGATTGTAAGGGTTCGAGGCTCAACGTCAATAA  
 -----  
 AGCAATTGGAATAAAG 2050

pUB110 H-O

1523 1  
ACCTCTCTTGATCTTTTATTATTGAGTGGTTTGTCCGTTACACTAGAAACCGAAGACAATAAAAAATT  
1 2  
< X  
TTATTCTGCTGAGCTGGCTTTCGGTAGCTAGACAAACGGACAAATAAAAAATTGGCAAGGTTAAAG  
2  
> < >  
GTGGAGAGATTTTCTGAGTGATCTCTCAAAAAATACTACTCTGCCTTGTGATTTTAAACGAGCACGAGAG  
3 3  
CAAAACCCCCCTTGTGAGGTGGCAGAGGGCAGGTTTTTTTGTCTCTTTTCTCGTAA 1246

PBA1 M-0

[illegible]

Palindromes are indicated by dashed lines and arrow heads with dots representing nonpalindromic bases.

Figure taken from Gruss and Ehrlich (1989).

Figure 5.19

A long inverted complimentary repeat sequence present within  
pHD2. A possible minus origin.

```

      1   2   3       4   5   6
>----->
GGAGATGCACATCGATAAACATAATGCGTAGTTGGTGGCTGAAGTTGCCGCCACCTA
CCTCTACGTGTAGCTATTTGATTACGCATCAACACCGACTTCAACGGGCGGTGGAT
      6   5

```

```

CTCATTAGAAATATCCGTGCATGGGTCCC
GAGTAAATCTTATAGGCACGTACCCAGGG
<-----
      4   3   2   1

```

Palindromes are indicated by dashed lines and arrow heads with dots  
representing nonpalindromic bases.

## 5.8 Summary and discussion

The *B. thuringiensis* plasmid pHD2 has been shown to contain two open reading frames designated ORF-A and ORF-B encoding presumptive polypeptides of Mr 26,447 and 9,122 respectively. ORF-A appears to be distantly related to the Rep proteins of the pT181 group of *S. aureus* plasmids and the presence of specific regions of conservation suggests that these may represent functional domains of the Rep protein. The presence of a putative plus origin and nick site just upstream of ORF-A is consistent with the observation that these generally occur upstream of or just within the Rep protein sequence. The presence of a long imperfect palindromic sequence is similar to that observed at the minus origins of other saDNA plasmids. The evidence based on DNA and amino acid sequence analysis and comparison suggests that the plasmid pHD2 may be a member of the family of saDNA plasmids found in other Gram-positive bacteria and appears to be most closely related to the pT181 group of *S. aureus* plasmids. Control of replication may involve a short 77 amino acid polypeptide, ORF-B, in a similar, although as yet uncertain, way to that observed with repA of pLS1.



## Chapter 6

### Replication of pHD2 chimeras in Gram-positive hosts

#### 6.1 Introduction

The study of a plasmid in any host is greatly facilitated if it confers a phenotype upon the host cell which can easily be identified or selected. Two chimeric plasmids, pDM100 and pDM200, were available consisting of the 2 kb *B. thuringiensis* plasmid, pHD2, cloned in different restriction sites in the *E. coli* vector pBR322 (Chapter 3). Assuming no sequences vital to the functioning of the pHD2 portion of the replicons have been inactivated during the cloning process, these chimeric replicons should be capable of functioning as in both *E. coli* and *B. thuringiensis* hosts. If either of these constructs were to be detected when transferred back into a *B. thuringiensis* strain, a selectable marker such as an antibiotic resistance gene would be required in order to identify recipient cells and would facilitate study of their segregational and structural stability in *B. thuringiensis* or other Gram-positive hosts.

A plasmid, pC194, which was originally isolated from *Staphylococcus aureus* by Iordaneanu (1976), was found to replicate in a wide range of hosts including *B. thuringiensis* (Martin *et al.*, 1981) and *B. subtilis* (Ehrlich, 1977) where it was found to confer resistance to chloramphenicol. Horinouchi and Weisblum (1982b) cloned portions of pC194 into the *E. coli* plasmid, pBR322, and identified a minimal (1035

bp) fragment which was capable of expressing resistance to chloramphenicol in *E. coli*. Whilst a composite plasmid consisting of this 1035 bp fragment in pBR322 was not found to be functional in *B. subtilis*, other constructs containing a larger fragment of pC194 were found to be capable of replication in this host and expressed resistance to chloramphenicol.

## 6.2 Cloning of a marker functional in *Bacillus*

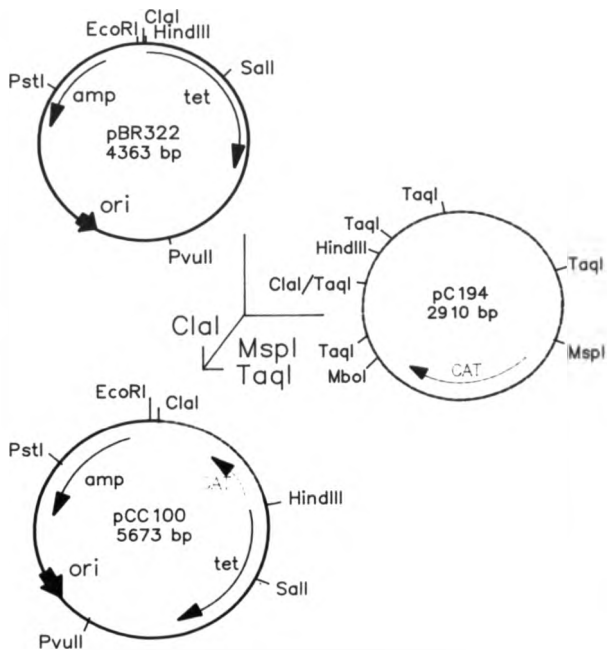
The minimal region of 1035 bp from pC194, expressing resistance to chloramphenicol, but not containing the pC194 origin of replication or any other open reading frames, was selected as a suitable candidate for an antibiotic resistance marker to be inserted into the constructs pDM100 and pDM200. This selectable marker could be excised from pC194 on an *HspI*/*TaqI* fragment. The 'sticky ends' generated by both of these restriction enzymes are compatible with those generated using *ClaI* and so fragments generated using these restriction enzymes could be ligated directly into a *ClaI* site in the recipient plasmid without any further modification of the DNA. The *ClaI* site would not be regenerated in this process without any further modification of the DNA.

The possibility of interrupting a sequence which is vital for the replication or maintenance of the plasmid pHD2 in any cloning experiment has been discussed previously (Section 3.5). It was decided to clone the chloramphenicol acetyl transferase (CAT) gene pC194 into the *ClaI* site of pBR322 rather than directly into the *ClaI* site of the *B.*

*thuringiensis* plasmid to avoid any such complication. Cleavage of either pDM100 or pDM200 with *Cla*I makes this process more complicated due to the presence of two *Cla*I sites. To overcome this problem, pC194 which had been restricted with *Msp*I and *Taq*I was ligated into pBR322 which had been linearised with *Cla*I. The product of this ligation should not contain any *Cla*I sites since this sequence is not regenerated by the fusion of *Taq*I or *Msp*I generated molecules with those generated using *Cla*I. Consequently, the 2 kb sequence from pDM100 could not be subcloned into this plasmid although this should be possible from pDM200. No purification of the donor fragment, using agarose gels or electroelution, or treatment of the vector with CIP, was carried out as positive selection was possible using chloramphenicol resistance. The products of this ligation were transformed into *E. coli* DH1 and chloramphenicol-resistant transformants were selected. Plasmids were purified from the chloramphenicol-resistant transformants using the alkaline lysis mini-preparation procedure. Restriction analysis of these plasmids showed the presence of an insert of approximately the correct size in the *Cla*I site which could be released using an *Eco*RI/*Hind*III double digest. It seemed likely that this construct was correct since it also conferred resistance to chloramphenicol. This construct will be referred to in future as pCC100.

Further restriction analysis of pCC100 by comparison of the insert size with that of the target sequence generated using an *Msp*I/*Taq*I double digest of pC194 revealed this insert to be slightly too large. A *Cla*I site was also found to be present which would not have been regenerated by the cloning route described above. It was thought that

Figure 6.1  
Derivation of pCC100



Lane 1 2 3 4 5 6 7 8



kb

▶ 23

▶ 9.4

▶ 6.6

▶ 4.4

▶ 2.3

▶ 2.0

▶ 0.56

Figure 6.2

## Restriction endonuclease analysis of pCC100 and pCC150

- Lane 1 pBR322 restricted with *EcoRI/HindIII*  
Lane 2 pCC100 restricted with *EcoRI/HindIII*  
Lane 3 pCC150 restricted with *EcoRI/HindIII*  
Lane 4 pC194 restricted with *MspI/TaqI*  
Lane 5 pCC100 restricted with *ClaI*  
Lane 6 pCC150 restricted with *ClaI*  
Lane 7 pCC100 restricted with *ClaI/HindIII*  
Lane 8 Lambda DNA restricted with *HindIII*

the initial *TaqI* digest used to prepare the CAT fragment had been partial and that cleavage had occurred at the *ClaI* site which was also a *TaqI* site. It turned out that the size of the cloned fragment corresponded to that which would have been generated had cleavage occurred at an adjacent *TaqI* site. The cloning was repeated and the originally intended clone generated and termed pCC150. Transformation of pCC150 into *E. coli* DH1 conferred resistance to chloramphenicol and restriction analysis showed that the intended restriction fragment had been cloned. The derivation of pCC100 is shown in Figure 6.1 and a restriction endonuclease analysis of this construct in Figure 6.2.

### 6.3 Construction of shuttle vectors

As previously discussed the CAT gene from pC194, which was to be used as a marker in *Bacillus* sp. had been cloned into pBR322 to give the constructs pCC100 and pCC150. Apart from the possibility of interrupting regions important in the replicon functions of the 2 kb sequence, it was not convenient to clone this fragment directly into pDM100 or pDM200 as these both contain two *ClaI* sites. It was consequently decided to subclone the 2 kb pHD2 plasmid sequences from pDM100 and pDM200 which had been excised using *ClaI* and *HindIII* as appropriate into the *ClaI* and *HindIII* sites of pCC100. The generation of the clone pCC100 with the larger than intended insert was advantageous in this procedure since a *ClaI* site was still available and so it was decided to base the shuttle vectors on this construct.

The loss of tetracycline resistance, which had been used previously to identify clones with inserts in the *Hind*III or *Cla*I sites of pBR322, could not be used to detect inserts in the derivative pCC100 since the tetracycline resistance gene had been inactivated by the cloning of the CAT gene from pC194. The *Cla*I and *Hind*III sites, into which the 2 kb, pHD2, plasmid sequence was to be inserted, were present within the sequence which had been cloned in the generation of the construct pCC100. No selective mechanism was available to detect the presence of inserts in these restriction sites in pCC100 although transformants could be detected using ampicillin or chloramphenicol resistance.

The cloned pHD2 plasmid sequences were released from pDM100 and pDM200 using *Cla*I and *Hind*III as appropriate and ligated into pCC100. In order to reduce the number of transformants needed to be screened, pCC100 which had been cut with the appropriate restriction enzyme was treated with CIP to prevent recircularization of the vector without an insert molecule. The efficiency of the CIP treatment was assessed by self-ligating the vector and comparing the efficiency of transformation into *E. coli* DH1 with that obtained in the cloning reaction.

The attempted subcloning of the pHD2 plasmid sequences, purified from pDM100 and pDM200 as described, gave rise to many ampicillin-resistant colonies upon transformation into *E. coli* DH1. It was concluded that the majority of these were likely to be the desired recombinants by comparison with the number of transformants obtained using self-ligated vector. A number of the putative clones were



Lane 1 2 3 4 5



kb

23

9.4

6.6

4.4

2.3

2.0

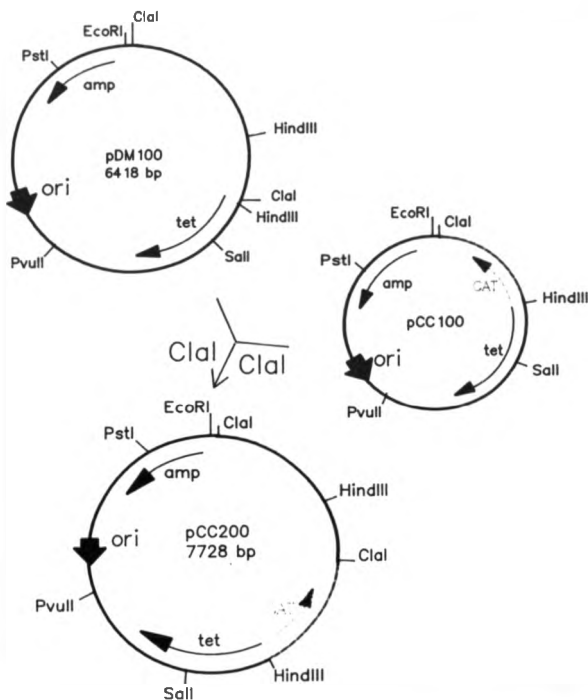
0.56

Figure 6.3

Restriction endonuclease analysis of pCC200

- Lane 1 pCC100 restricted with *EcoRI*/*HindIII*
- Lane 2 pDM200 restricted with *HindIII*
- Lane 3 pCC200 restricted with *EcoRI*/*HindIII*
- Lane 4 pCC200 restricted with *ClaI*
- Lane 5 Lambda DNA restricted with *HindIII*

Figure 6.4  
Derivation of pCC200



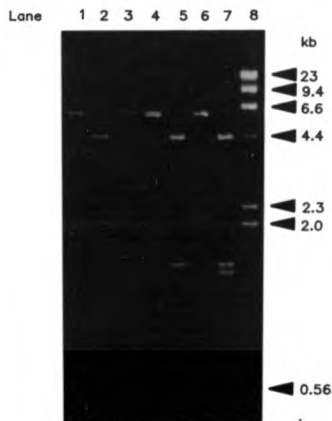
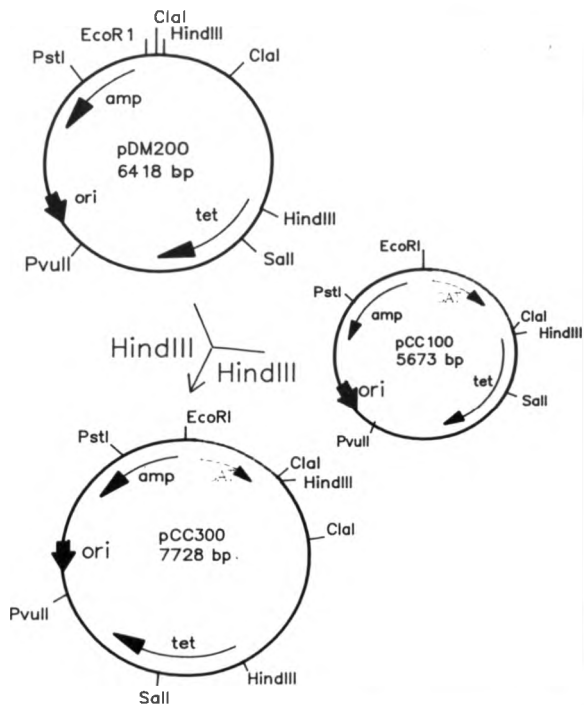


Figure 6.5

## Restriction endonuclease analysis of pCC300

- Lane 1 pCC100 restricted with *Clal*
- Lane 2 pDM100 restricted with *Clal*
- Lane 3 pCC300 restricted with *Clal*
- Lane 4 pCC100 restricted with *HindIII*
- Lane 5 pCC100 restricted with *Clal/HindIII*
- Lane 6 pCC300 restricted with *HindIII*
- Lane 7 pCC300 restricted *Clal/HindIII*
- Lane 8 Lambda DNA restricted with *HindIII*

Figure 6.6  
Derivation of pCC300



screened by restriction mapping of the plasmids, prepared using the mini-preparation procedure, for the presence of the 2 kb, pHD2, plasmid sequence. One example of each category of transformants, thought to contain the appropriate insert was prepared on a large scale and purified using a caesium chloride/ethidium bromide density gradient centrifugation. These plasmids were subjected to further restriction mapping (Figures 6.3 and 6.5) and were confirmed to contain the correct sequences. These constructs will be referred to as pCC200 (Figure 6.4) and pCC300 (Figure 6.6) in which the pHD2 plasmid sequence was obtained from pDM100 and pDM200 respectively.

The two constructs, pCC200 and pCC300, each have only one interruption in the pHD2 plasmid sequence and it was hoped that one or both still represented viable replicons based on pHD2 as well as pBR322. Both molecules are functional in *E. coli* where both chloramphenicol and ampicillin resistance can be used as selectable markers. Chloramphenicol resistance should also be available as a marker in *Bacillus* sp. if the replicons are still functional in such a host. If the constructs are functional in *B. thuringiensis*, they have the capability of being used to transfer cloned sequences between *E. coli* and *B. thuringiensis* and so may be considered shuttle vectors.

#### 6.4 Transformation into *B. subtilis* and *B. thuringiensis* subsp. *israelensis* IPS-78/11

A number of protocols are available for the transformation of plasmids into a number of *Bacillus* species including *Bacillus thuringiensis*. The reproducibility of some these transformation protocols in different laboratories appears to be variable and highly strain-dependent. Protocols for the transformation of *B. subtilis* appeared to be more reliable and, as it was uncertain whether the constructs, pCC200 and pCC300, would be functional in a *Bacillus* species, it was decided to transform them into *B. subtilis* in the first instance.

The plasmid, pC194, from which the chloramphenicol acetyl transferase gene to be used as a selectable marker had been derived, and which was known to be functional in *B. subtilis*, was used as a control to check the efficiency of transformation in the system being used. *B. subtilis* was generally found to be transformed by pC194 at a frequency of  $2 \times 10^2 \mu\text{g}^{-1}$  DNA although transformants could not be obtained when using either of the constructs pCC200 or pCC300.

The pC194 DNA being used as a positive control had been obtained from *B. subtilis* and it is possible that because of this it may transform at a far greater efficiency than either of the constructs which had been propagated in an *E. coli* host. As discussed in Chapter 1, multimeric forms of plasmids are required to transform *B. subtilis* and these may be more abundant when the plasmid DNA is purified directly



from this organism rather than from an intermediate *E. coli* host. Another possibility is that the pC194 DNA was unaffected by the restriction system of the *B. subtilis* recipient due to its previous modification in this host unlike the pCC200 and pCC300 plasmid DNA which had been propagated in *E. coli* DH1. Either of these factors could be responsible for a difference in the transformation efficiency between these plasmids of several orders of magnitude. In this case the transformation efficiency using the constructs pCC200 and pCC300 might be below the level at which they would be expected to be detected. An alternative strategy was therefore required in order to test whether either of these constructs were functional in *Bacillus*.

The electroporation of a wide variety of species have recently been reported. It was possible that the constructs were functional but were unable to replicate in *B. subtilis* because of a limitation in their host range. Electroporation provided an alternative to transformation by which the constructs could be introduced into strains of *B. thuringiensis* for which other transformation protocols appeared to be unreliable (B. Dancer, personal communication).

Work by Bone and Ellar (1989) showed that whilst the electroporation of certain subspecies of *B. thuringiensis* was highly efficient, giving up to  $10^5$  transformants  $\mu\text{g}^{-1}$  DNA, other subspecies, including *B. thuringiensis* subspecies *kurstaki* HD1, were found not to be transformed or transformed only poorly using this procedure. The transformation efficiency of the subspecies under investigation was also shown to be dependent upon the plasmid being used. Some hosts were

observed to be more readily transformed by a wider range of plasmids than others. Ellar and co-workers kindly supplied a derivative of *B. thuringiensis* subsp. *israelensis*, *B. thuringiensis* subspecies *israelensis* IPS-78/11 (Bone and Ellar, 1989) which was not only transformable by a variety of plasmids at high efficiency, but which had also been cured of many of its endogenous plasmids allowing easier recognition of transformed plasmids when the plasmids contained into the transformants were purified and displayed on agarose gels.

Attempts to transform *B. thuringiensis* subsp. *israelensis* IPS-78/11 using electroporation gave rise to chloramphenicol-resistant transformants with both pCC200 and pCC300. Both of these constructs were therefore shown to be functional in this host and to be capable of expressing the chloramphenicol resistance marker. The construct, pCC100, from which these two constructs had been derived was not found to be functional in this host. It was concluded from this that the functions required for replication and maintenance of the constructs pCC200 and pCC300 in *B. thuringiensis* subspecies *israelensis* IPS-78/11 were supplied by the pHD2 portion of the sequence rather than that which was derived from pC194. It was presumed that no important sequences had been interrupted during the cloning process since both pCC200 and pCC300 were both still functional in *B. thuringiensis*. The *Cla*I restriction site occurs downstream from ORF-B and the *Hind*III restriction site just downstream from ORF-A and neither interrupts either the proposed plus or minus origins (Chapter 5).

Numerous instances of deletion or recombination events occurring in plasmids upon transformation into *Bacillus* species have been reported (for review see Gruss and Ehrlich, 1989). In order to investigate whether any such events had taken place in this system the plasmid had to be re-isolated. The presence of other plasmids present in this strain made this slightly more awkward but was achieved by first transforming plasmids purified from the chloramphenicol-resistant *B. thuringiensis* isolates into *E. coli* DH1 thereby selecting for *E. coli* replicons. Plasmids were purified from chloramphenicol-resistant *E. coli* DH1 transformants on a large scale using the caesium chloride gradient procedure and investigated further by restriction mapping. It was not possible to obtain *E. coli* DH1 transformants when selecting for ampicillin resistance. Additionally, subculturing of chloramphenicol resistant transformants onto LB plates containing ampicillin ( $100 \mu\text{g ml}^{-1}$ ) was unsuccessful, suggesting that the ampicillin had in some way been inactivated. Restriction mapping of a number of such isolates revealed the presence of a consistent deletion (Figure 6.7). No *E. coli* transformants could be obtained using plasmids purified from *B. thuringiensis* transformed with pCC300. It is possible that a deletion occurring within this construct in *B. thuringiensis* has in some way inactivated the pBR322 portion of the replicon so that it is no longer functional in *E. coli*. It did not prove possible to visualize the presence of any plasmids within transformed isolates of *B. thuringiensis* subsp. *israelensis* IPS-78/11 using agarose gel electrophoresis despite the ability to recover functional replicons upon isolation and transformation into *E. coli* DH1. Similarly, it was not possible to detect single-stranded plasmid intermediates or high molecular weight



Figure 6.7

Restriction endonuclease analysis of pCC200 derivatives following electroporation into *Bacillus thuringiensis* subspecies *israelensis* IPS-78/11

- Lane 1 pCC100 restricted with *Cla*I
- Lane 2 pDM100 restricted with *Cla*I
- Lane 3 pCC200 restricted with *Cla*I
- Lane 4 pCC100 restricted with *Hind*III
- Lane 5 pCC100 restricted with *Cla*I/*Hind*III
- Lane 6 pCC200 restricted with *Hind*III
- Lane 7 pCC200 restricted *Cla*I/*Hind*III
- Lane 8 Lambda DNA restricted with *Hind*III
- Lane 9 pCC200 restricted with *Cla*I
- Lane 10 pCC200 restricted with *Hind*III
- Lane 11 pCC200 restricted with *Cla*I/*Hind*III
- Lane 12 pCC200 derivative 1 restricted with *Cla*I
- Lane 13 pCC200 derivative 1 restricted with *Hind*III
- Lane 14 pCC200 derivative 1 restricted with *Cla*I/*Hind*III
- Lane 15 pCC200 derivative 2 restricted with *Cla*I
- Lane 16 pCC200 derivative 2 restricted with *Hind*III
- Lane 17 pCC200 derivative 2 restricted with *Cla*I/*Hind*III
- Lane 18 pCC200 derivative 3 restricted with *Cla*I
- Lane 19 pCC200 derivative 3 restricted with *Hind*III
- Lane 20 pCC200 derivative 3 restricted with *Cla*I/*Hind*III

plasmid multimers the occurrence of which are both common features with ssDNA plasmids.

#### 6.5 Deletional events in ssDNA plasmids

Attempts to combine the advantages of the well-defined *E. coli* cloning system with that of the *Bacillus* host has led to the construction of hybrid vectors. Numerous instances of such fusions have been documented (Kreft and Hughes, 1982). Deletion products of such fusions are not uncommon and are thought to result from errors in the replication process. Almost all the steps in the replication of ssDNA plasmids have been shown or proposed to diverge from the usual process in the generation of genetic rearrangements. Work by Michel and Ehrlich (1986a) on hybrids between the plasmids pC194, pBR322 and the bacteriophage  $\phi$ 1 has implicated the erroneous recognition of nick sites in the generation of deletion products. Fusions between pC194, pBR322 and  $\phi$ 1 were found not to be stable in *E. coli* and underwent deletions. One end of this deletion was found to correspond to the nick site of pC194 whilst the other appeared to have occurred within regions resembling the termination region. Such a deletion could result from the erroneous termination of replication at sites resembling the correct termination site. Such a deletion event appears to be dependent upon the orientation of the  $\phi$ 1 phage replication fork. The deletion events appears to be 200 times more frequent when the orientation is such that the nicking event occurs on the same strand rather than opposite strands. Additionally, a cointegrate between pC194 and pUB110 which

decombines at high frequency (Rahn and Dubnau, 1985) has been shown to result from initiation of replication at one origin and correct termination at the second origin (Gros *et al.*, 1987).

The deletion observed upon transformation of the construct, pCC200, into *B. thuringiensis* subsp. *israelensis*, inactivates the *amp* gene and removes *Clal* site. This is consistent with a deletion event occurring between a region within the *amp* gene of pBR322 and a region within the first part of the pHD2 sequence. The proposed nick site for pHD2 is situated within the terminal region of the pHD2 sequence in the clone pCC200. A deletion event similar to that previously discussed (Michal and Ehrlich, 1986) in which erroneous termination of replication occurred at a site similar to the pHD2 origin within the *amp* gene could account for this result. Sequencing across the deletion end points would be required in order to confirm this hypothesis. All the cell lines produced on transformation contained only one type of plasmid suggesting that deletions arise in transformants at or before the onset of plasmid replication. It also appears that the deleted plasmid does not undergo further rounds of deletion since only a single deletion product may be isolated.

#### 6.6 Repeated sequences

The presence of directly repeated and inversely repeated sequences was screened for using the sequence analysis programme of Queen and Korn (1984). The sequence was observed to be rich in both repeated and

inverted repeated sequences with 211 of directly repeated and 75 inverted repeat sequences being found which satisfied certain minimal homology requirements.

It is possible that the generation of deletions in the construct pCC200 may be linked to these repeated sequences. The generation of deletions in recombinant plasmids by intramolecular rearrangement is a common phenomenon. Many of these deletions involve short direct repeats. Although some of them may arise by typical homologous recombination between directs, which in *E. coli* requires *recA* function, others occur in *recA* mutants apparently as a result of other mechanisms for recombining DNA (Albertini *et al.*, 1982). In the first kind of intramolecular recombination, endpoints of deletions occur within short direct repeats (Albertini *et al.*, 1982; Ballaster *et al.*, 1986). The plasmid replication mechanism may also be linked in this process. In *B. subtilis*, a hybrid plasmid composed of the *S. aureus* plasmids pUB110 and pSA2100 showed a systematic *recE*-independent, reciprocal, conservative, intramolecular recombination event giving rise to two plasmids. This event was shown to occur between 18 bp directly repeated sequences (Hahn and Dubnau, 1985). The presence of randomly distributed 22 or 29 base pair repeated sequences within a construct was shown not to be sufficient for such a recombination event to occur. Such an event appears to be associated with the origin of the plasmid and probably results from a mistake in the recognition of the termination sequence in plasmid replication. A similar process has been shown to occur in a derivative of pUB110, pBT33, which undergoes a deletion to form pBT82. Sequencing studies have revealed that this deletion occurs between 2



direct repeats of 6 bases of which, one was localized within the proposed origin sequence (Maciaz *et al.*, 1988). Structural instabilities in naturally occurring plasmids are unlikely to be observed. However, the manipulation of plasmids in which short direct repeats of sequences occurring within the origin are introduced, may play a role in the structural instability of some of these constructs based on mistakes in the replication process.

A role for the presence of inverted repeat sequences in the regulation of gene expression has been suggested (Morinouchi and Weisblum, 1982b). In the plasmid pC194, the presence of four inverted repeat sequences in close proximity to each other upstream from the translation start site on the mRNA coding for chloramphenicol acetyl transferase was noted. Five configurations of the mRNA were possible based on these four sequences, four of which masked the Shine-Dalgarno sequence. It was postulated that the mRNA was normally found in a structure in which the SD sequence was masked and proposed that induction of this gene caused a conformational realignment of the control region unmasking the SD sequence allowing translation to proceed.

The presence of many small inverted repeat sequences has also been reported within the *B. thuringiensis* subsp. *kurstaki* HD-73 crystal protein crystal gene. The longest of these were found to be in the 3'-flanking DNA approximately +100 and +200 bases from the termination codon comprising a 14 nucleotide and 17 nucleotide inverted repeat. It was suggested that such secondary stem-loop structure followed by T-rich

sequence could play a role in termination of transcription (Rosenberg and Court, 1979).

The majority of the inverted repeated sequences were fairly short being only 20 bases or less and contained a number of mismatches. The free energies of these structures suggested that these did not have a structural role in gene regulation and indeed, the examination of the sequences upstream of the open reading frames did not reveal the presence of any such structures which would mask the SD sequences. The only inverted repeat structure which seemed to have any significance occurred at the end of ORF-B. This long imperfect inverted repeat structure was similar to that observed at the minus origins of a number of ssDNA plasmids and was postulated to have a similar role in pHD2 (Section 5.7).

#### 6.7 Plasmid stability studies

The control of plasmid replication with respect to cell division and the partitioning of the plasmids between daughter cells during cell division are critical if a plasmid is to be stably maintained. The insertion of foreign DNA into a plasmid can disrupt the normal control mechanisms and result in segregational as well as genetic instability. This subject is discussed in detail in Chapter 1.

In order to assess its segregational stability, a *B. thuringiensis* subspecies *israelensis* strain IPS-78/11 containing the deleted form of

the construct pCC200 was cultured in SR (100 ml) broth containing chloramphenicol at a concentration of  $5 \mu\text{g ml}^{-1}$  at  $30^{\circ}\text{C}$  until an  $\text{OD}_{650}$  of approximately 1 was attained. This culture was then diluted into a fresh SR broth without the addition of chloramphenicol and incubation continued. The dilution factor was calculated to give 10 ten cycles of cell division for each dilution. Following each 10 rounds of cell division, the culture was plated out on SR agar without the addition of chloramphenicol and incubated at  $30^{\circ}\text{C}$  overnight. 100 of the colonies obtained were then replica-plated onto SR agar plates with and without selection using chloramphenicol at  $5 \mu\text{g ml}^{-1}$ . No loss of the chloramphenicol resistance phenotype was observed over 100 generations. This may suggest that a highly efficient partitioning system is functioning with pHD2 and that this process has not been disrupted during either the cloning steps in the generation of pCC200 or the formation of the deleted derivative upon transformation.

## 6.8 Summary and Discussion

The two constructs pDM100 and pDM200 had previously been generated by cloning the *B. thuringiensis* pHD2 plasmid in the *E. coli* vector pBR322. Assuming no vital sequences in the function of the pHD2 portion of the replicon had been interrupted in the cloning process, both of these constructs seemed likely to function in both *Bacillus* and *E. coli* hosts. However, the lack of a suitable marker for the selection of these replicons in a Gram-positive host meant that this could not be readily tested. Consequently, The chloramphenicol acetyl transferase

gene from the *S. aureus* plasmid, pC194, was subcloned into pBR322 to give the construct pCC100. The construct contained both *Cla*I and *Hind*III restriction sites as discussed and an antibiotic resistance marker which was known to be functional in a Gram-positive host. The construct pCC100 was demonstrated not to be functional in *B. thuringiensis* subspecies *israelensis* IPS-78/11. The pHD2 sequences from pDM100 and pDM200 were subcloned into pCC100 to yield to constructs pCC200 and pCC300. It was not possible to obtain transformants of *B. subtilis* using either of these replicons but both were capable of transforming *B. thuringiensis* subsp. *israelensis* IPS-78/11 to chloramphenicol resistance using the electroporation procedure. Plasmids could be purified from pCC200 transformed isolates and transformed into *E. coli* DH1 when selecting for chloramphenicol resistance. All the transformants obtained appeared to have a consistent deletion encompassing the terminal portion of the pHD2 portion of the pCC200 construct in which the *Cla*I restriction site was removed and the *Hind*III site retained and which also inactivated the ampicillin resistance gene in the pBR322 portion of the construct. One end of the deletion was in the region of the pHD2 sequence where the nick site was thought to be placed and it is possible that erroneous nick site recognition may be responsible for this deletional event as shown for a number of other ssDNA plasmid. It was not possible to obtain *E. coli* transformants from *B. thuringiensis* subsp. *israelensis* transformed with pCC300 and it is possible that a similar deletional event had inactivated the pBR322 portion of the replicon in this instance. It was not possible to identify the presence of plasmids in the form of monomers, high molecular weight multimers or single-stranded intermediates in the course of this study although the

presence of pCC200 in *B. thuringiensis* subsp. *israelensis* IPS-7-11 transformed to chloramphenicol resistance was confirmed by obtaining an *E. coli* replicon from this isolate. A *B. thuringiensis* subsp. *israelensis* IPS-78/11 isolate which had been transformed to chloramphenicol resistance using the construct pCC200 was shown to maintain the plasmid over a period of 100 generations in the absence of selective pressure and was therefore assumed to have a highly efficient plasmid partitioning system in operation. Whilst such a segregationally stable replicon could be of use in the cloning and expression of heterologous genes in a fermenter system where selective pressure could not be readily applied, the deletional events observed thought to result from pHD2 being a member of the ssDNA group of plasmids mean that vectors based on pHD2 are unlikely to be suitable in such studies.

## Chapter 7

### Overall conclusions and outlook

At the time of writing, though considerable work has been done to characterize plasmids from a range of Gram-positive bacteria, little is known concerning the plasmids of the industrially important species of organism collectively known as *Bacillus thuringiensis*. It was the primary aim of this project to characterize one or more of the low molecular weight plasmids from *Bacillus thuringiensis* subspecies *kurstaki* HD1-Dipel in terms of genetic organization and mode of replication.

As described in the preceding chapters, an investigation of a small plasmid of approximately 2 kb termed pH2 was performed using a number of constructs based upon this plasmid in both *E. coli* and *Bacillus* hosts. The majority of plasmids which have been studied from Gram-positive hosts appear to be members of a highly related family of plasmids which appear to replicate via a rolling circle mechanism involving a single-stranded DNA intermediate which are consequently termed ssDNA plasmids. Members of this family have been isolated from a number of different genera including *Staphylococcus*, *Streptococcus*, *Streptomyces*, *Lactobacillus*, *Clostridium* and *Bacillus* (see Gruss and Ehrlich, 1989).

The *Bacillus thuringiensis* plasmid, pKD2, was determined to be 2.055 kb in size and contained two open reading frames, termed ORF-A and ORF-B, with possible promoter and Shine-Dalgarno sequences upstream from the putative translational start site suggesting these open reading frames could be functional genes.

ORF-A encoded a polypeptide with a predicted molecular weight of 26.447 kDa which appears to be distantly related to the Rep proteins of the pTi81 group of *S. aureus* plasmids demonstrated to replicate in a mechanism involving a single-stranded DNA intermediate. This ssDNA family of plasmids has been shown to be divided into a number of groups based on sequence conservation around the nick site involved in the initiation of replication. A similar sequence was located just upstream of ORF-A which is consistent with the finding that the nick sites of the ssDNA group of plasmids are generally located just upstream of or towards the 5' end of the *rep* gene.

The comparison of sequences around the nick sites of the ssDNA group of plasmids has suggested that plasmids pTi81, pC221 and pS194 fall into one group, pC194, pUB110, pBAA1, pCB101 and pLP1 a second group whilst pE194, pLS1, pADB201 and pSH71 constitute a third (see Gruss and Ehrlich, 1989). To date, most comparisons between the plasmids have been at either a functional level whereby similarities in plasmid replication such as the presence of a plus origin, Rep protein, minus origin and the occurrence of a single-stranded intermediate as well as structural and segregational instabilities which appear to be a function of the type of replication exhibited by this group of plasmids.

Structural similarity has additionally been noted at the sequence level between relatively short sequences at the nick site between the plasmids of difference genera as well as longer regions of homology of the Rep proteins and copy control regions between the pT181 group of plasmids (Projan and Novick, 1988). Close sequence homology between the *S. aureus* plasmids and the comparison of other plasmids to the *S. aureus* plasmids has implied that the ssDNA group of plasmids is highly related. However, evidence presented here suggests that the family contains members much more distantly related to those members which have to date been extensively studied.

The mechanism of nicking in the initiation of the replication process where a rolling circle method is utilized has been suggested to be analogous to that of the *E. coli* phage  $\phi$ X174 due to the conservation of an amino acid motif within the Rep proteins utilizing the pC194 group of origins similar to that at the active site of the  $\phi$ X174 Rep protein. This is despite differences in host G+C content of between 27 % (pCB101 of *Clostridium butyricum*) and 72 % (pIJ101 of *Streptomyces lividans*). The fact that the Rep proteins and plus origins sequences from plasmids of Gram-positive bacteria are related is not therefore surprising since sequence conservation in this instance indicates a conservation of functional mechanism rather than strict relatedness.

The DNA sequence or ORF-B predicts a polypeptide with a molecular weight of 9.470 kDa which shows N-terminal homology to a small polypeptide of 5.1 kDa from a streptococcal plasmid pLS1. This polypeptide is termed RepA and is reported to have an  $\alpha$ helix-turn- $\alpha$ helix



structure characteristic of many DNA binding proteins (del Solar *et al.*, 1989). DNase I footprinting has shown that the RepA target is located in the region of the promoter for the repA and repB genes in pLS1 and is shown by *trans*-complementation analysis, *in vivo*, to function as a repressor by regulating plasmid copy number. It is proposed that the function of RepA is to regulate replication by repression of the synthesis of the initiator protein RepB. The polypeptide encoded by ORF-B is also predicted to form an  $\alpha$ helix-turn- $\alpha$ helix structure although this is towards the carboxy-terminus rather than in the region of sequence homology. Since plasmid replication is potentially autocatalytic, all plasmids have a negative repressor of replication. It is proposed that in the case of pHD2, replication is controlled at the level of production of the replication initiator protein ORF-A. However, DNA footprinting would be required to confirm if this was in fact the case. The control of replication of plasmids such as pLS1 and pHD2 would therefore appear to be entirely different than for the pT181 group of staphylococcal plasmids where control is exerted by the production of counter transcript RNAs which bind to the mRNA leader of the Rep protein. Such differences may suggest that plasmids from Gram-positive bacteria and which have to date been suggested to constitute one family may in fact constitute a number of families which despite replicating via a rolling circle mechanism, may utilize quite different mechanism replication control or indeed partitioning functions.

The ssDNA plasmids have often been shown to exhibit both structural and segregational instability under given circumstances. The segregational instability of most of the *S. aureus* plasmids in *B.*

*subtilis* is thought to result from a lack of interaction between host factors and the minus origins of these plasmids. The minus origin of the *S. aureus* plasmid, pUB110, is the exception to this and is functional in *B. subtilis*. Lack of minus origin function is additionally characterized by the accumulation of the single-stranded DNA intermediate since conversion to the double-stranded form is inefficient. In the case of pHD2, a construct containing pBR322, pHD2 and the chloramphenicol acetyl transferase gene from pC194 termed pCC200 was shown to be segregationally stable in *B. thuringiensis* subspecies *israelensis* IPS78/11 showing no loss when cultured for approximately 100 generation in the absence of selective pressure. It was additionally not possible to detect either the plasmid or the single-stranded intermediate from this host although functional replicons could be detected by transformation back into *E. coli*. This may suggest that pCC200 is fully functional in this host where it is maintained at low copy number with its segregational stability therefore requiring a functional minus origin. The minus origins of a number of ssDNA plasmids are characterized by the presence of long imperfect inverted repeat sequences. A similar structure is located in pHD2 and is proposed to be the minus origin in this instance. The function of the minus origin of a number of ssDNA plasmids has been shown to be strand-specific with complementation not being possible *in trans*. The confirmation of this sequence as the minus origin in pHD2 could be demonstrated by the generation of segregational derivatives following inactivation of this sequence by deletion or oligonucleotide mutagenesis. Whilst a functional minus origin could account for segregational stability of constructs based on pHD2, the presence of a partitioning

function involving membrane binding similar to that observed with the *E. coli* plasmids P1 and pSC101 involving cannot be ruled out without further investigation.

The presence of homologous sequences found between the plasmids of a number of *Bacillus thuringiensis* isolates has suggested that these may be divided into two categories consisting of low molecular weight and high molecular weight plasmids. Most of the work on plasmids from Gram-positive bacteria has concentrated on those of low molecular weight and which have been shown to be members of the ssDNA family of plasmids. As discussed, the mode of replication of these plasmids generates a single-stranded form which is a reactive intermediate in DNA recombination in which the presence of both direct and inverted repeat sequences is involved. The fidelity of plasmid arrays within *B. thuringiensis* subspecies implies a low frequency of recombinational events as selection of the original array is unlikely since the plasmids have been shown to be unessential for normal cell function by plasmid curing experiments. The presence of inverted repeat sequences on plasmids within the *B. thuringiensis* plasmid profile might be expected to yield recombinational forms of the plasmids at a greater frequency than that which is generally observed if replication involved a rolling circle mechanism. It is possible that replication of the high molecular weight does not occur by the same rolling circle method as is proposed here for those of low molecular weight and may account for the lack of cross-homology observed between the two classes of plasmids. Additionally, it appears that the low molecular weight plasmids or at least pHD2 is only distantly related to the ssDNA family of plasmids as exemplified by the

*S. aureus* plasmids. Further analysis of plasmids from *B. thuringiensis* and other genera would be required in order to determine whether this is in fact the case and whether the family of ssDNA plasmids has other more distantly related members than have so far been observed.

The work described within this thesis has helped to extend what is known of the replicative functions found in Gram-positive bacteria. The initial evidence presented from the *S. aureus* group of plasmids suggested that many of these plasmids were members of a highly related family of plasmids replicating via a rolling circle mechanism. Evidence presented here, regarding the *Bacillus thuringiensis* plasmid pND2, suggests a more distantly related set of members than that previously suggested by the *S. aureus* plasmids. It will be interesting to see if further work on plasmids, particularly from other genera, confirms the family of ssDNA plasmids to have a wide variety of members as is suggested here or whether the relationship of plasmids such as pND2 to this family is an exception.

Appendix 1

Amino acid sequences obtained for comparison with ORF-1 and ORF-2 of pHD2

REFM\$STAAU repM protein. - *Staphylococcus aureus*.

Species: *Staphylococcus aureus*.

Accession: P14490

Ehret, M., and Matzura, H. (1988).

Function: This protein is probably a specific topoisomerase involved in initiating replication. This protein is specifically required and may be rate-limiting for replication of the plasmid in vivo.

EMBL; X07371; SAPC223.

Keywords: DNA replication; plasmid.

Ala	A	15	Cys	C	2	Asp	D	28	Glu	E	23
Phe	F	17	Gly	G	7	His	H	10	Ile	I	15
Lys	K	32	Leu	L	31	Met	M	9	Asn	N	25
Pro	P	4	Gln	Q	9	Arg	R	17	Ser	S	22
Thr	T	16	Val	V	11	Trp	W	9	Tyr	Y	12

Mol. wt. (calc) = 37508      Residues = 314

```

1      M S K N N Y T N H S N H L E N H D S D N F S K T G Y S N S R
31     L D A N F V C T S N P K L S F D A M T I V G N L N K N S A K
61     K L S D F M S L D P Q I R L W D I L Q T K F K A K A L Q E K
91     V Y I E Y D K V K A D T W D R R N M R V E F N P N K L T H D
121    E N L W L K H N I I D Y M E D D G F T R L D L A F D F E D D
151    L S D Y Y A L S E K A L K R T V F F G T T G K A E T K Y F G
181    S R D S N R F I R I Y N K K K E R K E N A D V D V S A E H L
211    W R V E I E L K R D M V D Y W N N C F N D L H I L K P A W A
241    T L E S L K E Q A M V Y L L L H E E S K W G E L H R N S R R
271    K Y K Q I I Q E I S S I D L T D L M K S T L T D N E E N L O
301    K Q I N F W Q R K F E F W K

```

RQSAD2 repD protein - *Staphylococcus aureus* plasmid pC221

Species: *Staphylococcus aureus*

Accession: A03601

Brenner, D.G., and Shaw, W.V. (1985).

Sequence translated from the DNA sequence. This sequence is homologous with that of repC protein, which is essential for replication of the staphylococcal plasmid pT181.

Gene name: repD.

Superfamily: repC protein.

Keywords: gene replication.

Ala	A	11	Cys	C	1	Asp	D	29	Glu	E	26
Phe	F	17	Gly	G	9	His	H	5	Ile	I	15
Lys	K	35	Leu	L	28	Met	M	11	Asn	N	19
Pro	P	8	Gln	Q	13	Arg	R	16	Ser	S	15
Thr	T	18	Val	V	12	Trp	W	9	Tyr	Y	14

Mol. wt. (calc) = 37437      Residues = 311

```

1  M S T E N H S N Y L Q N K D L D N F S K T G Y S N S R L S G
31 N F F T T P Q F E L S F D A M T I V G N L N K T N A K K L S
61 D F M S T E P Q I R L W D I L Q T K F K A K A L Q E K V Y I
91 E Y D K V K A D S W D R R N M R V E F N P N K L T H E E M L
121 W L K Q N I I D Y H E D D G F T R L D L A F D F E D D L S D
151 Y Y A M T D K A V K K T I F Y G R N G K F E T K Y F G V R D
181 S D R F I R I V N K K Q E R K D N A D V E V M S E H L W R V
211 E I E L K R D M V D Y W N D C F D D L H I L K P D W T T P E
241 K V K E O A W V Y L L L N E E G T W G K L E R N A K Y K Y K
271 Q L I K E I S P I D L T E L M K S T L K E N E K Q L O K Q I
301 D F W Q R E F R F W K

```

RQSACT repC protein - *Staphylococcus aureus* plasmid pT181

Species: *Staphylococcus aureus*

Accession: A03600

Novick, R.P., Adler, G.K., Majumder, S., Khan, S.A.,

Carleton, S., Rosenblum, W.D., and Iordanescu, S. (1982)

Sequence translated from the DNA sequence. This protein is specifically required and may be rate-limiting for replication of the plasmid in vivo.

Superfamily: repC protein

Ala A 16	Cys C 2	Asp D 31	Glu E 21
Phe F 16	Gly G 7	His H 9	Ile I 19
Lys K 32	Leu L 26	Met M 11	Asn N 22
Pro P 6	Gln Q 12	Arg R 19	Ser S 19
Thr T 13	Val V 12	Tyr Y 12	

Mol. wt. (calc) = 37614      Residues = 314

```

1   M Y K N N H A N H S N H L E N H D L D N F S K T G Y S N S R
31  L D A H T V C I S D P K L S F D A M T I V G N L N R D N A Q
61  A L S K F M S V E P Q I R L W D I L Q T K F K A K A L Q E K
91  V Y I E Y D K V K A D S W D R R N M R I E F N P N K L T R D
121 E H I W L K Q N I I S Y M E D D G F T R L D L A F D F E E D
151 L S D Y Y A M S D K A V K K T I F Y G R N G K P E T K Y F G
181 V R D S N R F I R I Y N K K Q E R K D N A D A E V M S E H L
211 W R V E I E L K R D H V D Y W N D C F S D L H I L Q P D W K
241 T I Q R T A D R A I V F M L L S D E E W G K L H R N S R T
271 K Y K N L I K E I S F V D L T D L M K S T L K A N E K Q L Q
301 K Q I D F W Q H E F K F W K
  
```

REPASSTRPN repA protein (gene name repA). - *Streptococcus pneumoniae*, and *Streptococcus agalactiae*.

Species: *Streptococcus pneumoniae*, and *Streptococcus agalactiae*.

Accession: P13920

Lacks, S.A., Lopez, P., Greenberg, B., and Espinosa, M.J. (1986).  
van der Lelie, D., Bron, S., Venema, G., and Oskam, L. (1989).

EMBL; X15669; SAREPAB.

PIR; A25599; A25599.

Keywords: PLASMID; DNA REPLICATION.

Ala	A	3	Cys	C	0	Asp	D	0	Glu	E	6
Phe	F	0	Gly	G	2	His	H	0	Ile	I	2
Lys	K	7	Leu	L	6	Met	M	4	Asn	N	2
Pro	P	0	Gln	Q	1	Arg	R	2	Ser	S	5
Thr	T	2	Val	V	2	Trp	W	0	Tyr	Y	1

Mol. wt. (calc) = 5109

Residues = 45

1    M K K R L T I T L S E S V L E N L E K M A R E M G L S K S A  
31    M I S V A L E N Y K K G Q E K



REPESSTAAU REPE PROTEIN. - *Staphylococcus aureus*.

Species: *Staphylococcus aureus*.

Accession: P12053

Projan, S.J., Moghazeh, S., and Novick, R.P. (1988).

Function: This protein is probably a specific topoisomerase involved in initiating replication. This protein is specifically required and may be rate-limiting for replication of the plasmid in vivo. PS194 is a streptomycin-resistance plasmid.

EMBL: X06627; SAPS194.

Keywords: DNA replication; plasmid.

Ala A 13	Cys C 1	Asp D 26	Glu E 30
Phe F 15	Gly G 7	His H 7	Ile I 20
Lys K 36	Leu L 27	Met M 14	Asn N 19
Pro P 6	Gln Q 12	Arg R 19	Ser S 17
Thr T 13	Val V 10	Trp W 9	Tyr Y 13

Mol. wt. (calc) = 38043

Residues = 314

```

1  MSKKAEEIQAQKQSLKENS NFSKTYGSNSR
31 LNRRIIMYTPPEPKLHFDAMTIVGNLKNNAH
61 KLSEFMSIA PQIRLWDILQTKFKKAKALQEK
91 VYIEYDKVKADAWDRRNM RVEFNPNKLTHE
121 EMLWLKQNI IDYMEDDGGPTRRLDLAFDFEDD
151 LSDYYAMTDKSVKKTIIFYGRNGKPETKVFG
181 VRDSDRFRIRIYMKKQERKDNADIEVMSSEHL
211 WRVEIELKRDMVDYWNDCFNLDLHILKPDWS
241 SLEKVKDOAMIYMLIHEESTWQKLERRTKN
271 KYREMLKSISEIDLTL DLMKLT LKENEKQLQ
301 KOIEFWQREFRFE

```

### Bibliography

- Albertini, A.M., Hofer, M., Calos, M.P. and Miller, J.H. (1982). On the formation of spontaneous deletions: the importance of short sequence homologies in the generation of large deletions. *Cell* 29:319-328.
- Alonso, J.C. and Taylor, R.W. (1987). Initiation of plasmid pC194 replication and its control in *Bacillus subtilis*. *Mol. Gen. Genet.* 210:476-484.
- Anderson, S. (1981). Shotgun DNA sequencing using cloned DNase I - generated fragments. *Nucleic Acids Res.* 9:3015-3027.
- Angsuthanasombat, C., Chungjatupornchai, W., Kertbundit, S., Luxananil, P., Settasatian, C., Wilairat, P., and Panyim, S. (1987). Cloning and expression of 130-kd mosquito-larvicidal  $\delta$ -endotoxin gene of *Bacillus thuringiensis* var. *israelensis* in *Escherichia coli*. *Mol. Gen. Genet.* 208:384-389.
- Angus, T.A. (1954). A bacterial toxin paralysing silkworm larvae. *Nature* 173:545.
- Angus, T.A. (1965). Symposium on Microbial Insecticides, 1. Bacterial pathogens of insects as microbial insecticides. *Bacteriol. Rev.* 29:364-372.

Aronson, A.I., Beckman, W., and Dunn, P. (1986). *Bacillus thuringiensis* and related insect pathogens. Microbiol. Rev. 50:1-24.

Ataai, M.M., and Shuler, M.L., (1987). A mathematical model for prediction of plasmid copy number and genetic stability in *Escherichia coli*. Biotechnol. Bioeng. 30:389-397.

Austin, S. and Abeles, A. (1983a). Partition of unit-copy miniplasmids to daughter cells. I. P1 and F mini-plasmids contain discrete, interchangeable sequences sufficient to promote equipartition. J. Mol. Biol. 169:353-372.

Austin, S., and Abeles, A. (1983b). Partition of unit-copy miniplasmids to daughter cells. II. The partition region of miniplasmid P1 encodes an essential protein and a centromere-like site at which it acts. J. Mol. Biol. 169:373-387.

Austin, S., Ziese, M. and Sternberg, N. (1981). A novel role for site-specific recombination in maintenance of bacterial replicons. Cell 25:729-736.

Austin, S.J. (1988). Plasmid Partition. Plasmid 20:1-9.

Beas, P. and Jansz, H. (1988) Single-stranded DNA phage origins. Curr. Top. Microbiol. Immunol. 136:31-70.

Ballaster, S., Lopez, P., Alonso, J.C., Espinosa, M. and Lacks, S.A. (1986). Selective advantage of deletions enhancing chloramphenicol acetyltransferase gene expression in *Streptococcus pneumoniae* plasmids. Gene 41:153-163.

Bankier, A.T., Weston, K.M. and Barrell, B.G. (1986). In "MRC Sequencing Manual," ppl-37. Cambridge MRC.

Battisti, L., Green, B.D., Thorne, C.B. (1985). Mating system for transfer of plasmids among *Bacillus anthracis*, *Bacillus cereus*, and *Bacillus thuringiensis*. J. Bacteriol. 162:543-550.

Bazara, M. and Helinski, D.R. (1970). Replication of a bacterial plasmid and an episome in *Escherichia coli*. Biochemistry 9:399-406.

Bachtel, D.B., and Bulla, L.A. Jr. (1976). Electron microscope study of sporulation and parasporal formation in *Bacillus thuringiensis*. J. Bacteriol. 127:1472-1481.

Beck, E. and Bremer, E. (1980). Nucleotide sequence of the gene *ompA* coding the outer membrane II of *Escherichia coli*. Nucleic Acids Res. 8:3011-3024.

Bennetzen, J.L. and Hall, B.D. (1982). Codon selection in yeast. J. Biol. Chem. 257:3026-3031.

Beringer, J.G. (1974). R factor transfer in *Rhizobium leguminosarum*.

J. Gen. Microbiol. 84:188-198.

Birnboim, H.C. and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7:1513-1524.

Blake, R.D. and Hinds, P.W. (1984). Analysis of codon bias in *E. coli* sequences. J. Biomol. Struct. Dyn. 3:593-606.

Boe, L., Gros, M.-F., de Riele, H., Ehrlich, S.D. Gruss, A. (1989). Replication origins of single-stranded DNA plasmid pUB110. J. Bacteriol. 171:3366-3372.

Bolivar, F., Rodriguez, R., Greene, P.J., Betlach, M.C., Heyneker, H.L., Boyer, H.W., Crossa, J.H. and Falkow, S. (1977). Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene 2:95-113.

Bone, E.J. and Ellar, D.J. (1989). Transformation of *Bacillus thuringiensis* by electroporation. FEMS Microbiol. Lett. 58:171-178.

Bott, K.F. and Wilson, G.A. (1967). Development of competence in the *Bacillus subtilis* transformation system. J. Bacteriol. 94:562-570.

Brenner, D.G. and Shaw, W.V. (1985). The use of synthetic oligonucleotide primers with universal templates for rapid DNA sequencing: results with the staphylococcal replicon pC221. EMBO J. 4:561-568.

Brunier, D., Michel, B. and Ehrlich, S. (1988). Copy choice illegitimate recombination. Cell 52:883-892.

Bulla, L.A. Jr. Rhodes, R.A., and Julian, G.S.(1975). Bacteria as insect pathogens. Annu. Rev. Microbiol. 29:163-190.

Bulla, L.A., Bechtel, D.B., Kramer, K.J., Shethna, Y.I., Aronson, A.I., Fitz-James, P.C. (1980). Ultrastructure, physiology, and biochemistry of *Bacillus thuringiensis*. Crit. Rev. Microbiol. 8:147-202.

Bulmer, M. (1987). Coevolution of codon usage and transfer RNA abundance. Nature 325:728-730.

Carlton, S., Projan, S.J., Highlander, S.K., Mogazeh, S.M and Novick, R.P. (1984). Control of pT181 replication II. Mutational analysis. EMBO J. 3:2407-2414

Chang, S., Chang, S.Y., Gray, O. (1987). Structural and genetic analyses of the *par* locus that regulates plasmid partition in *Bacillus subtilis*. J. Bacteriol. 169:3952-3962.

Chardon-Loraiux, I., Charpentier, M., and Percheron, F. (1986). Isolation and characterisation of a linear plasmid from *Streptomyces rimosus*. FEMS Microbiol. Lett. 35:151-155.

Charnay, P., Mandert, E., Hampe, A., Fitoussi, F., Tiollais, P. and Galibert, F. (1979). Localization on the viral genome and nucleic acid sequence of the gene coding for two polypeptides of the hepatitis B surface antigen (H Bs Ag). Nucleic Acids Res. 7:335-346.

Clark, B.F.C., and Marcker, K.A. (1966). The role of N-formyl-methionyl mRNA in protein biosynthesis. J. Mol. Biol. 17:394-406.

Clewell, D.B. and Melnicki, D.R. (1969). Supercoiled circular DNA-protein complex in *E. coli*: purification and induced conversion to an open circular DNA form. Proc. Natl. Acad. Sci. USA 62:1159-1166.

Cohen, S.N. (1976). Transposable genetic elements and plasmid evolution. Nature 263:731-738.

Couturier, M., Bex, F., Bergquist, P.L. and Maas, W.K. (1988). Identification and classification of bacterial plasmids. Microbiol. Rev. 52:375-395.

Datta, N. and Hedges, R.W. (1972). Host range of R factors. J. Gen. Microbiol. 70:453-460.

Dayhoff, M.O. ed. (1972). "Atlas of protein sequence and structure," Vol. 5 Natl. Biomed. Res. Found., Washington, D.C.

Debakov, V.G. (1982). In "The molecular biology of the Bacilli" - The industrial use of Bacilli. pp331-371. Academic Press.

Debakov, V.G., Khlebalina, O.I., Galushka, F.P. and Sladkova, I.A. (1980). Plasmids of *Bacillus thuringiensis* var. *galleriae*, strain 612. J. Mol. Biol. 14:818-823.

Debro, L., Fitz-James, P.C. and Aronson, A. (1986). Two different parasporal inclusions are produced by *Bacillus thuringiensis* subsp. *finitimus*. J. Bacteriol. 165:258-268.

Deininger, P.L. (1983). Random subcloning of sonicated DNA: Application to shotgun DNA sequence analysis. Anal. Biochem. 129:216-223.

del Solar, G., Puyet, A. and Espinosa, M. (1987). Initiation signals for the conversion of single-stranded to double-stranded DNA forms in the streptococcal plasmid pLS1. Nucleic Acids Res. 15:5561-5580.

del Solar, G.W., de la Campa, A.G., Perez-Martin, J., Choli, T. and Espinosa, M. (1989). Purification and characterization of RepA, a protein involved in the copy number control of plasmid pLS1. Nucleic Acids Res. 17:2405-2420.



Devereaux, J., Haeberli, P. and Smithies, O. (1984). A comprehensive set of sequence analysis programmes for the VAX. *Nucleic Acids Res.* 12:387-395.

Devine, K.M., Hogan, S.T., Higgins, D. and McConnell, D.J. (1989). Replication and segregational stability of *Bacillus* plasmid pBAAl. *J. Bacteriol.* 171:1166-1172.

Doi, R.H. and Wang, L.-F. (1986). Multiple procaryotic ribonucleic acid polymerase sigma factors. *Microbiol. Rev.* 50:227-243.

Doi, R.H., Wong, S., and Kawamura, F. (1986). Potential use of *Bacillus subtilis* secretion and production of foreign proteins. *Trends Biotechnol.* 9:232-235.

Eckhardt, T. (1978). A rapid method for the identification of plasmid deoxyribonucleic acid in bacteria. *Plasmid* 1:584-588.

Ehret, M. and Matzura, H. (1988). Replication control of the *Staphylococcus aureus* chloramphenicol resistance plasmids pC223 and pUB112 in *Bacillus subtilis*. *Nucleic Acids Res.* 16:2045-2062.

Ehrlich, S.D. (1977). Replication and expression of plasmids from *Staphylococcus aureus* in *Bacillus subtilis*. *Proc. Natl. Acad. Sci. USA.* 74:1680-1682.

Garnier, T. and Cole, S.T. (1988). Complete nucleotide sequence and genetic organization of the bacteriogenic plasmid, pIP404, from *Clostridium parfringens*. Plasmid 19:134-150.

Gennaro, M.L. and Novick, R.P. (1986). *cmp*, a cis-acting plasmid locus that increases interaction between replication origin and inhibitor protein. J. Bacteriol. 168:160-166.

Gennaro, M. and Novick, R. (1988). An enhancer of DNA replication. J. Bacteriol. 170:5709-5717.

Gerdes, K., Bech, F.W., Jorgensen, S.T. and Lobner-Olesen, A. (1986). Mechanism of postsegregational killing by the *Hok* gene product of the *parB* system of plasmid R1 and its homology with the *relF* gene product of the *E. coli relB* operon. EMBO J. 5:2023-2029.

Gold, L. and Stormo, G. (1987). Translational initiation. In 'Escherichia coli and Salmonella typhimurium. Cellular and Molecular Biology' (F.C. Neidhardt, J.L. Ingraham, K.B. Low, B. Magasanik, M. Schaechter and H.E. Umberger, eds), pp 1302-1307. American Society for Microbiology, Washington, D.C.

Goldberg, L.J. and Margalit, J. (1977). A bacterial spore demonstrating rapid larvicidal activity against *Anopheles sergentii*, *Uranotaenia unguiculata*, *Culex univittatus*, *Aedes aegypti* and *Culex pipiens*. Mosquito News 37:355-358.

Gonzalez, J.M. Jr., Brown, B.J., and Carlton, B.C. (1982). Transfer of *Bacillus thuringiensis* plasmids for  $\delta$ -endotoxin among strains of *B. thuringiensis* and *B. cereus*. Proc. Natl. Acad. Sci. USA 79:6951-6955.

Gonzalez, J.M. and Carlton, B.C. (1980). Patterns of plasmid DNA in crystalliferous and acrySTALLIFEROUS strains of *Bacillus thuringiensis*. Plasmid 3:92-98.

Gonzalez, J.M. and Carlton, B.C. (1984). A large transmissible plasmid is required for crystal toxin production in *Bacillus thuringiensis* variety *israelensis*. Plasmid 11:28-38.

Gonzalez, J.M., Dulmage, H.T., and Carlton, B.C. (1981). Correlation between specific plasmids and  $\delta$ -endotoxin production in *Bacillus thuringiensis*. Plasmid 5:351-356.

Goxe, A. and Ehrlich, S.D. (1980). Replication of plasmid from *Staphylococcus aureus* in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 77:7333-7337.

Grantham, R., Gautier, C., Gouy, M., Mercier, R. and Pavé, A. (1980). Codon catalogue usage and the genome hypothesis. Nucleic Acids Res. 8:r40-r60

Gros, M.F., teRiele, H., and Ehrlich, S.D. (1987). Rolling circle replication of single-stranded DNA plasmid pC194. EMBO J. 6:3863-3869.

Grosschedl, R. and Hobom, G. (1979) DNA sequences and structural homologies of the replication origins of lambdoid bacteriophages. *Nature* 277:621-627.

Gruss, A., Ross, H. and Novick, R. (1987). Functional analysis of a palindromic sequence required for normal replication of several staphylococcal plasmids. *Proc. Natl. Acad. Sci. USA* 84:2165-2169.

Gryczan, T.J., Contente, S., and Dubnau, D. (1978). Characterization of *Staphylococcus aureus* plasmids introduced by transformation into *Bacillus subtilis*. *J. Bacteriol.* 134:318-329.

Gustafsson, P., Nordstrom, K. and Perram, J.W. (1978). Selection and timing of replication of plasmids R1drd-119 and F'lac in *Escherichia coli*. *Plasmid* 1:187-203.

Gustafsson, P., Wolf-Watz, H., Lind, L., Johansson, K.-E. and Nordstrom, K. (1983). Binding between the *par* region of plasmid R1 and pSC101 and the outer membrane fraction of the host bacteria. *EMBO J.* 2:27-32.

Hahn, J., and Dubnau, D. (1985). Analysis of plasmid deletional instability in *Bacillus subtilis*. *J. Bacteriol.* 162:1014-1023.

Haider, M.Z., Knowles, B.H. and Ellar, D.J. (1986). Specificity of *Bacillus thuringiensis* var. *colmeri* insecticidal  $\delta$ -endotoxin is

determined by differential proteolytic processing of the protoxin by larval gut proteases. Eur. J. Biochem. 156:531-540.

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. 166:557-580.

Hannay, C.L. (1953). Crystalline inclusion in aerobic spore-forming bacteria. Nature 172:1004.

Hannay, C.L. and Fitz-James, P., (1955). The protein crystals of *Bacillus thuringiensis* Berliner. Can. J. Microbiol 1:694-710.

Hansen, F.G. and von Meyerburg, K. (1979). Characterization of the *dnaA*, *gyrB* and other genes in the *dnaA* region of the *Escherichia coli* chromosome on specialized transducing phages lambda-tna. Mol. Gen. Genet. 175:135-144.

Hardy, K. (1981). Bacterial Plasmids. Aspects of Microbiology 4. Thomas Nelson and Sons Ltd.

Harris, J.D., Heilig, J.S., Martinez, I.I., Calendar, R. and Isaksson, L.A. (1978). Temperature-sensitive *Escherichia coli* mutant producing a temperature sensitive  $\sigma$  subunit of DNA-dependent RNA polymerase. Proc. Natl. Acad. Sci. USA 75:6177-6181.

Harvey, W.R., Cioffi, M., Dow, A.T., and Wolfersberger M.G. (1983). Potassium ion transport ATPase in insect epithelium. J. Exp. Biol. 106:91-117.

Heimpel, A.M. and Angus, T.A. (1958). The taxonomy of insect pathogens related to *Bacillus cereus* Frankland and Frankland. Can. J. Micro. 4:531-541.

Heimpel, A.M. and Angus, T.A. (1960). Bacterial insecticides. Bacteriol. Rev. 24:266-288.

Henikoff, S. (1984). Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene 28:351-359.

Higgins, D.G. and Sharp, P.M. (1988) CLUSTAL: a package for performing multiple sequence alignments on a microcomputer. Gene 73:237-244.

Higgins, D.G. and Sharp, P.M. (1989) Fast and sensitive multiple sequence alignments on a microcomputer. CABIOS 5:151-153.

Highlander, S.K. and Novick, R.P. (1990). Mutational and physiological analyses of plasmid pT181 functions expressing incompatibility. Plasmid 23:1-15.

Highlander, S.K. and Novick, R.P. (1987). Plasmid repopulation kinetics in *Staphylococcus aureus*. Plasmid 17:210-221.

Hilton, M.D., and Whiteley, H.R. (1985). UV cross-linking of the *Bacillus subtilis* RNA polymerase to DNA in promoter and nonpromoter complexes. J. Biol. Chem. 260:8121-8127.

Hofmann, C., Luthy, P., Hutter, R. and Pliska, V. (1988b) Binding of the delta endotoxin from *Bacillus thuringiensis* to brush-border membrane vesicles of the cabbage butterfly (*Pieris brassicae*). Eur. J. Biochem. 173:85-91.

Hofmann, C., Vanderbruggen, H., Hofte, H., van Rie, J., Jansens, S. and van Mellaert, H. (1988a). Specificity of *Bacillus thuringiensis*  $\delta$ -endotoxins is correlated with the presence of high-affinity binding sites in the brush border membrane of target insect midguts. Proc. Natl. Acad. Sci. USA 85:7844-7848.

Hofte, H., de Greve, H., Seurinck, J., Jansens, S., Mahillon, J., Ampe, C., Vandekerckhove, J., Vanderbruggen, H., van Montagu, M., Zabeau, M. and Vaeck, M. (1986). Structural and functional analysis of a cloned delta endotoxin of *Bacillus thuringiensis* berliner 1715. Eur. J. Biochem. 161:273-280.

Hofte, H., Whiteley, H.R. (1989). Insecticidal crystal proteins of *Bacillus thuringiensis*. Microbiol. Rev. 53:242-255.

Hopwood, D.A., Bibb, M.J., Chater, K.F., Janssen, G.R., Malpartida, F. and Smith, C.P. (1986). Regulation of gene expression in antibiotic-

producing *Streptomyces*. In "Regulation of gene expression - 25 years on" (I. Booth and C. Higgins, eds) pp 251-176. Cambridge University Press.

Morinouchi, S., and Weissblum, B. (1982b). Nucleotide sequence and functional map of pC194, a plasmid that specifies inducible chloramphenicol resistance. *J. Bacteriol.* 150:815-825.

Morinouchi, S., and Weissblum, B. (1982a). Nucleotide sequence and functional map of pE194, a plasmid that specifies inducible resistance to macrolide, lincosamide, and streptogramin type B antibiotics. *J. Bacteriol.* 150:804-814.

Hyde, E.F., Hilton, M.H., and Whiteley, H.R. (1986). Interactions of *Bacillus subtilis* RNA polymerase with subunits determining the specificity of initiation. *J. Biol. Chem.* 261:16565-16570.

Ikemura, T. (1981a). Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *J. Mol. Biol.* 146:1-21.

Ikemura, T. (1981b). Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *J. Mol. Biol.* 151:389-409.



Iordanescu, S. (1976). Three distinct plasmids originating in the same *Staphylococcus aureus* strain. Arch. Roum. Path. Exp. Microbiol. 35:111-118.

Iordanescu, S., (1986). Effect of the deletion of a fragment dispensable for the autonomous maintenance of plasmid pT181 on the competition between incompatible plasmids. Plasmid 15:191-198.

Iordanescu, S., Surdeanu, M., Della-Latta, P., and Novick, R. (1978). Incompatibility and molecular relationships between small staphylococcal plasmids carrying the same resistance marker. Plasmid 1:468-479.

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

Janniere, L. and Ehrlich, S. (1987). Recombination between short repeat sequences is more frequent in plasmids than in the chromosome of *Bacillus subtilis*. Mol. Gen. Genet. 210:116-121.

Josson, K., Scheirlinck, T., Michels, F., Platteeuw, C., Stanssens, P., Joom, H., Dhaese, P., Zabeau, M., and Mahillon, J. (1989). Characterisation of a Gram-positive broad-host range plasmid isolated from *Lactobacillus hilgardii*. Plasmid 21:9-20.

Khan, S., Murray, R. and Koepsel, R. (1988) Mechanism of plasmid pT181 replication. *Biochim. Biophys. Acta* 951:375-381.

Khan, S.A., Carlton, S.M., and Novick, R. (1981). Replication of plasmid pT181 *in vitro*: requirement for a plasmid-encoded product. *Proc. Natl. Acad. Sci. USA* 78:4902-4906.

Khan, S.A., Novick, R.P. (1983). Complete nucleotide sequence of pT181, a tetracycline-resistance plasmid from *Staphylococcus aureus*. *Plasmid* 10:251-259.

Klier, A., Bourguin, C. and Rapoport, G. (1983). Mating between *Bacillus subtilis* and *Bacillus thuringiensis* and transfer of cloned crystal genes. *Mol. Gen. Genet.* 191:257-262.

Kline, B.C. and Miller, J.R. (1975). Detection of nonintegrated plasmid deoxyribonucleic acid in the folded chromosome of *Escherichia coli*: Physicochemical approach to studying the unit of segregation. *J. Bacteriol.* 121:165-172.

Knowles, B.H. and Ellar, D.J. (1987). Colloid-osmotic lysis is a general feature of the mechanism of action of *Bacillus thuringiensis*  $\delta$ -endotoxins with different insect specificities. *Biochim. Biophys. Acta* 924:509-518.

Koepsel, R.R., Murray, R.W., Rosenblum, W.D. and Khan, S.A. (1985). The replication initiator protein of plasmid pT181 has sequence specific

endonuclease and topoisomerase-like activities. Proc. Natl. Acad. Sci. USA 82:6845-6849.

Kozak, M. (1983). Comparison of initiation of protein synthesis in prokaryotes, eukaryotes, and organelles. Microbiol. Rev. 47:1-45.

Kronstad, J.W., Schnepf, H.E., and Whiteley, H.R. (1983). Diversity of locations for *Bacillus thuringiensis* crystal protein genes. J. Bacteriol. 154:419-428.

Kronstad, J.W., Whiteley, H.R. (1984). Inverted repeat sequences flank a *Bacillus thuringiensis* crystal protein gene. J. Bacteriol. 160:95-102.

Krywienczyk, J., Dulmage, H.T., and Fast, P.G. (1978). Occurrence of two serologically distinct groups within *Bacillus thuringiensis* serotype 3ab var. *kurstaki*. J. Invert. Pathol. 31:372-375.

Kyte, J. and Doolittle, R.F. (1982). A simple method for displaying the hydropathic character of a protein. J. Mol. Biol. 157:105-132.

Lacks, S.A., Lopez, P., Greenberg, B. and Espinosa, M. (1986). Identification and analysis of genes for tetracycline resistance and replication functions in the broad-host-range plasmid pLS1. J. Mol. Biol. 192:753-765.

Lederberg, J. (1952). Cell genetics and heredity symbiosis. *Physiol. Rev.* 32:403-430.

Lereclus, D., Klier, A., Menou, G. and Lecadet, M.M. (1984). A transposon-like structure related to the  $\delta$ -endotoxin gene of *Bacillus thuringiensis*. *EMBO J.* 3:2561-2567.

Lereclus, D., Lecadet, M.M., Ribier, J., and Dedonder, R. (1982). Molecular relationships among plasmids of *Bacillus thuringiensis*: conserved sequences through 11 crystalliferous strains. *Mol. Gen. Genet.* 186:391-398.

Lereclus, D., Mahillon, J., Menou, G., and Lecadet, M.M. (1986). Identification of Tn4430, a transposon of *Bacillus thuringiensis* functional in *Escherichia coli*. *Mol. Gen. Genet.* 204:52-57.

Lereclus, D., Menou, G. and Lecadet, M.M. (1983). Isolation of a DNA sequence related to several plasmids from *Bacillus thuringiensis* after mating involving the *Streptococcus faecalis* plasmid pAM81. *Mol. Gen. Genet.* 191:307-313.

Light, J., and Molin, S. (1982). The sites of action of the two copy number control functions of plasmid R1. *Mol. Gen. Genet.* 187:486-493.

Losick, R. and Pero, J. (1981). Cascades of sigma factors. *Cell* 25:582-584.

Losick, R., Youngman, P. and Figgot, P.J. (1986). Genetics of endospore formation in *Bacillus subtilis*. Annu. Rev. Genet. 20:625-669.

Low K.B. (1972) *Escherichia coli* K12 F-Prime factors, old and new. Bacteriol. Rev. 36:587-607.

Luthy, P., Cordier, J.L., and Fischer, H.M. In "Microbial and viral pesticides" *Bacillus thuringiensis* as a bacterial insecticide: basic considerations and application. (Kurstak, E. ed.) Marcel Dekker Inc. 1982.

Maciag, I.E., Viret, J.-F. and Alonzo (1988). Replication and incompatibility properties of plasmid pUB110 in *Bacillus subtilis*. Mol. Gen. Genet. 212:232-240.

Mahillon, J. and Seurinck, J. (1988). Complete nucleotide sequence of pGI2, a *Bacillus thuringiensis* plasmid containing Tn4430. Nucleic Acids Res. 16:11827-11828.

Mahillon, J., Hespel, F., Pierssens, A.-M. and Delcour, J. (1988). Cloning and partial characterization of three small cryptic plasmids from *Bacillus thuringiensis*. Plasmid 19:169-173.

Mandel, M. and Higa, A. (1970). Calcium dependent bacteriophage DNA infection. J. Mol. Biol. 53:159-162

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982). Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.

Martin, K.A., Friedman, S.A. and Austin S.J. (1987). Partition site of the P1 plasmid. Proc. Natl. Acad. Sci. USA 84:8544-8547.

Martin, P.A.W., Lohr, J.R., and Dean, D.H. (1981). Transformation of *Bacillus thuringiensis* protoplasts by plasmid deoxyribonucleic acid. J. Bacteriol. 145:980-983.

McLaughlin, J.R., Murray, C.L., and Rabinowitz, J.C. (1981). Unique features of the ribosome binding site sequence of the Gram-positive *Staphylococcus aureus*  $\beta$ -lactamase gene. J. Biol. Chem. 256:11283-11291.

McGaughy, W.H. (1985). Insect resistance to the biological insecticide *Bacillus thuringiensis*. Science 229:193-195.

McLaughlin, J.R. Murray, C.L. and Rabinowitz, C. (1981). Unique features in the ribosome binding site sequence of the Gram-positive *Staphylococcus aureus*  $\beta$ -Lactamase gene. J. Biol. Chem. 256:11283-11291.

Meacock, P.A. and Cohen. S.N (1980). Partitioning of bacterial plasmids during cell division: a *cis* acting locus that accomplishes stable plasmid inheritance. Cell 20:529-542.

Meijer, M., Beck, E., Hansen, F.G., Bergmans H.E., Messer, W., von Meyenburg, K. and Schaller, H. (1979). Nucleotide sequence of the origin of replication of the *Escherichia coli* K12 chromosome. Proc. Natl. Acad. Sci. USA 76:580-584.

Meselson, M.S. and Radding, C.M. (1975). A general model for genetic recombination. Proc. Natl. Acad. Sci. USA 72:358-361.

Messing, J., Gronenborn, B., Muller-Hill, B. and Hofschneider, P.H. (1977). Filamentous coliphage M13 as a cloning vehicle: insertion of a *Hind*III fragment of the lac regulatory region in the M13 replicative form *in vitro*. Proc. Natl. Acad. Sci. USA 74:3642-3646.

Meyer, R. and Helinski, D.R. (1977). Unidirectional replication of the P-group plasmid RK2. Biochim. Biophys. Acta 478:109-113.

Meyer, R., Figurski, D. and Helinski, D.R. (1977). Physical and genetic studies with restriction endonucleases on the broad host range plasmid RK2 Mol. Gen. Genet. 152:129-135.

Michel, B. and Ehrlich, S. (1986a). Illegitimate recombination occurs between the replication origin of plasmid pC194 and a progressing replication fork. EMBO J. 5:3691-3696.

Michel, B. and Ehrlich, S. (1986b). Illegitimate recombination at the replication origin of bacteriophage M13. Proc. Natl. Acad. Sci. USA 83:3386-3390.

Minnich, S.A. and Aronson, A.I. (1984). Regulation of protoxin synthesis in *Bacillus thuringiensis*. J. Bacteriol. 158:447-454.

Miteva, V.I., and Grigorova, R.T. (1988). Construction of a bifunctional genetically labelled plasmid for *Bacillus thuringiensis* subsp. *israelensis*. Arch. Microbiol. 150:496-498.

Murai, M., Miyashita, H., Araki, T., Seki, and Oshima, Y. (1987). Molecular structure of replication origin of a *Bacillus amyloliquefaciens* plasmid pFTB14. Mol Gen. Genet. 210:92-100.

Murray, N.E., and Kelly, W. (1979). Characterisation of *lambda polA* transducing phages; effective expression of the *E. coli polA* gene. Mol. Gen. Genet. 175:77-87.

Needleman, S.B. and Wunsch, C.D. (1970). A general method applicable for the search for similarities in the amino acid sequence of two proteins. J. Mol. Biol. 48:443-453.

Nordstrom, K., Molin, S., and Light, J. (1984). Review. Control of replication of bacterial plasmids: genetics, molecular biology, and physiology of the plasmid R1 system. Plasmid 12:71-90.

Norris, J.R. (1964). The classification of *Bacillus thuringiensis*. J. Appl. Bacteriol. 27:439-447.



Novick, J.P., Adler, G.K., Majumder, S., Khan, S.A., Carleton, S., Rosenblum, W.D. and Iordanescu, S. (1982). Coding sequence for the pT181 *repC* product: A plasmid-coded protein uniquely required for replication. Proc. Natl. Acad. Sci. USA 79:4108-4112.

Novick, R., Gruss, A., Highlander, S., Gennaro, M., Projan, S. and Ross, H. (1986). Host-plasmid interactions affecting plasmid replication and maintenance in Gram-positive bacteria. Banbury Report 24:225-254.

Novick, R.P. (1987). Plasmid incompatibility. Microbiol. Rev. 51:381-395.

Novick, R.P., Adler, G.K., Projan S.J., Carlton, S., Highlander, S., Gruss, A., Khan, S.A. and Iordanescu, S. (1984). Control of pT181 replication. 1. The pT181 copy control function acts by inhibiting the synthesis of a replication protein. EMBO J 3:2399-2405.

Novick, R.P., Adler, G.K., Majumder, S., Khan, S.A., Carleton, S., Rosenblum, W.D., and Iordanescu, S. (1982). Coding sequence for the pT181 *repC* product: A plasmid-coded protein uniquely required for replication. Proc. Nat. Acad. Sci. USA 79: 4108-4112.

Novick, R.P., Clowes, R.C., Cohen, S.N., Curtiss, R., Datta, N. and Falkow, S. (1976). Uniform nomenclature for bacterial plasmids: a proposal. Bacteriol. Rev. 40:168-189.

Novick, R.P., Projan, S.J., Kumar, C.C., Carleton, S., Gruss, A., Highlander, S.K. and Kornblum, J. (1985). In "Plasmids in Bacteria"

Replication for pT181, an indirectly regulated plasmid, p299-320. D.R. Helinski, S.M. Cohen, D.B. Clewell, D.A. Jackson and A. Hollaender (eds.), Plenum Publishing Corp., New York.

Oertel, W., Kollek, R., Beck, E. and Goebel, W. (1979). The nucleotide sequence of a DNA fragment from the replication origin of the antibiotic resistance factor R1drd19. Mol. Gen. Genet. 171:277-285.

Oka, A.N., Nomura, N., Morita, M., Sugisaki, H., Sugimoto, K. and Takanami, M. (1979). Nucleotide sequence of small ColE1 derivatives: structure of the regions essential for autonomous replication and colicin E1 immunity. Mol. Gen. Genet. 172:151-159.

Osawa, T. and Yura, T. (1981). Effects of reduced amount of RNA polymerase sigma factor on gene expression and growth of *Escherichia coli*: studies of the *rpd40*(amber) mutation. Mol. Gen. Genet. 184:166-173.

Pabo, C.O. and Sauer, R.T. (1984). Annu. Rev. Biochem. 53:293-321. Protein-DNA recognition.

Peeters, B.P.H., de Boer, J.H., Bron, S. and Venema, G. (1988). Structural plasmid instability in *Bacillus subtilis*: effect of direct and inverted repeats. Mol. Gen. Genet. 212:450-458.

Pendleton, I.R., and Morrison, R.B. (1966). Analysis of the crystal antigens of *Bacillus thuringiensis* by gel diffusion. J. Appl. Bacteriol. 29:519-528.

Pigac, J., Vajacklija, D., Toman, Z., Gamulin, V. and Schrepf. (1988). Structural instability of a bifunctional plasmid pZG1 and single-stranded DNA formation in *Streptomyces*. Plasmid 19:222-230.

Poncz, M., Solowiejczyk, D., Ballantine, M., Schwartz, E., and Surrey, S. (1982). Nonrandom DNA sequence analysis in bacteriophage M13 by the dideoxy chain-termination method. Proc. Natl. Acad. Sci. USA 79:4298-4302.

Porter, D. (1984). Protein synthesis during differentiation of *Rhodospirillum rubrum* swarmer cells. PhD thesis, University of Warwick Coventry.

Pritchard, R.H., Barth, P.T. and Collins, J. (1969). Control of DNA synthesis in bacteria. Symp. Soc. Gen. Microbiol. 19:263-297.

Projan, S.J. and Novick, R.P. (1984). Reciprocal intrapool variation in plasmid copy numbers: a characteristic of segregational incompatibility. Plasmid 12:52-60.

Projan, S.J., and Novick, R. (1988). Comparative analysis of five related staphylococcal plasmids. Plasmid 19:203-221.

Projan, S.J., Carleton, S. and Novick, R.P. (1983) Determination of plasmid copy number by fluorescence densitometry. *Plasmid* 9:183-190

Projan, S.J., Moghazeh, S., and Novick, R. P. (1988). Nucleotide sequence of pSI94, a streptomycin-resistance plasmid from *Staphylococcus aureus*. *Nucleic Acids Res.* 16:2179-2187.

Puyet, A., del Solar, G.H. and Espinosa, M. (1988). Identification of the origin and direction of replication of the broad-host-range plasmid pLS1. *Nucleic Acids Res.* 16:115-133.

Queen, C. and Korn, L.J. (1984). A comprehensive sequence analysis program for the IBM personal computer. *Nucleic Acids Res.* 12:581-599.

Ray, D.S., Himes, J.C., Kim, M.H., Imber, R., and Nomura, N. (1982). M13 vectors for selective cloning of sequences specifying initiation of DNA synthesis on single-stranded templates. *Gene* 18:231-238.

Reznikoff, W.S., Siegele, D.A., Cowing, D.W., and Gross, C.A. (1985). The regulation of transcription initiation in bacteria. *Annu. Rev. Genet.* 19:355-387.

Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977). Labeling deoxyribonucleic acid to high specific activity in vitro by nick translation with DNA polymerase I. *J. Mol. Biol.* 113:237-251.

Rodier, F., Gabarro-Arpa, J., Ehrlich, R., and Reiss, C. (1982). Key for protein coding sequence identification: Computer analysis of codon strategy. *Nucleic Acids Res.* 10:391-402.

Rosen, J., Ohtaubo, N. and Ohtaubo, E. (1979). The nucleotide sequence of the region surrounding the replication origin of an R100 resistance factor derivative. *Mol. Gen Genet.* 171:287-293.

Rosenberg, M. and Court, D. (1979). Regulatory sequences involved in the promotion and termination of RNA transcription. *Annu. Rev. Genet.* 13:319-353.

Rownd, R. (1969). Replication of a bacterial episome under relaxed control. *J. Mol. Biol.* 44:387-402.

Sancar, A., Hack, A.M. and Rupp, W.D. (1979). Simple method for identification of plasmid-coded proteins. *J. Bacteriol.* 137:692-693.

Sanger, F., Air, G.M., Barrell, B.G., Brown, N.L., Coulson, A.R., Fiddes, J.C., Hutchinson, C.A., Slocombe, P.M. and Smith, M. (1977). Nucleotide sequence of bacteriophage  $\phi$ X174 DNA. *Nature* 265:687-695.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain termination inhibitors. *Proc. Nat. Acad. Sci. USA* 74:5463-546.

Schinkel, A.H., Marian, J.A., Koerkamp, G., Touw, E.P.W. and Tabak, H.F. (1987). Specificity factor of yeast mitochondrial RNA polymerase: purification and interaction with core RNA polymerase. J. Biol. Chem. 262:12785-12791.

Schnepf, M.E., and Whiteley, H.R. (1985). Delineation of a toxin-encoding segment of a *Bacillus thuringiensis* crystal protein gene. J. Biol. Chem. 260:6273-6280.

Scott, J.R. (1984) Regulation of Plasmid Replication. Microbiol. Rev. 48:1-23.

Sharp, P.M., Tuohy, T.M.F., and Mozurki, K.R. (1986). Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. Nucleic Acids Res. 14:5125-5143.

Shindoh, Y., Urabe, H., Nakano, M.M. and Ogawara, H. (1987). Identification of the minimal replication region of the multicopy *Streptomyces* plasmid pSL1. Plasmid 17:149-157.

Shine, J. and Dalgarno, L. (1974). The 3' terminal sequence of *Escherichia coli* 16S ribosomal RNA: complementarity to nonsense triplets and ribosome binding sites. Proc. Natl. Acad. Sci. USA 71:1342-1346.

Sinma, J. and Dressler, D. (1978). Site-specific initiation of a DNA fragment: Nucleotide sequence of the bacteriophage G4 negative strand initiation site. *Proc. Natl. Acad. Sci. USA* 75:3094-3098.

Singer, S., and Rogoff, M.H. (1968). Inhibition of growth of *Bacillus thuringiensis* by amino acids in defined media. *J. Invert. Pathol.* 12:98-104.

Somerville, H.J. and Jones, M.L. (1972). DNA competition studies with the *Bacillus cereus* group of Bacilli. *J. Gen. Microbiol.* 73:257-265.

Sompayrac, L. and Maaloe, O. (1973). Autorepressor model for control of replication. *Nature New Biol.* 241:133-135.

Southern, E. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol.* 98:503-517.

Staden, R. (1984). Graphic methods to determine the function of nucleic acid sequences. *Nucleic Acids Res.* 12:521-538.

Stalker, D.M., Kolter, R., and Helinski, D.R. (1979). Plasmid R6K DNA replication. I. Complete nucleotide sequence of an autonomously replicating segment. *J. Mol. Biol.* 161:33-43.

Stambaugh, K. and Blakesley, R. (1988). Extended DNA sequencing with Klenow fragment: the kilobase sequencing system. *Focus* 10:29-31.

Steitz, J.A. and Jakes, K. (1975). How ribosomes select initiator regions in mRNA: base pair conformation between the 3'-terminus of 16S rRNA and the mRNA during initiation of protein synthesis in *Escherichia coli*. Proc. Natl. Acad. Sci. USA 72:4734-4738.

Stoker, N.G., Fairweather, N.F., and Spratt, G. (1982). Versatile low-copy-number plasmid vectors for cloning in *Escherichia coli*. Gene 18:335-341.

Stormo, D.G. Schneider, T.F. and Gold, L.M. (1982). Characterization of translational initiation sites in *E. coli*. Nucleic Acids Res. 10:2971-2996.

Summers, D.K. and Sherratt, D.J. (1984). Multimerization of high copy number plasmids causes instability: ColE1 encodes a determinant essential for plasmid monomerization and stability. Cell 36:1097-1103.

Tanaka, T., and Sueoka, N. (1983). Site-specific in vitro binding of plasmid pUB110 to *Bacillus subtilis* membrane fraction. J. Bacteriol. 154:1184-1194.

TeRiele, H. Michel, B., and Ehrlich, S.D. (1986). Single-stranded plasmid DNA in *Bacillus subtilis* and *Staphylococcus aureus*. Proc. Natl. Acad. Sci. 83:2541-2545.



Tessier, L.H., Sondermeyer, P., Faure, T., Dreyer, D., Benavente, A., Villeval, D., Courtney, M. and Lecoq, J.P. (1984). The influence of mRNA primary and secondary structure on human IFN-gamma gene expression in *E. coli*. Nucleic acids Res. 12:7663-7675.

Thomas, C.M. (1988). Recent studies on the control of plasmid replication.. Biochim. Biophys. Acta 949:253-263.

Thomas, C.M., Meyer, R. and Helinski, D.R. (1980). Regions of the broad host range plasmid RK2 essential for replication and maintenance. J. Bacteriol. 141:213-222

Thorne, L., Garduno, F., Thompson, T., Decker, D., Zounes, M., Wild, M., Walfield, A.M., and Pollock, T.J. (1986). Structural similarity between the lepidoptera- and diptera-specific insecticidal endotoxin genes of *Bacillus thuringiensis* subsp. *kurstaki* and *israelensis*. J. Bacteriol. 166:801-811.

Tinoco, I., Borer, P.N., Dengler, B. and Levine M.D. (1973). Improved estimation of secondary structure in ribonucleic acids. Nature New Biol. 246:40-41.

Tomizawa, J.I., Ohmori, M. and Bird, R.E. (1977). Origin of replication of colicin E1 plasmid. Proc. Natl. Acad. Sci. USA 74:1865-1869.

Vaeck, M., Reynaerts, A., Hofte, H., Jansens, S., De Beuckeleer, M., Dean, C., Zabeau, M., Van Montagu, M., and Leemans, J. (1987). Transgenic plants protected from insect attack. *Nature* 328:33-37.

van Manneveld, A., van Teeffelen, H., Baas, P. and Janse, H. (1986). Two juxtaposed tyrosyl-OH groups participate in  $\phi$ X174 catalyzed cleavage and ligation of DNA. *Nucleic Acids Res.* 14:4229-4238.

Vieira, J. and Messing, J. (1982). The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* 19:259-268.

Villafane, R., Bechhofer, D.H., Narayanan, C.S. and Dubnau, D. (1987). Replication control genes of plasmid pE194. *J. Bacteriol.* 169:4822-4829.

Ward, E.S., and Ellar, D.J. (1983). Assignment of the  $\delta$ -endotoxin gene of *Bacillus thuringiensis* var. *israelensis* to a specific plasmid by curing analysis. *FEMS Microbiol. Lett.* 58:45-49.

Watanabe, T. (1963). Infectious heredity of drug resistance in bacteria. *Bacteriol. Rev.* 27:635-643.

Wiggs, J.L., Bush, J.W. and Chamberlin, M.J. (1979) Utilization of promoter and terminator sites on bacteriophage T7 DNA by RNA polymerases from a variety of bacterial orders. *Cell* 16:97-109.

Williams, S.T., Sharpe, M.S. and Molt. (1989). Bergy's Manual of Systematic Bacteriology. Williams and Wilkins.

Winston, S., and Sueoka, N. (1980). DNA-membrane association is necessary for initiation of chromosomal and plasmid replication in *Bacillus subtilis*. Proc. Natl. Acad. Sci. USA 77:2834-2838.

Wolf-Watz, H. and Masters, M. (1979). Deoxyribonucleic acid and outer membrane: Strains diploid for the *orfC* region show elevated levels of deoxyribonucleic acid binding protein and evidence for specific binding of the *orfC* region to outer membrane. J. Bacteriol 140:50-58.

Zykind, J.W. and Smith, D.W. (1980). Nucleotide sequence of the *Salmonella typhimurium* origin of DNA replication. Proc. Natl. Acad. Sci. 77:2460-2464.