

**Characterisation of a high
copy number mutant
pAL5000
origin of replication**

Yvette Jansen

The crest of the University of Stellenbosch is centered behind the author's name. It features a shield with various symbols, including a book and a scale, topped with a crown and a banner. The crest is rendered in a light, semi-transparent style.

A thesis submitted in partial fulfilment of the requirements for the degree of Master of Science in Medical Sciences (Medical Biochemistry) at the University of Stellenbosch

Promoter: Dr. W.R. Bourn

Co-Promoter: Prof. P.D. van Helden

DECEMBER 2001

Declaration

I, the undersigned, hereby declare that the work contained in this thesis is my own original work, and has not, to my knowledge, previously in its entirety or in part been submitted at any university for a degree.

\u2713. Jansen (student number 9772901)

(Date)

Summary

The plasmid pAL5000 is a mycobacterial plasmid isolated from *Mycobacterium fortuitum*. It is a low copy number plasmid, which replicates in both fast growing (e.g. *M. smegmatis*) and slow growing (e.g. *M. bovis* BCG) mycobacteria. Most mycobacterial-*E. coli* shuttle vectors utilise the pAL5000 origin of replication. The minimum replicon consists of ORF1 (RepA), ORF2 (RepB) and the origin of replication.

Dr W.R. Bourn created an *E. coli*-mycobacterial vector based on the pAL5000 origin of replication (pORI) and then subjected it to semi-random mutagenesis. A high copy number mutant was identified (pHIGH) and the causative mutation was tentatively identified as a 3bp deletion situated just upstream of *repB*. This work describes the further characterisation of the mutant plasmid.

Firstly, it was shown by retransforming *M. smegmatis* with both the original and mutant plasmids (pORI and pHIGH), that the mutation causing the increased copy number was plasmid-encoded and not on the chromosome. Following this, it was demonstrated by simple subcloning of the region that carries the 3bp deletion, that other pAL5000-based vectors could be converted to high copy number. In addition to this, the subcloned region was sequenced and the nature of the mutations was confirmed. The subcloning experiment confirmed that the 3bp deletion caused the high copy number phenotype.

Following this, the exact copy number of pHIGH and the relative increase in copy number was determined. From this, the copy number of pORI could also be determined. The plasmid pHIGH has a copy number of approximately 54, compared to the 8 of pORI (a relative increase by a factor of 7).

Because it is important for researchers to know the characteristics of the vectors that they use, especially the influence it will have on its host, stability tests and growth curves were also performed. It was seen that the higher copy number did not markedly increase the stability, however, this is because pORI is already extremely,

and unexpectedly, stable in the host *M. smegmatis*. According to the growth curves, the increased copy number has little effect on the growth of the host *M. smegmatis*.

Possible mechanisms for the increased copy number were then investigated. By using a promoter probe vector, the possible existence of a promoter situated between the two open reading frames of pAL5000 (*repA* and *repB*) was investigated. It was thought that the mutation might have created, or changed an existing promoter, situated between *repA* and *repB*. The results showed, however, that in both pORI and pHIGH there might be a very weak promoter upstream of *repB*, but the mutation did not cause any change that was measurable by the method that was used.

A further possibility was that the mutation caused a change in the RNA secondary structure, which might then have an effect on the translational efficiency of RepB. It was found that the 3bp deletion in pHIGH causes a change in the local RNA secondary structure around the ribosomal binding site and the start codon, when compared to pORI (wild type). This change may cause the translation initiation rate of RepB to be different between pHIGH and pORI. Ultimately it would lead to a different ratio of RepA and RepB in the cell.

Opsomming

Die plasmied pAL5000 is 'n mikobakteriële plasmied wat vanuit *Mycobacterium fortuitum* geïsoleer is. Dit is 'n lae kopie-getal plasmied wat in beide vinnig groeiende (bv. *M. smegmatis*) en stadig groeiende (bv. *M. bovis* BCG) mikobakterieë kan repliseer. Die meeste mikobakteriële-*E. coli* shuttle vektore gebruik die pAL5000 oorsprong van replisering. Die minimum replikon bestaan uit ORF1 (RepA), ORF2 (RepB) en die oorsprong van replisering.

Dr. W.R. Bourn het 'n *E. coli*-mikobakteriële vektor gemaak wat gebaseer is op die pAL5000 oorsprong van replisering (pORI), en dit onderwerp aan semi-random mutagenese. 'n Hoë kopie-getal mutant is geïdentifiseer (pHIGH) en die mutasie hiervoor verantwoordelik was tentatief geïdentifiseer as 'n 3bp deleisie, net stroomop van *repB*. Die projek beskryf die verdere karakterisering van die mutante plasmied.

Eerstens, deur *M. smegmatis* te hertransformeer met die plasmied DNA (pORI en pHIGH), is dit bewys dat dit mutasie wat die toename in kopie-getal veroorsaak, deur die plasmied gekodeer word, en dat dit nie 'n mutasie op die chromosoom is nie. Hierna is dit deur eenvoudige subklonering bewys dat die gedeelte wat die 3bp deleisie dra, ander pAL5000-gebaseerde vektore ook kan verander in 'n hoër kopie-getal. Die sub-klonering eksperiment het ook bewys dat die 3bp deleisie die oorsaak is vir die hoë kopie-getal fenotipe.

Volgende is die presiese kopie-getal van pHIGH en die relatiewe toename in kopie-getal bepaal. Die kopie-getal van pORI kon vanaf hierdie data bepaal word. Die plasmied pHIGH het 'n kopie-getal van ongeveer 54 in *M. smegmatis*, in vergelyking met die 8 van pORI ('n relatiewe toename met 'n faktor van 7).

Aangesien dit vir navorsers belangrik is om die eienskappe van die vektore wat hulle gebruik, te ken, en veral die invloed wat dit op die gasheer sal hê, is stabiliteits toetse, en groeikurwes gedoen. Die hoër kopie-getal het nie die stabiliteit werklik verbeter nie, maar dit is omdat pORI alreeds uiters stabiel is in die gasheer *M. smegmatis*.

Volgens die groeikurwes het die toename in kopie-getal 'n minimale effek op die groei van die gasheer *M. smegmatis*.

Moontlike meganismes vir die hoër kopie-getal is ook ondersoek. Die moontlike bestaan van 'n promoter tussen die twee oop-leesrame van pAL5000 (*repA* en *repB*) is ondersoek deur gebruik te maak van 'n "promoter probe" vektor. Die mutasie kon moontlik 'n promoter geskep het, of 'n bestaande een tussen *repA* en *repB* verander het. Die resultate het gewys dat daar in beide pORI en pHIGH moontlik 'n baie swak promoter stroomop van *repB* is, maar die mutasie het nie enige veranderinge veroorsaak wat meetbaar was met die metode wat gebruik is nie.

'n Verdere moontlikheid was dat die mutasie 'n verandering in die RNA sekondêre struktuur kon veroorsaak het, en dit mag 'n effek hê op die translasië effektiwiteit van RepB. Daar is gevind dat, in vergelyking met pORI, het die 3bp delesie in pHIGH 'n verandering in die lokale RNA sekondêre struktuur rondom die ribosomale bindings posisie en die begin-kodon veroorsaak. Die verandering mag veroorsaak dat die translasië inisiasie tempo van RepB verskillend is vir pORI en pHIGH. Uiteindelik sal dit lei tot 'n heeltemal ander verhouding van RepA en RepB in die sel.

Acknowledgements

I would like to take the opportunity to thank the following people, who have helped me during the course of this project.

My promoter, William, without you this whole project would not have been possible. Thank you for all the advice, all the help, and all the moral support you gave me over the past years.

Prof. Paul van Helden, my co-promoter and head of department, thank you for your support and advice.

Ek wil dankie sê vir my ma en pa, julle het regtig baie vir my beteken. Dankie vir al die ondersteuning en die moed inpraat wanneer ek moedeloos geword het. Uiteindelik hoef julle nie meer te vra: “Is jy al klaar?” nie!

Eugene, al kon jy ook net via email ondersteuning gee, dankie daarvoor, en vir al die raad oor die skryfwerk!

Francois, dankie vir al die kere wat jy kos aangedra het, koffie gemaak het, en wanneer jy net saam met my by die lab was, sodat ek nie alleen werk nie. Dankie dat jy het my help ontspan het wanneer ek dit nodig gehad het. Ek sal nooit genoeg kan dankie se vir al jou hulp nie!

Nora, though you actually failed to be the “bitch from hell”, you did succeed in helping me through tough times. When things kept going wrong, you would be there with a box of tissues (or a roll of paper-towel!). Thank you for talking me through those tears and also for all the advice on my write-up.

Sam, I really appreciate that you read my thesis while you were busy with your write-up. Also, thank you for the help with those primers!

Mireille, ek wil baie graag ook vir jou dankie se vir die morele ondersteuning wat ek konstant van jou gekry het! Ek waardeer al die kuiertjies en koffie regtig baie.

Rob Warren, thank you for all the advice and help you gave me. Thanks also for the times you have helped me in the lab, I really appreciate it!

Ian Wiid and Cedric Werely, thank you also for the assistance and advice you gave me with some of the experiments.

This work was supported financially by Glaxo-Smithklein as part of the Action TB initiative.

Table of Contents:

CHAPTER 1: Introduction

1.1	General overview of the mycobacteria	2
1.1.1	Mycobacterial diseases in the modern age	2
1.1.2	Mycobacteria: classification and physiology.....	3
1.1.3	General genetics of the <i>Mycobacteria</i>	4
1.2	Plasmids	7
1.2.1	General	7
1.2.2	Replication of plasmids	8
1.2.2.1	Theta type replication	9
1.2.2.2	Strand displacement replication.....	9
1.2.2.3	Rolling circle replication.....	10
1.2.3	Control of plasmid replication.....	10
1.2.3.1	Control by iterons	11
1.2.3.2	Control by ctRNAs	12
1.2.3.3	Dual regulation by ctRNA and inhibitory protein	13
1.2.3.4	Non-iteron, non-ctRNA control	13
1.2.4	Mycobacterial plasmids	14
1.2.4.1	Plasmid-encoded functions	15
1.3	The mycobacterial plasmid pAL5000	16
1.3.1	Introduction	16
1.3.2	Sequence analysis of pAL5000	16
1.3.3	Identification of the origin of replication and essential proteins of pAL5000	21
1.3.4	DNA-protein interactions.....	26
1.3.5	Functional analysis of H-site mutations.....	32
1.3.6	Identification of promoter elements in the <i>ori</i> region	33
1.3.7	Study of <i>rap</i> and <i>repA/B</i> promoter mutants.....	34
1.3.8	Studies on Rap	37

1.4	Creating a high copy number mycobacterial vector	37
1.4.1	Making of pORI and pHIGH	37
1.4.2	Tentative identification of the mutation causing high copy number	38
1.5	Aims of project	43

CHAPTER 2: Materials and Methods

2.1	Materials	45
2.1.1	Antibiotic stock solutions	45
2.1.2	Buffers and Solutions	45
2.1.3	Culture media	48
2.1.4	Bacterial strains and plasmids.....	50
2.1.5	Biochemical kits	51
2.1.6	Enzymes used for DNA modification.....	51
2.1.7	Electrophoresis gels	52
2.1.8	Radiochemicals	52
2.1.9	Zhiel-Nielsen staining.....	52
2.1.10	Oligonucleotides.....	53
2.2	Methods	54
2.2.1	Culture conditions, monitoring and culture storage	54
2.2.1.1	<i>Escherichia coli</i>	54
2.2.1.2	<i>Mycobacterium smegmatis</i>	54
2.2.2	Zhiel-Nielsen Staining and light microscopy.....	55
2.2.3	DNA isolation and purification methods.....	55
2.2.3.1	Total DNA extraction from mycobacteria.....	55
2.2.3.2	Small-scale isolation of plasmid DNA from <i>E. coli</i> cultures.....	56
	i) Classic alkaline lysis.....	56
	ii) Wizard Miniprep kit (Promega)	57
2.2.3.3	Large-scale plasmid isolation from <i>E. coli</i> cultures.....	57
	i) Isopycnic CsCl-Ethidium bromide density gradient centrifugation.....	57
	ii) Wizard Midi-prep kit (Promega)	58

2.2.3.4	Phenol extraction	58
2.2.3.5	Isopropanol precipitation	58
2.2.3.6	Ethidium bromide / high-salt extraction protocol for plasmid DNA.	59
2.2.3.7	Removal of unwanted enzymes and EDTA.....	59
2.2.3.8	Wizard PCR preps	59
2.2.4	DNA quantification: Spectrophotometry	60
2.2.5	Enzymatic treatment of DNA	60
2.2.5.1	Restriction endonuclease digests	60
	i) Single restriction digests.....	60
	ii) Double restriction digests	60
	iii) Partial restriction digests	61
2.2.5.2	Ligation reactions.....	61
2.2.5.3	Removal of 5' and 3' overhanging DNA ends by T4 DNA polymerase	61
2.2.5.4	Dephosphorylation of linearised plasmid DNA using shrimp alkaline phosphatase	61
2.2.5.5	Radioactive labelling of probes	62
2.2.6	Polymerase chain reaction	62
2.2.7	Electrophoresis and visualisation of the DNA in the gel.....	62
2.2.7.1	Agarose gel electrophoresis	62
2.2.7.2	Polyacrylamide gel electrophoresis	63
2.2.8	Southern transfer and hybridisation	63
2.2.9	Stripping blots.....	64
2.2.10	Production of competent cells.....	65
2.2.10.1	Heat-shock competent cells (<i>Escherichia coli</i>).....	65
2.2.10.2	Electro-competent cells (<i>Escherichia coli</i>).....	65
2.2.10.3	Electro-competent cells (<i>Mycobacterium smegmatis</i>)	66
2.2.11	Transformation of competent cells	66
2.2.11.1	Heat-shock (<i>Escherichia coli</i>).....	66
2.2.11.2	Electro-transformation (<i>Escherichia coli</i>)	67
2.2.11.3	Electro-transformation (<i>Mycobacterium smegmatis</i>)	67
2.2.12	Mycobacterial protein assay	68
2.2.13	Sequencing.....	68

CHAPTER 3: Characterisation of mutation and mutant plasmids

3.1	Introduction.....	70
3.2	Experimental approach.....	71
3.2.1	Retransformation of <i>M. smegmatis</i> with pORI and pHIGH.....	71
3.2.2	Sequencing of the region containing the mutation.....	71
3.2.3	Subcloning of the mutation into a kanamycin resistance encoding vector.....	72
3.2.4	Analysis of copy number phenotype of pJCX.....	73
3.2.5	Comparing antibiotic resistance levels of pJC86 and pJCX.....	73
3.2.5.1	Initial tests – solid media.....	73
3.2.5.2	Checking for chromosomal mutants – liquid culture.....	73
3.2.6	Direct determination of the copy number.....	74
3.2.6.1	Initial blotting approach.....	74
3.2.6.2	Determining the relative change in copy number.....	74
3.2.6.3	Constructing pYJ200.....	75
3.2.6.4	Probe 1, using pYJ200.....	75
3.2.6.5	Probe 2, using pYJ200.....	76
3.2.6.6	First approach, using pYJ200.....	76
3.2.6.7	Second approach, using pYJ200.....	76
3.2.7	Stability tests.....	77
3.2.8	Growth curves.....	78
3.2.9	Subcloning the mutation into a promoter probe vector.....	78
3.2.10	Testing the chloramphenicol-acetyl transferase activity.....	79
3.3	Results and discussion.....	81
3.3.1	Confirmation that mutant high copy phenotype is plasmid encoded.....	81
3.3.2	Conversion of another vector to high copy number.....	83
3.3.3	Plasmid copy number determination.....	88
3.3.3.1	Single cell resistance test.....	88
3.3.3.2	Initial blotting approach.....	92
3.3.3.3	Determination of relative change in copy number.....	95
3.3.3.4	Determining the exact copy number.....	101

3.3.4	Further characterisation of pORI and pHIGH	108
3.3.4.1	Growth curves.....	108
3.3.4.2	Stability tests.....	110
3.3.5	Investigation of possible mechanisms for increased copy number	113
3.3.5.1	Promoter probe vector – pTV102	113

CHAPTER 4: Conclusions

5.	Bibliography	128
-----------	---------------------------	------------

List of Tables:

CHAPTER 2: Materials and Methods

Table 2.1	Bacterial cells and strains used.....	49
Table 2.2	Vectors.....	49
Table 2.3	Oligonucleotides.....	52

CHAPTER 3: Characterisation of mutation and mutant plasmids

Table 3.1	Plate counts using different diluents.....	88
Table 3.2a	Calculation of generation number of <i>M. smegmatis</i> [pORI].....	111
Table 3.2b	Calculation of generations of <i>M. smegmatis</i> [pHIGH].....	112

List of Figures

CHAPTER 1: Introduction

Figure 1.1	Schematic map of the plasmid pAL5000.....	17
Figure 1.2	The overlap between ORF1 and ORF2 of pAL5000.....	20
Figure 1.3	Linear representation of pAL5000 as it was cloned into pYUB12...	23
Figure 1.4	DNA sequence of the <i>inc</i> region from pAL5000.....	27
Figure 1.5	Schematic map of the minimal replicon of pAL5000.....	28
Figure 1.6	Mutational analysis of the H-site.....	31
Figure 1.7	Results of the mutational assay of the <i>rap-repA/B</i> promoters.....	36
Figure 1.8	Construction of pORI and pHIGH.....	39
Figure 1.9a	Construction of pOSEQ2 and pHSEQ2.....	40
Figure 1.9b	Construction of pOSEQ2 and pHSEQ2.....	41
Figure 1.10	Part of the sequence of pORI showing the 3bp deletion in pHIGH...	42

CHAPTER 3: Characterisation of mutation and mutant plasmids

Figure 3.1	Ethidium gel photograph of <i>M. smegmatis</i> [pORI] and <i>M. smegmatis</i> [pHIGH].....	82
Figure 3.2a	Ligation of pJC86 and pHSEQ2, making pJCX.....	84
Figure 3.2b	Making of pJCX.....	85
Figure 3.3	Ethidium gel photograph of <i>M. smegmatis</i> [pJCX] and <i>M. smegmatis</i> [pJC86].....	87
Figure 3.4	Single cell resistance tests on <i>M. smegmatis</i> cells carrying the plasmids pJC86 and pJCX.....	90
Figure 3.5	Autoradiograph of first approach and ethidium gel photograph of <i>M. smegmatis</i> [pORI] and <i>M. smegmatis</i> [pHIGH].....	93
Figure 3.6	Making of pYJ100 and pYJ200.....	98
Figure 3.7a	Determining the correction factor for the amount of chromosomal DNA loaded on the gel.....	99
Figure 3.7b	The ratio between pORI and pHIGH copy number.....	100
Figure 3.8	Comparisons of the <i>M. smegmatis</i> chromosomal fragment, pADM4 and pYJ200.....	102
Figure 3.9	Autoradiograph of <i>M. smegmatis</i> [pYJ200] and the plot generated by ScionImage.....	104
Figure 3.10	Determination of the actual copy number.....	106
Figure 3.11	Growth curves of <i>M. smegmatis</i> [pORI] and <i>M. smegmatis</i> [pHIGH] compared to the growth curve of <i>M. smegmatis</i>	109
Figure 3.12a	Making of pTVO.....	115
Figure 3.12b	Making of pTVO.....	116
Figure 3.13a	Making of pTVH.....	117
Figure 3.13b	Making of pTVH.....	118
Figure 3.14	Details of the <i>NcoI-EcoRI</i> fragment inserted into pTV102.....	119
Figure 3.15	Results of the chloramphenicol acetyl transferase assay on pTVO and pTVH in liquid culture.....	121
Figure 3.16	Comparison of the mRNA secondary structures of the <i>repA/B</i> transcript.....	124

Figure 3.17 Close-up of the ribosomal binding sites and transcription start sites of *repB* on the mutant and wild type structures..... 125

Abbreviations Used In The Text

µg	microgram
µl	microlitre
/	per
°C	degrees Celcius
³² P	phosphorous radioisotope P-32
BSA	bovine serum albumin
CaCl ₂	calcium chloride
cm	centimetre
CsCl	cesium chloride
dCTP	deoxycytosine triphosphate
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
EDTA	ethylene diamine tetra-acetic acid
g	gram
H ₂ O	water
HCl	hydrochloric acid
http	hyper text transfer protocol
KAc	potassium acetate (CH ₃ COOH)
KCl	potassium chloride
kV	kilovolt
l	litre
LB	Luria Bertani medium
M	molar
MIC	maximum inhibitory concentration
mg	milligram
MgCl ₂	magnesium chloride
ml	millilitre
mM	millimolar
mm	millimetre
MSG	sodium L-glutamate monohydrate (monosodium glutamate)
MTB	monosodium glutamate – tris – EDTA extraction buffer

MW	molecular weight
NaAc	sodium acetate (CH ₃ COOH)
NaCl	sodium chloride
NaOH	sodium hydroxide
NCBI	National Centre for Biotechnology Information
ng	nanogram
NH ₄ Ac	ammonium acetate (CH ₃ COONH ₄)
nm	nanometre
OD	optical density
OD ₂₆₀	optical density at 260 nanometres
pmol	picomol
RNA	ribonucleic acid
RNAse	ribonuclease
rpm	revolutions per minute
SDS	sodium dodecyl sulphate
SOB	culture medium
SOC	culture medium
ssDNA	single-stranded deoxyribonucleic acid
SSPE	sodium chloride-sodium phosphate-ethylene diamine tetracetic acid buffer
STE	sodium chloride-tris-ethylene diamine tetracetic acid buffer
TBE	tris-boric acid-ethylene diamine tetracetic acid buffer
TE	tris-ethylene diamine tetracetic acid buffer
TEMED	N, N, N', N', tetramethylene diamine
Tris	Tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
U	units
URL	Uniform Resource Location
V	volts
wt	wild type
w/v	weight per volume
w/w	weight per weight
www	world wide web

CHAPTER1: Introduction

1.1 General overview of the mycobacteria

1.1.1 Mycobacterial diseases in the modern age

The members of the genus *Mycobacterium* are of interest to researchers, mainly because certain of them cause disease. Of all the mycobacteria, *Mycobacterium tuberculosis* and *Mycobacterium leprae* are the two best-known human pathogens.

M. leprae is the causative agent of leprosy and is an obligatory intracellular pathogen that infects a variety of cells, including macrophages, Schwann cells, muscle cells and endothelial cells. The bacterium also affects skin and peripheral nerves, causing disfigurement and loss of sensation. Leprosy is still a serious health problem in developing countries (92% of cases are detected in 16 of these countries) mostly in Brazil and India (WHO, 1998). Although the registered cases of leprosy decreased from 5.4 million world-wide in 1985, to below 1 million in 1998, 685 000 new cases were registered in 1997, (WHO, 1998). Further information on Leprosy and *M. leprae* can be found in reviews by Young and Cole (1993), Freedman *et. al.* (1999), Jacobson and Krahenbuhl (1999) Rambukkana (2000), Spierings *et. al.* (2000), Visschedijk *et. al.* (2000) and Cole *et. al.* (2001).

In 1882 Robert Koch discovered that tuberculosis (TB) is caused by *M. tuberculosis* and despite the existence of effective treatment for the disease, tuberculosis remains a world-wide problem. In 1997, an estimated 7.96 million new cases of TB were reported, in addition to the 16.2 million existing cases and in that year an estimated 1.87 million people died of the disease. About 32% of the global population (1.86 billion people) are infected with *M. tuberculosis*, and 10.7 million people are co-infected with both *M. tuberculosis* and HIV (Dye *et. al.*, 1999).

In the developed countries, there has been a long-term decline in TB incidence since the introduction of isoniazid and streptomycin in the middle of the last century. However, in the low- and middle-income countries, reported cases are increasing rapidly (Murray, 1998). TB has re-emerged as a serious public health threat for a number of reasons. There has also been a significant increase in multiple-drug-resistant TB (MDR-TB),

probably because TB treatment requires a 6 to 12 month, multiple antibiotic regime. Failure to complete the full course of therapy is one reason for the development of MDR-TB. Furthermore, the synergism between the Human Immunodeficiency Virus (HIV) and *M. tuberculosis* infection accelerates the development, worsens the severity and increases the transmission of TB (Murray, 1998).

M. tuberculosis is arguably one of the most successful human pathogens ever known, but little is known about the molecular basis of its pathogenicity. It is clear that much of the disease pathology results from excessive host immune responses (Cole *et al.*, 1998). However, further research aimed at the development of effective vaccines and short course treatments is required. Further information on TB can be found in reviews by Heifets (1997), Salfinger (1997), Lowry (1999) Jamieson and Chedore (1999) and Hirsch *et al.* (1999).

Many other non-tuberculous mycobacteria that are found commonly in the environment (soil, water and animal hosts) can also cause human disease. These diseases are often compounded by AIDS. Causative agents include *Mycobacterium bovis*, *Mycobacterium xenopi*, *Mycobacterium avium* complex (MAC), *Mycobacterium kansasii* and *Mycobacterium scrofulaceum* (Guay, 1996; French *et al.*, 1997; Opravil, 1997; Salfinger, 1997; Pozniak and Bull, 1999).

1.1.2 Mycobacteria: classification and physiology

The genus *Mycobacterium* is a member of the *Mycobacteriaceae* family, *Actinomycetales* order and in the *Actinomycetes* class. The *Actinomycetes* class comprise the high GC-content, Gram-positive eubacteria (Shinnick and Good; 1994). Within the genus, *Mycobacteria* are divided into two major groups on the basis of growth rate. The fast growing mycobacteria, such as *Mycobacterium smegmatis* and *Mycobacterium parafortuitum* require less than 7 days to form visible colonies on a solid medium, while slow growing mycobacteria, such as *M. tuberculosis* and *M. bovis*, require longer periods (Shinnick and Good; 1994). *M. smegmatis* is often

used as a model organism for research on mycobacteria because of its rapid growth and the fact that it is non-pathogenic.

The mycobacteria can be distinguished from other bacteria under a light microscope by the *Ziehl-Neelsen* (ZN) staining method. Mycobacteria are acid-alcohol-fast and after ZN staining they appear red while the non-acid-alcohol-fast organisms appear blue. This unique staining is due to a group of complex, branched-chain hydroxy lipids termed mycolic acids, which are found on the surface of mycobacterial cells (Kolattukudy *et. al.*, 1997). The mycolic acids are complexed to the peptidoglycan layer of the cell wall and ultimately cause the cell wall to be hydrophobic. The slow growth of these organisms may in part be due to the hydrophobic character of the cell wall, which is thought to cause the cells to be poorly permeable to nutrients. Mycobacteria have very simple nutritional requirements *in vitro* and can be grown in mineral salts with ammonia as the nitrogen source and glycerol or acetate as the carbon source and electron donor (Brock *et. al.*, 1994).

1.1.3 General genetics of the *Mycobacteria*

Research on mycobacterial genetics has, until recently, lagged far behind that of many other bacteria. This lack of progress has been for two reasons. Firstly, there was a perception that mycobacterial diseases would ultimately be defeated by vaccination programs and antibiotic treatment, which caused interest in research on mycobacteria to dwindle (Kolattukudy *et. al.*, 1997). Such confidence has proved to be premature (see section 1.1.1). Secondly, mycobacteria are difficult to work with because mycobacteria grow very slowly (and as yet *M. leprae* cannot be cultured *in vitro*). Mycobacteria are hydrophobic and tend to grow in clumps, hindering the purification of individual cells for genetic analysis. There is no known naturally occurring genetic exchange in mycobacteria and few genetic markers, which could be exploited for development of molecular biological tools, have been identified.

In order to develop recombinant DNA technology for mycobacteria, it was necessary to devise a method to introduce foreign DNA into the cells. Early attempts failed because

of the inability to transfer DNA across the lipid cell wall of the mycobacteria and the inability to regenerate viable cells from protoplasts (cells with an incomplete cell wall) (Falkinham and Crawford, 1994).

Two very important tools used in genetic research on mycobacteria are mycobacteriophages and plasmids. Initially mycobacteriophages were used, because they proved to be easier to introduce into mycobacteria than plasmids (Hatfull and Jacobs; 1994). The first mycobacteriophage was described in 1947, and now more than 250 have been identified (Falkinham and Crawford, 1994). Mycobacteriophages have played an extremely important role in the development of molecular genetic systems for mycobacteria, in the same way that bacteriophages have played a role in genetic research on other bacteria. The first successful introduction of recombinant DNA into mycobacterial cells was performed by using mycobacteriophages. Jacobs *et. al.* (1987) created the first mycobacteriophage cloning vectors, using an *E. coli*-bacteriophage λ cosmid and the mycobacteriophage TM4. They utilised protoplasts, created by using the enzymes lysozyme and lipase on *M. smegmatis* cells.

The first plasmids were introduced into mycobacteria by making use of protoplasts or spheroplasts, but this was not very successful when applied to *M. smegmatis* (Gicquel-Sanzey *et. al.*, 1984). Since then, mycobacterial research utilising plasmid-based vectors has increased immensely, due to the application of recent methodological advances. These include the application of electroporation to mycobacteria (Snapper *et. al.* 1990), the identification of selectable markers expressed in mycobacteria (Snapper *et. al.*, 1988), and the isolation of high frequency transformation mutants (Snapper *et. al.*, 1990).

The development of temperate shuttle phasmids (shuttle vectors that have a phage encoded ability to integrate site-specifically into the chromosome of mycobacteria) by Snapper *et. al.* (1988) extended the technology even further, and selectable markers that were expressed in mycobacteria were identified. Further advances were made with the use of electroporation, which uses a strong electrical field to drive DNA across the cell wall and membrane into the cell, allowing recombinant DNA to be introduced into both the fast growing *M. smegmatis* and the slow growing *M. bovis* and *M. tuberculosis* with

greater efficiency (Snapper *et. al.*, 1988). Following this, the molecular genetics of mycobacteria expanded rapidly. Genetic tools that enabled the introduction and stable expression of foreign genes, isolation and characterisation of promoters, and the cloning of large pieces of DNA were developed (Snapper *et. al.*, 1990 and Jacobs *et. al.*, 1991).

Recently the whole genome of *M. tuberculosis* H37Rv was sequenced (Cole *et. al.*, 1998). The genome consists of 4 411 529 base pairs, and has an average G+C content of 65.6%. Some areas of the genome have exceptionally high G+C contents of >80%. Fifty genes encoding stable RNA species, and 3 924 genes encoding proteins have been identified, accounting for more than 91% of the potential coding capacity (Cole *et. al.*, 1998). Using database comparisons, the protein-coding genes were divided into 11 broad groups, although functions could be attributed to only 40% of the ORF's with any confidence (Cole, 1999).

For further information on the molecular biology of mycobacteria the reader is referred to reviews by Shinnick *et. al.* (1995), Kolattukudy *et. al.* (1997), Parsons *et. al.* (1998), De Voizz *et. al.* (1999), Vergne *et. al.* (2000) and Holland (2001).

1.2 Plasmids

1.2.1 General

Plasmids are extrachromosomal DNA elements (small double stranded DNA molecules) that replicate independently of the host chromosome. Plasmids have characteristic copy numbers within their hosts, and have been found in species of *Archaea*, *Bacteria* and *Eukarya* (Del Solar *et. al.*, 1998), ranging in size from 300bp to 2400kbp. These replicons exist both as covalently closed circular DNA molecules and as linear plasmids, capped with protein at both ends (Hayakawa *et. al.*, 1979; Hinnebusch *et. al.*, 1990, Sakaguchi, 1990).

All plasmids have an essential region that contains the genes and loci involved in control of their own replication. The essential region includes the *ori*, which is characteristic for each replicon, genes encoding proteins involved in the initiation of replication (Rep proteins), and other genes involved in the control of plasmid replication. In addition to this essential region, plasmids frequently carry genes that confer a selective advantage on the host organism. For example, in environmental microorganisms, plasmid-encoded functions include resistance to heavy-metal ions or other toxic compounds (Brock *et. al.*, 1994), and the ability to utilise uncommon nutrient sources. In pathogenic microorganisms, plasmids often carry genes conferring resistance to antibiotics, or encoding virulence factors such as toxins, hemolysins and proteases.

Low copy-number plasmids also require a specific partitioning mechanism (encoded by the *par* locus) to ensure faithful segregation of the plasmids at cell division (Gerdes *et. al.*, 2000). All known plasmid-encoded *par* loci contain three components: one *cis*-acting centromere-like site and two *trans*-acting proteins that form a nucleoprotein complex at the centromere-like site. The two proteins are encoded by two genes in an operon that is autoregulated by the Par proteins themselves. The upstream *par* gene encodes an ATPase while the downstream one encodes a protein that binds the centromere-like region (Gerdes *et. al.*, 2000).

Another system of stable inheritance used by some plasmids, is post-segregational-killing, or plasmid addiction (Jensen and Gerdes, 1995). Whereas other systems increase the likelihood of daughter cells receiving a plasmid during cell division, this system selectively kills any plasmid-free segregant. Post-segregational killing has been described in the IncFII plasmids R1 and F, in plasmids of the incompatibility group P and in bacteriophage P1 (Holcík and Iyer, 1997). Post-segregational killing is always based on a toxin-antidote principle. The system consists of a stable toxin, and an unstable factor that either prevents expression of the toxin, or functions as an antidote to it. The killer toxin is a protein in all cases, but the antitoxin is either a protein or an antisense RNA (Holcík and Iyer, 1997).

Plasmids can incorporate and deliver genes by recombination or transposition and in this way facilitate genetic exchange in bacterial populations. Plasmids play a central role in recombinant DNA technology and are effective cloning vectors because they are small and therefore easy to manipulate. Circular plasmids are very stable during chemical isolation, the replication is independent of the chromosome, and they often carry selectable markers (e.g. antibiotic resistance) allowing for selection of plasmid carrying clones.

1.2.2 Replication of plasmids

Linear plasmids have two types of structures: one type has a hairpin at each end, and replicates via concatemeric intermediates, while the other type has protein covalently bound at the 5'ends, and replicates by a protein-priming mechanism, similar to that of bacteriophage ϕ 29 (Del Solar *et. al.*, 1998). Circular plasmids, on the other hand, have three general replication mechanisms: theta type, strand displacement and rolling circle (RC). These methods of replication by plasmids are discussed in the following section.

1.2.2.1 Theta type replication

Upon initiation of theta type replication the parental DNA strands are melted and a primer RNA is then synthesised. DNA synthesis is initiated by covalent extension of the primer RNA. (Del Solar *et. al.*, 1998). The replication can start at a single point, or several origins of replication, and can proceed in a uni- or bi-directional fashion. DNA synthesis is continuous on the leading strand and discontinuous on the lagging strand. A plasmid-encoded replication initiator protein is required, with few exceptions.

Some plasmids that replicate via this mechanism may also require host DNA polymerase I in the early stages of leading strand synthesis. The replication intermediates are seen (under electron microscopy) as typical θ -shaped molecules, hence the name “theta-type” replication (Del Solar *et. al.*, 1998).

1.2.2.2 Strand displacement replication

The best-known examples of strand displacement replication are found in the inc Q plasmid family, and the prototype for these plasmids is RSF1010. Three plasmid-encoded proteins termed RepA, RepB and RepC, are required for initiation of replication in this plasmid family. Plasmid RSF1010 has a complex *ori*, where initiation of replication occurs. The minimal *ori* contains three identical 20bp iterons, a 31bp AT-rich stretch, and two small palindromic sequences containing the *ssiA* and *ssiB* sites on opposite strands (Del Solar *et. al.*, 1998).

RepC interacts with the iterons and promotes the exposure of the *ssi* sites in a single stranded DNA (ssDNA) configuration, while RepA (a helicase) binds to the AT-rich region. Melting of the duplex by subsequent 5' → 3' translocation of the RepA helicase (bound to the leading strand) exposes the *ssi* sites for RepB primase to initiate DNA synthesis. Replication then proceeds continuously, while RepA helicase displaces the non-replicated parental strand. The double stranded (ds) end product is supercoiled by DNA gyrase. The single stranded (ss) DNA molecules are converted to dsDNA through initiation of DNA synthesis at the *ssi* sequences (Del Solar *et. al.*, 1998).

1.2.2.3 Rolling circle replication

The third type of circular plasmid replication is the rolling circle mechanism. This type of replication is unidirectional and asymmetrical (the leading and lagging strand synthesis is not coupled). One of the most characteristic features of this type of replication is that the leading plus-strand stays covalently bound to the displaced parental plus-strand at its 5'-end. The parental strand is only released at the end of replication of the leading strand.

Replication is initiated by a plasmid-encoded Rep protein that induces a site-specific nick on the plus strand at the double strand origin (*dso*). The primer for leading strand synthesis is the 3'-hydroxyl end left by the nick. DNA synthesis involves several host proteins such as DNA polymerase III, single strand binding protein (SSB) and a helicase (Del Solar *et. al.*, 1998; Khan, 1997 and Khan, 2000).

The replication fork proceeds around the circle until the replisome reaches the regenerated *dso*. Here DNA strand transfer reactions take place and result in the termination of leading strand replication. The end products of the rolling circle replication are a dsDNA molecule (consisting of the parental minus strand and the newly synthesised plus strand) and an ssDNA intermediate (the parental plus strand). The latter is then converted to dsDNA by host proteins, initiating at the single strand origin (*sso*).

1.2.3 Control of plasmid replication

Any particular plasmid has a characteristic copy number in a given host and under fixed growth conditions. Plasmid-encoded control systems that regulate the initiation of replication, are used to maintain the specific copy number.

A plasmid has two stages in its life cycle: establishment (when a plasmid enters a new permissive host) and steady state (when the characteristic copy number is reached). During establishment the plasmid has to replicate rapidly before the host cell divides.

However, in the steady state it is advantageous for the plasmid to control its copy number in order to minimise the load it puts on the host cell.

There are three general types of copy number control systems that have been well characterised, and they depend upon the negative control element used. Firstly there is the iteron system, in which short repeated sequences (iterons) complex with related replication initiator proteins. The second system involves counter transcribed (ct) RNAs, which are antisense RNAs that hybridise to a complementary region of an essential RNA. And finally there are systems in which ctRNA is used in conjunction with a protein (Del Solar and Espinosa, 2000). Another method of copy number control, in which it seems that neither iterons nor ct-RNA are involved, has been discovered recently.

1.2.3.1 Control by iterons

Replicons controlled by iterons have a similar organization. All have two essential functional units: the origin and an adjacent gene encoding a replication initiator protein (Rep). Much of the *ori* sequence consists of iterons, which are arrays of approximately 20bp repeats. The iteron sequences are characteristic for each plasmid and specifically bind the related Rep protein. This protein is important for initiation of replication, but also functions as an autoregulator of the *rep* gene and plays a role in negatively regulating the rate of plasmid replication (Chattoraj, 2000).

Iteron-carrying replicons usually replicate via the theta type mechanism (1.2.2.1). Binding of RepA to the *ori*-iterons is required for melting of the origin, and then the host factor, DnaA (1.2.2.1), initiates replication of the plasmid. In this way host factors are used, while initiation of replication remains under plasmid control.

The rate of replication is determined by iteron concentration, rather than the level of Rep. Iterons can be thought of as the negative regulators, because when the iteron concentration is increased in *cis* or in *trans*, the copy number always decreases. Thus, during normal replication, when the copy number increases, the total iteron concentration also increases. This will in turn inhibit further replication.

It was initially thought that iterons titrate initiators (Rep), making it limiting for replication, but artificially increasing the Rep protein concentration had little effect on the copy number of several different plasmids (Chattorja, 2000). Therefore “handcuffing” had been proposed as a more reasonable model for the control mechanism by iterons.

The “handcuffing” model (Del Solar, *et. al.*, 1998) states that initiation of replication would take place if the copy number is low and Rep proteins bind to and saturate the *ori*-iterons. As the number of plasmid copies increases, the Rep molecules, bound to the iterons of one *ori*, would interact with similar complexes generated at the origins of the other plasmid copies. This would create plasmid pairing through Rep-Rep interactions, which would lead to steric hindrance of the functions of both origins. Plasmid pairs are thought to separate during subsequent cellular growth, due to the increase in cell volume, thus restoring the initiation potential of the individual plasmids.

The Rep protein can be thought of as a positive regulator of the copy number, because once Rep is reduced below the physiological level, the copy number of the plasmid decreases. The availability of active (monomeric) Rep is crucial for initiation, and the protein acts as a positive regulator by rescuing situations that reduce copy number. Handcuffing could limit the initiator by preventing synthesis through autorepression and preventing its binding to the origin through steric hindrance.

1.2.3.2 Control by ctRNAs

Plasmids that control their copy number by using ctRNAs have a specific genetic structure in the control region. This consists of two oppositely orientated promoters directing the synthesis of a RNA essential for replication and the inhibitor ctRNA respectively. The ctRNA is usually complementary to a region near the 5' end of the essential RNA. The important feature of this type of control system is that the rate of synthesis of the ctRNA is much higher than that of the essential RNA, the ctRNA is synthesised from a constitutive promoter and the ctRNA has a short half-life. This leads to ctRNA intracellular concentration staying nearly proportional to that of the copy number. Inhibition of replication is then achieved when the copy number is above

normal, by RNA-ctRNA pairing, which abolishes the activity of the essential RNA (Del Solar & Espinosa, 2000).

1.2.3.3 Dual regulation by ctRNA and inhibitory protein

In the case of dual regulators, expression of the essential *rep* gene is not constitutive as is the case with plasmids controlled by only ctRNAs (1.2.3.2). Instead, this gene is directed by a strong promoter, which is regulated by a protein termed Cop such that there is a high rate of *rep* transcription when the regulatory protein does not operate.

In the case of plasmid pMV158, the *cop* gene is transcribed with the *rep* gene from the same promoter. The Cop protein binds to the DNA region that includes this promoter, and this then inhibits transcription of the *cop-rep* operon (Del Solar and Espinosa, 2000). In this way, by repressing its own promoter, Cop keeps the level of synthesis of *cop-rep* mRNA within narrow limits.

The ctRNA has a short half-life and inhibits *rep* expression by pairing directly with the translation initiation signals of *rep*. Because of its constitutive expression, the concentration of the ctRNAs depends upon the copy number of the plasmid within the cell, and this varying level of ctRNAs would then act upon the practically constant levels of *rep* mRNA.

Variations on this theme can occur, for example in members of the Inc18 family and in plasmid pAM β 1 (Del Solar and Espinosa, 2000).

1.2.3.4 Non-iteron, non-ctRNA control

In this type of copy number control mechanism, the Rep protein is the only plasmid-encoded factor involved in initiation and repression of replication (Burian *et. al.*, 1999).

The copy number control (*cop*) region of plasmids using this type of control (small cryptic plasmids of *E. coli*) comprises of an *inc* (incompatibility) determinant, the *rep* promoter and two Rep binding sites (BD1 and BD2). BD1, which is close to the replication initiation region, binds Rep protein monomers/dimers preferentially. BD2, which is next to the *rep* promoter region, on the other hand, favours binding of Rep oligomers.

It is thought that binding of Rep to both BD1 and BD2 is necessary for initiation of replication. Possibly an increase in the copy number of the plasmid, leading to an increase in the Rep concentration, would favour the formation of Rep oligomers. The oligomers might displace Rep monomers/dimers from BD2 and autorepress *rep* transcription. Therefore in this system, replication control may be exerted by the Rep monomer-multimer equilibrium (Del Solar and Espinosa, 2000).

1.2.4 Mycobacterial plasmids

In the case of mycobacteria, plasmids have been shown to exist both in common free-living species and in important opportunistic pathogens. Naturally occurring plasmids have been found in *M. avium* complex (*M. avium* and *M. intracellulare*), *M. scrofulaceum* and in *M. fortuitum* complex. Many strains in these species carry multiple plasmid types and single plasmid types are a rarity. Most of these plasmids are cryptic i.e. the function of the genes they carry is unknown. To date there are no reports of plasmids identified in *M. tuberculosis*.

Plasmids are commonly found in *M. avium* complex isolates from infected humans. These plasmids range in size from small (< 30kb) to very large (> 150kb), (Falkinham and Crawford, 1994). Two groups of small plasmids have been identified. Group 1 is represented by pVT2 (12.9kb), a plasmid carried by an *M. avium* isolate (Jucker and Falkinham, 1990), and group 2 is represented by pLR7 (15.3kb), which is carried by a serotype 4 *M. avium* strain (Crawford and Bates, 1984).

Related plasmids have been identified in different strains of *M. avium*, *M. intracellulare* and *M. scrofulaceum*, suggesting that mycobacterial plasmids are capable of horizontal transfer (Falkinham and Crawford, 1994). It has been demonstrated by using restriction endonuclease mapping and hybridisation that the group 1 plasmids share discrete regions that hybridise under stringent conditions. In the larger plasmids, these regions of similarity are separated by inserted sequences. This suggested that the *M. avium* plasmids have evolved by insertion of large DNA sequences of unknown origin (Falkinham and Crawford, 1994).

Other mycobacterial plasmids that have been isolated are pMSC262 from *M. scrofulaceum*, and pAL5000 from *M. fortuitum*. The plasmid pMSC262 was used to create an *E. coli*-mycobacterial shuttle vector (pYT937) by inserting a 2.3kb replication region of pMSC262 into the *E. coli* plasmid pACYC177 (Qin *et. al.*, 1994). This shuttle plasmid was unable to replicate in *M. smegmatis*, but was viable in *M. bovis*, *M. phlei* and *M. fortuitum*. The plasmid pAL5000 and vectors based on it are discussed below (section 1.3).

1.2.4.1 Plasmid-encoded functions

In determining the functions encoded by a plasmid, the characteristics of isogenic plasmid-free strains and plasmid-carrying strains are compared. The number of plasmid-associated characteristics identified in mycobacteria, thus far, is limited, because naturally occurring mycobacterial plasmids are extremely stable, and there are very few instances of cured strains (Falkinham and Crawford, 1994).

The plasmid-encoded functions that have been determined are: restriction and modification in *M. avium* complex (Crawford and Falkinham, 1990), mercury and copper resistance in a *M. scrofulaceum* strain (Meissner and Falkinham, 1984; and Erardi *et. al.*, 1987) and growth at 43°C and growth without oleic acid in *M. avium*, *M. intracellulare* and *M. scrofulaceum* (Fry *et. al.*, 1986).

1.3 The mycobacterial plasmid pAL5000

1.3.1 Introduction

The best-characterised mycobacterial replicon is the plasmid pAL5000 (Fig.1.1), which was isolated from *M. fortuitum* (Labidi *et. al.*, 1984). Many *E. coli*-mycobacterial shuttle vectors have been created using the pAL5000 origin of replication. Ainsa *et. al.* (1996) have created a whole family of shuttle vectors derived from pAL5000 and pACYC184. Another series of shuttle vectors (pMC1 to pMC10) were constructed by Chawla and das Gupta (1999), by inserting different sized fragments of the pAL5000 replication region into the vector pBC1 (a modified pUC19 vector). Many other vectors have been created for specific purposes and used in particular roles (for example Raney *et. al.*, 1990; Hatfull, 1993; Lim *et. al.*, 1995). The usefulness of the pAL5000 *ori* lies in the fact that it replicates within both fast growing (such as *M. smegmatis* and *M. fortuitum*) as well as slow growing (such as *M. bovis* and *M. tuberculosis*) mycobacteria.

1.3.2 Sequence analysis of pAL5000

There are two published reports of the pAL5000 sequence, one by Rauzier *et. al.* (1988), and the other by Labidi *et. al.* (1992). Stolt & Stoker (1996a) have re-sequenced pAL5000 over the region (encoding ORF1 and ORF2; see fig. 1.1) where these two reports differ significantly, and found no errors in the data reported by Rauzier *et. al.* (1988). Stolt and Stoker (1996a) have since proven that two particular open reading frames (ORF1 and ORF2) are necessary for replication of the plasmid (section 1.3.3), and yet these two open reading frames are not identified on the sequence reported by Labidi *et. al.* (1992). Therefore in this work, the sequence reported by Rauzier *et. al.* (1988) was assumed to be correct (GenBank accession no. M23557).

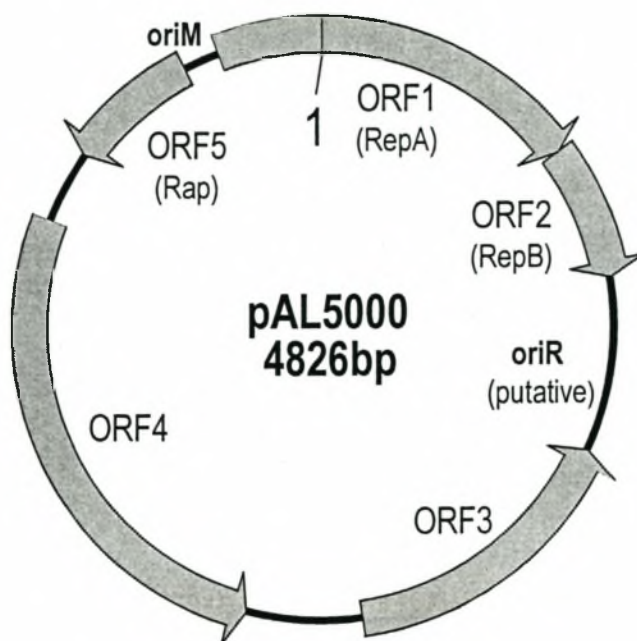


Figure 1.1 Schematic map of the plasmid pAL5000. The map is based on the sequence data by Rauzier *et. al.* (1988). The arrows represent open reading frames. Both the putative *ori* (oriR) and the *ori* (oriM) are indicated.

Published reports concerning the position of the origin of replication are confusing, because this region was originally incorrectly identified within the sequence of pAL5000. Rauzier *et. al.* (1988) noted a region between nucleotides (nt) 1083 and 1543 that had a lower G + C content than the rest of the plasmid. This region also contains several inverted and direct repeats, and no large or obvious open reading frames. The authors described this as the putative origin of replication (situated between ORF2 and ORF3; see Fig. 1.1). Later work proved this to be incorrect (section 1.3.3).

Rauzier *et. al.* (1988) identified, on the basis of codon bias, five large ORF's in the pAL5000 sequence that are likely to represent protein-coding sequences (Fig. 1.1). All of these ORF's start with ATG, except for ORF1, which starts with GTG, and all are preceded by typical Shine Delgarno (SD) sequences.

ORF1 and ORF2 overlap by one base pair. Based on the sequence analysis, these two ORF's code for proteins of 307 amino acids (aa) and 119aa respectively (Stolt and Stoker 1996b). Within ORF1 is a perfect palindrome of 20nt and it was suggested at the time that it might be a recombination site important in plasmid replication. To date there has been no evidence for this.

Hiraga *et. al.* (1994) has compared ORF1 and the replication protein of ColE2-related plasmids, and found approximately 30% identical amino acids through almost the entire region. This implies that the ORF1 product plays a role in replication.

Subsequent sequence homology searches revealed similarity of ORF1 to many replicase proteins, including the replication protein of *Rhodococcus rhodochrous*, RepA from *Brevibacterium linens*, RepA protein from *Thiobacillus sp.* and the Rep protein from *Shigella sonnei*. There was also strong homology to ColE2 and ColE3 Rep proteins (Yasueda *et. al.*, 1989)

The Rep protein of ColE2 appears to act as a primase, but it is unique in the fact that it binds double stranded DNA. Rep binds to the plasmid *ori* and synthesises a unique primer RNA, which is then used by DNA polymerase I to initiate the leading-strand DNA synthesis (Takechi and Itoh, 1995; and Takechi *et. al.*, 1995).

A potential SD sequence of ORF 2 constitutes part of the C-terminal of ORF1 (Fig. 1.2). The deduced protein sequence of ORF2 revealed three repetitions of the motif SGDY at the N-terminus of the protein (Rauzier *et. al.*, 1988). Recent re-analysis of the sequencing data revealed that there were in fact only two repeats (Brigitte Giquel, personal communication). Within ORF2 there is a characteristic helix-turn-helix motif, typical of DNA binding regulatory proteins. The predicted helix 1 shares homology with the Cro protein of phage λ , while helix 2 is similar to CAP, the cAMP receptor protein (Rauzier *et. al.*, 1988).

The N terminus of the ORF3 protein (255aa) possesses a typical signal sequence; two basic residues followed by a 20aa hydrophobic region (Watson 1984). The ORF3 product may therefore be exported. No homology with other proteins has been reported.

ORF4 encodes a protein of 431aa, containing several tandem polypeptide repeats (in the middle of the protein there is a duplication of the amino acid motif: DDTIVGGGA). A smaller amino acid motif, usually containing a GC doublet, is often repeated throughout the sequence. No homology with other proteins has been reported (Rauzier *et. al.*, 1988).

ORF5 is orientated in the opposite direction to ORF1 and ORF2 and encodes a 115aa protein. No convincing similarity to any other gene product has been reported.

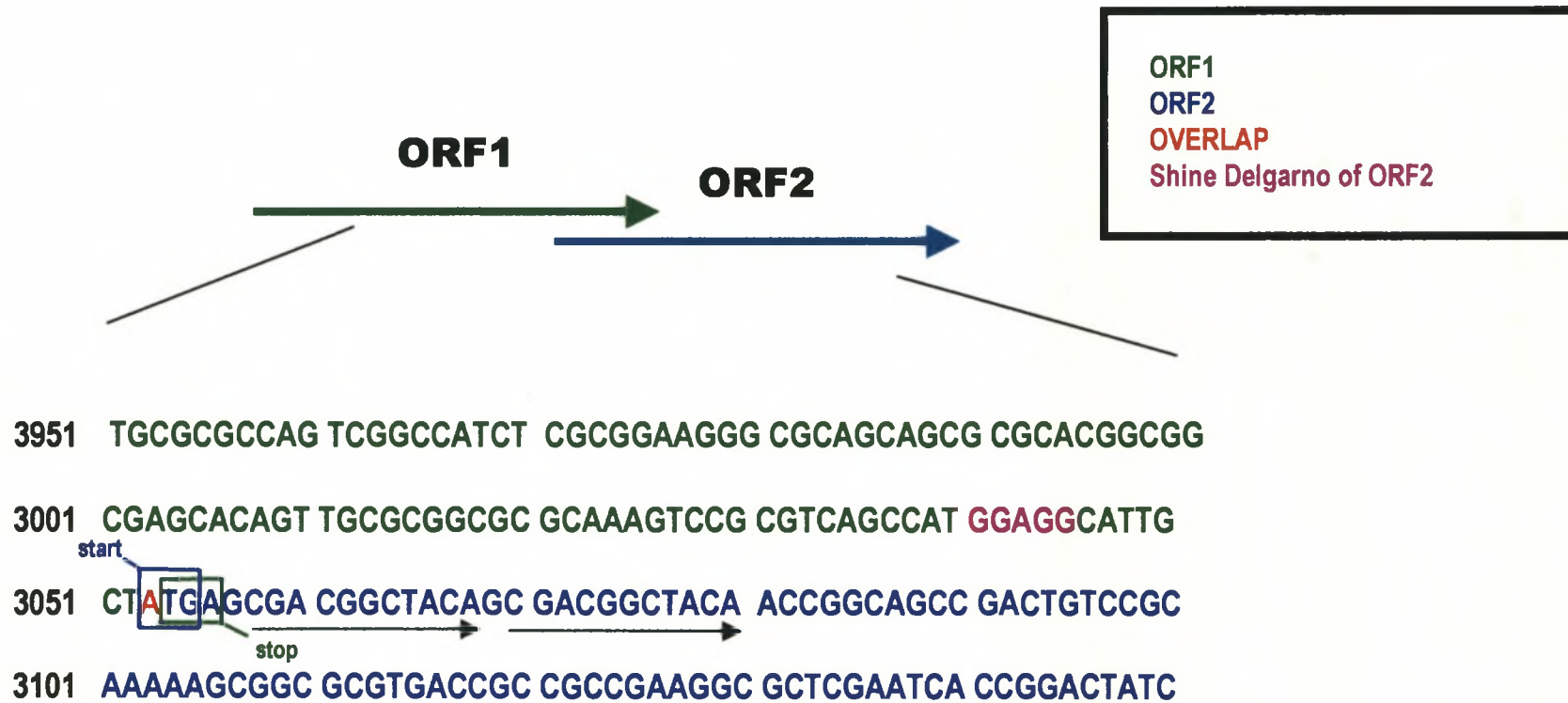


Figure 1.2 The overlap between ORF1 and ORF2 of pAL5000. The potential ribosomal binding site of ORF2 is indicated in pink. Both the stop and start codons of ORF1 and ORF2 are also indicated. The arrows indicate direct repeats.

1.3.3 Identification of the origin of replication and essential proteins of pAL5000

The sequence features of ORF3 and ORF4 suggested to early researchers that they encode proteins that are not necessary for replication. Therefore these genes were chosen as insertion sites during the construction of mycobacterium-*Escherichia coli* shuttle vectors. *M. bovis* BCG and *M. smegmatis* were successfully transformed with plasmids carrying a kanamycin resistance gene (*aph* gene from Tn5) inserted at either of these sites (Ranes *et. al.*, 1990). One of these plasmids was subjected to Southern blot analysis and it was shown that the plasmid was maintained in *M. smegmatis* at a low copy number of about 3 per cell.

Ranes *et. al.* (1990) also showed that a deletion that removed ORF4, ORF5 and the N-terminal regions of ORF1 and ORF3 created a non-viable plasmid. This indicated that ORF1 and/or ORF5 must remain intact for the plasmid replication machinery to function. Alternatively, it was possible that the replication defect was due to non-expression of ORF1 and ORF2 as a result of the deletion of the ORF1 upstream region. Based on these results a mycobacterial-*E. coli* shuttle vector containing the “mini” pAL5000 replicon (consisting of the putative ori, ORF1 and ORF2) and also ORF5 was constructed (Ranes *et. al.*, 1990). This plasmid (pRR3) was successfully introduced into *M. Bovis* BCG and into *M. smegmatis* with very high efficiency (10^4 transformants/ μ g of DNA).

Villar and Benitez (1992) transformed both *M. fortuitum* and *M. smegmatis* with pAL5000-based vectors in order to compare the length of the pAL5000 fragment necessary for autonomous replication of vectors in these two systems. It was shown that ORF1 and ORF5 are not required for replication and stable maintenance of pAL5000 vectors in a *M. fortuitum* host, and only ORF2 is essential. It is possible that proteins encoded by *M. fortuitum* participate in the replication and partitioning of pAL5000 and that ORF1 and ORF5 provide these factors in heterologous hosts that do not encode the necessary factors.

The pAL5000 sequence published by Rauzier *et. al.* (1988) shows that ORF1 and ORF2 are flanked by sets of repeats. Stolt and Stoker (1996a) have further investigated whether these ORF's encode proteins necessary for replication, and whether the repeats have an influence on copy number control and incompatibility. In their study, two pAL5000-derived plasmids were constructed. These consisted of pAL5000 with an *E. coli* origin of replication and either a kanamycin resistance gene from transposon Tn5 (pYUB12) or a hygromycin resistance gene from *Streptomyces hygroscopicus* (pUH4) inserted into ORF4 (Stolt and Stoker, 1996a). It was found that when ORF2 and half of ORF1 were deleted from pYUB12, the resulting construct could not replicate in *M. smegmatis*. In co-transformations with pUH4, however, double transformants were obtained and both plasmids could be recovered intact. This indicated that one or both products from ORF1 and ORF2 are indeed necessary for replication, and that it acts *in trans*.

Constructs in which ORF2 was deleted, leaving only ORF1, could not replicate autonomously in *M. smegmatis*, irrespective of how much of the upstream sequence was included (including ORF5). This showed that ORF2 is essential for replication. If the whole of ORF1 and ORF2 were included, only a few hundred base pairs upstream of ORF1 were necessary to create a viable plasmid. The upstream region with the high incidence of repeats (region 1, Fig. 1.3) was always necessary for a viable plasmid. Stolt and Stoker (1996a) have designated the products from ORF1 and ORF2 as RepA and RepB respectively.

At the time, it was argued that RepA and RepB might not be two distinct proteins, but a single protein produced by ribosomal slipping. Stolt and Stoker (1996b) demonstrated the presence of two proteins by co-electroporation of *M. smegmatis* with two derivative plasmids of pAL5000. One plasmid lacked *repB* and the other had a deletion in *repA*. Neither of these could replicate on its own. In co-transformation experiments, both plasmids replicated in the transformants, proving the existence of two proteins, both of which act *in trans*.

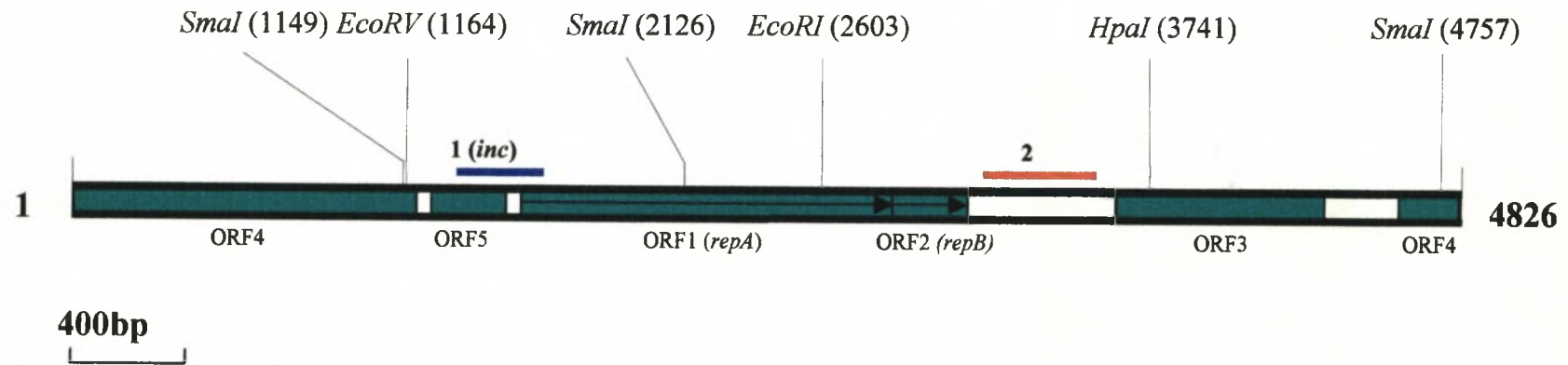


Figure 1.3. Linear representation of pAL5000 as it was cloned into pYUB12 (Stolt & Stoker, 1996a)
 The regions with high frequency of repeated motifs are labeled 1 and 2 and are marked by blue and orange lines respectively. ORF4 was split by the cloning procedure. The arrows indicate the orientations of ORF1 and ORF2 (based on figure by Stolt and Stoker, 1996a).

Stolt and Stoker (1997) also tested the effect of *repA* and *repB* expression levels on plasmid copy number. As there is evidence that sufficient levels of RepA and/or RepB in the cells are important for establishing a plasmid population (see below, section 1.3.5) it was thought that by raising the expression of these genes the copy number could be increased. This proved not to be the case, and it was established that a replicon with high expression of *repA/B* is not stable in *M. smegmatis*. Further co-transformation experiments proved that only high levels of RepB are deleterious for the cells, but that high levels of RepA could be tolerated.

Additional copies of regions 1 and 2 (Fig. 1.3) were cloned into pYUB12 to test whether the repeated sequences in pAL5000 have an influence on incompatibility and copy number (Stolt and Stoker 1996a). *M. smegmatis* cells were then transformed with these constructs together with pUH4, and differences in co-transformation efficiency were scored as an indication of incompatibility. No noticeable effect on the co-transformation ability or stability of the pYUB12 derivative plasmid was observed when the number of repeats in region 2 was doubled. Doubling region 1 however, caused a marked decrease in the co-transformation efficiency between the two constructs. The transformation efficiency of each pYUB12-derived construct on its own was not much different from that of pYUB12. Therefore the differences in co-transformation ability were attributed to the effects of region 1 and this region was designated *inc*. The active region (*inc*) was narrowed down to a 435bp region (Fig. 1.4). All the functional replicons that had been created contained this region, implying that *inc* is also necessary for the replication of pAL5000 (Stolt and Stoker, 1996a).

The decrease in co-transformation efficiency was only observed with plasmids where the extra *inc* regions were cloned in the same orientation. Plasmids with two *inc* regions cloned in the opposite orientation behaved like the wild type. The authors argued that this indicated that the sets of direct repeats (and not the inverted repeats) in the region are responsible for the observed effects, because inverted repeats would be orientation independent. The *inc* region was also shown to confer incompatibility in an unrelated mycobacterium-*E. coli* shuttle vector (Stolt and Stoker, 1996a).

Stolt and Stoker (1996a) argued that, if the *inc* region was the target for replication proteins acting in *trans*, this region might be the origin of replication for pAL5000. To

test this hypothesis, they cloned the 1kb *SmaI* fragment (Fig. 1.3) containing *inc*, into a plasmid vector that could not replicate, but could be selected for in *M. smegmatis*. The final construct could only replicate autonomously in *M. smegmatis*, when it was co-transformed with pUH4, which encodes RepA and RepB. The authors concluded that all the *cis*-acting elements necessary for replication are encoded in the *inc* region (bases 3961 to 4620; Fig. 1.2). Thus the true pAL5000 *ori* is the *inc* region.

1.3.4 DNA-protein interactions

In *vitro* DNA binding assays have been performed with purified RepA and RepB proteins that had been expressed as maltose binding protein (MBP) fusion proteins in *E. coli* (Stolt and Stoker, 1996b). The intact RepA-MBP fusion protein had to be used in these assays, because digestion with factor X (used to cleave the MBP from the fusion product after purification) gave rise to several bands on SDS-PAGE despite the fact that RepA should have no factor X recognition site. No such problems were experienced with RepB.

The 435-bp *ori* segment (Fig. 1.4; from nt 4175 to 4620) was used as a probe to test the recombinant proteins for DNA-binding activity. This probe was ³²P-end labelled and used in gel retardation assays. Despite the similarities of RepA to other replication proteins (see section 1.3.2), no DNA-binding activity was detected with the RepA-MBP fusion protein, even when tested under conditions of varied pH (5 to 9) in the presence or absence of ATP or in the presence of different cations (Mg²⁺, Ca²⁺, Zn⁺, Mn²⁺ and K⁺). Possible explanations for this are that the MBP fusion causes RepA to be inactive or that RepA might need host proteins for its activity.

In contrast, it was shown that RepB binds specifically to the 435bp area (Fig. 1.4) that contains the *ori* region. DNase I footprinting experiments were used to determine the exact sequence where RepB binds to *ori*. Two regions where bases were protected from DNase I cleavage by bound RepB protein were identified and the footprinting gels suggested that RepB binds with different affinities to the two regions. These two sites (about 57bp apart) were designated the H-site (high affinity), and the L-site (low affinity; Fig. 1.5). The affinity of RepB for the H-site was shown to be approximately ten times higher than for the L-site (Stolt and Stoker, 1996b). The DNA sequence of the L-site (between bases 4534 and 4564) does not display any obvious pattern or symmetry, and binding of RepB to this site is to one side of the DNA helix only. The exact role of the L-site and its interaction with RepB is unknown.

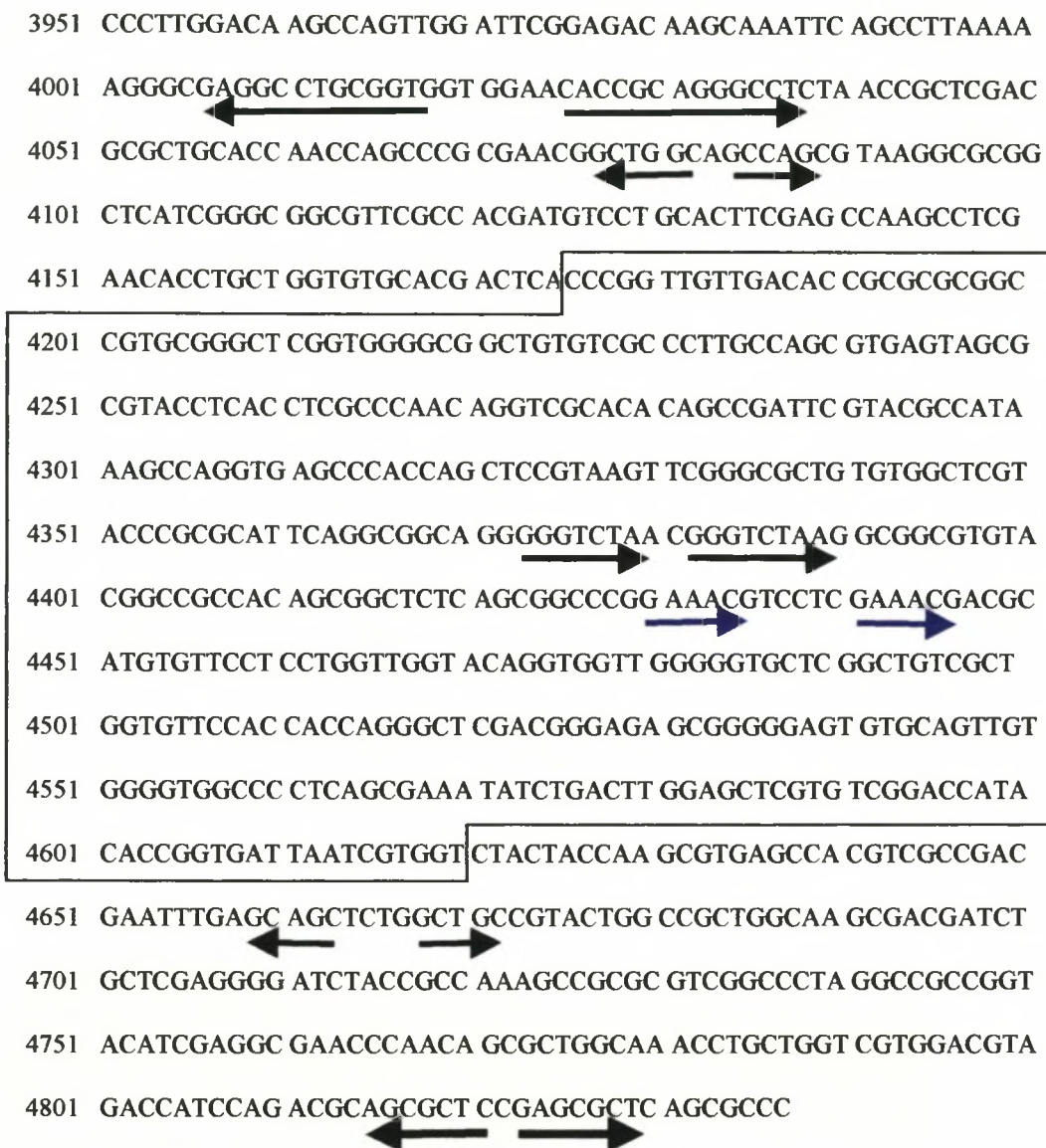


Figure 1.4 DNA sequence of the *inc* region (region 1) from pAL5000. The boxed sequence is the minimum defined *inc* region. The arrows indicate direct and inverse repeats in the sequence. Numbering refers to the sequence in GenBank M235557 (based on figure by Stolt and Stoker, 1996a).

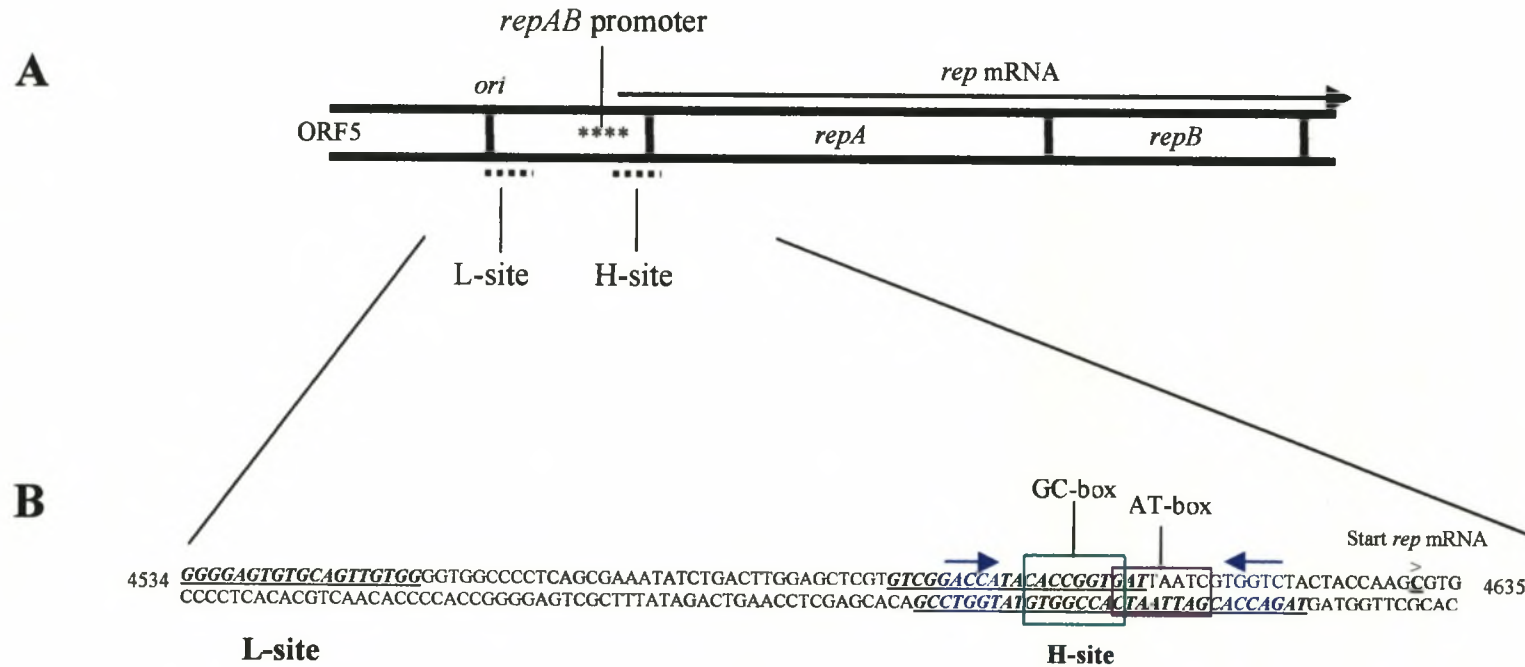


Figure 1.5 (A) Schematic map of the minimal replicon of pAL5000. The dotted lines mark the binding sites for RepB, and the *rep* promoter region is indicated with asterisks.

- (B) Sequence of the RepB binding region. The numbering refers to the sequence of pAL5000 in GenBank (Accession No. M23557). The bases protected by DNaseI cleavage by RepB binding are boldfaced and underlined. The start point of *rep* mRNA transcription is indicated with an arrowhead above an underlined bold base. The AT-box is boxed in purple; the GC-box is boxed in green. The inverted repeat structure in the H-site is marked in blue. This figure is derived from that of Stolt and Stoker (1996b).

The H-site (between bases 4590 and 4623) has an intricate structure. It consists of two interlocking 8bp palindromes, one of which is G+C rich and the other A+T rich, bracketed by an inverted repeat of 5bp (Fig. 1.5). RepB binding to the H-site occurs on both sides of the DNA helix (Stolt and Stoker, 1996b). It appears that RepB binds here co-operatively in two copies, although it is not known whether it binds as monomers or as a dimer. The H-site is immediately (approximately 9bp) upstream of the *repA/B* transcription start site and partially occludes the putative promoter structures, so it has been suggested that RepB acts as an autoregulator of its own expression (Stolt and Stoker, 1996b). The authors suggest that the L-site might be the *ori* itself, based on the assumption that it is far from the promoter region and is therefore unlikely to have any influence on the regulation of *rep* expression.

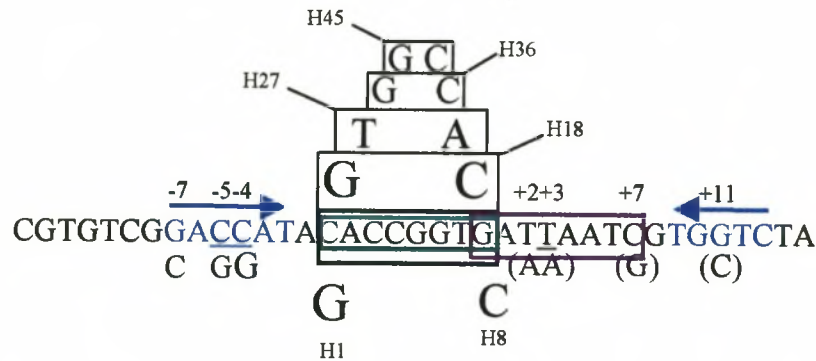
Based on their findings, Stolt and Stoker (1996b) suggested a model for the triggering of replication of pAL5000. They postulate that RepB will be titrated by the H-sites on plasmids in the cell, and RepB will not bind to the *ori* while there are still free H-sites available. Only when all the H-sites are occupied will RepB bind the low-affinity *ori*, which will trigger replication and generate a new set of unoccupied binding sites.

It was further suggested by Stolt & Stoker (1997) that the palindrome GATTAATC (AT-box) might be a feature of the promoter because of the high (A+T)-content, while the GC-box (CACCGGTG) could be a recognition sequence for RepB. To test this, the authors constructed a set of oligonucleotides with specific mutations in one or the other of the palindromes or in the 5bp repeats (Fig. 1.6). These oligonucleotides were then used in PCR reactions to produce 118bp double-stranded target DNA templates, used in gel retardation assays with purified RepB protein.

The binding affinities ($K_{d_{DNA}}$) of these mutant H-sites to RepB were compared to that of the wild-type H-site. The results are summarised in Fig 1.6. Changing the first C of the G+C rich palindrome (GC-box) to a G (mutation H1) reduced the affinity of RepB for this site almost 100 times. When both the first C and last G were changed, to preserve the palindrome, binding was also abolished (mutation H18). A less pronounced effect was observed when the final G of the GC box was changed to a C (mutation H8). Mutations of the four central bases of the palindrome had a small effect on RepB binding, reducing the binding affinity about 5-fold. It was shown that for

RepB binding, the AT-box was much less important than the GC-box, as three different, single base pair changes (mutations +2, +3 and +7) had no significant effect. Mutation +11 produced a wild-type binding pattern, although it destroyed the inverted repeat structure. Base changes in the upstream sequence of the inverted repeat (mutations -7, -5 and -4) reduced binding about 10-25-fold. This indicates that the upstream sequence of the inverted repeats is more important for binding of RepB than the downstream sequence. Mutations at bases -7, -5, -4 (in the upstream sequence of inverted repeat) and H8 also caused RepB to bind in only one copy, in contrast to the wild-type, which binds in two copies (Stolt and Stoker, 1997).

A



B

Mutation	Relative $K_{d_{DNA}}$	Replication
H1	<2%	-
H18	<2%	-
-4	2-10%	-
H27	2-10%	-
H8	2-10%	-
-5	2-10%	-
-7	2-10%	+
H36	10-99%	-
H45	10-99%	-
+2	100%	+
+3	100%	-
+7	100%	+
+11	100%	+

Figure 1.6 Mutational analysis of the H-site.

- A. The GC-box is boxed in green and the AT-box is boxed in purple. The blue arrows mark the inverted repeat structure. The different mutations introduced are shown above and below the sequence. The boxed bases are mutations where pairs of bases were changed in order to preserve the palindromic structure of the GC-box. The numbering of the mutations are H1-H8 for bases inside the GC-box, positive numbers for bases downstream of the GC-box and negative numbers for bases upstream of the GC-box. The sizes of the mutated bases are roughly proportional to the impact of the base-change on DNA binding.
- B. The effect of the mutations on the binding of RepB to its target DNA, in terms of the determined relative binding constants ($K_{d_{DNA}}$).

This figure is derived from that of Stolt and Stoker (1997).

1.3.5 Functional analysis of H-site mutations

In order to test the ability of the H-site variants to replicate in *M. smegmatis* Stolt and Stoker (1997) also created plasmids carrying the mutations described in section 1.3.4. The results of this assay are summarised in Fig. 1.6. None of the plasmids which had changes made to the GC-box could replicate. In contrast, mutations at +2 and +7 to the AT-box produced viable plasmids, while the -4 and -5 mutations in the upstream sequence of the inverted repeat, abolished replication. The mutation +11, which destroys the inverted-repeat motif downstream of the AT-box, also yielded a viable plasmid. In later work, further mutations introduced at +4 and +6 also yielded viable plasmids, but the construct carrying mutation +14 could not replicate (Stolt *et al.*, 1999). The -7 mutation yielded a viable plasmid, but its transformation efficiency was consistently ten times lower than that of the other constructs. The other viable mutants transformed cells with efficiency comparable to that of the wild type. There was no difference in growth rate and copy number between cells carrying the -7 mutant and those carrying the wild type (Stolt and Stoker, 1997).

The fact that mutations in the AT-box did not affect binding of RepB *in vitro*, but did abolish the replication ability of the plasmid in some cases, supports the conclusion of Stolt and Stoker (1997) that this box is part of the promoter structure and not a recognition motif for RepB.

It was also noted that cells transformed with any of the viable plasmids containing mutations in the H-site, took 1-2 days longer to form colonies on kanamycin plates than cells transformed with pYUB12, which contains all of pAL5000. The authors suggest that this indicates that the initial expression of *repA/B*, immediately upon transformation, was less efficient in the manipulated plasmids (Stolt and Stoker, 1997).

In co-transformation experiments using a plasmid that cannot support replication on its own, it was found that none of the viable, manipulated constructs could support replication of this plasmid *in trans*. Only cells already containing the manipulated constructs, and thus already expressing *repA/B*, could be successfully transformed with this plasmid (Stolt and Stoker, 1997).

1.3.6 Identification of promoter elements in the *ori* region

Prior to the fine detail mapping of the start point of transcription for the *repA* mRNA, reverse transcription-PCR (RT-PCR) was used to narrow down the region to be examined (Stolt and Stoker, 1996b). Using pairs of primers it was shown that the start point of transcription is downstream of bp 4420 in the pAL5000 sequence. By using RT-PCR, Stolt and Stoker (1996b) also demonstrated that there is a common transcription unit for *repA* and *repB*.

Following this, a single signal, attributable to a 5' end of the mRNA, was detected by S1 nuclease protection assays. The start point of the *rep* mRNA was thus defined to be at the nucleotide C, number 4632 in the pAL5000 sequence. Probes derived from oligonucleotide sequences found upstream of this locus, produced no signals attributable to 5' ends. This indicates that the single 5' end is not a processing site from an RNA species that starts further upstream.

Of the five open reading frames on the pAL5000 sequence, only RepA and RepB are necessary for replication in *M. smegmatis* (section 1.3.3). It had been noticed, however, that *M. smegmatis* cells recover faster upon electroporation when the minimal replicon contains additional upstream sequences. Stolt and Stoker (1996a) have noted that the minimal replicon containing the pAL5000 region 4327-1093 (*ori*, *repA* and *repB*) took 5-6 days to form colonies on plates while the replicon containing the pAL5000 region 3875-1093 (*ori*, *repA*, *repB* and *orf5*) took only 3-4 days to form colonies.

Stolt *et al.* (1999) suggested two possibilities: either that ORF5, which lies in this upstream region (nt 5542-4102), encodes an auxiliary factor, or that the nucleotide sequence upstream of the L-site was itself responsible for the improved replication ability. They investigated this using various constructs carrying all or parts of ORF5. When 203bp was deleted from the upstream region of the previously defined minimal replicon (pAL5000 region 4327-1093) the resulting plasmid was still viable (Stolt *et al.*, 1999). This construct contained only *repA*, *repB*, the H-site, the L-site and 3bp upstream of the L-site and needed ~5 days to form colonies of 0.5mm in diameter. Another construct that included additional repeated sequence motifs, and only half of

ORF5 also needed ~5 days to form colonies. However, in the case of constructs where the whole of ORF5 was present, the transformants only took 3 days to form colonies of similar size. The authors concluded that the improved replication ability was due to the gene product of ORF5, which they termed Rap, and not due to a *cis*-acting DNA structure.

No effect on recovery of transformants was observed when additional sequences upstream of *rap* were included, or when the repeated sequence features between *rap* and the L-site were deleted. The role of the structural features in this region remains undefined, but does not seem to be related to transformation (Stolt *et al.*, 1999). Replicons without the *rap* gene are still functional, which indicates that the product is either not essential or that Rap can be substituted for by a host factor. No genes with significant similarity to *rap* could be found in the GenBank database and therefore any similarities must be on a higher level than the amino acid sequence (Stolt *et al.*, 1999).

Using RT-PCR and primer extension experiments, the start point of the Rap ORF was shown to be at base 4452, and the start point of the *rap* mRNA at base 4595. This is within the H-site and only 37bp from the start point of *repA/B* transcription, implying that the promoters overlap to a great extent. The H-site contains several palindromes and repeats, and it is possible to place a line of symmetry in the middle of the AT-box, showing many of the sequences on each side to be mirror images of each other. The entire dual promoter can thus be considered as an imperfect palindrome. The close proximity and positions of the two transcription start sites imply that a bi-directional promoter exists rather than back-to-back or face-to-face divergent promoters (Stolt *et al.*, 1999).

1.3.7 Study of *rap* and *repA/B* promoter mutants

Divergent promoters are a common form of gene organization (Beck and Warren, 1988), but the extensive symmetry seen in the *repA/B* and *rap* promoters is unusual, indicating that the shared sequence motifs are important for functioning of both promoters (Stolt *et al.*, 1999).

Stolt *et al.* (1999) postulated that a mutation to the proximal element of one promoter would lead to a mutation in the distal element of the other. These authors performed a study in order to test the importance of the shared sequence motifs for the two overlapping promoters. The *xylE* gene was used as a reporter gene for the study, and the mutant promoters were inserted in front of the reporter gene in one or the other orientation. The mutations that were created involved changing only A to T and C to G and vice versa, in order to keep the melting energy constant, because promoter activity is dependent on both binding affinity and the energy needed to open the DNA helix. Figure 1.7 gives a summary of the results.

The expected down-regulation mutation in one promoter accompanied by an up-regulation in the other was only observed in the case of mutations -5 and +14. With these mutations there was an up-regulation in the *rap* promoter with a simultaneous down-regulation of the *repA/B* promoter. Stolt *et al.* (1999) came to the conclusion that both the proximal and the distal regions are important for both promoters. The influence of changes made to the AT-box was not as conclusive, but some changes to the proximal and distal boxes (e.g. -5 or +10 in *repA/B* or -2 or +10 in *rap*) almost abolished the promoter activity (Fig. 1.7).

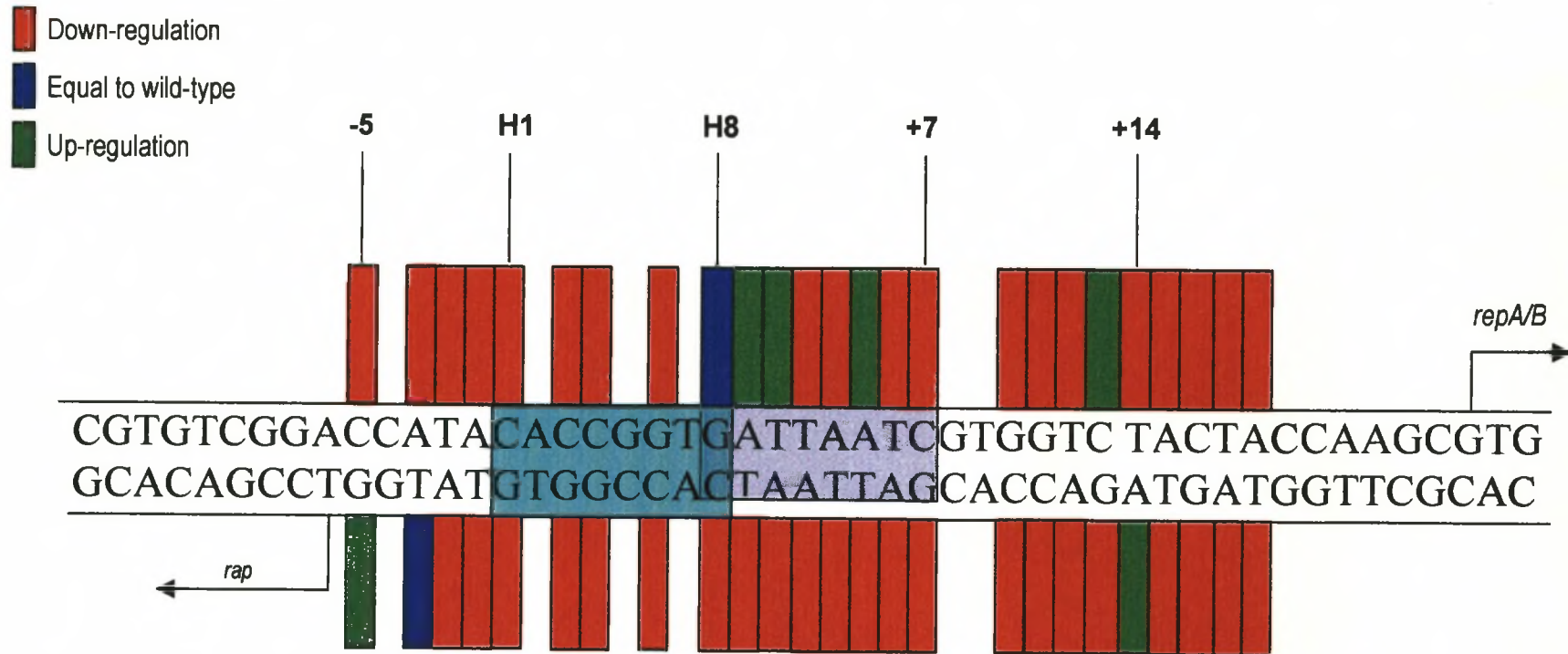


Figure 1.7 Results of the mutational assay of the *rap-repA/B* promoters. H1 is the first base in the GC-box, H8 is the last base in the GC-box; +1 is the first base downstream of this box and -1 is the first base upstream of the box. The GC-box is highlighted in green and the AT-box in purple. All the mutations were single base pair substitutions – A was changed to T and G to C and vice versa, in order to keep the melting energy constant. This figure is derived from that of Stolt and Stoker (1999).

1.3.8 Studies on Rap

No DNA-binding activity was observed for the Rap protein in gel retardation assays, using both the *ori* region (section 1.3.4) and the *rap* encoding DNA upstream of the L-site as templates.

In gel retardation assays with RepB, it was noted that RepB had a higher affinity for the *ori* region when Rap was present in the mixture. However, the same effect was observed when the same concentration of BSA was used in the assay (Stolt *et al.*, 1999)

1.4 Creating a high copy number mycobacterial vector

Many *E. coli*-mycobacterial shuttle vectors have been constructed, based on pAL5000 (1.3.1), but none of these are high in copy number. The usefulness of a high copy number vector is discussed in depth later (3.1), but briefly, the advantages include greater ease of isolation and manipulation of the DNA, higher expression of cloned genes and increased stability. A high copy number *E. coli*-mycobacterial shuttle vector would therefore be of great use in mycobacterial research. A vector based on pAL5000 would be even more useful, because the pAL5000 *ori* is active in both fast- and slow-growing mycobacteria. Furthermore, the many pAL5000-based vectors that already exist could be converted to high copy number phenotype by simple recombinant DNA construction methods.

1.4.1 Making of pORI and pHIGH

In order to create a high copy number shuttle vector, a plasmid (pORI) derived from pSMT3, a derivative of p16R1 (Garbe *et al.*, 1994) that carries the heat shock promoter Hsp60 (Dr. W.R. Bourn; personal communication), was subjected to semi-random mutagenesis (Dr. W.R. Bourn, personal communication). This involved partial digestion of pORI, with a mixture of 4bp-recognising restriction enzymes to linearise

the plasmid, removal of overhanging ends and re-ligation to create a library of plasmids with different small insertions and deletions. A high copy number mutant was identified after a series of consecutive plasmid isolations and transformations designed to enrich the culture for those clones carrying high copy number vectors. Total DNA was isolated from these cultures to test for the presence a high copy number plasmid. As a quick indication, 3 μ g of total DNA from the culture containing the putative high copy number plasmid, and 3 μ g of total DNA from the culture with the original plasmid, was digested with *EcoRI* and the digests run on a 0.7% agarose gel. It was seen on these gels, from the increase in the relative intensity of the bands, that there was a definite increase in the copy number of the plasmid. The putative high copy number plasmid isolated was called pHIGH (Fig. 1.8).

1.4.2 Tentative identification of the mutation causing high copy number

The origins of replication from pORI and pHIGH were subcloned for the purpose of sequencing. The resulting plasmids were called pHSEQ2 and pOSEQ2 (Fig. 1.9). These plasmids were sequenced on one strand only and this revealed a 3bp deletion (CGC) in the pHIGH sequence, between ORF1 and ORF2 (nt. 647 to 649 in pAL5000 sequence, and nt. 2981 to 2983 in pORI sequence). This 3bp deletion is out of frame (in that it overlaps two separate codons), but effectively leads to the elimination of only one amino acid: an alanine (Fig. 1.10).

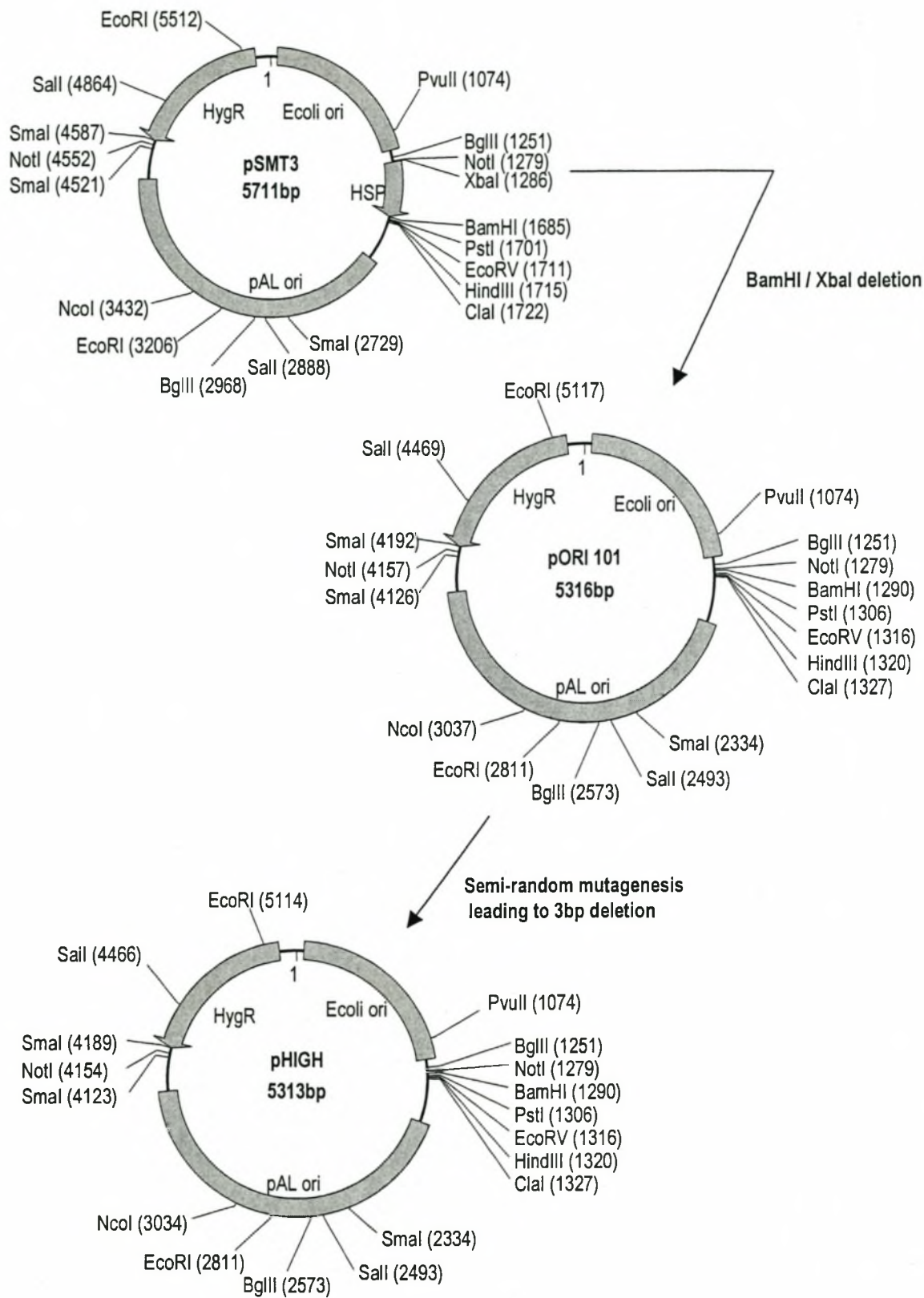


Figure 1.8 Construction of pORI and pHIGH as done by Dr. WR Bourn (personal communication).

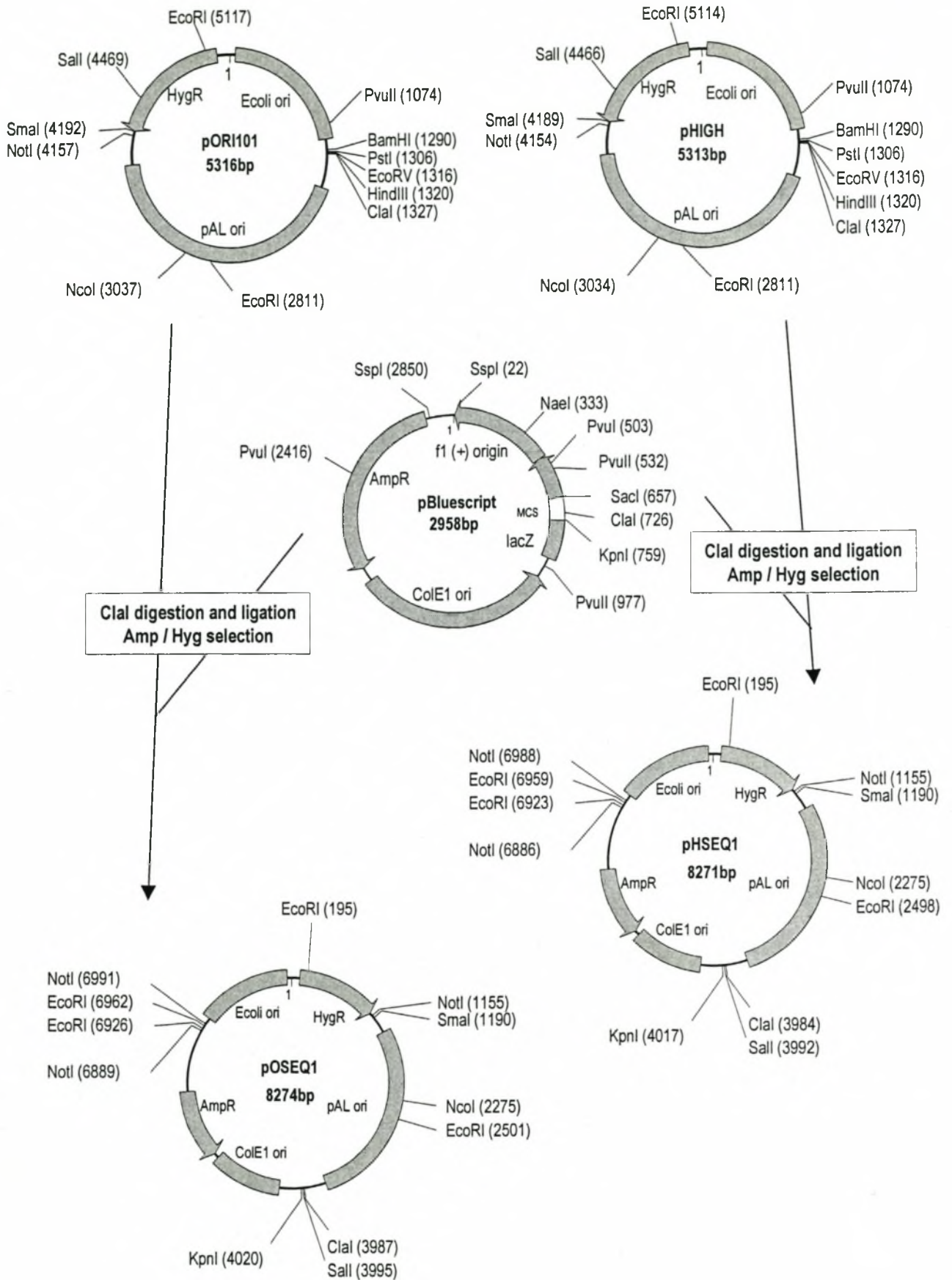


Figure 1.9(a) Construction of pOSEQ2 and pHSEQ2 as done by Dr. WR Bourn (personal communication).

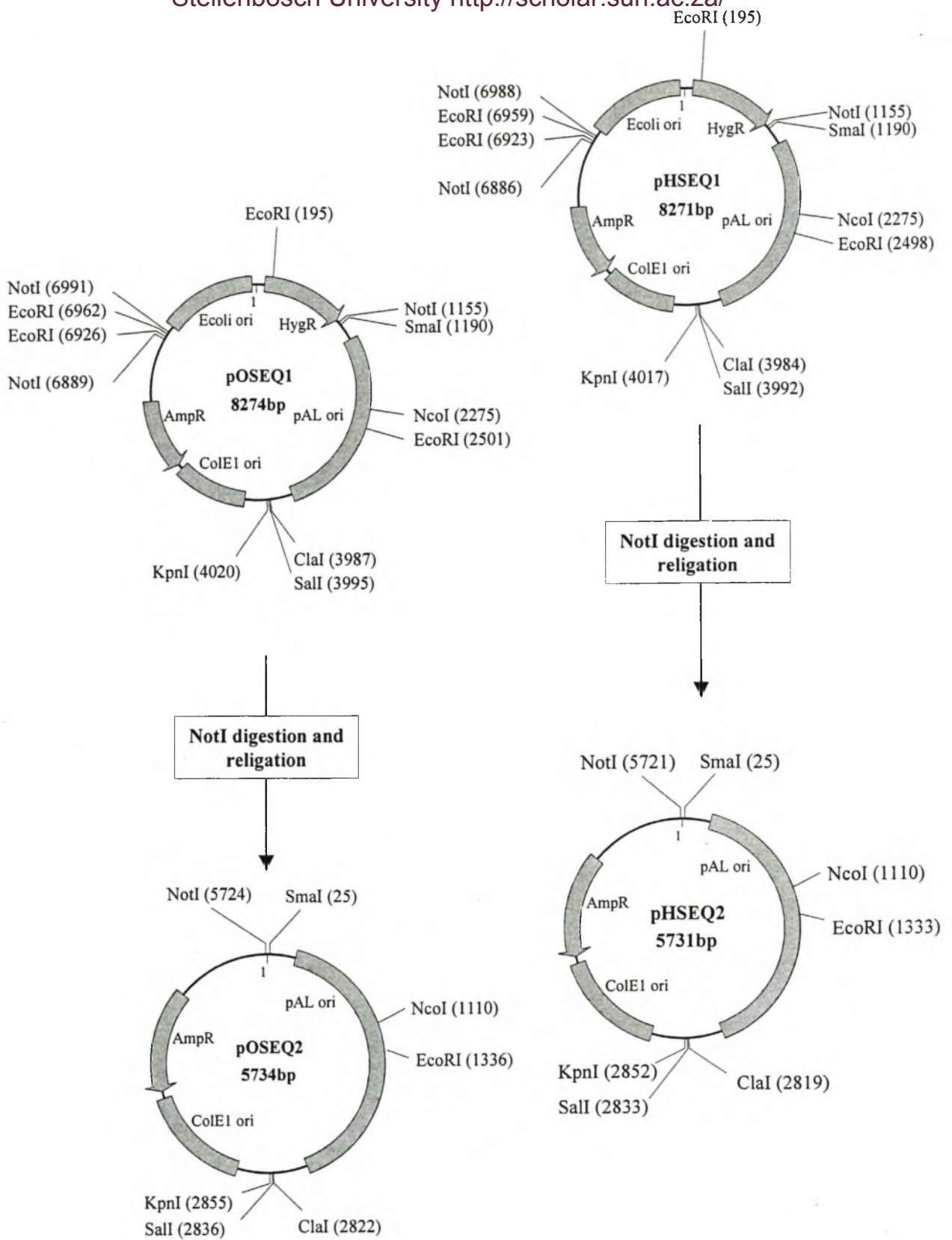


Figure 1.9(b) Construction of pOSEQ2 and pHSEQ2 as done by Dr. WR Bourn (personal communication).

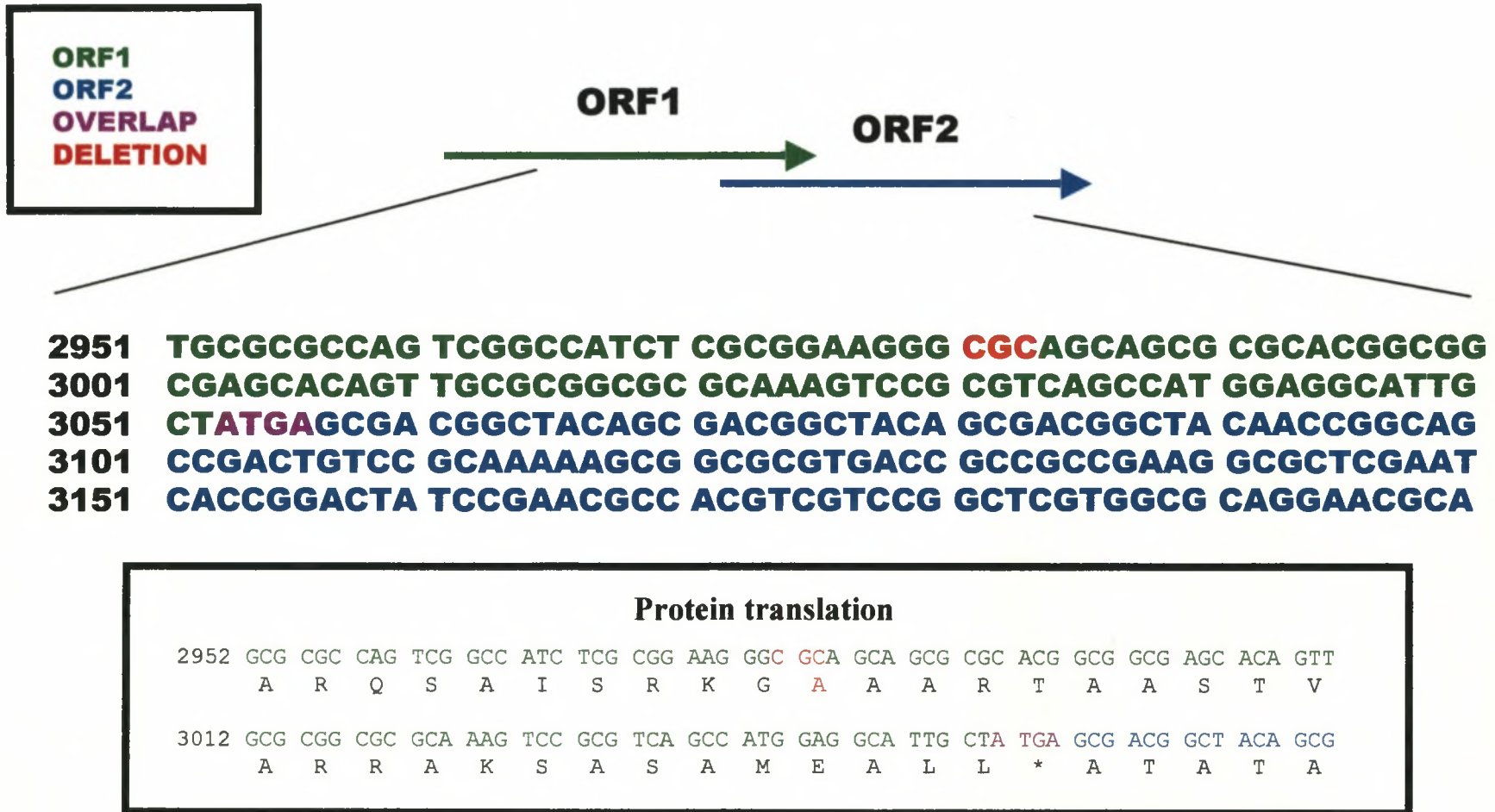


Figure 1.10 Part of the sequence of pORI, showing the 3bp deletion in pHIGH. The CGC in red is not present in the pHIGH sequence (the 3bp deletion). In the protein translation it can be seen that this deletion causes the loss of only 1 amino acid (an alanine).

1.5 Aims of project

The aims of this project were as follows:

1. To confirm that the 3bp deletion shown in the sequencing of Dr. Bourn (1.4.2) was not a sequencing artefact.
2. To confirm that the apparent increase in copy number was plasmid-encoded and not chromosomal.
3. To confirm that the 3bp deletion is the cause of the increase in copy number.
4. To confirm that other pAL5000-related plasmids can be converted to high copy number vectors.
5. To determine the relative increase in copy number.
6. To determine the exact copy number of the plasmid.
7. To determine whether this high copy number plasmid is stably inherited.
8. To conduct a preliminary investigation of the possible mechanisms for the increased copy number.

CHAPTER 2: Materials and Methods

2.1 Materials

2.1.1 Antibiotic stock solutions

Ampicillin (50mg/ml)

Purchased from Roche, dissolved in water, filter-sterilised and stored at -20°C.

Chloramphenicol (34mg/ml)

Purchased from Roche, dissolved in absolute ethanol and stored at -20°C.

Hygromycin (50 mg/ml)

As supplied by Roche, stored at 4°C.

Kanamycin (50mg/ml)

Purchased from Sigma, dissolved in water, filter-sterilised and stored at -20°C.

2.1.2 Buffers and Solutions

Denatured Herring sperm DNA

Herring sperm DNA was dissolved at a concentration of 5mg/ml in TE buffer (pH 8.0), sonicated in a Misonic Ultrasonic Processor XL, and denatured at 100°C for 10min followed by rapid cooling.

100 x Denhardt's Solution

1% BSA, 1% Ficoll, 1% polyvinyl pyrrolidone, filtered through a 0.45µM filter.

Developing solution (poly-acrylamide gels)

3% sodium carbonate, 0.056% formaldehyde and 8µM sodium thiosulphate.

Formamide hybridisation solution

50 % formamide, 5X SSPE, 0.2 % SDS, 0.1mg/ml denatured herring sperm DNA, 5X Denhardt's solution.

6X Loading buffer for agarose gels

0.25 % Bromophenol Blue, 15% Ficoll 400 in TE buffer; pH 8.0.

Lysozyme buffer

0.3M sucrose, 25mM EDTA, 25mM Tris; pH7.2.

MTB extraction buffer

267mM MSG, 50mM Tris-HCl, 25mM EDTA; pH 7.4.

Mycobacterial lysis buffer

2% SDS and 0.3M NaOH.

Neutralising solution (membrane stripping)

0.1X SSC, 0.1 % SDS, 0.2M Tris-HCl; pH 7.5.

Phenol/chloroform/isoamylalcohol

Phenol was dissolved in chloroform/isoamylalcohol (24:1), 3.44mM hydroxyquinoline was added and the mixture equilibrated with an equal volume 1M Tris (pH 8.0). After two changes of buffer, the phenol was equilibrated with water.

Phosphate buffered saline solution (PBS)

0.8% NaCl, 0.02% KCl, 0.144% Na₂HPO₄, 0.024% KH₂PO₄ dissolved in water, pH adjusted to 7.0 with HCl.

10X Proteinase K buffer

100mM Tris-HCl, 50mM EDTA, 5 % SDS; pH 7.8.

RNAse 10mg/ml (DNAse-free)

Purchased from Roche, dissolved in 10mM Tris-HCl (pH 7.5), 15mM NaCl, heated to 100°C for 15min, cooled slowly to room temperature, and stored as 1ml aliquots at -20 °C.

Sephadex G50 (medium)

Sephadex G50 (medium) was added to distilled sterile STE pH 8.0 (10g of dry powder yields 160ml of slurry). Autoclaved and stored at RT.

20X SSC (Sodium chloride-sodium citrate) buffer

3.0M NaCl, 0.3M trisodium citrate; pH 7.0.

20X SSPE (Sodium chloride-sodium phosphate-EDTA) buffer

3.0M NaCl, 0.2M NaH₂PO₄, 20mM EDTA; pH 7.4.

STE buffer (Sodium chloride-Tris-EDTA)

10mM Tris-HCl, 1mM EDTA, 0.1M NaCl; pH 8.0.

Solution 1

50mM Glucose, 25mM Tris-Cl (pH 8.0), 10mM EDTA.

Solution 2

0.2M NaOH, 1% SDS.

Solution 3

3M Kac (pH 5.2) and 11.5% acetic acid.

Staining solution (poly-acrylamide gels)

0.1% silver nitrate, 0.056% formaldehyde (added 5min prior to use).

10X TBE buffer (Tris-borate-EDTA)

0.89M Tris-HCl, 0.89M Boric acid, 20mM EDTA; pH 8.3.

TE buffer (Tris-EDTA)

10mM Tris-HCl, 1mM EDTA; pH 8.0.

2.1.3 Culture media

ADC (NOTE: Difco recipe)

5% BSA fraction V, 2% glucose, 40µg/ml catalase, 0.85% NaCl. Dissolved for 2h with stirring.

Filtered through pre-filter millipore AP25, 0.45µM Millipore HA filter, 0.22µM Millipore GS. Final filtration through 0.22µM Millipore GS filter for sterilisation.

LB (Luria-Bertani) broth

1% Biolab tryptone, 0.5% Biolab Yeast Extract, 1% NaCl, autoclaved.

LB (Luria-Bertani) agar

1.5 % Biolab agar bacteriological, in LB medium, autoclaved.

Middlebrooks broth (MB7H9)

Purchased from Difco and made according to the manufacturer's directions, with the addition of Tween 80 (0.024%) before autoclaving. After autoclaving (just before use), 10% sterile ADC was added.

Middlebrooks agar (MB7H11)

Purchased from Difco and made up according to the manufacturers directions, with the addition of glycerol (0.55%) before autoclaving. After autoclaving (just before use), 10% sterile OADC was added.

SOB medium

2.0 % Biolab tryptone, 0.5 % Biolab yeast extract, 0.05% NaCl and 2.5mM KCl (pH 7.0), autoclaved.

SOC medium

SOB medium to which 20mM (sterile) glucose and 10mM (sterile) MgCl₂ was added, immediately prior to use.

OADC (NOTE: Difco recipe)

0.05% (w/v) Oleic acid, 5% BSA fraction V, 2% glucose, 40µg/ml catalase, 0.85% NaCl. Dissolved for 2h with stirring.

Filtered through pre-filters Millipore AP25, 0.45µM Millipore HA filter, 0.22µM Millipore GS. Final filtration using 0.22µM Millipore GS filter for sterilisation.

2.1.4 Bacterial strains and plasmids

Table 2.1 Bacterial cells and strains used

Species	Strain / ATCC	Genotype	Source
<i>E. coli</i>	XL1-Blue	<i>recA1, endA1, gyrA96, thi-1, hsdR17, supE44, relA1, lac</i> [F' <i>proAB, lacI^fZΔM15Tn10</i> (Tet ^R)]	Purchased from Stratagene, California, USA
<i>E. coli</i>	JM109	<i>e14-(McrA-), recA1, endA1, gyrA96, thi-1, hsdR17(r_K⁻ m_K⁺), supE44, relA1, Δ(lac-proAB)</i> [F' <i>traD36, proAB, lacI^fZΔM15</i>]	Purchased from Stratagene, California, USA
<i>Mycobacterium smegmatis</i>	Cornell 3 ATCC 19420	-	Department of Medical Microbiology, University of Cape Town
<i>Mycobacterium bovis</i>	BCG Japanese ATCC 35737	-	Department of Medical Microbiology, University of Cape Town

Table 2.2 Vectors

Vector	Genotype	Source	Reference
pADM4	<i>XhoI</i> -fragment of <i>M.smegmatis</i> <i>polA</i> gene (inserted in <i>Sall</i> -site of pGEM3Z-f), Amp ^R	Valerie Mizrahi Department of Hematology, University of the Witwatersrand Medical School, Johannesburg, SA	Gordhan <i>et. al.</i> , 1996
pBluescript	ColE1 origin, fl origin, <i>lacZ</i> , Amp ^R , T3 and T7 promoters	Purchased from Stratagene, California, USA	Alting-Mees and Short (1988); Short <i>et. al.</i> (1989)
pJC86	<i>E.coli</i> origin of replication, pAL5000 origin of replication, kanamycin resistance	Gift from, Department Medical Microbiology, University of Cape Town, SA	Derived from pJC85; Beggs <i>et. al.</i> (1995)
pTV102	Multiple cloning site, promoterless chloramphenicol acetyl transferase gene, stop codons in all 3 reading frames (upstream of CAT gene), Hyg ^R	Gift from F. Da Silva- Tatley Department of Medical Biochemistry, University of Cape Town, SA	(unpublished, personal communication)
pUC19	ColE1 origin, <i>lacZ</i> , Amp ^R	Gibco BRL	Yanisch-Perron <i>et. al.</i> (1985)

2.1.5 Biochemical kits

Prime-It RmT random primer labelling kits

Purchased from Stratagene (California, USA).

Wizard PCR preps

Purchased from Promega (Wisconsin, USA).

Wizard plus mini- and midipreps DNA purification kits

Purchased from Promega (Wisconsin, USA).

2.1.6 Enzymes used for DNA modification

Shrimp alkaline phosphatase (SAP), T4 DNA ligase, T4 polymerase, and RNase A

Purchased from Roche (Indianapolis, USA).

Lysozyme (50mg/ml)

Purchased from Roche (Indianapolis, USA), dissolved in water and stored at -20°C.

Proteinase K (10mg/ml)

Purchased from Roche (Indianapolis, USA), dissolved in water and stored at -20°C.

Restriction endonucleases were purchased from Amersham (Buckinghamshire, UK), Roche (Indianapolis, USA) and Promega (Wisconsin, USA).

Polymerase chain reaction

Taq DNA polymerase was purchased from GibcoBRL (Carlsbad, California, USA).

2.1.7 Electrophoresis gels

Agarose gels

0.8 % to 2 % agarose in 1 X TBE.

Poly-acrylamide gels

5% Acrylamide ([29:1] Acrylamide:Bis-acrylamide), 0.125% Ammonium persulphate (APS) and 0.125% Temed in 1 X TBE.

2.1.8 Radiochemicals

[α^{32} P] dCTP (Specific activity ~ 3000Ci/mmol) was purchased from Amersham Life Sciences (Buckinghamshire, UK).

2.1.9 Zhiel-Nielsen staining

ZN carbol fuchin

1% Basic fuchin (pararosaniline chloride), 5% phenol, 10% absolute ethanol.

3% acid alcohol

3% HCl in 95% ethanol.

Methylene blue

0.3% methylene blue in water.

2.1.10 Oligonucleotides

Oligonucleotides were purchased from Integrated DNA Technologies, Inc. (Coralville, Iowa, USA) and are described in Table 2.3.

Table 2.3 Oligonucleotides

Name	Sequence	T_m^*
pHRe	5'-TAG TCC GGT GAT TCG AGC-3'	48.05°C
pHFo	5'-GAT CTA TGC CGA GTG CCA-3'	48.87°C
pBFo	5'-TGG TGA TGT CGA TCT GCC C-3'	57.9°C
pBRe	5' CGC CGA GAA CTT CAA GAC G-3'	56.6°C

*Theoretical $T_m = 4(G + C) + 2(A + T)$

2.2 Methods

The media, solutions and other materials used in the following procedures are described in section 2.1.

2.2.1 Culture conditions, monitoring and culture storage

2.2.1.1 *Escherichia coli*

For broth cultures, *E. coli* cells were grown in LB broth (2.1.3) with shaking at 37°C. Antibiotic solutions were added to the required concentration (given in text).

For cultures on solid medium, *E. coli* were grown on LB-agar plates (2.1.3), incubated at 37°C until colonies appeared. Antibiotic solutions were added to the required concentration (given in text).

Stock solutions of the cultures were made by aliquoting 200-300µl of culture into sterile eppendorf tubes. The solution was centrifuged briefly to create a soft pellet. The supernatant was removed and 200µl sterile 15% glycerol was added and the cells resuspended. The stocks were stored at -70°C.

2.2.1.2 *Mycobacterium smegmatis*

For broth cultures, *M. smegmatis* cells were grown in Middelbrooks 7H9 (2.1.3), shaking at 37°C. Antibiotic solutions were added to the required concentration (given in text).

The growth of the cultures was monitored by measuring absorbency at $\lambda=600\text{nm}$ using a MILTON ROY Spectronic 1201 spectrophotometer.

For cultures on solid medium, *M. smegmatis* was grown on MB7H11 agar plates (2.1.3) incubated at 37°C until colonies appeared. Antibiotic solutions were added to the required concentration (given in text).

Stock solutions of the cultures were made by aliquoting 200-300µl of culture into sterile eppendorf tubes. The solution was centrifuged briefly to create a soft pellet. The supernatant was removed and 200µl sterile 15% glycerol was added and the cells resuspended. The stocks were stored at -70°C.

2.2.2 Zhiel-Nielsen Staining and light microscopy

A single drop (10-20µl) of mycobacterial culture was smeared on a slide and allowed to air-dry. The cells were fixed by passing the slide three or four times through the blue cone of a Bunsen-burner flame. After cooling, the slide was flooded with ZN carbol fuchin (2.1.9) and heated with a flame until steam came off the surface. It was then left to cool (2-3min) and the heating step was repeated. The slide was rinsed with tap water and decolourised with 3% acid-alcohol (2.1.9) until no more stain came off. The slide was again rinsed with tap water and counter-stained with methylene blue (2.1.9) for 1-2min. Finally, the slide was rinsed with tap water, blotted dry and examined under a light microscope using the oil-immersion objective (1000x magnification). Under the microscope, the mycobacteria appear as red/pink stained rods or cocco-bacilli; non-acid-fast contaminants appear blue.

2.2.3 DNA isolation and purification methods

2.2.3.1 Total DNA extraction from mycobacteria

Mycobacterial broth cultures grown to stationary phase ($OD_{600} > 1$) were centrifuged at 10 000g for 10min at 4°C. The pellet was resuspended in 20ml MTB extraction buffer (2.1.2) and centrifuged at 12 000g for 10min at 4°C, after which the pellet was resuspended in 12ml of MTB extraction buffer (2.1.2). One hundred milligrams of

lysozyme powder was added and the cells were incubated for 5-6 hours at 37°C. Following this, 10µl Rnase (10mg/ml), 400µl ProteinaseK (10mg/ml) and 1.4ml ProteinaseK buffer (2.1.2) were added and the incubation was continued for a period of 14-18h at 45°C.

After the 45°C incubation, an equal volume of phenol/chloroform/isoamylalcohol [25:24:1] (2.1.2), was added, gently mixed, and centrifuged for 5min at 12 000g at 4°C. The aqueous phase was transferred to a new tube, and the DNA was precipitated with isopropanol (2.2.3.5). The sample was centrifuged for 5min at 12 000g (4°C), the supernatant removed and, without drying, the DNA pellet was resuspended in 1ml TE buffer. An equal volume of phenol/chloroform/isoamylalcohol [25:24:1] (2.1.2) was added, gently shaken by hand for 1min and then centrifuged at 13 000g for 5min. The aqueous phase was re-extracted with an equal volume of chloroform/isoamylalcohol [24:1]. This step was repeated before the DNA was precipitated with isopropanol (2.2.3.5). After removing the supernatant, the DNA pellet was washed with 100µl 70% ethanol, centrifuged at 13 000g for 5min, and the pellet air-dried for 5-10min before resuspending the DNA in 400µl TE buffer (2.1.2).

2.2.3.2 Small-scale isolation of plasmid DNA from *E. coli* cultures

i) Classic alkaline lysis

(Protocol adapted from that described by Sambrook *et. al.*, 1989)

A sample of 400µl of an overnight broth culture of *E. coli* (2.2.1) was centrifuged for 2min at 13 000g and the supernatant discarded. The cells were resuspended in 200µl Solution 1 (2.1.2), before 400µl freshly made Solution 2 (2.1.2) was added. The suspension was mixed by gently turning it upside down 3-6 times, and stored on ice for 5min. Following this, 300µl of Solution 3 (2.1.2) was added, mixed by gentle shaking, and the sample was stored on ice for 5min. The tube was centrifuged at 13 000g for 2min.

An equal volume of isopropanol was added to 600 μ l of the supernatant that had been transferred to a fresh tube. The solution was centrifuged at 13 000g for 10min after which the supernatant was removed. The DNA pellet was washed with 80 μ l of 70% ethanol, and centrifuged at 13 000g for 5min. The sample was air dried for 5-10min and then resuspended in 50 μ l TE buffer.

ii) Wizard Miniprep kit (Promega)

The kit was used as described by the manufacturer

2.2.3.3 Large-scale plasmid isolation from *E. coli* cultures

i) Isopycnic CsCl-Ethidium bromide density gradient centrifugation

(Method adapted from that described by Sambrook *et. al.*, 1989)

An overnight broth culture (100ml) of *E. coli* (2.2.1) containing the correct concentration of antibiotic (specified in text), was centrifuged at 15 000g for 10min at 4°C, and the pellet was resuspended in 5ml of Solution 1 (2.1.2), after 5min, 10 ml of Solution 2 (2.1.2) was added and the sample was placed on ice for 10min. Following this, 7.5ml of Solution 3 (2.1.2) was added and the sample was kept on ice for a further 10min. The sample was then centrifuged for 10min at 15 000g at 4°C. The supernatant was removed and subjected to isopropanol precipitation (2.2.3.5). The resulting pellet was resuspended in 4.5ml TE buffer.

To this was added 5.25g of CsCl and 250ml of ethidium bromide (10mg/ml). The refractive index was then adjusted to 3.692 by the addition of CsCl or TE buffer. The sample was loaded into a Beckman Quickseal centrifuge tube and centrifuged at 55 000rpm for 10-16h at 18°C in a Beckman Vti65 rotor. The plasmid DNA band was visualised under UV light ($\lambda=365\text{nm}$) and removed with a syringe and needle. The sample was shaken with NaCl saturated isopropanol and the organic phase was removed after the phases had separated. This was repeated until all the ethidium

bromide was removed from the aqueous phase. Two volumes of TE buffer was then added to the aqueous phase. The sample was subjected to isopropanol precipitation (2.2.3.5). The pellet was immediately resuspended (without air drying) in 500µl of TE buffer and again subjected to isopropanol precipitation (2.2.3.5). The resulting pellet was resuspended in 100µl of TE buffer.

ii) Wizard Midi-prep kit (Promega)

The kit was used as described by the manufacturer.

2.2.3.4 Phenol extraction

An equal volume of phenol/chloroform/isoamylalcohol [25:24:1] was added to the sample. It was mixed for 1min and then centrifuged at 13 000g for 5min. This was repeated until no impurities at the interface were visible. To the aqueous phase was added an equal volume of chloroform/isoamylalcohol, the sample mixed for 1min and then centrifuged for 3min at 13 000g. This chloroform wash was repeated twice.

2.2.3.5 Isopropanol precipitation

To aqueous solutions of DNA, a one-tenth volume of 3M sodium acetate was added and then an equal volume isopropanol. After mixing, the sample was centrifuged for 10min at 13 000g. The liquid was removed, and 70% ethanol added. After centrifuging for 5min at 13 000g the ethanol was removed and the sample left to air dry for a 5-10min. The DNA pellet was then dissolved in the appropriate volume TE buffer.

2.2.3.6 Ethidium bromide / high-salt extraction protocol for plasmid DNA

This method is based on that used by Stemmer (1991). DNA samples were made up to 250µl with TE buffer and transferred to a 2ml-epENDORF microfuge tube. Following this, 10µl ethidium bromide (10mg/ml) and 140µl 7.5M ammonium acetate were added to the DNA and mixed. An equal volume of phenol/chloroform/isoamylalcohol [25:24:1] was then added. The solution was vortexed for 20 seconds before centrifugation at 13 000g for 5min. The phenol extraction was repeated, following which an equal volume of chloroform/isoamylalcohol [24:1] was added. The sample was vortexed and centrifuged for 5min at 13 000g. After the chloroform/isoamylalcohol wash was repeated, the DNA was precipitated using isopropanol (2.2.3.5).

2.2.3.7 Removal of unwanted enzymes and EDTA

To remove unwanted enzymes and salts from a DNA sample/restriction reaction, it was made up to a final volume of 100µl with TE-buffer, and 16µg of tRNA was added as a carrier for the DNA. The sample was then subjected to phenol extraction (2.2.3.4) and isopropanol precipitation (2.2.3.5).

2.2.3.8 Wizard PCR preps

The Wizard PCR preps kit was used to isolate DNA fragments from agarose gels, and to purify PCR products from the PCR reaction mixture. The kit was used as described by the manufacturer.

2.2.4 DNA quantification: Spectrophotometry

Determination of DNA concentration was by measuring OD₂₆₀ in a MILTON ROY Spectronic 1201 spectrophotometer:

$$\text{OD}_{260} \times \text{dilution factor} \times 50\text{mg/ml} = [\text{DNA}] \text{ mg/ml}$$

(Sambrook *et. al.*, 1989)

2.2.5 Enzymatic treatment of DNA

2.2.5.1 Restriction endonuclease digests

i) Single restriction digests

For restriction digests utilising only one restriction enzyme, unless otherwise stated in the text, the procedure was as recommended by the enzyme manufacturer, except that 5U restriction enzyme was used per 1 µg DNA.

ii) Double restriction digests

In cases where the DNA was digested with more than one enzyme, both enzymes were added to the digest, only when both were active in the same buffer. If not, the digestion with the enzyme needing the lowest salt concentration was performed first, following which an appropriate volume of 0.1M NaCl, the second buffer and the second enzyme were added. All other conditions were as recommended by the manufacturer.

iii) Partial restriction digests

Digestions were as described in 2.2.5.1 (i), except that a range of enzyme: DNA ratios of 4U/ μ g to 0.0625U/ μ g in halving concentrations was used. All other conditions were as recommended by the manufacturer.

2.2.5.2 Ligation reactions

Prior to ligation, enzymes or salts were removed (2.2.3.4-2.2.3.5). Ligations were carried out in 20 μ l, using 5U ligase/ ligation reaction. Vector: insert ratios were 1:1 (where it was possible to establish DNA concentrations) and the final DNA concentration varied from 5ng/ μ l to 42.5ng/ μ l. Ligations were carried out at room temperature for 14-18h. All other conditions were as recommended by the manufacturer.

2.2.5.3 Removal of 5' and 3' overhanging DNA ends by T4 DNA polymerase

After digestion (2.2.5.1) and removal of unwanted enzymes and salts (2.2.3.4-2.2.3.5) the reaction was performed as described by the manufacturer. After incubation the reaction was again subjected to phenol extraction (2.2.3.4) and isopropanol precipitation (2.2.3.5).

2.2.5.4 Dephosphorylation of linearised plasmid DNA using shrimp alkaline phosphatase

The procedure was a variation of that recommended by the manufacturer. After restriction digestion the DNA was subjected to a phenol extraction (2.2.3.4) and isopropanol precipitation (2.2.3.5). The DNA was then resuspended in 48 μ l water. To this, SAP 6 μ l Buffer and 6 μ l shrimp alkaline phosphatase (6U) was added. The sample was incubated at 37°C for 1h. Then the sample was transferred to a new tube, 3 μ l (3U)

of shrimp alkaline phosphatase added, and again incubated for 30min at 37°C. After this, the sample was transferred to a new tube, and heat inactivated at 65°C for 30min.

2.2.5.5 Radioactive labelling of probes

All probes used were labelled using the Prime-It random primer labelling kits (2.1.5). The kit was used essentially as described by the manufacturer, with the exception that 250ng of DNA was used instead of 50ng, and the incubation period was extended from 10min to 1h. Immediately before adding the probe to the hybridisation solution, the probe was incubated at 95-100°C for 10min, in order to denature the DNA.

2.2.6 Polymerase chain reaction

Taq DNA polymerase (2.1.6) was used as described by the manufacturer with the following specific conditions: 5ng template DNA (pYJ200) was used in a total volume of 100µl reaction mix, containing 0.2mM dNTPs, 2mM MgCl₂, 1U Taq polymerase, and 50pmol of each primer (pBRe and pBFo).

The reaction mixture was cycled 30 times in a Perkin Elmer Gene Amp PCR system (2400) as follows: 95°C for 3min and then 56°C for 20s. After this, 30 cycles of 94°C for 30s, 56°C for 20s and 72°C for 30s were completed. This was followed by a 7min extension step at 72°C.

2.2.7 Electrophoresis and visualisation of the DNA in the gel

2.2.7.1 Agarose gel electrophoresis

Electrophoresis of DNA was in 1X TBE in a 0.7% - 2.0% (specified in text) agarose gel at 60V (approximately 2V/cm), except where otherwise specified in text. The gel was

stained for 20min with gentle agitation in 400ml 1X TBE solution containing 0.5µg/ml ethidium bromide. The DNA was visualised under UV illumination and photographed.

2.2.7.2 Polyacrylamide gel electrophoresis

Electrophoresis of DNA was in 1X TBE in a 1-1.5mm thick 5% polyacrylamide gel (2.1.7), at approximately 5V/cm.

Following the electrophoresis, the DNA was fixed by washing the gel, with gentle agitation, in 7.5% acetic acid at 37°C for 15min. It was then rinsed twice in water for 3min at 37°C. The silver staining solution (2.1.2) was pre-heated to 37°C. The gel was stained with this solution for 45min at 37°C. Before developing, the gel was rinsed briefly with water at ambient temperature (not longer than 10 seconds). The developing solution (2.1.2) was pre-chilled to -20°C prior to use. The gel was stained for 5-10min at ambient temperature. The reaction was stopped when optimal band development was obtained by rinsing the gel in cold 7.5% acetic acid for 5min at ambient temperature. The gel was then rinsed for 5min with distilled water at ambient temperature. Following this, the gel was either dried or sealed in a plastic sleeve (to prevent desiccation).

2.2.8 Southern transfer and hybridisation

Following agarose gel electrophoresis (2.2.7.1) the DNA in the gel was denatured by soaking it with gentle agitation in 0.4M NaOH for 20min, after which the gel was rinsed with distilled water.

The denatured DNA was Southern transferred by capillary action onto a nylon membrane (Hybond-N⁺, Amersham) using at least 500ml 20xSSPE (2.1.2), as described by Sambrook *et. al.* (1989). On completion of transfer (14-18h) the membrane was rinsed with 2X SSC (2.1.2), and the DNA was fixed to the membrane

by baking it for 2h at 80°C. The membrane was rehydrated with distilled water prior to pre-hybridisation.

To prevent non-specific binding, membranes were pre-hybridised at 42°C for 6 hours with 0.1ml per cm² of formamide hybridisation solution (2.1.2). After 6 hours, the pre-hybridisation solution was discarded. Fresh formamide hybridisation solution (0.1ml/cm²) containing 5-20ng of heat denatured ³²P-labelled probe (specified in text) was added, and hybridisation was allowed to continue for 16 hours at 42°C.

On completion of hybridisation the membranes were washed twice at room temperature in 2X SSC (2.1.2) containing 0.1% SDS and then twice at 55-65°C with 2X SSC containing 0.1% SDS, for 30min at a time. The membranes were then sealed in plastic bags and exposed to autoradiographic film overnight at -70°C. Intensifying screens were not used, in order to maintain a linear relationship between radioactive dose and photographic exposure.

If a higher stringency wash was required to reduce the background signal, the blot was incubated again in 0.2X SSC containing 0.1% SDS, at 65°C.

2.2.9 Stripping blots

In order to remove bound probe, the Hybond N⁺ membranes were washed at 45 °C in 250 ml 0.4M NaOH for 45min. The stripping solution was then decanted and the membranes were neutralised by incubation in neutralising solution (2.1.2) for 15min at 45 °C with gentle shaking.

2.2.10 Production of competent cells

2.2.10.1 Heat-shock competent cells (*Escherichia coli*)

A culture of *E. coli* was inoculated at 0.1% from frozen stocks into 10ml LB broth and incubated shaking for 14-18h at 37°C such that the culture was in stationary phase. Two 100ml pre-warmed, pre-aerated LB broth flasks were inoculated at 1% from the stationary phase culture and incubated shaking at 37°C until mid-log phase (OD₆₀₀ = approximately 0.5) was reached. The flasks were chilled on ice for 20min, and all subsequent steps were performed at 4°C. The cultures were centrifuged at 1 200g for 5min in a pre-cooled rotor. The resulting pellets were pooled and resuspended in 20ml ice cold 0.1M MgCl₂. The suspensions were centrifuged again at 1 200g for 5min, and the pellets were resuspended in 20ml ice cold 0.1M CaCl₂. The centrifugation step was repeated and the pellets resuspended in 1ml ice cold CaCl₂. The cells were left on ice for 20min before 100µl aliquots of the cell suspension were dispensed into pre-cooled eppendorf microfuge tubes.

These cells were used immediately (2.2.11.1).

2.2.10.2 Electro-competent cells (*Escherichia coli*)

E. coli (XL1-Blue) was inoculated at 0.1% into 50ml LB broth containing tetracycline (50µg/ml) and incubated with shaking for 14-18h at 37°C. Two 500ml pre-warmed, pre-aerated LB broth flasks were inoculated at 1% from the stationary phase culture and incubated at 37°C (with shaking) until OD₆₀₀ = 0.5-0.8.

The flasks were chilled on ice for 15-20min and all subsequent steps were performed at 4°C. Each culture was centrifuged in pre-cooled centrifuge bottles for 20min at 4 000g. The supernatant was removed, the pellets resuspended in 250ml ice cold sterile distilled water and again centrifuged for 20min at 4 000g. The resulting pellets were resuspended in 125ml ice-cold sterile distilled water, pooled, and centrifuged again (20min at 4 000g). Each pellet was resuspended in 10ml ice cold, sterile 10% glycerol.

These were pooled and then centrifuged again as before. The resulting pellet was resuspended to a final volume of 2-3ml in ice cold sterile 10% glycerol. Aliquots of 100 μ l were frozen at -70°C.

2.2.10.3 Electro-competent cells (*Mycobacterium smegmatis*)

M. smegmatis was inoculated at 0.1% from frozen glycerol stocks (2.2.1.2) into 100ml Middlebrooks 7H9+ADC, and grown to stationary phase (approximately 2 days). Four 200ml cultures were inoculated at 0.1% from this starter culture (100 μ l inoculum/100ml media), and grown for 14-18h shaking at 37°C to an OD₆₀₀ of approximately 0.5 (mid-log phase). The flasks were then chilled on ice for 1h, and all subsequent steps were performed at 4°C. The cultures were centrifuged at 10 000g for 5min and resuspended in 30ml ice-cold 10% glycerol (w/v). This step was repeated twice. The resulting pellet was resuspended in 4ml 10% glycerol. Aliquots of 220 μ l were frozen at -70°C.

2.2.11 Transformation of competent cells

2.2.11.1 Heat-shock (*Escherichia coli*)

The DNA sample (not > 10 μ l total volume) was added to 100 μ l freshly prepared heat-shock competent cells (2.2.10.1) and left on ice for 30min. The cells were heat shocked on a heating block at 37°C for 4min, following which 600 μ l LB broth was added. This was incubated for 45min at 37°C after which 100 μ l volumes were spread on LB-agar plates containing the appropriate antibiotic(s). The plates were incubated 16-18h at 37°C.

2.2.11.2 Electro-transformation (*Escherichia coli*)

Sterile 0.2cm electroporation cuvettes, solutions and the electroporation chamber slide were pre-chilled on ice. The electro-competent cells (2.2.10.2) were also placed on ice to thaw. The DNA used for the procedure was phenol-washed (2.2.3.4), precipitated with isopropanol (2.2.3.5), and dissolved in less than 5 μ l sterile water.

The DNA was added to 40 μ l of the competent cells and placed on ice for 1min. Electroporation was with a Bio-Rad Gene Pulser at a voltage of 2.5kV, a capacitance of 125 μ F, a secondary capacitance of 25 μ F and a resistance of 200 Ω . This gave a time constant of approximately 3-4. One millilitre of SOC medium (2.1.2) was added to the contents of the cuvette and mixed. The transformed cells were transferred to sterile 2ml eppendorf microfuge tubes and incubated with shaking at 37°C for 1h. Samples of 100 μ l of the transformed cells were plated on LB-agar plates containing the appropriate antibiotic(s).

2.2.11.3 Electro-transformation (*Mycobacterium smegmatis*)

Sterile 0.2cm electroporation cuvettes, solutions and the electroporation chamber slide were pre-chilled on ice. The electro-competent cells (2.2.10.3) were also placed on ice to thaw. The DNA used for the procedure was phenol-washed (2.2.3.4), precipitated with isopropanol (2.2.3.5), and dissolved in less than 5 μ l sterile water.

The DNA was added to 40 μ l of the electro-competent cells, transferred to an ice-cold electroporation cuvette and placed on ice for 1min. Electroporation was with a BioRad Gene Pulser set at a voltage of 2.5kV, a capacitance of 125 μ F, a secondary capacitance of 25 μ F and a resistance of 1 000 Ω . This gave a time constant of approximately 17-19. One millilitre of Middlebrooks 7H9+ADC (2.1.3) was added and the transformed cells were transferred to a sterile 2ml eppendorf microfuge tube. This was incubated with shaking at 37°C for 2-3h. Samples of 100 μ l of the transformed cells were plated on Middlebrooks 7H11 agar plates + OADC, containing the appropriate antibiotic. The plates were incubated at 37°C for 3-5 days.

2.2.12 Mycobacterial protein assay

Mycobacterial protein assays were based on the method described by Meyers *et. al.* (1998). Duplicate 1ml samples of mycobacterial cultures were taken and centrifuged at 13 800g for 5min. The samples were then rotated once through 180° and centrifuged again for 5min to get a compact cell pellet. The pellets were washed in 1ml phosphate buffered saline (PBS) (2.1.2), without resuspending the cells, and again centrifuged for 5min. The liquid was poured off and the samples frozen at -20°C until analysis.

The cell pellets were resuspended in 100µl 1M NaOH, sealed in 1.5ml eppendorf microfuge tubes, and heated to 95-100°C for 10min. The samples were neutralised by the addition of 20µl 5M HCl, and the volume adjusted to 1ml by adding 880µl PBS (pH 7.0). After centrifugation for 30min at 13 800g, 800µl of the supernatant was removed and used for the determination of protein concentration. The absorbance of each sample was measured at both 230nm and 260nm and the protein concentration (µg/ml) determined as follows:

$$[\text{PROTEIN}] = (183 \times A_{230}) - (75.8 \times A_{260})$$

The assay is linear over the range of 6 to 225µg of protein per ml. Heavily turbid cultures were diluted in PBS until the measurements were in the linear range.

2.2.13 Sequencing

Sequencing was performed at the Core DNA sequencing facility of the University of Stellenbosch. Dye terminator sequencing with BigDye Ver 2 was done, using a ABI3100 genetic analyser by Applied Biosystems.

CHAPTER 3: Characterisation of mutation and mutant plasmids

3.1 Introduction

There are many advantages to having a high copy number plasmid vector in bacterial cells. One advantage is greater ease of manipulation, because the plasmid DNA is present in a much higher concentration. Higher expression of a gene cloned into such a high copy number plasmid is also usually achieved. The stability of the plasmid is also likely to be increased, since it would probably be partitioned into both daughter cells during bacterial cell division. This is because a plasmid that is present in a high copy number is present throughout the whole bacterial host cytoplasm. Upon cell division, the probability that plasmids will be in both daughter cells is much higher than with a low copy number plasmid, independent of a specific partitioning system. For these reasons, the mutant high copy number pAL5000-based plasmid, pHIGH, created by Dr. Bourn (1.4.1) is a potentially useful molecular biological tool.

It is also important for researchers to know the characteristics of the molecular biological tools that they are using. These include firstly, the copy number and stability of the plasmid, because certain procedures (such as expression of cloned genes for commercial production) may require a stable high copy number vector (for reasons mentioned earlier) rather than a low copy number vector that is unstable. For experimental considerations, it is also important to know whether a high copy number is putting a load on the cell, causing the bacterial cells to grow slower.

The mutant plasmid (pHIGH) has, however, only been partially characterised. It is therefore important to determine the exact mutation causing the increase in copy number of the plasmid, the stability of the mutant in its host (*M. smegmatis*), how this mutation works, whether the mutation can be used to convert other low copy number plasmids to a higher copy number, and what exactly the increase in copy number is.

3.2 Experimental approach

All materials and standard methods used are described in Chapter 2.

3.2.1 Retransformation of *M. smegmatis* with pORI and pHIGH

E. coli [pHIGH] and *E. coli* [pORI] (supplied by Dr. Bourn) were grown in LB liquid culture containing 100µg/ml hygromycin (2.2.1.1). Plasmid DNA was isolated from these cultures (2.2.3.3.ii) and the resulting plasmid DNA used to transform electrocompetent *M. smegmatis* cells (2.2.10.3) by electroporation (2.2.11.3). Transformants were selected on Middlebrooks 7H11 agar (2.1.3) containing 100µg/ml hygromycin. A single transformant was selected in each case and inoculated into a 100ml liquid culture containing 100µg/ml hygromycin (2.2.1.2). From these cultures total DNA was isolated (2.2.3.1) and the concentration assayed (2.2.4). A 3µg sample of each was digested with *EcoRI* (2.2.5.1) and then halving dilutions were made with 1X restriction buffer. In addition, a 250ng sample of the original plasmids pORI and pHIGH, was digested with *EcoRI* (2.2.5.1). The samples were subjected to electrophoresis on a 0.7 % agarose gel (2.2.7.1). The gel was stained in 1X TBE solution containing 0.5µg/ml ethidium bromide and the DNA visualised under UV illumination and photographed (2.2.7.1).

3.2.2 Sequencing of the region containing the mutation

E. coli [pJC86] and *E. coli* [pJCX] were cultured in LB-broth (2.2.1.1) containing 100µg/ml hygromycin, and plasmid isolations were performed using a Wizard midi-prep kit (2.2.3.3.ii). This DNA was used to sequence both strands (2.2.13) of the region containing the mutation using the primers pHFo and pHRe (2.1.10).

3.2.3 Subcloning of the mutation into a kanamycin resistance encoding vector

The plasmid pJC86 (2.1.4) was partially digested with restriction endonuclease *EcoRI* (2.2.5.1). The digested DNA was separated by electrophoresis on a 0.7% agarose gel (2.2.7.1) and the linearised fragment was purified from the gel using Wizard PCR prep kit (2.2.3.8). The plasmid pHSEQ2 (1.4.2) was totally digested with *EcoRI* (2.2.5.1) and the two fragments were ligated (2.2.5.2) and the ligation mixture was used to transform heat-shock competent *E. coli* (JM109) (2.2.10.1 & 2.2.11.1). The transformants were selected on LB agar (2.1.3) with 100µg/ml ampicilin and 50µg/ml kanamycin (pJC86 carries a kanamycin resistance gene and pHSEQ2 carries an ampicilin resistance gene). Twelve colonies were selected and the plasmids were isolated (2.2.3.2). Restriction digests with *KpnI* and *NcoI* were conducted (2.2.5.1) and the DNA was separated on a 0.7% agarose gel (2.2.7.1). The desired construct, pJC104, was identified on that basis. Plasmid DNA was isolated on large scale using the Wizard midi-prep kit (2.2.3.3) and further restriction mapping confirmed the correct ligation of the two fragments.

Plasmid pJC104 DNA was then partially restriction endonuclease digested with *NcoI* (2.2.5.1) and the DNA separated on a 0.7% agarose gel (2.2.7.1). Using Wizard PCR prep kit, the 5.2kb fragment was isolated from the gel and the DNA purified (2.2.3.8). A ligation reaction was set up with this DNA in order to re-circularise the plasmid (2.2.5.2), and the mix was then used to transform heat-shock competent *E. coli* (JM109) cells (2.2.10.1 & 2.2.11.1). The cells were plated on LB containing 50µg/ml kanamycin. Transformants were subjected to small-scale plasmid isolation (2.2.3.2). The plasmids were mapped with restriction endonuclease enzymes: *EcoRI*, *XhoI* and *BglII* (2.2.5.1). The correct plasmid was called pJCX and was isolated on large scale using Wizard midi-prep kit (2.2.3.3). *M. smegmatis* was transformed with both pJC86 and pJCX by electroporation (2.2.10.3 & 2.2.11.3) for further analysis.

3.2.4 Analysis of copy number phenotype of pJCX

Total chromosomal DNA was isolated from *M. smegmatis* [pJC86] and *M. smegmatis* [pJCX] cultures (2.2.3.1), grown in MB7H9 broth containing 100µg/ml kanamycin (2.2.1.2). The DNA was digested with *EcoRI* (2.2.5.1) and then purified using the ethidium bromide clean-up method (2.2.3.6), after which the concentration was determined (2.2.4). Samples of 3µg each of total chromosomal *M. smegmatis* [pJC86] and *M. smegmatis* [pJCX] DNA were again restriction digested with *EcoRI* (2.2.5.1), halving dilutions were made in TE-buffer, and the DNA was then separated on a 0.7% agarose gel (2.2.7.1). As controls, 250ng samples of pure pJC86 and pJCX DNA were digested with *EcoRI* (2.2.5.1) and also separated on the agarose gel.

3.2.5 Comparing antibiotic resistance levels of pJC86 and pJCX

3.2.5.1 Initial tests – solid media

M. smegmatis [pJC86] and *M. smegmatis* [pJCX] were cultured in MB7H9 broth containing 100µg/ml kanamycin up to stationary phase (2.2.1.2). This was then used to reinoculate fresh MB7H9 broth (also containing 100µg/ml kanamycin). These cultures were grown to mid-log phase (2.2.1.2).

Dilutions of 10^{-5} of these cultures were made in MB7H9 and spread on MB7H11 plates (2.1.3) containing a range of kanamycin concentrations from 0µg/ml up to 2000µg/ml.

3.2.5.2 Checking for chromosomal mutants – liquid culture

M. smegmatis cells were freshly transformed with pJCX and pJC86 (2.2.10.3 & 2.2.11.3) and selected on MB7H11 plates containing 20µg/ml kanamycin. Single colonies were isolated and used to inoculate liquid cultures (MB7H9) containing 10µg/ml kanamycin (2.2.1.2). Then 200µl of these cultures were used to inoculate a second series of 20ml liquid MB7H9 cultures containing kanamycin at different

concentrations between 0 and 1000µg/ml. This was also done for plasmid free *M. smegmatis*.

3.2.6 Direct determination of the copy number

3.2.6.1 Initial blotting approach

Total chromosomal DNA was isolated from cultures of *M. smegmatis* [pORI] and *M. smegmatis* [pHIGH] (2.2.3.1) grown in MB7H9 media (containing 100µg/ml hygromycin in the cultures with pHIGH and pORI). As before (3.2.4), the DNA was first digested with *EcoRI* (2.2.5.1), and then purified using the ethidium bromide method (2.2.3.6). The concentration was determined and then a 0.375µg sample of *M. smegmatis* [pHIGH] and *M. smegmatis* [pORI] was aliquoted and digested again with *EcoRI* (2.2.5.1). Halving dilutions were made in 1x restriction buffer. A total of 0.6ng of pure pHIGH plasmid DNA was digested with *EcoRI* (2.2.5.1) and $1/3$ dilutions were made in 1 x restriction buffer. All the samples were separated on a 0.7% agarose gel (2.2.7.1).

The agarose gel was subjected to Southern transfer (2.2.8) and the Hybond-N⁺ membrane was probed using the plasmid pUC19 (2.1.4), which had been radioactively labelled with $\alpha^{32}\text{P}$ dCTP by random priming (2.2.5.5).

3.2.6.2 Determining the relative change in copy number

Total DNA was isolated (2.2.3.1) from cultures of *M. smegmatis* [pORI] and *M. smegmatis* [pHIGH], grown in MB7H9 containing 100µg/ml hygromycin. A sample of 5µg DNA of each was restriction endonuclease digested with *SmaI* (2.2.5.1.). Halving dilutions of these digests were made in 1x restriction buffer and the DNA separated on a 0.7% agarose gel (2.2.7.1). The gel was subjected to Southern transfer (2.2.8) and the first probe used was the 600bp gel-purified fragment of pYJ200

(3.2.6.3) labelled with $\alpha^{32}\text{P}$ dCTP by random priming (2.2.5.5). The blot was then stripped (2.2.9) and probed with radioactively labelled pBluescript (2.2.5.5).

3.2.6.3 Constructing pYJ200

The plasmid pHIGH was linearised with *EcoRV*, and pADM4 was digested with *SmaI* (2.2.5.1). The ends of the pHIGH fragments were dephosphorylated (2.2.5.4). The linearised pHIGH was then ligated with the *SmaI* fragments from pADM4 (2.2.5.2) and used to transform *E. coli* (2.2.10.2 & 2.2.11.2). Transformants were selected on LB agar (2.1.3) containing 100 $\mu\text{g/ml}$ hygromycin. Plasmid isolations were performed (2.2.3.2) and the DNA was digested with *SmaI* (2.2.5.1) and separated on a 1.5% agarose gel (2.2.7.1) in order to confirm successful insertion of pADM4 *SmaI* fragments. A control *SmaI* restriction endonuclease digest of pHIGH plasmid DNA was also included. The resulting plasmid (pYJ200) was used to transform *M. smegmatis* by electroporation (2.2.10.3 & 2.2.11.3).

3.2.6.4 Probe 1, using pYJ200

E. coli [pYJ200] was cultured in LB (2.1.3) containing 100 $\mu\text{g/ml}$ hygromycin (2.2.1.1). A large-scale plasmid isolation, using isopycnic CsCl-ethidium bromide density gradient centrifugation (2.2.3.3.i), was performed and the resulting plasmid DNA was used to make the probe for the Southern blot:

A sample of 50 μg of pYJ200 DNA was restriction endonuclease digested with *SmaI* (2.2.5.1) and separated on a 1.5% agarose gel (2.2.7.1). The smallest band (approximately 600bp), corresponding to the *M. smegmatis* chromosomal fragment insert, was purified from the gel using the Wizard PCR prep kit (2.2.3.8). This fragment was radio-labelled by random priming using $\alpha^{32}\text{P}$ -dCTP and Prime-It RmT-kit (2.2.5.5), and used as a probe in subsequent Southern blotting.

3.2.6.5 Probe 2, using pYJ200

Using DNAMAN sequence analysis computer software (Lynnon Biosoft), primers (2.1.10) were designed for PCR. The primers were such that upon PCR a 214bp fragment homologous to bp1361-bp1574 on pYJ200 would be generated. The primers (pBRe and pBFo) were used in a PCR, as described in section 2.2.6, using the isopycnic purified pYJ200 (3.2.6.4) as target DNA. The 100µl sample was purified using the Wizard PCR prep kit (2.2.3.8). A sample of 200ng of the purified sample was radio-labelled (2.2.5.5) and used as probe for the Southern blotting.

3.2.6.6 First approach, using pYJ200

Total chromosomal DNA was isolated from *M. smegmatis* [pYJ200] cultures (2.2.3.1), grown in MB7H9 media containing 100µg/ml hygromycin and also from *M. smegmatis* cultures, grown in MB7H9 media without selection. A sample of 8µg of each was restriction endonuclease digested with *SmaI* (2.2.5.1). Halving dilutions of the DNA containing pYJ200 were made in 1 x restriction buffer, and separated, together with the *M. smegmatis* sample, on a 1.5 % agarose gel (2.2.7.1). As controls, *SmaI* digests of pHIGH (200ng), pADM4 (200ng) and pYJ200 (200ng) DNA were also separated on the gel. This gel was subjected to Southern transfer (2.2.8). The probe used on this blot is described in section 3.2.6.4.

3.2.6.7 Second approach, using pYJ200

Total chromosomal DNA was isolated from *M. smegmatis* [pYJ200] cultures (2.2.3.1), grown in MB7H9 media containing 100µg/ml hygromycin and also from *M. smegmatis* cultures, grown in MB7H9 media without selection. A sample of 8µg of each was double digested with *BamHI* and *BglII* (2.2.5.1). Halving dilutions of the restriction endonuclease digested *M. smegmatis* [pYJ200] DNA were made in 1 x restriction buffer, and separated, together with the restriction digested *M. smegmatis* DNA, on a 0.7 % agarose gel (2.2.7.1). As controls, pADM4 (200ng) and pYJ200 (200ng) DNA,

double digested with *Bam*HI and *Bgl*III (2.2.5.1), were also separated on the gel. After electrophoresis, the gel was subjected to Southern blotting (2.2.8). The probe used on this blot is described in section 3.2.6.5.

3.2.7 Stability tests

M. smegmatis [pORI] and *M. smegmatis* [pHIGH] were inoculated at 0.2% from frozen glycerol stocks (2.2.1.2) into MB7H9 containing 100µg/ml hygromycin. The cultures were grown up to stationary phase (2.2.1.2) and used as starter cultures. From the starter cultures 100ml cultures of MB7H9 (without hygromycin) were inoculated to an approximate OD₆₀₀ of 3 x 10⁻⁸. This figure was calculated from the dilution of the inoculum required to result in an OD₆₀₀ of 1 (stationary phase) after approximately 25-30 generations. The cultures were then incubated until stationary phase was reached.

Fresh cultures (MB7H9 without hygromycin) were inoculated from these, and again incubated for approximately 25-30 generations. This process was repeated until the bacterial cells had passed through approximately 200 generations. Zhiel-Nielsen staining (2.2.2) was performed with every inoculation to make sure there was no contamination. The OD₆₀₀ was determined to calculate the inoculum for the next round of growth.

The following formula was used with both the protein concentration and OD₆₀₀ readings:

$$\text{Generation number} = \frac{(\ln x - \ln y)}{\ln 2}$$

x = end concentration (culture after ±25 generations); y = start concentration (freshly inoculated culture)

From every culture that was used, 1ml samples were taken in duplicate as samples to be used for protein assays (2.2.12) to determine the generation number by an alternative method. The final cultures (after approximately 200 generations) were used to make 10⁶-fold dilutions in MB7H9 (2.1.3), and spread onto MB7H11 plates (2.1.3) with

(hygromycin at 100µg/ml) and without selection. The colony counts of these plates were used to determine the percentage of plasmid-carrying cells.

3.2.8 Growth curves

Starter cultures of *M. smegmatis*, *M. smegmatis* [pORI] and *M. smegmatis* [pHIGH] were inoculated into MB7H9 media (2.1.3). The media for *M. smegmatis* [pORI] and *M. smegmatis* [pHIGH] contained 100µg/ml hygromycin. These were grown to stationary phase ($OD_{600} \geq 1.0$) and then used to inoculate 200ml cultures of MB7H9 (no selection) to an approximate OD_{600} of 0.1. The cultures were incubated (2.2.1.2) for 10h and then samples were taken. Samples of 1ml were taken hourly thereafter for OD_{600} reading. This was repeated until the OD_{600} reached approximately 1.0.

3.2.9 Subcloning the mutation into a promoter probe vector

The plasmids pOSEQ2, pHSEQ2 (1.4.2) and pTV102 (2.1.4) were used to construct the promoter test vectors. The plasmids pOSEQ2 and pHSEQ2 carry the original pAL5000 origin of replication and the high copy number mutation thereof, respectively. These plasmids were originally constructed in order to sequence the origin of replication in search of the mutation (Bourn, personal communication).

Plasmid pTV102 was restriction endonuclease digested with *HindIII* and pOSEQ2 and pHSEQ2 were restriction endonuclease digested with *NcoI* (2.2.5.1). The 5' overhanging ends of the linearised plasmids were converted to blunt ends (2.2.5.3) and ligated (2.2.5.2), pTV102 to pOSEQ2 and pHSEQ2 respectively. *E. coli* XL1-Blue cells were transformed with the ligation mixture (2.2.10.2 & 2.2.11.2), and transformants were selected on LB-agar (2.1.3) containing 100µg/ml hygromycin and 50µg/ml ampicillin. Ten colonies were selected and the plasmids isolated (2.2.3.2). The plasmid DNA was double digested with *PvuII* and *KpnI* (2.2.5.1), and separated on a 0.7% agarose gel (2.2.7.1). The restriction maps of the different plasmid isolated were used to determine which of the constructs had the inserts in the correct orientation.

The constructs with the insert in the correct orientation were partially digested with *EcoRI* (2.2.5.1) and separated on a 0.7% agarose gel (2.2.7.1). A Wizard PCR prep kit (2.2.3.8) was used to purify the band of 9.9kb, containing the pTV102 plasmid ligated to the desired fragments of pOSEQ2 and pHSEQ2, from the gel. The purified DNA was used in recircularisation ligation reactions (2.2.5.2) that were then used to transform *E. coli* XL1-Blue (2.2.10.2 & 2.2.11.2). Transformants were selected on LB agar (2.1.3) containing 100µg/ml hygromycin. The plasmid consisting of pTV102 with the insert from pOSEQ2 was called pTVO and the one with the insert from pHSEQ2 was called pTVH. The plasmids were isolated (2.2.3.2) and then restriction endonuclease mapped with *EcoRI* (2.2.5.1). The digests were run on a 0.7% agarose gel (2.2.7.1), and on a 5% polyacrylamide gel (2.2.7.2).

These plasmids were then further restriction mapped using the following enzymes in single digests: *DraI*, *EcoRI*, *KspI*, *PstI*, *PvuII*, *SacI*, *XhoI* (2.2.5.1), as well as double digests with *DraI* & *PstI*, *KspI* & *SacI* and *EcoRI* & *XhoI* (2.2.5.1). Digests were run on polyacrylamide gels in cases where very small fragments needed to be mapped (2.2.7.2).

A positive control for the CAT-activity test was constructed by restriction endonuclease digestion of total *M. smegmatis* chromosomal DNA with *MboI* (2.2.5.1). The plasmid pTV102 was digested with *BamHI* (2.2.5.1), and ligated to the chromosomal DNA fragments (2.2.5.2). The ligation reaction was then electroporated into *M. smegmatis* cells (2.2.10.3 & 2.2.11.3), and spread onto MB7H11 plates (2.1.3) containing 100µg/ml hygromycin and 40µg/ml chloramphenicol. Twenty of these clones were toothpicked onto a MB7H11 plate containing 150µg/ml chloramphenicol. A single clone of those that grew on the plate was selected and called pTVC. This plasmid was restriction mapped to check for the presence of an insert.

3.2.10 Testing the chloramphenicol-acetyl transferase activity

M. smegmatis [pTV102], *M. smegmatis* [pTVO], *M. smegmatis* [pTVH] and *M. smegmatis* [pTVC] were inoculated from frozen glycerol stocks at 0.2% into

MB7H9 (2.2.1.2) containing 100 μ g/ml hygromycin and grown up to stationary phase. This was used as a starter culture to inoculate new cultures. A series of 10ml cultures (MB7H9) with a range of increasing chloramphenicol concentration was inoculated to the same concentration for each of the starter cultures. These were incubated until the chloramphenicol-free cultures reached stationary phase (2.2.1.2). Samples of 1ml were taken for OD₆₀₀ readings to determine the amount of growth in each culture.

3.3 Results and discussion

3.3.1 Confirmation that mutant high copy phenotype is plasmid encoded

From the results of Dr Bourn (1.4.1) it would appear that a high copy number mutant pAL5000 plasmid had been isolated. Furthermore, the mutation that caused this increase in copy number had been tentatively identified (1.4.1). However, it was formally possible that the apparent increase in copy number might be a result of a mutation on the bacterial chromosome. In order to prove that the mutation causing the increase in copy number is indeed plasmid encoded, fresh, plasmid free *M. smegmatis* cells were transformed with pORI and pHIGH plasmid DNA. Total DNA from the resulting transformants was restriction endonuclease digested and separated by electrophoresis (3.2.1).

The results are shown in figure 3.1. From this figure it can be seen that the high copy number phenotype is maintained in the newly transformed *M. smegmatis* [pHIGH]. On the basis of light intensity of the background (chromosomal DNA) lanes containing similar amounts of DNA can be compared. It can be seen that in lane 2 (*M. smegmatis* [pORI]) the chromosomal DNA is approximately equal to that of lane 8 (*M. smegmatis* [pHIGH]). There is obviously a higher concentration of plasmid DNA in the lanes containing *M. smegmatis* [pHIGH] than in the lanes with *M. smegmatis* [pORI]. It was therefore proven that the mutation that caused the increase in copy number was on the plasmid.

A relative increase in copy number of between 8 and 16 can be estimated from the ethidium bromide stained agarose gel. The band intensity of the pORI DNA fragments in lane 2 appears to be equivalent to the band intensity of the equivalent pHIGH DNA fragments between lanes 11 and 12. However, on the agarose gel, the chromosomal background cannot be eliminated, and this is particularly problematic in estimating band intensity for low copy number plasmids. In order to get a more accurate copy number estimate, a Southern blot, in which chromosomal background is largely eliminated, is required.

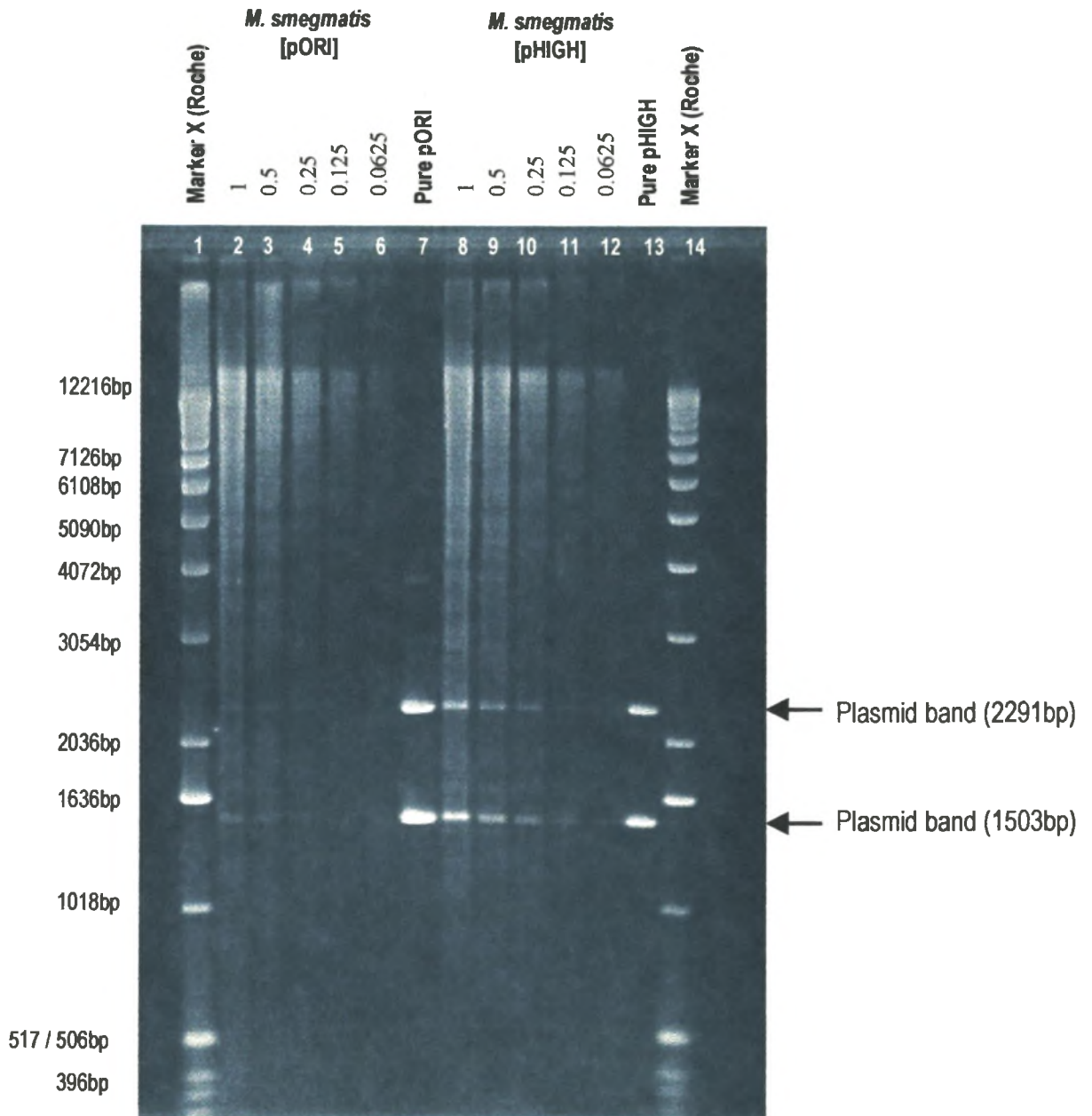


Figure 3.1 Ethidium gel photograph of *M. smegmatis* [pORI] and *M. smegmatis* [pHIGH] (dilution series). The total chromosomal DNA and the plasmid DNA have been restriction digested with *EcoRI*. The dilutions of each lane is given above the gel photograph. Experimental details are given in the text (3.2.1).

3.3.2 Conversion of another vector to high copy number

While it was shown above that the increased copy number phenotype is plasmid encoded (3.3.1), the identity of the mutation that caused this had yet to be confirmed. While the work of Dr Bourn had suggested that the causative mutation is a 3bp deletion (1.4.1), it remained possible that this was a sequencing artefact as only one strand of the DNA had been sequenced. Furthermore, only a limited region on the replicon was thus sequenced, so the possibility that there were active mutations outside of this region could not be excluded.

In order to confirm the nature of the mutation and to demonstrate that it was indeed the cause of the increased copy number, the following approach was taken. Plasmid pJC86 was used because it is totally different from pORI and pHIGH, apart from the fact that it also carries the pAL5000 origin of replication (Fig. 3.2). It also has a different selective marker (kanamycin). In order to subclone the 3bp deletion into pJC86, a two-step cloning strategy had to be implemented (section 3.2.3) to replace the wild-type *NcoI-EcoRI* fragment with the mutant *NcoI-EcoRI* fragment of pHIGH.

The plasmids pJC86 and pHSEQ2 (1.4.2) were linearised and the fragments ligated. The transformants were selected on kanamycin and ampicillin, and by restriction mapping, the clone that had the fragments in the correct orientation (pJC104), was selected. This clone was then partially digested, recircularised and selected on kanamycin (Fig. 3.2). Again restriction mapping was used to select the correct clone (pJCX).

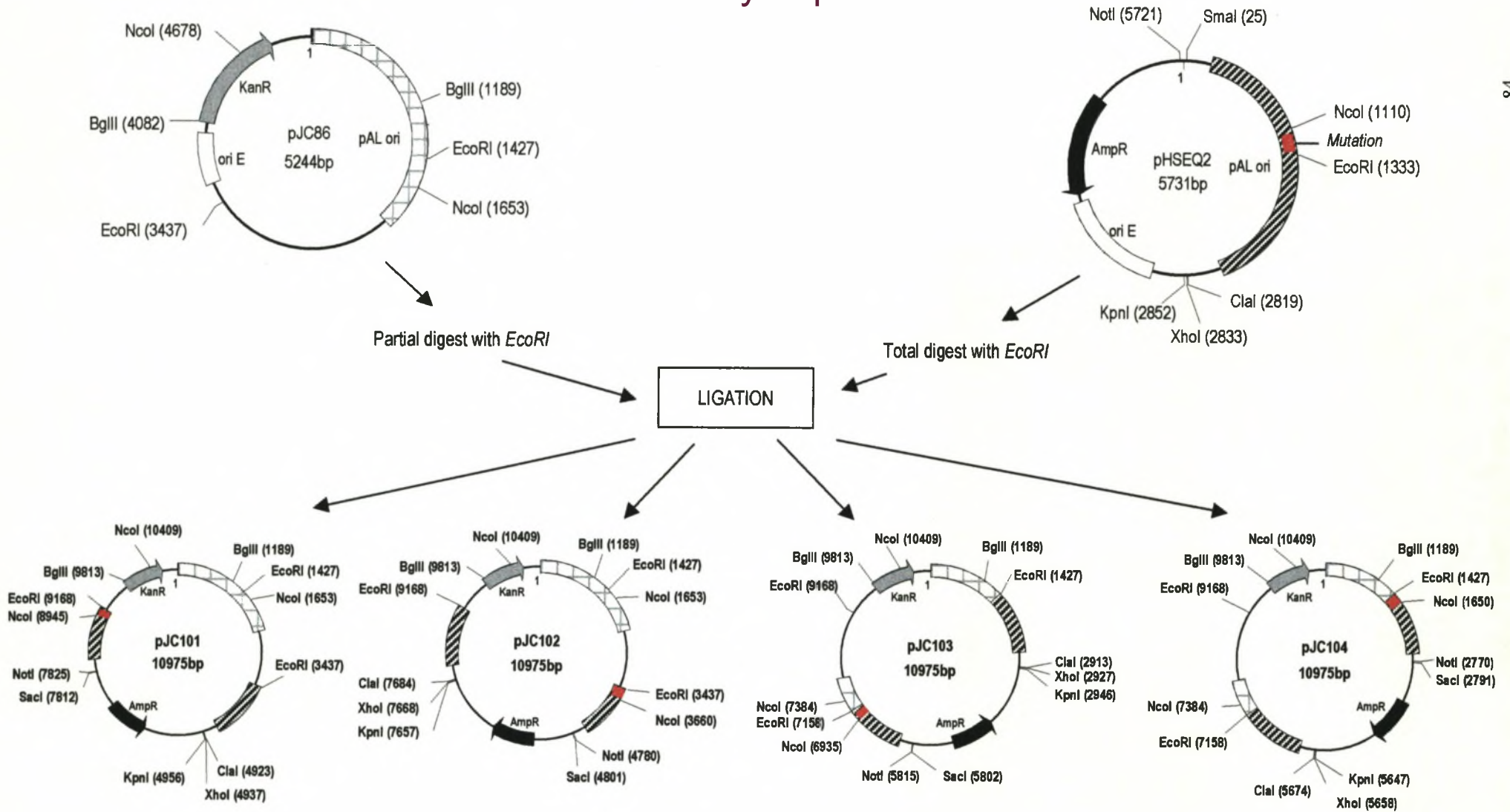


Figure 3.2(a) Ligation of pJC86 and pHSEQ2, making pJCX. Experimental details are given in the text (3.2.3).

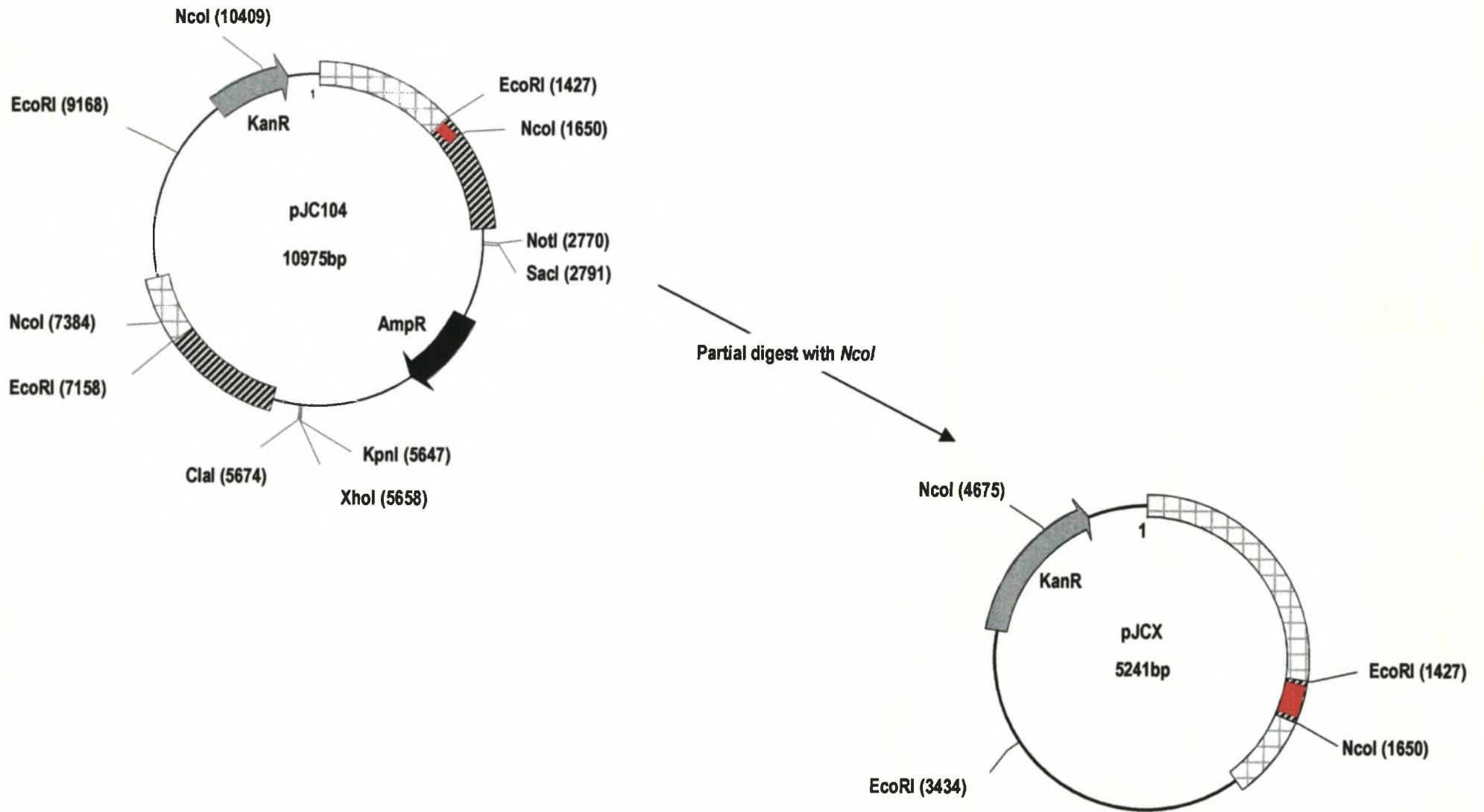


Figure 3.2(b) Making of pJCX. Experimental details are given in the text (3.2.3).

The reasons for this cloning strategy being adopted were threefold. Firstly, by subcloning the mutation into a different vector that carried a different selectable marker, it was ensured that cross-contamination during the cloning was not possible (as opposed to cloning it into pORI). Secondly, the two-step cloning strategy removes the possibility of the parental pJC86 being recreated during cloning. This is important in view of the fact that pJC86 and pJCX are indistinguishable on the basis of restriction mapping. Finally, by creating a kanamycin resistant high copy number vector, it was thought that it would be possible to measure the copy number using the method of Stolt and Stoker (1996a) and then compare the results with previously published work.

Following the subcloning, both strands of the *NcoI-EcoRI* fragment were sequenced on pJC86 and on pJCX. (3.2.2). This confirmed two things. Firstly, it was shown that the mutation did indeed consist of the 3bp deletion described by Dr Bourn. Secondly it was shown that the mutation was successfully subcloned into pJCX. The only difference between the plasmids pJC86 and pJCX is therefore the 3 bp deletion and with this confirmed, *M. smegmatis* was transformed with the two plasmids by electroporation (2.2.10.3 & 2.2.11.3) for further analysis.

Total DNA was isolated from cultures of *M. smegmatis* [pJC86] and *M. smegmatis* [pJCX], restriction endonuclease digested and separated on an agarose gel (3.2.4). The results are shown in Figure 3.3. On this gel it can be seen that the lanes containing *M. smegmatis* [pJCX] has a higher concentration of plasmid DNA than the lanes containing *M. smegmatis* [pJC86], in cases where the chromosomal background in the lanes are approximately the same (compare lanes 2 and 9). Again, from this, a copy number increase of greater than 8-fold can be estimated from the gel. This not only proved that the high copy number is maintained, but also demonstrated that the mutation could convert another plasmid with the same origin of replication into a higher copy number vector. This implies that one can convert other cloning vectors (promoter traps, expression vectors, shuttle vectors, etc) to a high copy phenotype by a simple cloning step. The mutation-carrying fragment should therefore prove to be extremely useful to other researchers.

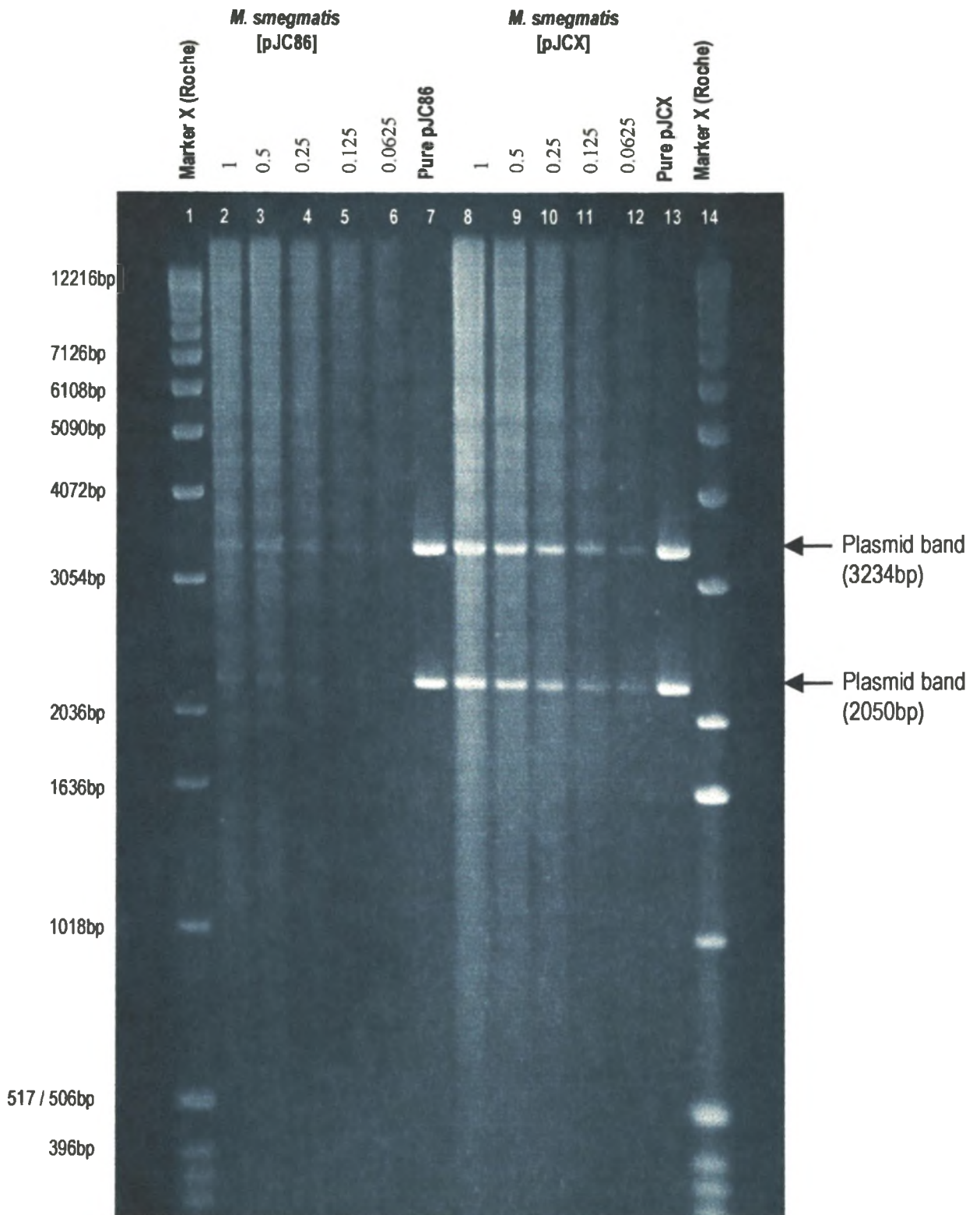


Figure 3.3 Ethidium gel photograph of *M. smegmatis* [pJCX] and *M. smegmatis* [pJC86] (dilution series). The total chromosomal DNA and plasmid DNA have been restriction digested with *EcoRI*. The dilutions of each lane is given above the gel photograph. Experimental details are given in the text (3.2.4).

3.3.3 Plasmid copy number determination

3.3.3.1 Single cell resistance test

A secondary advantage in the construction of pJCX was that the new high copy number vector was kanamycin resistant. It was expected that this would prove useful in the next stage of the investigation, involving a single cell resistance test. The single cell resistance test is a method in which the copy number of a plasmid that carries an antibiotic resistance gene, is estimated on the basis that higher numbers of plasmid will generate resistance to higher concentrations of antibiotic. Having the mutation in a kanamycin resistant plasmid, it would also be possible to compare the new copy number test results with those of Stolt and Stoker. Stolt and Stoker (1996a) used single-cell resistance (SCR) to kanamycin (as described by Nordström, 1993) to compare the relative copy numbers of their plasmid pYUB12 (see section 1.3.3) and certain mutants thereof. It was decided to repeat the SCR experiment of Stolt and Stoker in order to compare the relative copy numbers of pJC86 (wild type) and pJCX (mutant) in *M. smegmatis*.

Cultures of *M. smegmatis* [pJC86] and *M. smegmatis* [pJCX] in mid-log phase were diluted 1:10⁵ and spread on MB7H11 plates containing kanamycin in a range of concentrations (3.2.5.1). In order to relate the resistance level to the copy number, it is assumed that there is a linear gene dosage effect. Thus, an increase in resistance would be due to an increase in the number of kanamycin resistance genes, which in turn reflects the increase in number of plasmids. If the gene dosage effect is linear, then the copy number increase can be calculated.

Initially, colony counts obtained when using 0.9% (w/v) NaCl as a diluent, as described by Nordström (1993) and used by Stolt and Stoker (1996a), were erratic. In view of this, a comparison test was performed between using 0.9% NaCl and MB7H9 as diluent.

Two 1ml samples of a culture of *M. smegmatis* [pJC86] were diluted 10⁴ times, one in MB7H9 and the other in 0.9% (w/v) NaCl. Three MB7H11 plates (containing

100µg/ml kanamycin) were spread with 100µl of each sample. The dilutions were then left at ambient temperature for 1h and then 100µl of each sample was again spread on another three MB7H11 plates (containing 100µg/ml kanamycin). The plates were incubated at 37°C for 3 days, after which the colonies on each were counted. The results obtained are shown in Table 3.1.

Table 3.1 Plate counts using different diluents

Time	Middlebrooks 7H9		0.9% (w/v) NaCl	
	0h	1h	0h	1h
Plate #1	1100	1263	815	493
Plate #2	1390	1305	839	556
Plate #3	1208	1328	853	476
Average	1233±147	1299±34	836±19	508±42

From this it can be seen that not only are the initial (time = 0) counts different but in the case of NaCl as a diluent, the decrease continued over time. It was concluded that, as predicted, the 0.9% (w/v) NaCl solution caused the mycobacterial cells to clump, especially after prolonged incubation. This was probably because no detergent such as Tween 80 was included in the solution. Alternatively the NaCl could somehow be making the cells non-viable. Thereafter MB7H9 was used as the dilution media for all cultures. The SCR test was thus repeated.

Initially, a preliminary test to determine the range of kanamycin concentration to be used was performed. Therefore only a single series of plates containing different concentrations of kanamycin was used. It was found that even at a kanamycin concentration of 2mg/ml (Fig. 3.4), the cells still survived. Furthermore there was no evidence of any decrease in viability of the cells with increasing kanamycin concentration.

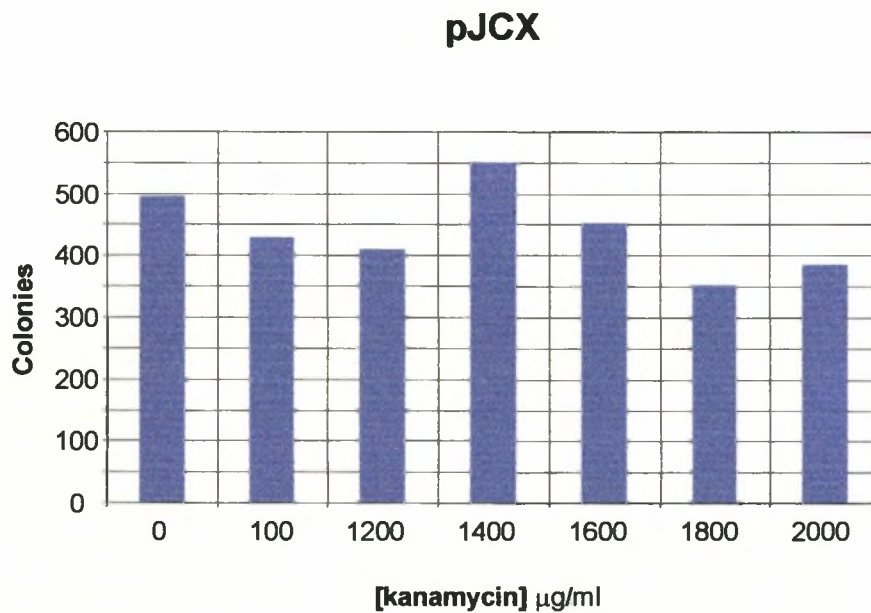
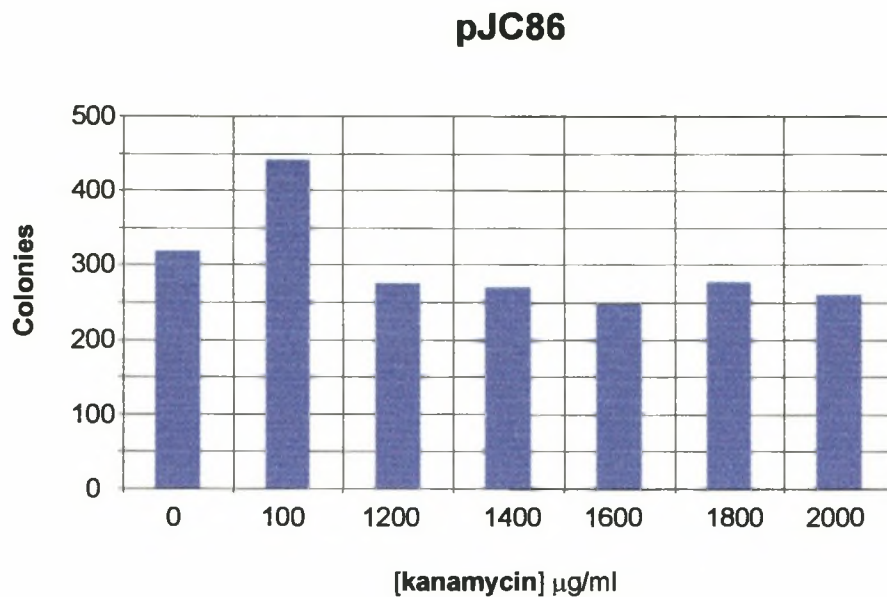


Figure 3.4 Single cell resistance tests on *M. smegmatis* cells carrying the plasmids pJC86 and pJCX.

In the case of the SCR tests performed by Stolt and Stoker (1996a), the *M. smegmatis* cells failed to survive on plates containing kanamycin above a concentration of approximately 600µg/ml. Although the same kanamycin resistance gene (from Tn5) that is carried by the plasmids used by Stolt and Stoker is also carried by pJC86 and pJCX, it must be remembered that there are differences between pJC86 and the plasmids used by Stolt and Stoker (1996a). It is possible that there are unknown promoters upstream of the kanamycin resistance gene in pJC86 and pJCX and this would cause an increase in expression of the kanamycin resistance gene, resulting in a higher resistance levels.

An alternative possibility was that the kanamycin concentration in the starter cultures was too high. If growth was inhibited (despite the presence of a kanamycin resistant plasmid), it could result in the selection of kanamycin resistant chromosomal mutants. To investigate this possibility, *M. smegmatis* cells were freshly transformed with pJC86 and pJCX, and cultured from single colonies. These cultures were grown in media containing only 10µg/ml kanamycin (3.2.5.2), and were then used to determine the MIC of the plasmids in liquid culture (3.2.5.2).

It was shown that the MIC for plasmid-free *M. smegmatis* was 2µg/ml, whereas *M. smegmatis* [pJC86] and *M. smegmatis* [pJCX] continued to grow in concentrations of up to 1000µg/ml.

It has subsequently been shown (Dr. W. Bourn, personal communication) that different kanamycin resistant plasmids generate very different MIC's, even if they carry the same selective marker and origin of replication. The reason for this is unknown. Therefore, in order to compare results with that of Stolt and Stoker, it would be necessary to introduce the mutation into the plasmid pYUB12 (Stolt and Stoker, 1996a), which was used by them.

As this method was unsuccessful in determining the change in copy number, a different approach was decided upon.

3.3.3.2 Initial blotting approach

As *M. smegmatis* [pJC86] and *M. smegmatis* [pJCX] had an extremely high resistance to kanamycin, no useful information could be obtained from the SCR test (3.3.3.1). Therefore it was decided to use a system whereby the copy number could be determined directly. A method involving Southern blotting was chosen, because densitometric scanning of an autoradiograph to compare band intensities is more accurate than when using an ethidium bromide stained agarose gel. This is for two reasons. Firstly, when Southern blots are used, only the band of interest gives a signal and the background chromosomal DNA signal is largely eliminated. This is particularly important when measuring band intensities for low copy number plasmids. Secondly the signal to band intensity ratio is linear (up to a point at which the signal becomes saturating) for a radioactive signal (provided intensifier screens are not used) whereas this is not the case for light signals (information from KODAK, X-ray film manufacturers).

In the experiment the size of the *M. smegmatis* chromosome (Baess, 1984), the size of the plasmids pORI and pHIGH, and the amount of DNA loaded on the gel are all known. Therefore it is theoretically possible to determine the copy number of pHIGH and pORI by densitometric scanning of an autoradiograph on which plasmid bands appear. In order to do this, known quantities of total DNA (chromosomal and plasmid) and known quantities of pure plasmid are separated on the gel. By comparing plasmid band intensities and from the knowledge of the quantity of DNA present in each lane, the copy number can be determined.

Total chromosomal DNA from *M. smegmatis* [pORI] and *M. smegmatis* [pHIGH] were restriction endonuclease digested, halving dilutions were made and the DNA fragments were separated on a 0.7% agarose gel. Pure pHIGH DNA was digested with the same enzyme and $\frac{1}{3}$ dilutions of the sample were run on the same gel. The Southern blot of this gel was then probed with pUC19, which hybridised with the plasmid bands in all the lanes (3.2.6.1). This was repeated several times with erratic results. A typical example is shown in Figure 3.5A.

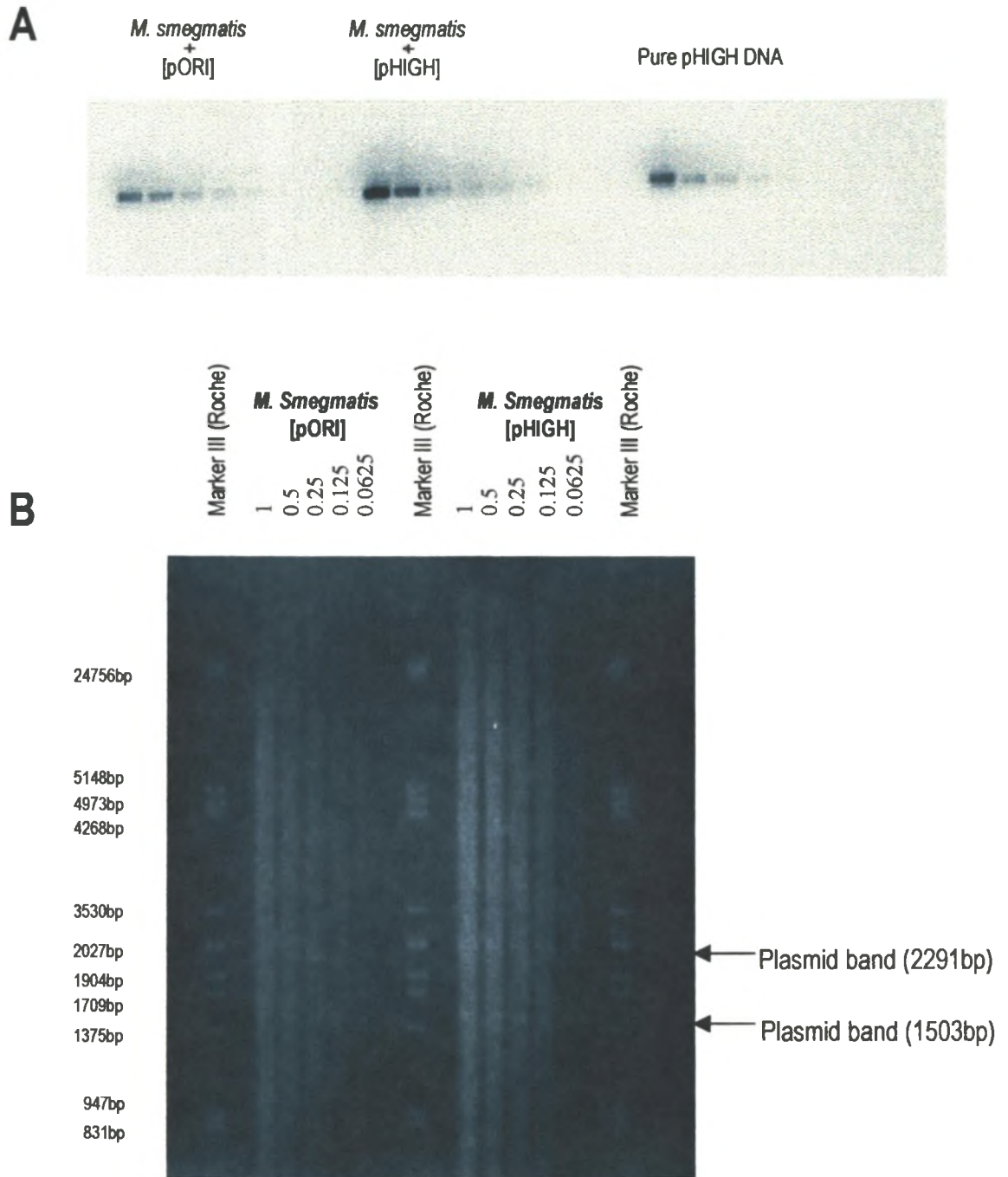


Fig. 3.5 A. Autoradiograph of the first approach in determining the copy number of pORI and pHIGH.

B. Ethidium gel photograph of *M. smegmatis* [pORI] and *M. smegmatis* [pHIGH] (dilution series). The total chromosomal DNA has been restriction endonuclease digested with *EcoRI*.

In Figure 3.5A it can be seen that the results on the autoradiograph are completely different from that expected from the gel in Figure 3.1. The band intensities for pORI (in this case) appear to be roughly equivalent to those of pHIGH. In other cases different results were obtained. No photograph of the gel is shown, because the low concentrations of DNA used in these Southern blots were not clearly visible on the gel photographs.

Subsequently, further gels, using higher concentrations of DNA (as in section 3.2.1), were run and the DNA visualised by ethidium bromide staining. A typical example can be seen in Figure 3.5B. In this approach, the total amount of DNA in each lane containing *M. smegmatis* [pORI] and *M. smegmatis* [pHIGH] has to be accurately known. From Figure 3.5B it can be seen that the intensity of the total DNA in the lanes containing *M. smegmatis* [pORI] and *M. smegmatis* [pHIGH] differed, although the same amount of 3µg total DNA of each were digested and used to make the dilution series. The DNA concentration was determined using spectrophotometry (2.2.4), and according to this, the same amounts of DNA were digested and loaded in the lanes. However, on the gel-photograph (Fig.3.5B) it is clear that different amounts of total DNA were loaded on the gel. This can also be seen in the case of *M. smegmatis* [pJC86] and *M. smegmatis* [pJCX] (Fig. 3.3), and again, to some degree, for *M. smegmatis* [pORI] and *M. smegmatis* [pHIGH] as shown in Figure 3.1.

Even in cases of gels where the concentrations in the corresponding lanes appeared the same, it was felt that the DNA quality contributed a significant variable. Even in cases of the same intensity on the gel, there is still no guarantee that the absolute measurement of absorbency is correct, as derived by spectrophotometry.

In order to try and rectify this, different methods were used:

Extra phenol and Rnase steps were incorporated to further clean the DNA, because the A_{260}/A_{280} measurements were variable, and any contaminants would interfere with the determination of concentration when using spectrophotometry (2.2.4). Ethidium bromide quantitation of DNA was also attempted (Sambrook *et. al.*, 1989). As this was not successful an ethidium bromide cleaning method (2.2.3.6) was used in order to remove any extra proteins that might still be bound to the DNA. Another approach was to restriction endonuclease digest a large amount of the relevant total DNA with *EcoRI*

before determining the concentration by spectrophotometry. The correct amount of DNA was then aliquoted, and then again digested with *EcoRI* before separating the DNA on the gel. This was to both ensure total digestion of the DNA, and to minimize pipetting errors due to using too viscous, high concentration solutions, consisting of large fragments of DNA (the examples shown in Figure 3.5 were processed in this manner). Instead of spectrophotometry, ethidium bromide staining was used to quantify the DNA (Sambrook *et. al.*, 1989). Other members of this laboratory attempted to clean the DNA using CETAB, but this was also unsuccessful (W. Bourn, personal communication). Similar difficulties in quantification of chromosomal DNA from *M. tuberculosis* (although not as marked) have been reported by other members of this laboratory (Dr. R. Warren, personal communication). It has also been reported that mycobacterial DNA samples are often contaminated with free lipoglycans and polysaccharides, which can make them unsuitable for particular molecular biological techniques (Belisle and Sonnenberg, 1998).

Thus a new approach, in which it would not be necessary to know the exact amount of DNA in each lane, had to be utilised.

3.3.3.3 Determination of relative change in copy number

By performing a Southern transfer of total chromosomal DNA and using a probe that is able to hybridise with a *M. smegmatis* chromosomal fragment, it is theoretically possible, by densitometry, to determine the ratio of amounts of *M. smegmatis* [pORI] and *M. smegmatis* [pHIGH] total DNA that were loaded on the gel. From this, a correction factor can be calculated. By then using a probe that is able to hybridise with the plasmid on the same Southern blot, the plasmid bands can be compared and the correction factor used to determine the relative change in copy number between pORI and pHIGH.

In order to pursue this approach, and conduct subsequent experiments (see below, section 3.3.3.4), a vector had to be constructed that contained a fragment of *M. smegmatis* chromosomal DNA cloned in the pHIGH vector. The plasmid pADM4 was used as source of a *M. smegmatis* chromosomal fragment. This plasmid was

originally constructed by insertion of an *XhoI* fragment that carries the *polA* gene of *M. smegmatis* (Gordhan *et. al.*, 1996) into the *SalI* site of the pGEM-3Zf (Promega) cloning vector (V. Mizrahi, personal communication).

The cloning of the fragment is described in section 3.2.6.3. Plasmid pADM4 DNA was restriction endonuclease digested with *SmaI*, and pHIGH DNA was linearised at a unique *EcoRV* site. The *SmaI* fragments of pADM4 were ligated to the linearised pHIGH DNA (Fig. 3.6). The resulting clones were restriction endonuclease digested with *SmaI*, and separated on a 0.7% agarose gel. One plasmid gave a restriction pattern where the only difference with that of pHIGH was an extra band of ± 600 bp (pYJ200). This fragment carried a *SmaI* fragment from pADM4, and this was confirmed by subsequent restriction mapping and Southern blotting (see below and Fig. 3.9).

Total *M. smegmatis* [pORI] and *M. smegmatis* [pHIGH] DNA were digested with *SmaI*, separated on a 0.7% agarose gel and subjected to Southern transfer (3.2.6.2). The first probe used was a 600bp gel-purified *SmaI* fragment of pYJ200 (3.2.6.4). Initially this gave very high background on Southern blots, probably due to contamination of the gel-purified fragment with chromosomal DNA (although this was deliberately minimized by using CsCl gradient prepared plasmid, rather than a commercial kit), and fragments from other parts of the plasmid (data not shown). In order to rectify this, the experiment was repeated using a probe that was generated by PCR (3.2.6.5).

This probe hybridised with the *M. smegmatis* chromosomal *polA* fragment, present in each lane containing total chromosomal DNA and very little background was observed. The autoradiograph was densitometrically scanned and, using ScionImage, the band intensities were plotted. The amount of background on the autoradiograph was determined and subtracted from the area under each curve (Fig. 3.7). Two graphs were subsequently drawn with the area under the curve (one for *M. smegmatis* [pORI] and one for *M. smegmatis* [pHIGH]) plotted against the dilution of the sample.

The graphs were used to determine a correction factor for the amount of chromosomal DNA loaded on the gel. In order to correctly compare the pORI plasmid DNA with that of the pHIGH plasmid DNA, the amount of total chromosomal DNA in that lane has to

be equal. As this is clearly not the case (Fig. 3.5), a correction factor is necessary to correct the ratio calculated for the pORI and pHIGH plasmid DNA. The correction factor is determined by the ratio of the slopes of the graphs (Fig. 3.7a).

$$\begin{aligned} M. smegmatis \text{ [pORI]} : M. smegmatis \text{ [pHIGH]} &= 1070 : 790 \\ &= 1 : 1.4 \end{aligned}$$

This means that 1.4 times more total DNA was used in *M. smegmatis* [pORI] lanes than in the *M. smegmatis* [pHIGH] lanes.

Following this, the blot was stripped and then the second probe (pBluescript) was used on the Southern blot. In this case, the hybridisation was with the *E. coli* origin on pORI and pHIGH. Therefore the bands on the resulting autoradiograph were representative of the amount of plasmid in each lane.

Again the autoradiograph was densitometrically scanned and a graph drawn in the same manner as for the chromosomal bands. The background was again subtracted as with the chromosomal bands. The ratio between the slopes of the graphs for pORI and pHIGH gives the ratio between the copy number of pORI and pHIGH (Fig. 3.7b):

$$M. smegmatis \text{ [pORI]} : M. smegmatis \text{ [pHIGH]} = 1500 : 13600$$

However, the slope of *M. smegmatis* [pORI] has to be corrected first, by multiplying it by 1.4. Therefore the correct ratio between pORI and pHIGH = 2100 : 13600

$$= 7$$

This means that pHIGH has a copy number of 7 times higher than pORI in *M. smegmatis*.

These figures are, by nature of the experiment, an estimation, as are all estimations of copy number. This is because experiments to determine copy number are all based on certain assumptions. In this case, it is assumed that the plasmid and chromosomal DNA are extracted with equal efficiency.

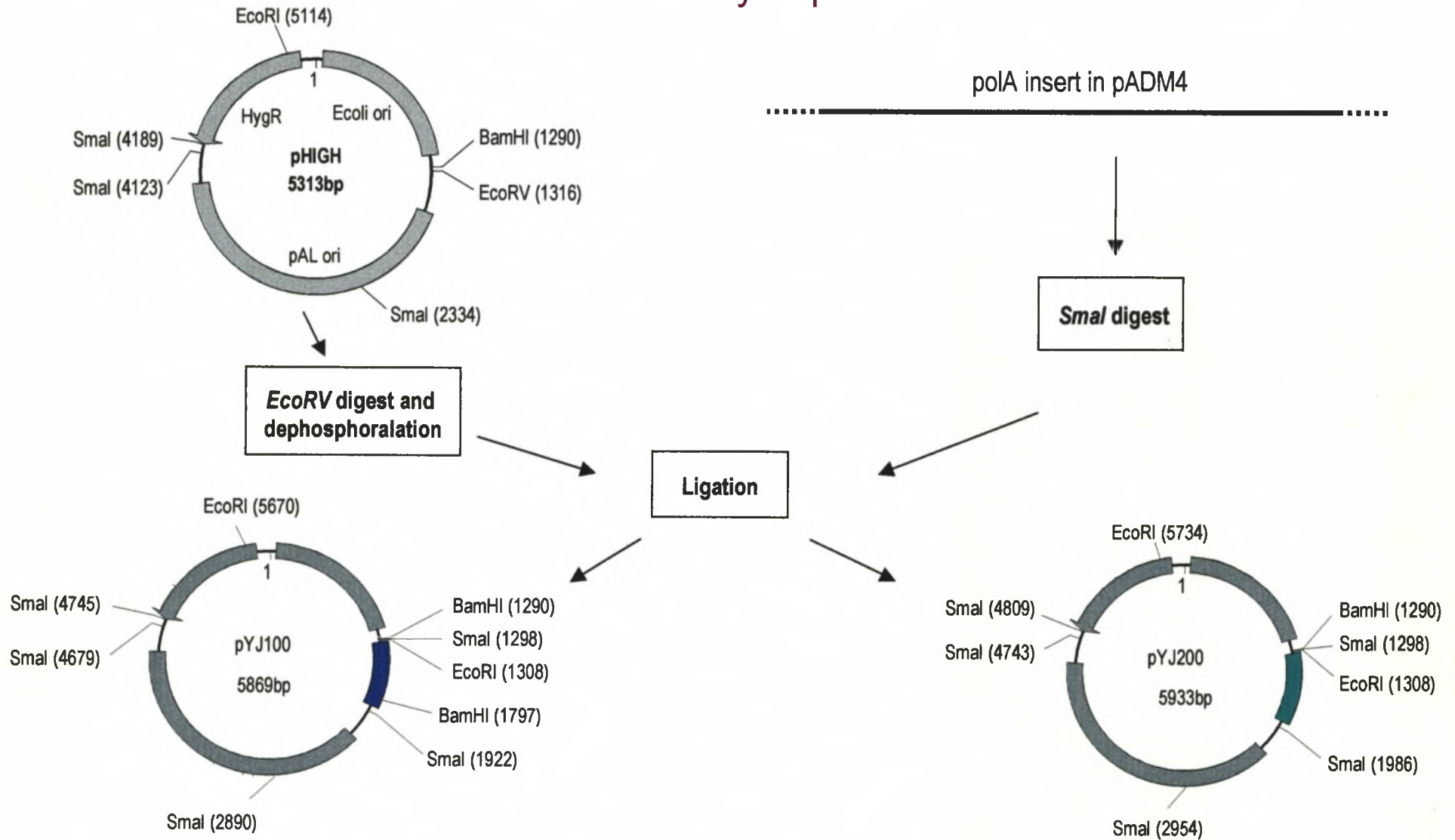
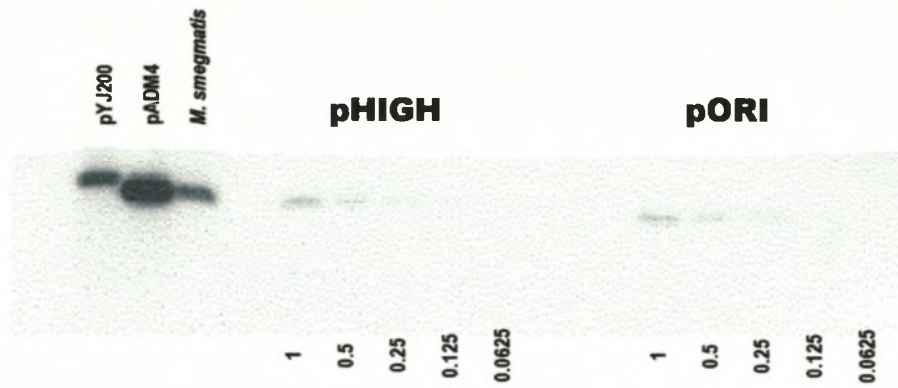
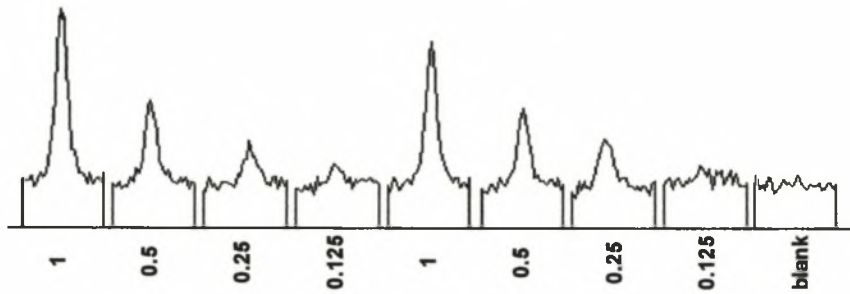


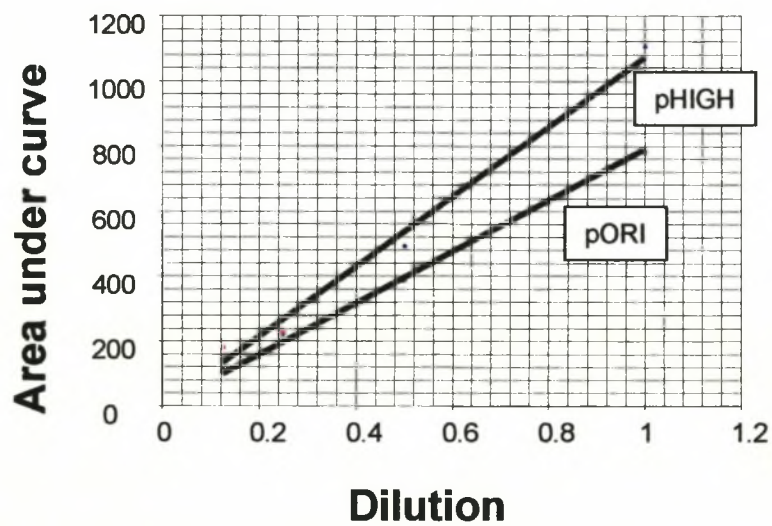
Figure 3.6 Making of pYJ100 and pYJ200. Experimental details are given in the text (3.2.6.3).



1. Autoradiograph

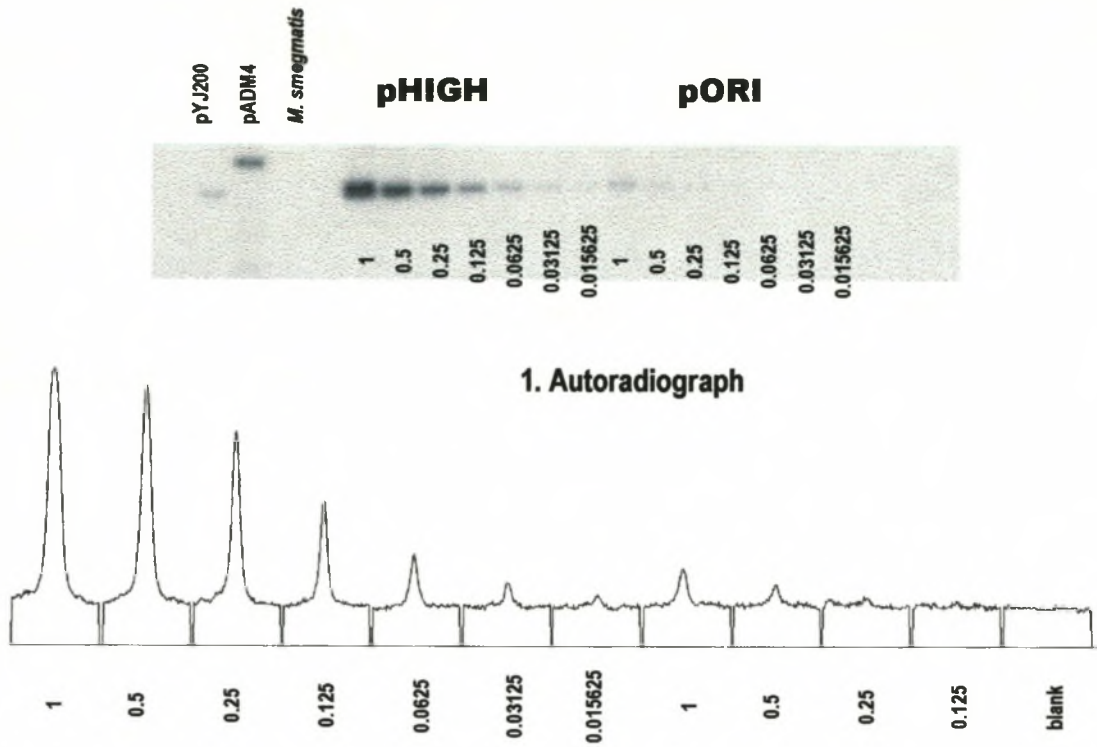


2. Plots generated by ScionImage

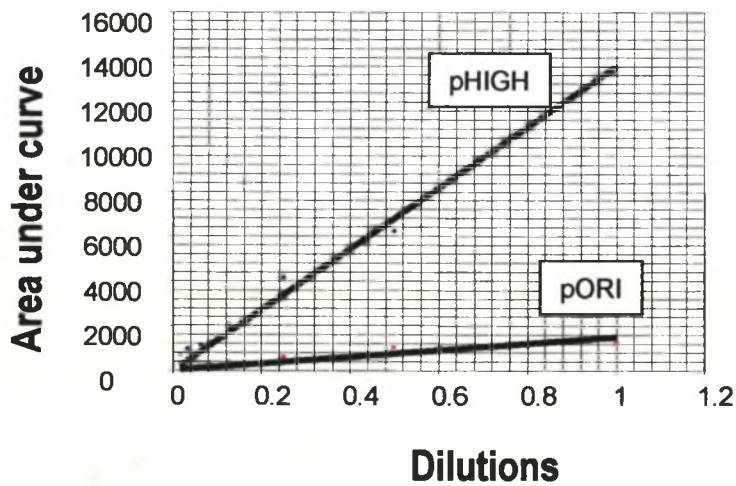


3. Graph with area under curve of (2) plotted against dilutions

Figure 3.7(a) Determining the correction factor for the amount of chromosomal DNA loaded on the gel.



2. Plots generated by ScionImage



3. Graph with area under curve of (2) plotted against dilutions

Figure 3.7(b) The ratio between pORI and pHIGH copy number.

3.3.3.4 Determining the exact copy number

(i) First approach

From the previous work (3.3.3.3), the ratio of the increase in copy number had been determined, however, the exact copy number remained unknown. It is only necessary to determine the exact copy number of one of the plasmid pair (pORI and pHIGH) as the ratio between the two is known, and the copy number of the second plasmid can therefore be determined mathematically. As discussed previously, difficulties had been encountered in loading equal amounts of total DNA on the agarose gel (3.3.3.2), so it was decided to use a different method, one in which knowledge of the exact amount of DNA loaded on the gels would not be necessary. Using this method it would be possible to determine the exact copy number by comparing the amount of the plasmid DNA directly to that of the chromosomal DNA in the same lane.

For this method a probe was designed that would hybridise to both a chromosomal and a plasmid fragment in the same sample of total DNA. The experiment was designed such that the chromosomal band and the plasmid band would be of different sizes upon restriction digestion. The region of homology with the probe, however, was exactly the same, in order to correctly compare the band intensities on the autoradiograph. On the autoradiograph two bands are expected to appear: one corresponding to the plasmid and the other representing the chromosome (Fig 3.8). Determining the ratio of the intensities between the two bands would then give the ratio of plasmid copy number per chromosome in the bacterial cell.

In the first approach (3.2.6.6), total chromosomal *M. smegmatis* [pYJ200] DNA was digested with *SmaI*, halving dilutions made and then separated on a 1.5% agarose gel. The gel was subjected to Southern transfer and the probe used on this blot was the gel-purified *SmaI* fragment of pYJ200 (3.2.6.4).

It was expected that the probe would hybridise with the chromosomal *polA* fragment, as well as with the plasmid pYJ200, both present in the total chromosomal DNA run on the gel. On the autoradiograph two bands were expected: the plasmid band (800bp), and a smaller chromosomal band (620bp) (Fig. 3.8).

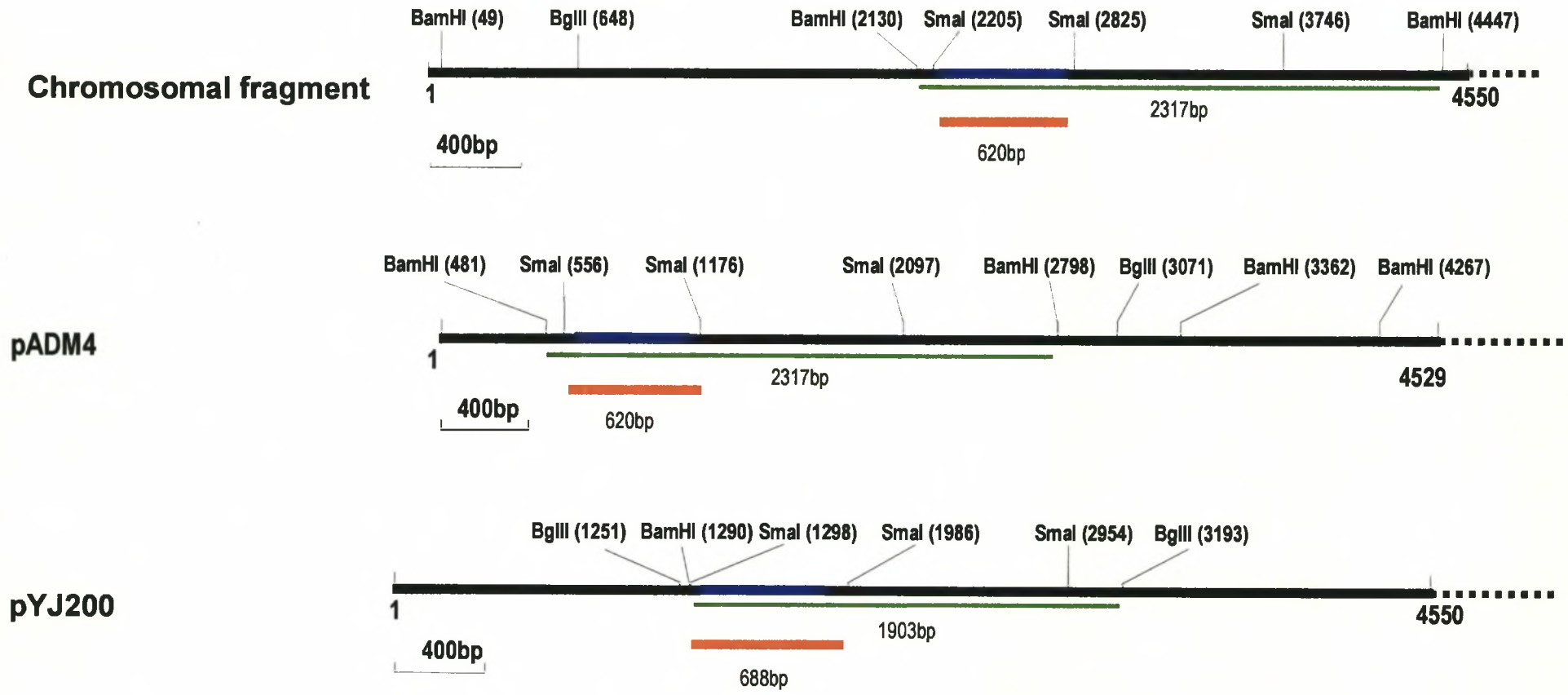


Fig. 3.8 Comparisons of the *M. smegmatis* chromosomal fragment, pADM4 and pYJ200. The blue lines indicate the fragments to which the PCR probe hybridise. The orange lines indicate the fragments the probe would hybridise to in a *SmaI* restriction digest, and the green lines indicate the fragments to which the probe would hybridise in a *BamHI*-*BglIII* digestion. The sizes of the fragments are indicated underneath the lines.

The results can be seen in Figure 3.9A. As can be seen in this figure, the two expected bands were present. However, there was a high background in the region between the plasmid band and the chromosomal band. Furthermore, the two bands were quite close to each other. Consequently densitometric scanning of the autoradiograph did not give two separate peaks (Fig. 3.9B). Because of these factors, the area under each curve could not be measured and consequently it was impossible to determine the copy number of pYJ200 from this blot.

(ii) Second approach

In the first approach, the plasmid fragment and the chromosomal fragment were too close in size, and did not separate well on the gel. As can be seen in Figure 3.9, on the autoradiograph the region in front of the plasmid band has a very high background. This is probably due to the unavoidable breakage of a fraction of the plasmid DNA fragments during plasmid isolation and restriction endonuclease digestion. It is crucial that the chromosomal band should lie in a position that has a low background signal on the autoradiograph, as the band itself will be of relative low intensity.

Using the gel purified fragment as a probe, also caused a high general background on the autoradiograph. This was because small fragments of the plasmid (isolated from the gel together with the desired fragment) would also be radioactively labelled and hybridisation of these fragments to the DNA on the Southern blot, makes for very high background. To eliminate this problem, it was decided to use a probe generated by PCR, which would be free of any smaller, unwanted, fragments.

By using a restriction digest that would generate a chromosomal band that is bigger than the plasmid band, a more accurate reading would be possible for the chromosomal band. Therefore, it was necessary to analyse the sequence of the *polA* insert in pADM4, and the insert in pYJ200, in order to decide on a different restriction endonuclease digestion strategy. By doing this it was possible to choose a different restriction endonuclease digestion that would give bands with a bigger difference in size and with the chromosomal band bigger than the plasmid band.

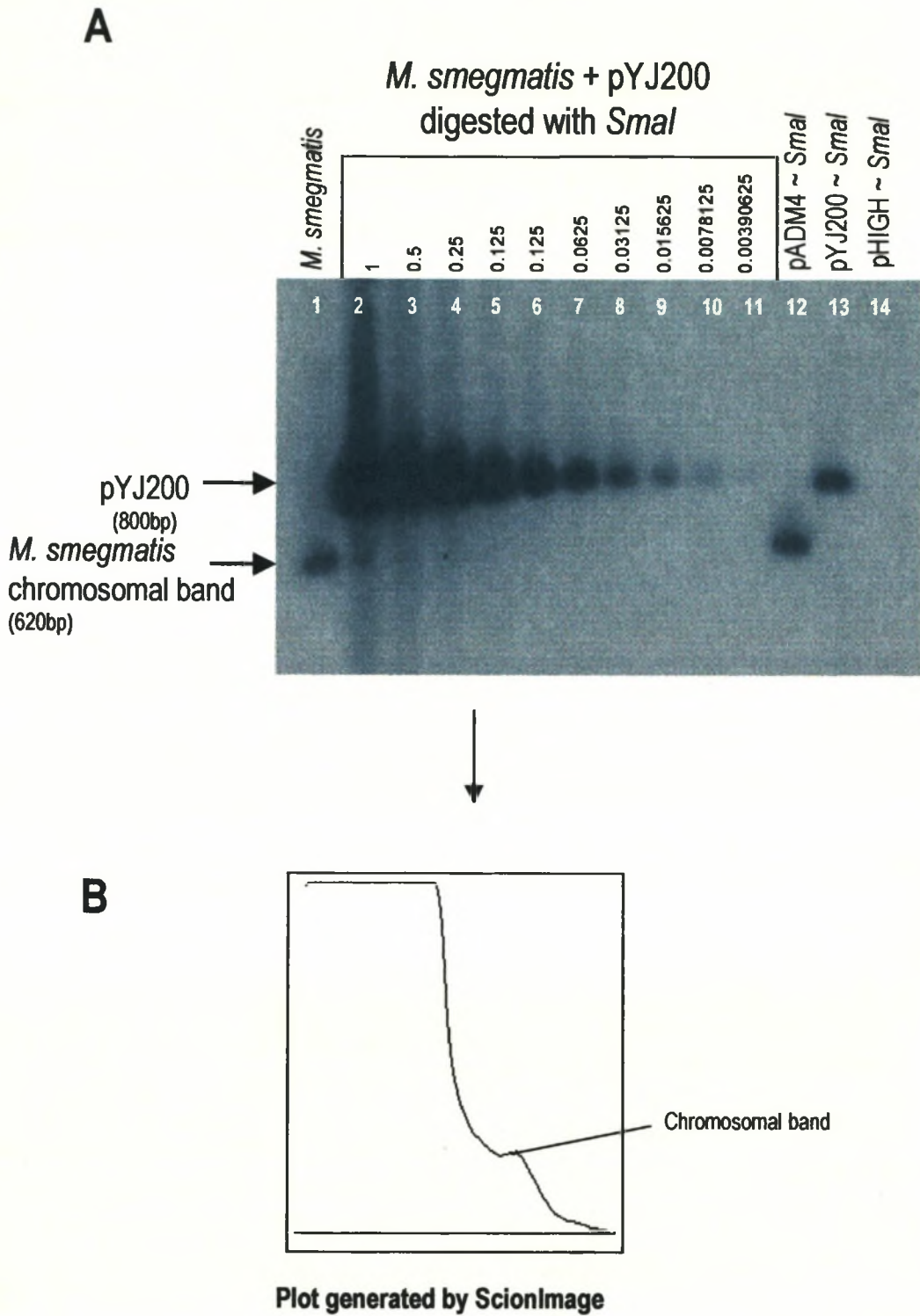


Figure 3.9 A. Autoradiograph of *M. smegmatis* [pYJ200].

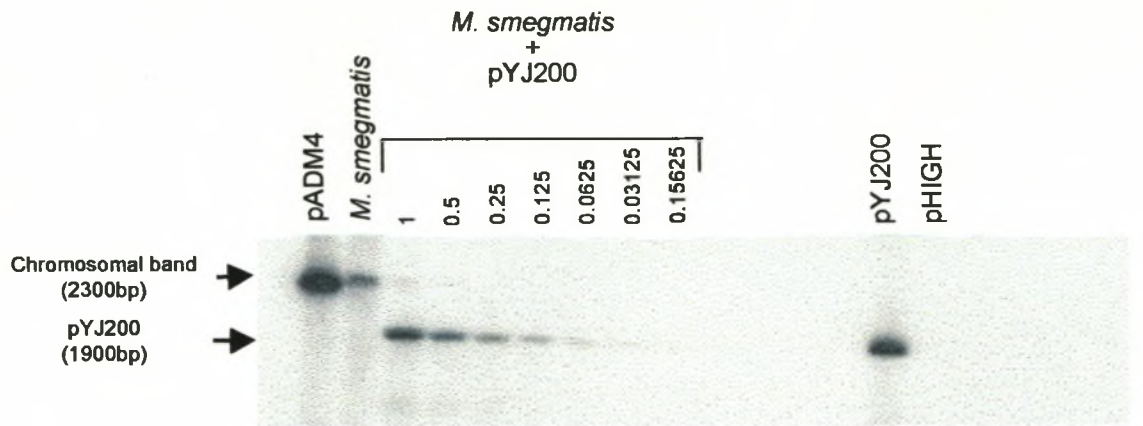
B. Plot generated by ScionImage (lane 2 on the autoradiograph).

Figure 3.8 shows a restriction map of part of the *polA* gene that was inserted into pADM4 as it is in the chromosome (sequence of *M. smegmatis* mc²155, obtained from the TIGR Microbial Database) and as it is in pYJ200. A *Bam*HI-*Bgl*II double digest was chosen as the best option for the *M. smegmatis* [pYJ200] DNA. This would generate a plasmid DNA fragment of approximately 1900bp and a chromosomal DNA fragment of approximately 2300bp, which would both hybridise with the probe in a Southern blot (Fig 3.8). It was thought that the separation between the chromosomal and plasmid band would be large enough to compare the band intensities using ScionImage.

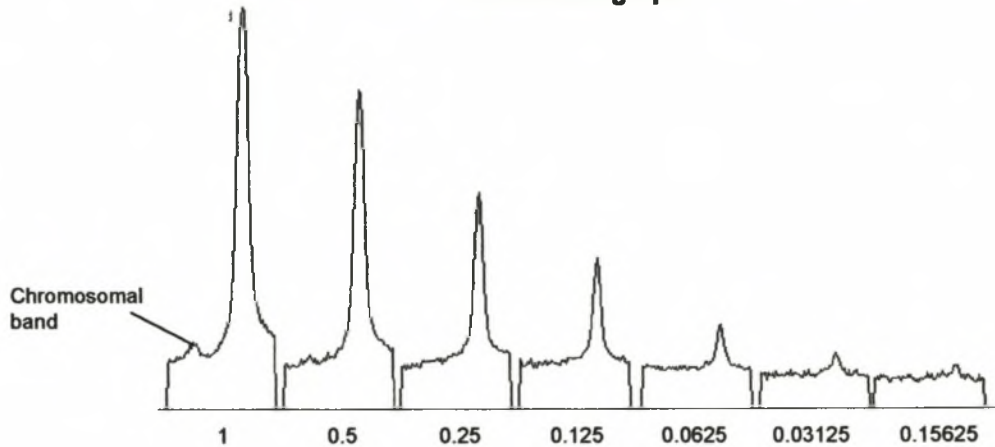
Total chromosomal *M. smegmatis* [pYJ200] DNA was double restriction endonuclease digested with *Bam*HI and *Bgl*II and halving dilutions made in 1X restriction buffer. Controls of *M. smegmatis*, pADM4, pHIGH and pYJ200 also restriction endonuclease digested with *Bam*HI and *Bgl*II, were separated along with the halving dilutions of the chromosomal digest, on a 0.7% agarose gel (3.2.6.7). The gel was then subjected to Southern transfer and the blot was hybridised with the PCR generated probe, using pYJ200 as the target for the PCR reaction (3.2.6.5).

The results can be seen in Figure 3.10. It can be seen that the separation between the plasmid band and the chromosomal band was now large enough to be distinguished as 2 different peaks. As mentioned earlier, higher background is visible in front of the plasmid band, caused by fragments of the plasmid.

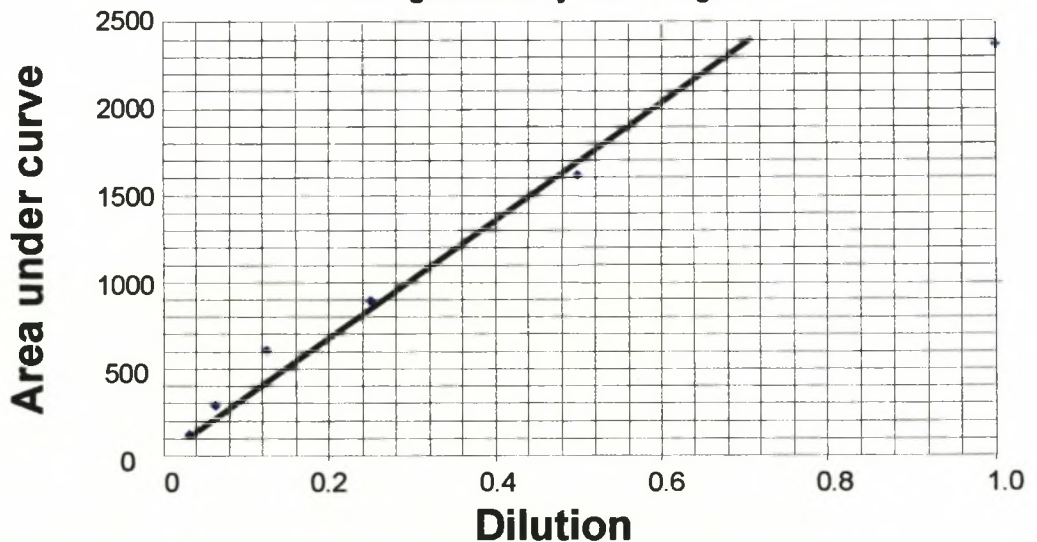
As in section 3.3.3.3 a graph was drawn using the area under the curve of the plots generated by ScionImage (Fig. 3.10). On the graph it can be seen that the plasmid band of undiluted sample (1, lane 3), is of such high intensity that the signal is in the non-linear range. This is due to over exposure of the autoradiograph for that band. Drawing a straight line using the other points on the graph (thus excluding the last point), it is possible to determine the ratio between the chromosomal band and the plasmid band.



1. Autoradiograph



2. Plots generated by ScionImage



3. Graph with area under curve of (2) plotted against dilutions

Fig. 3.10 Determination of the actual copy number.

1. Photo of the autoradiograph that was densitometrically scanned.
2. Plot generated by ScionImage for each band on the autoradiograph.
3. Graph generated using the area under the curves of the ScionImage plot in B, plotted against the dilution of each band. The equation of the straight line in C was used to determine the exact copy number (3.3.3.4).

By extrapolation, the area under the curve for dilution 1 was determined to be 3395. The area under the curve of the chromosomal band is 63. The ratio between chromosomal band and plasmid band is then given as $63 : 3395 = 1 : 54$

Therefore the copy number of pHIGH in *M. smegmatis* is approximately 54, and using the ratio of 7, determined in section 3.3.3.3, the copy number of pORI is determined as approximately 8.

3.3.4 Further characterisation of pORI and pHIGH

3.3.4.1 Growth curves

The presence of a recombinant plasmid can place stress on the metabolism of its host and this can manifest itself in a reduced growth rate. This reduction in growth rate is an important experimental consideration when plasmid-bearing strains are used. For example, it is likely to affect the virulence of pathogenic mycobacteria and is thus an important experimental consideration. As mycobacteria are slow growers, compared to *E. coli*, it is also a significant practical laboratory consideration. Furthermore, by placing a load on the bacteria, a selective advantage for plasmid free cells is created, thus causing an apparent increase in instability (Proctor, 1994). It was therefore necessary to determine whether the increased copy number of pHIGH would put any load on the mycobacterial cell, causing it to grow slower than it would with a lower copy number plasmid.

The growth curves were obtained by hourly measurements of the OD₆₀₀ of actively growing cultures of *M. smegmatis*, *M. smegmatis* [pORI] and *M. smegmatis* [pHIGH] (see section 3.2.8). The OD₆₀₀ measurements were then plotted against time and the curves of *M. smegmatis* [pORI] and *M. smegmatis* [pHIGH] were compared to that of *M. smegmatis*.

The results are shown in Figure 3.11. From these graphs it can be seen that, by this method, there is no discernible difference between the growth curves of *M. smegmatis*, *M. smegmatis* [pORI] and *M. smegmatis* [pHIGH]. This argues that neither the high copy number, nor the low copy number plasmid exerts a significant load on its host, and that there is no strong selective pressure on the cells containing either pORI or pHIGH.

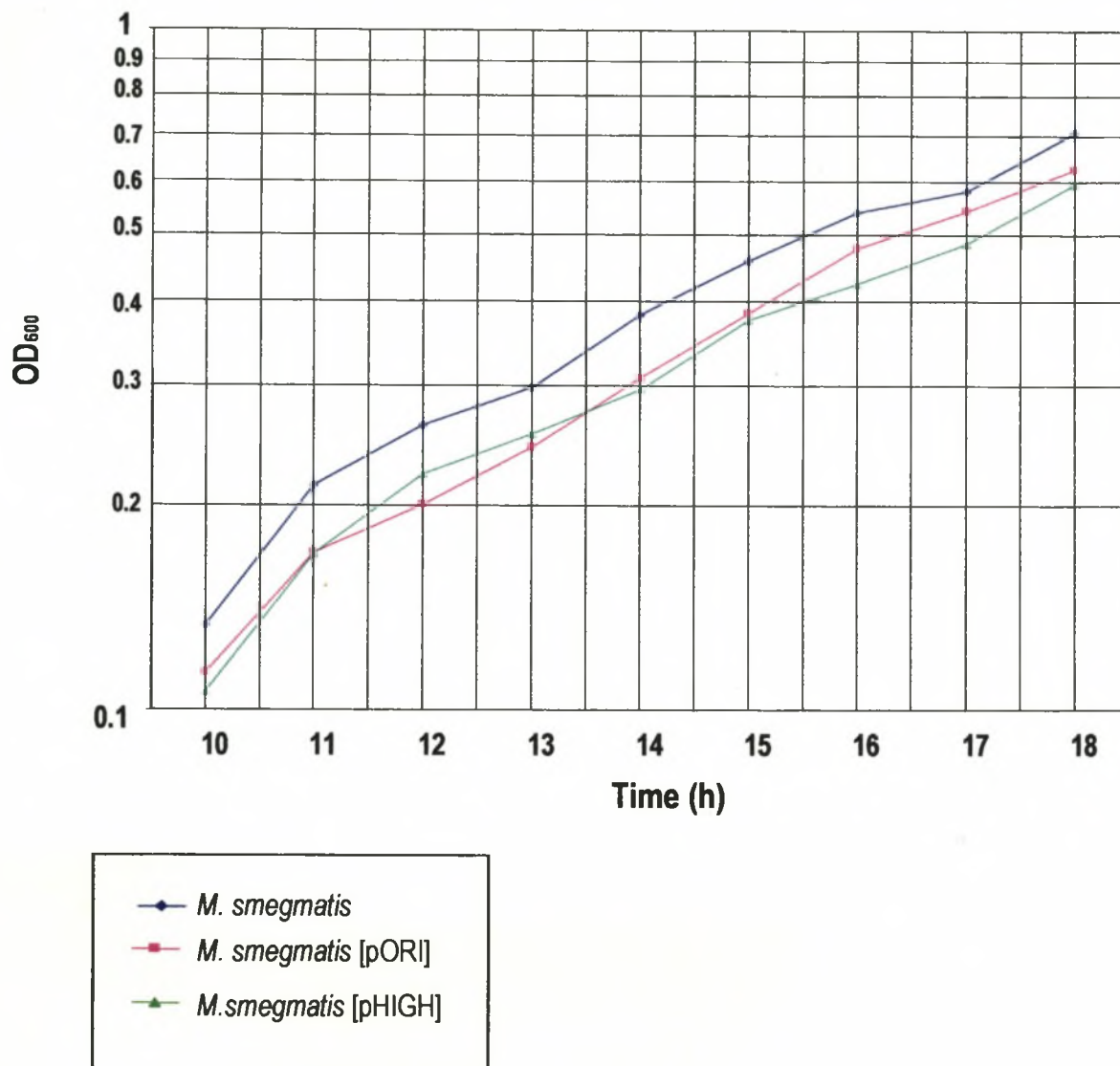


Figure 3.11 Growth curves of *M. smegmatis* [pORI] and *M. smegmatis* [pHIGH] compared to the growth curve of *M. smegmatis*. The OD₆₀₀ of the cultures were taken hourly from time=10h and plotted against time. The experimental detail is given in the text (3.2.8).

It should be noted, however, that this need not always be the case. If a plasmid expresses a factor that is toxic or inhibitory to the host, an increase in copy number is likely to worsen that effect. The experiment described above proves only that the presence of the mutant origin and consequent increase in plasmid DNA do not themselves affect the growth rate of the host.

3.3.4.2 Stability tests

The only formal study of the stability of a pAL5000-based vector involved the use of plasmid pYUB12 (Stolt and Stoker, 1996a; see section 1.3.3), which was shown to be relatively unstable in *M. smegmatis*. The frequency of *M. smegmatis* cells carrying pYUB12, decreased by 10^5 over 120 generations in the stability tests carried out by Stolt and Stoker (1996a). The plasmids pORI and pHIGH contain only part of pAL5000 (the origin, ORF1 and ORF2), and it was thought that therefore they might be less stable than pYUB12, which contained the whole of pAL5000. Low copy number plasmids usually have specific partitioning mechanisms to ensure stable inheritance of the plasmid (1.2.1). No partitioning mechanism has been described for pAL5000, however, because pORI and pHIGH consist of only part of pAL5000, it is possible that a partitioning mechanism of pAL5000 would not be present on these plasmids, causing them to be unstable.

The ability of a plasmid to remain in its host from one generation to the next, in the absence of selective pressure, is used as a measure of the plasmid stability, and the stability of a recombinant DNA vector in its host, dictates its usefulness. As it had already been established that pHIGH has a copy number of approximately 54 in *M. smegmatis*, it was postulated that this increased copy number would result in a higher stability.

The experimental procedure for determining the stability of pORI and pHIGH in *M. smegmatis* is described in section 3.2.7. Cultures of *M. smegmatis* [pORI] and *M. smegmatis* [pHIGH] were grown in MB7H9 media containing no hygromycin, over approximately 200 generations. The last cultures were then used to determine the fraction of *M. smegmatis* cells still carrying the plasmids.

Two methods were used to determine the amount of growth, in order to calculate the generation number: the first used was the absorbance (OD_{600}) and the second, protein concentration (Meyers *et. al.*, 1998).

Passage 9 of both *M. smegmatis* [pORI] and *M. smegmatis* [pHIGH] were used to spread 100 μ l samples (diluted 1×10^{-7}) on Middlebrooks 7H11 plates, half of which contained 100 μ g/ml hygromycin. The colony counts from these plates were used to determine the amount of cells still carrying plasmids, by comparing the hygromycin resistant cells (on the MB7H11 + hygromycin plates) with the total amount of cells (on the MB7H11 plates with no hygromycin).

This experiment was a pilot study to determine whether there is a difference in stability between pORI and pHIGH, and also whether there is a difference in stability when compared with pYUB12 of Stolt and Stoker. When determining stability in cultures that clump as excessively as mycobacteria, one will get variable results with colony counts because one colony is not equal to one cell, and clumping differs with each culture. The cultures used in this assay were studied under the microscope and it was determined that one colony represented between 1 and 10 cells.

After more than 200 generations (or 9 passages), in media containing no selection, for pORI an average of 664 colonies per plate were counted on 8 plates with no antibiotic selection, whereas on a total of 10 plates containing 100 μ g/ml hygromycin, an average of 147 colonies were counted per plate. For pHIGH, on 8 plates with no selection, an average of 745 colonies per plate were counted, whereas on a total of 10 plates, containing 100 μ g/ml hygromycin, an average of 64 colonies per plate were counted.

Using these numbers, it was determined that 22.1% of cells still carried pORI and 8.6% of cells still carried pHIGH. It can be seen from Tables 3.2a and 3.2b that the cultures went through 200+ generations, according to both the protein assay and OD_{600} readings.

This showed pORI and pHIGH both to be much more stable in *M. smegmatis* than pYUB12. A reason for this might be that the plasmid pYUB12 puts a load on the host cell, causing it to grow slower. ORF4 was destroyed when creating pYUB12 (the

kanamycin resistance gene was inserted here), but ORF3 still encodes a factor that could cause a load on the host cell. If this is the case, it would then result in an apparent instability because the faster growing plasmid-free cells would take over the culture.

The increased copy number of pHIGH did not seem to affect the stability of the plasmid, because the plasmid pORI was already unexpectedly stable. Intriguingly pHIGH appeared, if anything, to be marginally less stable than pORI. Perhaps this is because (and it is expected) the high copy number places a slight extra metabolic load on the bacterial cells. This may be too small to be noticeable in the growth curves (Fig. 3.12), but may exert an effect over numerous passages.

The question of why pORI is so stable still remains. Possibly the origin of replication contains an unidentified toxin-antidote system (Jensen and Gerdes, 1995). Such systems are extremely difficult to identify on the basis of sequence data alone, as the genes involved are usually small and may be unrelated to other proteins.

Table 3.2a Calculation of generation number of *M. smegmatis* [pORI]

In the tables the OD₆₀₀ and protein assay concentration of each passage is given as determined at the beginning and end of growth of the approximated 25-30 generations. The number of generations was determined using the formula given in section 3.2.7. The figures are rounded off to simplify the tables.

Passage number	Calculated OD ₆₀₀ at inoculation	Final OD ₆₀₀	Generation number (OD ₆₀₀)	Calculated protein concentration at inoculation	Final protein concentration (µg/ml)	Generation number (protein)
1	2.20 x 10 ⁻⁹	3.20	30	7.54 x 10 ⁻⁷	345.70	29
2	3.42 x 10 ⁻⁸	2.70	26	3.70 x 10 ⁻⁶	302.59	26
3	2.45 x 10 ⁻⁸	3.66	27	2.74 x 10 ⁻⁶	459.00	27
4	4.50 x 10 ⁻⁸	2.26	26	5.65 x 10 ⁻⁶	428.87	26
5	1.71 x 10 ⁻⁸	3.52	28	3.25 x 10 ⁻⁶	479.55	27
6	4.15 x 10 ⁻⁸	2.50	26	5.66 x 10 ⁻⁶	372.98	26
7	2.10 x 10 ⁻⁸	2.34	27	3.13 x 10 ⁻⁶	339.46	27
8	1.84 x 10 ⁻⁸	3.28	27	2.66 x 10 ⁻⁶	425.86	27
Total generations			217	Total generations		215

Table 3.2b Calculation of generations of *M. smegmatis* [pHIGH]

Passage number	Calculated OD ₆₀₀ at inoculation	Final OD ₆₀₀	Generation number (OD ₆₀₀)	Calculated protein concentration at inoculation	Final protein concentration (µg/ml)	Generation number (protein)
1	3.62 x 10 ⁻⁹	2.72	29	9.54 x 10 ⁻⁷	340.61	28
2	2.48 x 10 ⁻⁸	3.26	27	3.11 x 10 ⁻⁶	310.90	27
3	3.55 x 10 ⁻⁸	2.20	26	3.39 x 10 ⁻⁶	239.35	26
4	1.62 x 10 ⁻⁸	1.89	27	1.77 x 10 ⁻⁶	169.33	27
5	1.20 x 10 ⁻⁸	1.75	27	1.07 x 10 ⁻⁶	379.43	28
6	1.03 x 10 ⁻⁸	2.45	29	2.23 x 10 ⁻⁶	236.65	27
7	2.01 x 10 ⁻⁸	2.80	27	1.95 x 10 ⁻⁶	349.50	27
8	2.63 x 10 ⁻⁸	3.26	27	3.28 x 10 ⁻⁶	342.80	27
Total generations			219	Total generations		217

3.3.5 Investigation of possible mechanisms for increased copy number

There are three obvious possible causes for the increase in copy number caused by the 3bp deletion in pHIGH. Firstly the mutation could have caused a change in the function of the replication protein (RepA). This possibility is difficult to investigate as the function for pAL5000 RepA has not been determined and can therefore not be assayed. Secondly the mutation could have caused a change in the mRNA secondary structure leading to an increased or decreased expression of RepB. Alternatively it could also possibly have created a promoter between ORF1 and ORF2, or changed an existing one, because the mutation is found at a site where an internal promoter, if it existed, is likely to be positioned.

3.3.5.1 Promoter probe vector – pTV102

The S1 nuclease protection assays, performed by Stolt and Stoker (1996b) identified the transcription start site upstream of RepA. However, the design of this experiment was such that it would not have identified any promoters situated just upstream of ORF2.

Therefore it was decided to use a promoter probe vector system to test for any promoter activity in the region containing the mutation.

The promoter probe vector pTV102 was used for this. This plasmid has a multiple cloning site (Fig 3.12) where the pAL5000 region containing the mutation (*NcoI-EcoRI* fragment) and the mutation-free homologue were inserted. Downstream of this site, there is a chloramphenicol acetyl transferase (CAT) gene, without a promoter but with a ribosomal binding site. Upstream of this gene there are stop codons in all three reading frames to prevent any translational read-through.

Again a two-step cloning strategy was adopted to ensure that the insert would be in the correct orientation. This was because the insert was too small to easily check the orientation if a different method of cloning was used. The subcloning of the *NcoI-EcoRI* fragment into this site is described in detail in section 3.2.9 (Fig. 3.12 and Fig. 3.13). In short, the plasmids pOSEQ2 and pTV102 were linearised and ligated (the same procedure was followed for pHSEQ2). The transformants were selected on hygromycin and ampicillin, and the clones were then restriction mapped in order to determine which had the insert in the correct orientation. The correct clones were then partially digested, recircularised and the transformants selected on hygromycin. A close-up of the insert and the surrounding area of pTVO and pTVH can be seen in Figure 3.14.

The CAT-activity of the constructs in *M. smegmatis* was then tested (3.2.10). The CAT-activities of *M. smegmatis* [pTVO] (carrying the wild-type fragment) and *M. smegmatis* [pTVH] (carrying the mutant fragment) were compared to that of *M. smegmatis* [pTV102], as well as the positive control *M. smegmatis* [pTVC], which had a *M. smegmatis* promoter cloned into the MCS (multiple cloning site) of pTV102. The CAT activities were tested in liquid culture. A series of 10ml cultures with a range of increasing chloramphenicol concentration were inoculated with *M. smegmatis* [pTVO], *M. smegmatis* [pTVH], *M. smegmatis* [pTVC] and *M. smegmatis* [pTV102]. The amount of growth in each culture was then determined when the chloramphenicol-free cultures had reached stationary phase.

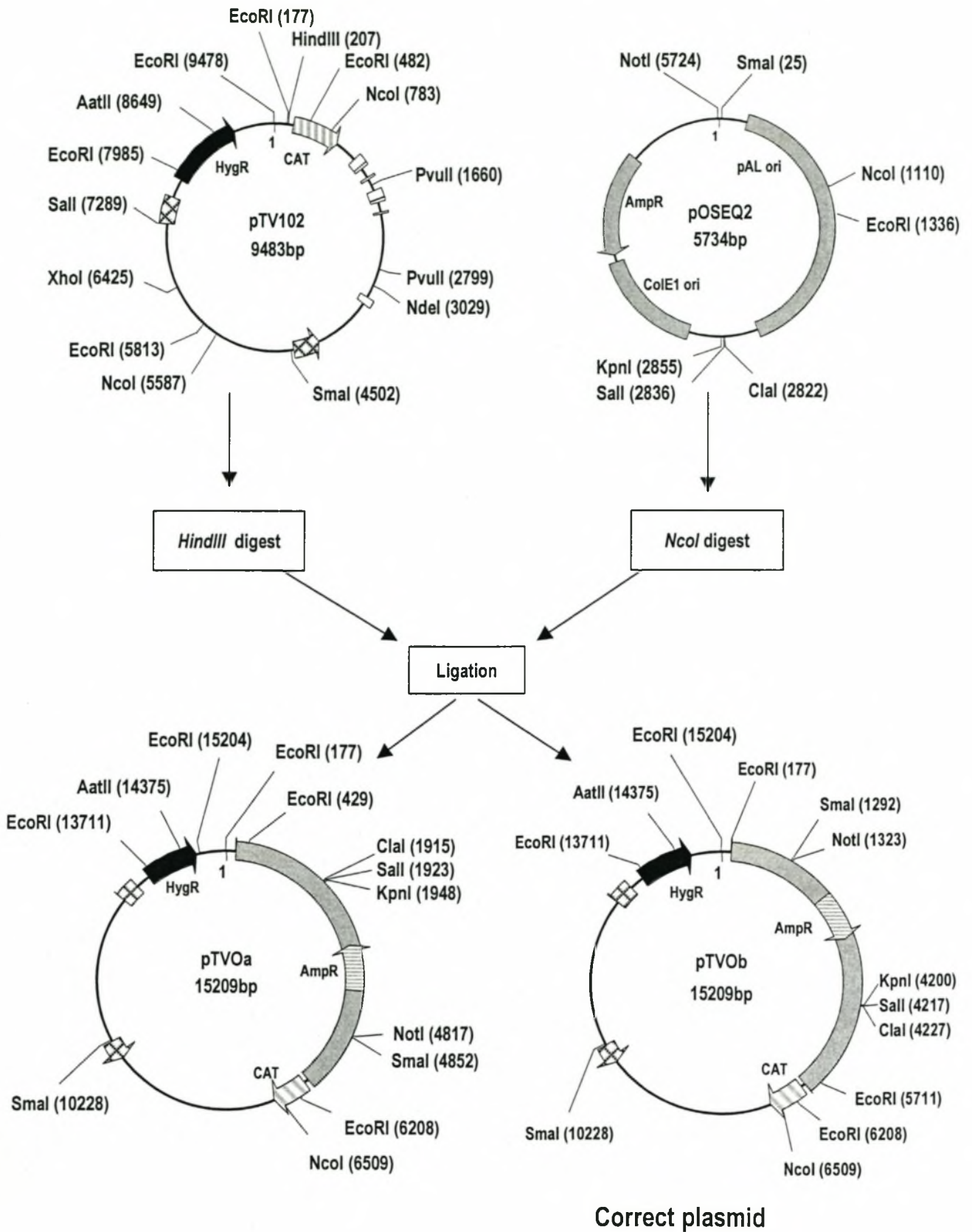


Figure 3.12(a) Making of pTVO. Experimental details in text (3.2.9).

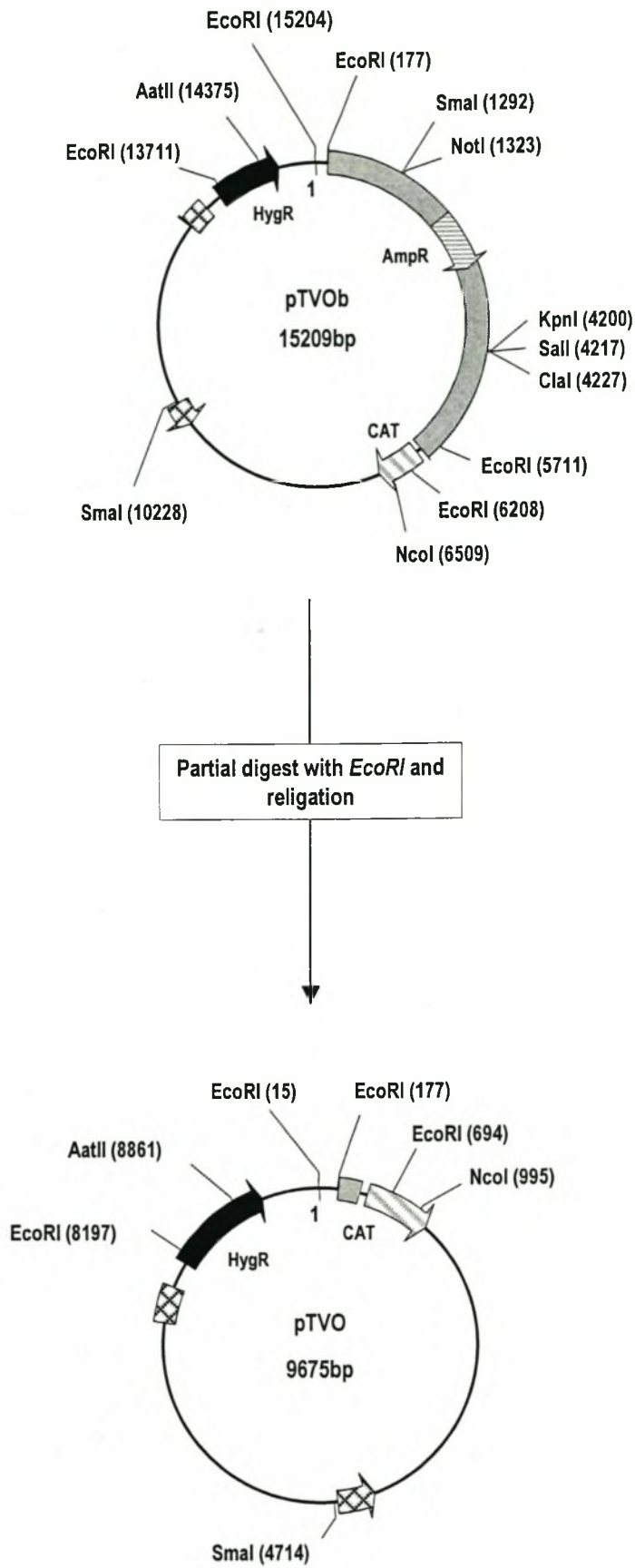


Figure 3.12(b) Making of pTVO. Experimental details in text (3.2.9)

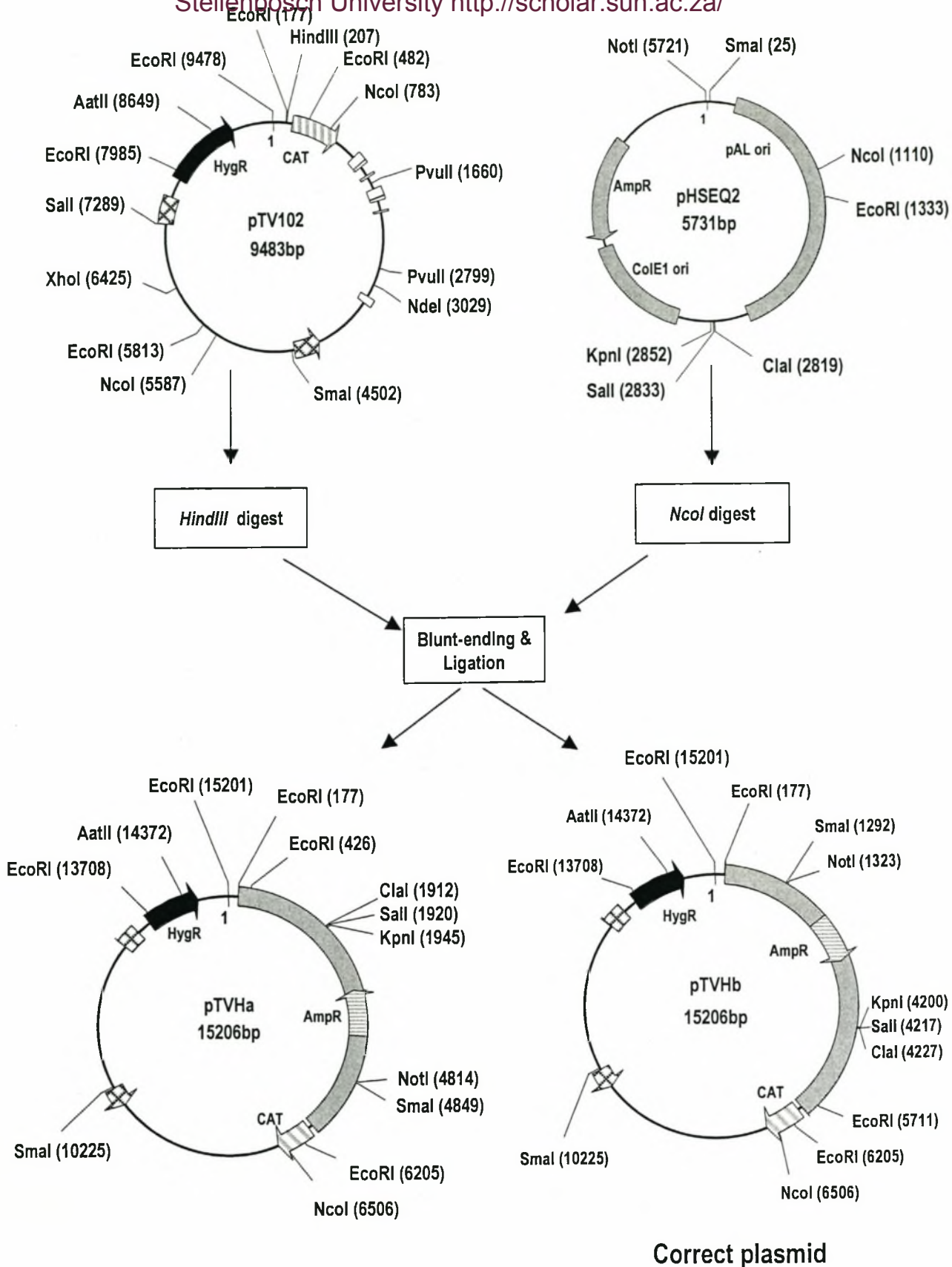


Figure 3.13(a) Making of pTVH. Experimental details in text (3.2.9).

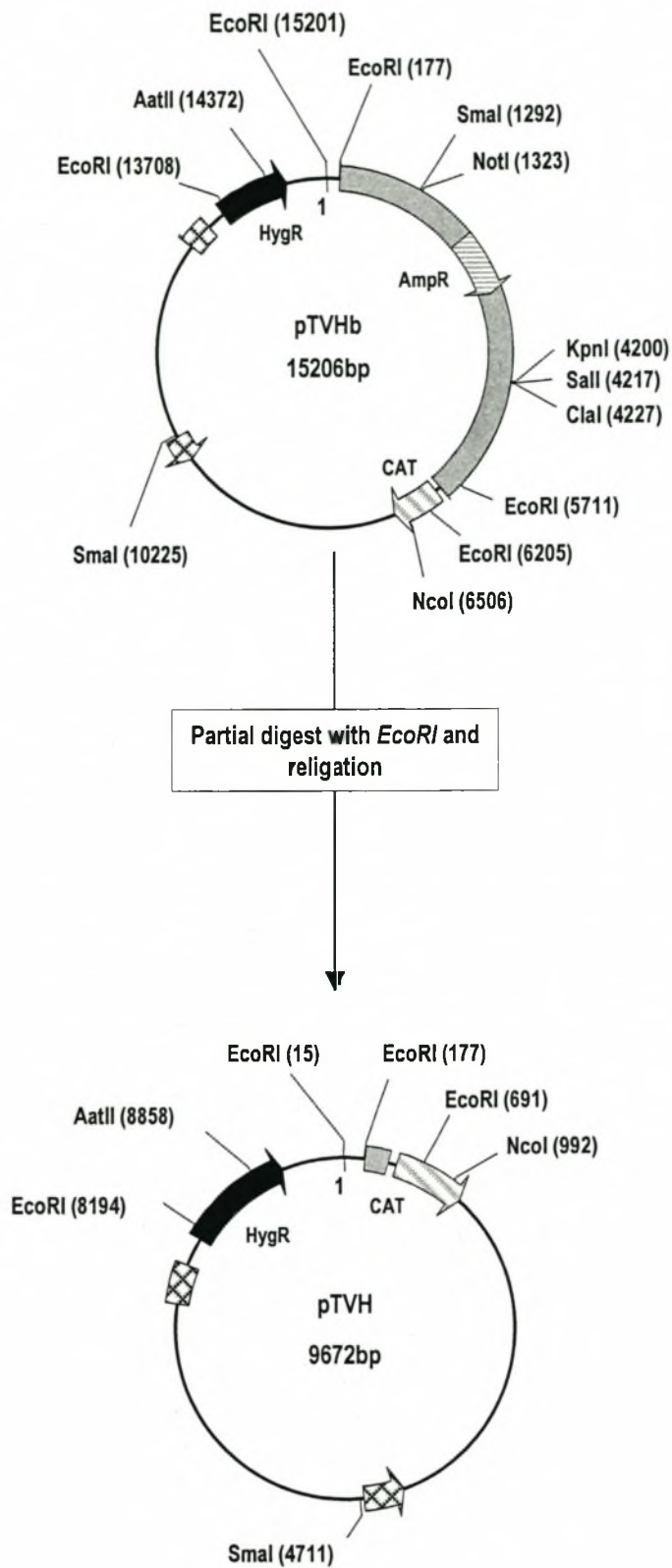
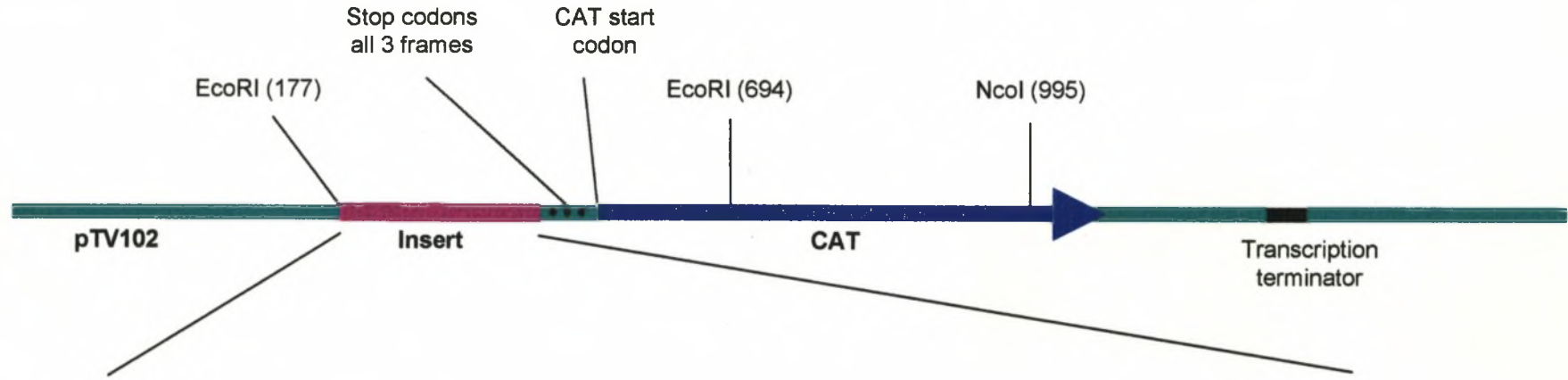


Figure 3.13(b) Making of pTVH. Experimental details in text (3.2.9).



A

2801- GCGAAACGCC GAATTCCCGT GCAACGACGT GTGTCCCGGA CCGCTACCGG ACAGCGAGGT CCGCGCCATC GCCAACAGCA

2881- TTTGGCGTTG GATCACAACC AAGTCGCGCA TTTGGGCGGA CGGGATCGTG GTCTACGAGG CCACACTCAG TGC GCGCCAG

3 BP DELETION

2961- TCGGCCATCT CGCGGAAGGG CGCAGCAGCG CGCACGGCGG CGAGCACAGT TGC GCGGCGC GCAAAGTCCG

3031- CGTCAGCCAT GG

B

pORI 101 2970- TCG CGG AAG GGC GCA GCA GCG CGC ACG -2996

S R K G A A A R T

pHIGH 100 2970- TCG CGG AAG GGA GCA GCG CGC ACG GCG -2996

S R K G A A R T A

Figure 3.14 Details of the *NcoI-EcoRI* fragment inserted into pTV102.

The results of this assay are shown in Figure 3.15. A very small difference was found between the CAT-activities of pTVO and pTVH. There is a slight increase in activity above that of the wild-type pTV102, but not as big an increase as one would expect should there be a strong promoter (compare with the CAT-activity of the control pTVC1).

The activities of both pTVO and pTVH are higher than that of the negative control pTV102 (Fig. 3.15). This indicates that there might exist a very weak promoter between ORF1 and ORF2 in both pTVO and pTVH. The increase in activity between that of pTVO and pTVH however, is too small to give a concrete answer on a relative change as a result of the 3bp mutation present in pTVH.

It is interesting to note that in every case there is a steady increase in the OD₆₀₀ as the chloramphenicol concentration increases, until a lethal concentration is reached. The reason for this is unknown, but may be due to a different morphology or metabolism in the face of antibiotic induced stress.

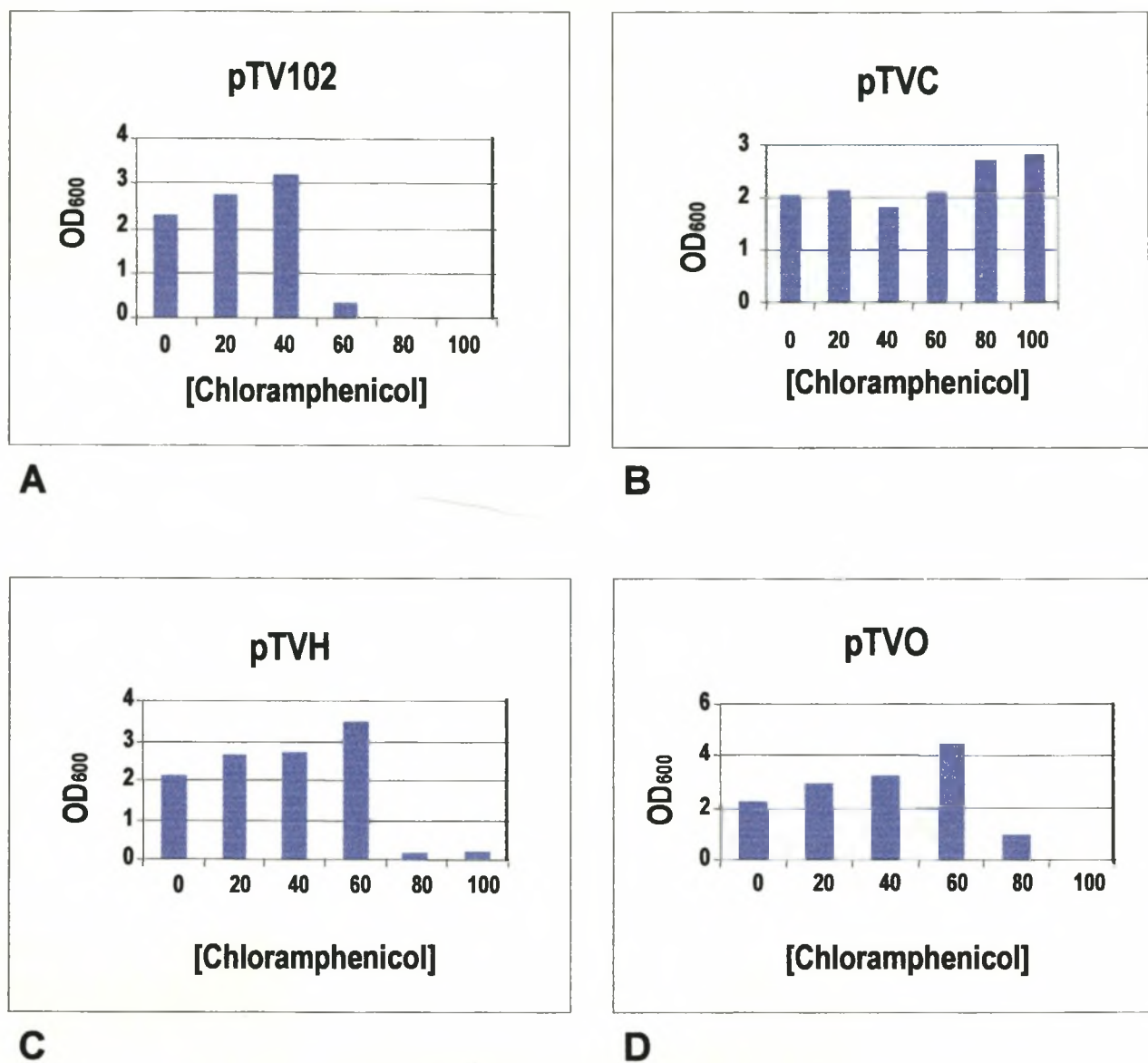


Figure 3.15 Results of the chloramphenicol acetyl transferase assay on pTVO and pTVH in liquid culture.

- A. OD₆₀₀ of *M. smegmatis* [pTV102] cultures grown in different concentrations of chloramphenicol (µg/ml)
- B. OD₆₀₀ of *M. smegmatis* [pTVC] cultures grown in different concentrations of chloramphenicol (µg/ml)
- C. OD₆₀₀ of *M. smegmatis* [pTVO] cultures grown in different concentrations of chloramphenicol (µg/ml)
- D. OD₆₀₀ of *M. smegmatis* [pTVH] cultures grown in different concentrations of chloramphenicol (µg/ml)

RNA secondary structure

The second possibility that could cause an increase in plasmid copy number was that the RNA secondary structure of the two open reading frames was affecting the translation efficiency of RepB by affecting ribosomal binding. The predicted secondary structure of the *RepA-RepB* mRNA of pORI (wild type) and pHIGH (mutant) was compared to determine the influence of the 3bp deletion on its secondary structure. The predicted mRNA secondary structures were obtained by using the mfold server (<http://bioinfo.math.rpi.edu/~zukerm/>) (Zuker *et. al.*, 1999; Mathews *et. al.*, 1999).

The DNA sequence used for the mRNA folding was from nucleotide C (4632 in pAL5000 sequence) that was defined by Stolt and Stoker (1996b) as the start point of the *rep* mRNA (1.3.6) up to the stop codon of *repB*. In the latest sequence of pAL5000 (Brigitte Giquel - personal communication) this is nt 4621. It should be noted that the transcription termination point of RepAB has not been identified. For this reason a representation of a true transcript could not be used.

The results from this folding can be seen in Figures 3.16 and 3.17. For pORI 21 structures with different energies were obtained and 26 for pHIGH, but only the two with the lowest free energies (ΔG) are shown as examples, because the structures with the lowest energies are the most stable ones.

In the case of pORI and pHIGH the RNA secondary structure around the potential SD sequence and initiation codon of *repB* has changed (Fig. 3.17). In pORI (which represents the wild type) the last 3nt of the SD sequence (-AGG) are exposed in a loop, while the start codon of *repB* is base-paired. In the mutant (pHIGH) only two of the five nucleotides in the SD sequence are exposed, but the start codon (AUG) of *repB* is exposed in a loop as well.

The mRNA secondary structure may be used to regulate the expression of certain genes (Looman *et. al.*; 1986), because the local secondary structure can determine whether the ribosomal binding site (RBS) is available for initiation of translation by the ribosome. Looman *et. al.* (1986) compared the efficiencies of RBS's, isolated from their original

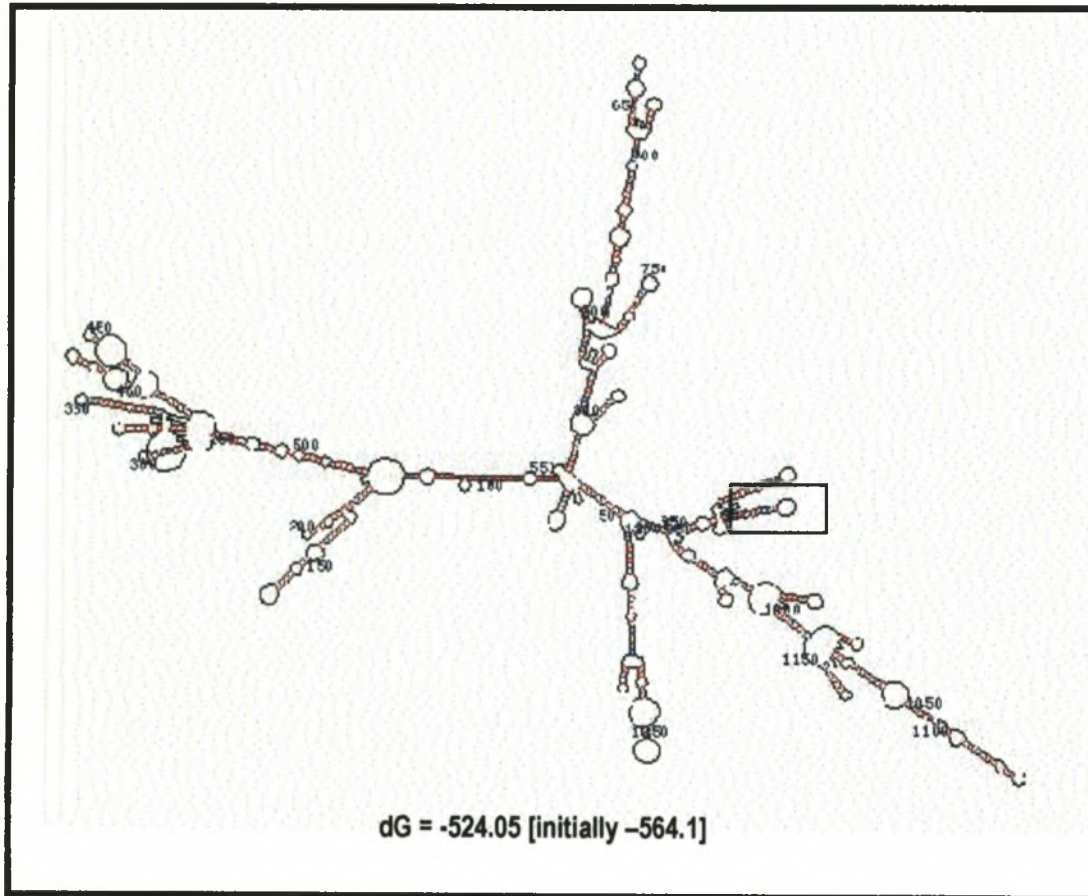
context, and concluded that the activity of the RBS is primarily determined by its accessibility for initiating ribosomes. The main factor that has an influence on this accessibility is the local secondary structure of the mRNA, which can either shield or expose the Shine-Delgarno and/or initiation codon.

There is a very close relationship between translational efficiency and stability of the secondary RNA structure in the initiation region. According to De Smit and Van Duin (1990) an increase in the free energy of -1.4kcal/mol , causes a factor of 10 reduction in the initiation rate and expression was decreased by a factor of 500 when a single nucleotide substitution turned a mismatch into a match (De Smit and Van Duin, 1990).

From the RNA secondary structure of pORI and pHIGH it can thus be assumed that, if the putative RBS shown in Figure 1.2 is active, then the initiation of RepB translation may vary between pORI and pHIGH. In order to draw absolute conclusions, however, one would have to analyse all the possible structures.

It can be argued that this may have two effects. Firstly it would change the ratio of the proteins RepA and RepB, which may have profound effects, particularly if they interact in some manner. Secondly, as RepB is an autoregulator (probably a repressor of its own expression), it is expected that the absolute level of both RepA and RepB present in the cell, would be affected.

pORI (wild-type)



pHIGH (mutant)

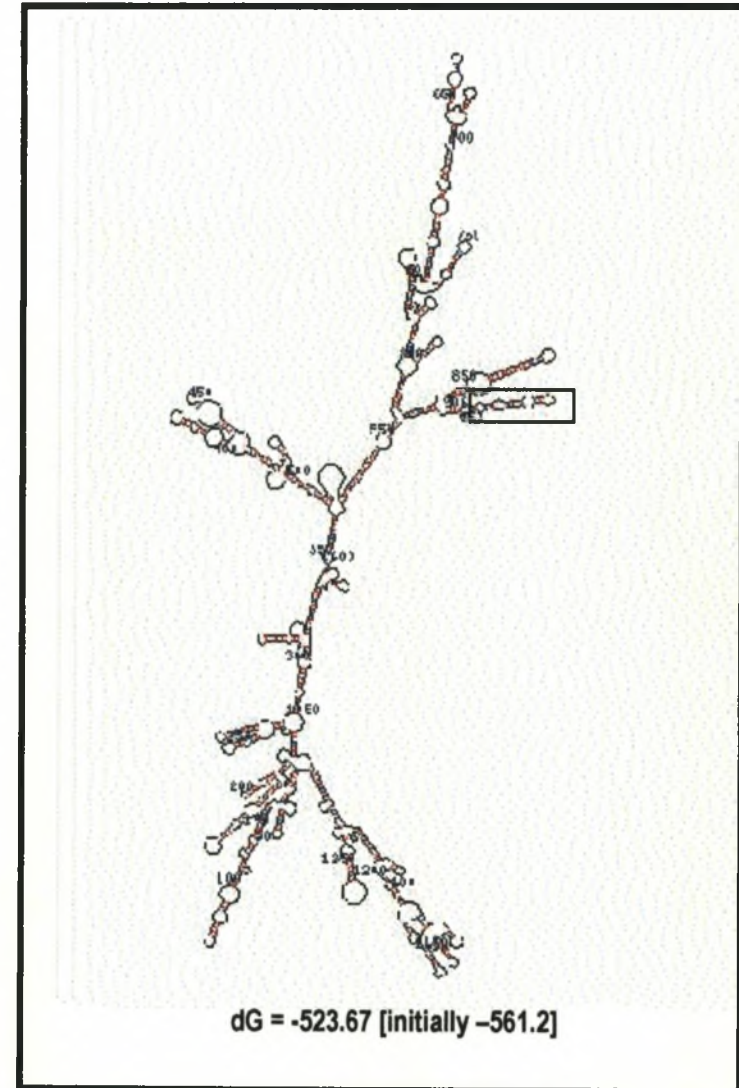


Figure 3.16 Comparison of the mRNA secondary structures of the *repA/B* transcript. The structures with the lowest energies are shown here. The regions containing the *repB* start codon are boxed and shown in more detail in Fig. 3.17.

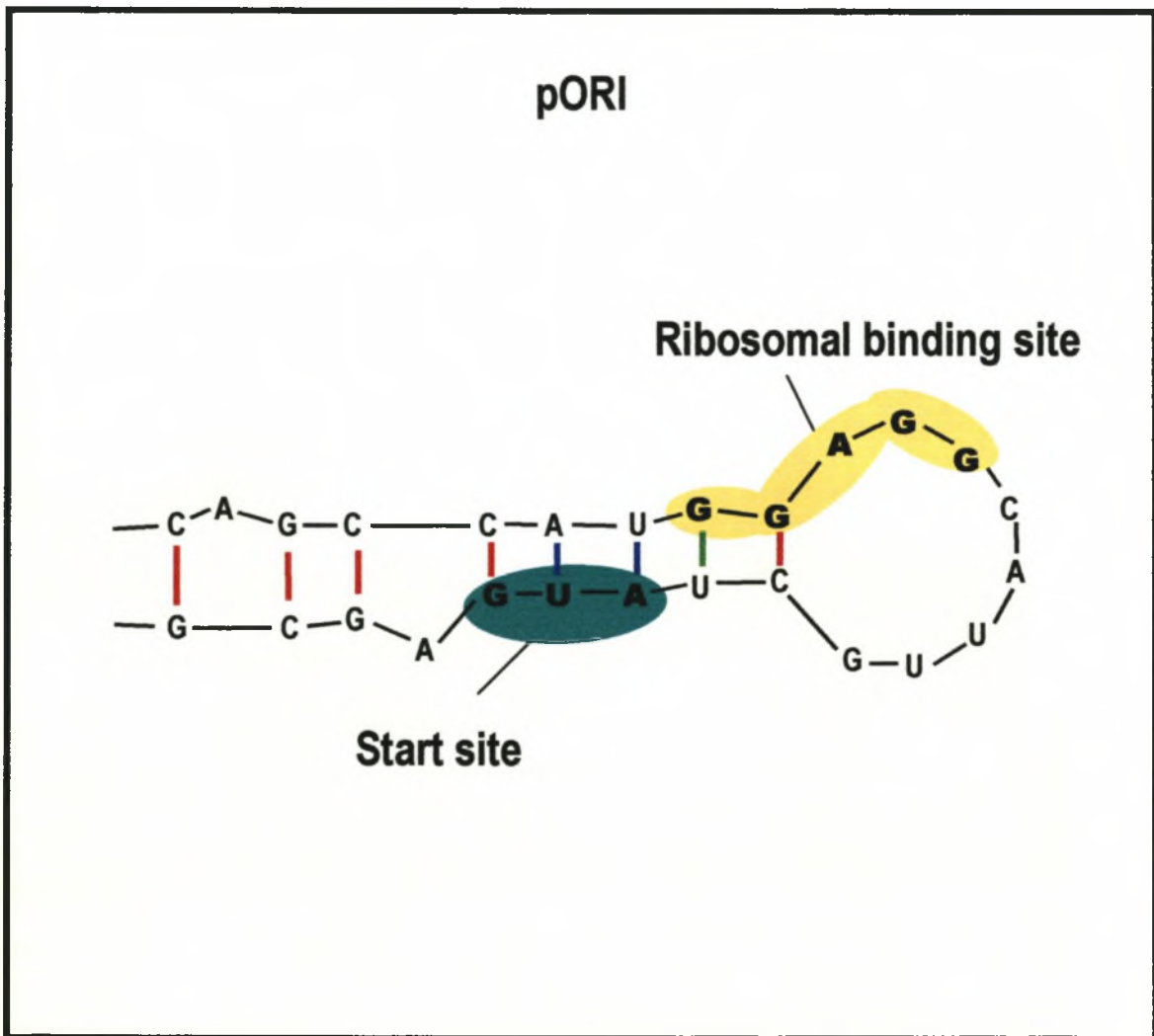
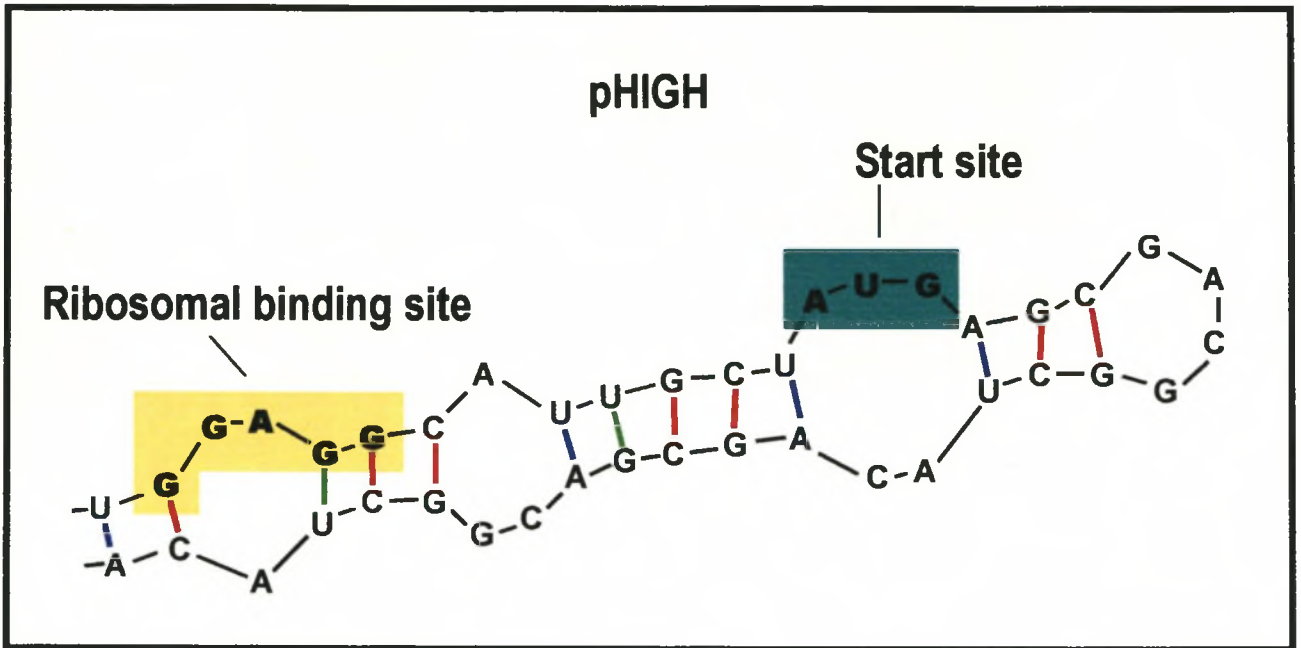


Figure 3.17 Close-up of the ribosomal binding sites and transcription start sites of *repB* on the mutant and wild type structures.

CHAPTER 4: Conclusions

The work here describes the further characterisation of a previously constructed high copy number mutant of the pAL5000 origin of replication.

By sequencing both strands of the region containing the mutation, it was proved that the 3bp deletion shown by the sequencing of Dr Bourn (1.4.2) was not a sequencing artefact.

To confirm that the apparent increase in copy number was plasmid-encoded and not a chromosomal mutation, fresh competent *M. smegmatis* cells were retransformed with both pORI and pHIGH. Restriction digests of total DNA extractions from these cultures that were run on agarose gels, showed a definite increased copy number in the case of pHIGH.

The fragment containing the mutation was subcloned into a pAL5000-related plasmid (pJC86). The resulting plasmid (pJCX) was converted to a high copy number. This proved in one step that the 3bp deletion, detected by sequencing, is the cause of the increase in copy number. It also proved that other pAL5000-related plasmids could be converted to high copy number by simple subcloning.

The relative increase in copy number as a result of the 3bp mutation was determined to be 7. The actual copy number of pHIGH was determined as approximately 54, by using Southern blots and densitometric scanning methods. Therefore the copy number of pORI is approximately 8.

This high copy number plasmid was further characterised by firstly comparing growth curves of *M. smegmatis* with *M. smegmatis* [pORI] and *M. smegmatis* [pHIGH], and secondly determining the stability of the plasmids pORI and pHIGH. This was necessary to determine the practical usefulness of the vector in the laboratory.

The growth curves showed no significant difference between the three cultures, proving that neither of the two plasmids exerts a significant load on its host, and there is no selective pressure on the cells containing either pORI or pHIGH.

The stability tests proved that both pORI and pHIGH are extremely stable in its host *M. smegmatis*. It was determined that 56% of cells still carried pORI and 17% of cells still carried pHIGH after it was cultured in media without selection, for over 200 generations. This shows pHIGH to be marginally less stable than pORI, perhaps because the high copy number places a slight extra load on the host cells.

Preliminary investigation of possible mechanisms for the increased copy number was also done. By using a promoter probe vector, the possible existence of a promoter just upstream of ORF2 was investigated. The results suggested the existence of a weak promoter between ORF1 and ORF2 in both pORI and pHIGH, but no increase or decrease in promoter activity between the two could be detected by the method used.

Another possibility was that the RNA secondary structure could be affecting the translation of RepB. It was shown that the 3bp deletion in pHIGH caused a change in the predicted RNA secondary structure of the *RepA-RepB* mRNA. Especially of importance was the change in local secondary structure around the SD sequence and the initiation codon. The 3bp deletion therefore may affect initiation of RepB translation in pHIGH, leading to a different ratio of RepA to RepB in the wild type and mutant.

In conclusion, the high copy number mutant pAL5000 origin of replication described here, should prove to be a useful tool for molecular biologists involved in research upon mycobacteria.

5. Bibliography

1. **Ainsa J.A., C. Martìn, M. Cabeza, F. De la Cruz and M.V. Mediola.** (1996) Construction of a family of *Mycobacterium/Escherichia coli* shuttle vectors derived from pAL5000 and pACYC184: their use for cloning an antibiotic-resistance gene from *Mycobacterium fortuitum*. *Gene* **176**: 23-26.
2. **Alting-Mees M.A. and J.M. Short.** (1988) pBluescript II: gene mapping vectors. *Nucleic Acids Res.* **17**(22): 9494.
3. **Baess I.** (1984) Determination and re-examination of genome sizes and base ratios on deoxyribonucleic acid from mycobacteria. *Acta Pathol Microbiol Immunol Scand [B]*. **92**(4):209-11.
4. **Beck C.F. and R.A.J. Warren.** (1988) Divergent promoters, a common form of gene organization. *Microbiol. Rev.* **52**: 318-326.
5. **Beggs M.L, J.T. Crawford and K.D. Eisenach.** (1995) Isolation and sequencing of the replication region of *Mycobacterium avium* plasmid pLR7. *Jnl. Bacteriol.* **177**: 4836-4840.
6. **Belisle J.T. and M.G. Sonnenberg.** (1998) *In* Mycobacteria Protocols, p. 31-44. T. Parrish and N.G. Stoker (eds), Humana Press, New Jersey.
7. **Brock T.D., M.T. Madigan, J.M. Martinko and J. Parker.** (1994) The Bacteria, p. 718-814 and p. 261-266. *In* D.K. Brake (ed), *Biology of Microorganisms (Seventh Edition)*. Prentice-Hall International (UK) Ltd. London.
8. **Brunel C., P. Romby, C. Sacerdot, M. de Smit, M. Graffe, J. Dondon, J. van Duin, B. Ehresmann, C. Ehresmann and M. Springer.** (1995) Stabilised secondary structure at a ribosomal binding site enhances translational repression in *E. coli*. *J. Mol. Biol.* **253**: 277-290.

9. **Burian J., S. Stuchlík and W.W. Kay.** (1999) Replication control of a small cryptic plasmid of *Escherichia coli*. *J. Mol. Biol.* **294**: 49-65.
10. **Chattoraj D.K.** (2000) Control of plasmid DNA replication by iterons: no longer paradoxical. *Mol. Microbiol.* **37**(3): 467-476.
11. **Chawla M. and S. Kr. das Gupta.** (1999) Transposition-induced structural instability of *Escherichia coli*-mycobacteria shuttle vectors. *Plasmid* **41**: 135-140.
12. **Cole S.T., R. Brosch, J. Parkhill, T. Garnier, C. Churcher, D. Harris, S.V. Gordon, K. Eiglmeier, S. Gas, C.E. Barry III, F. Tekaiia, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M.A. Quail, M.A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J.D. Sulston, K. Taylor, S. Whitehead and B.G. Barrell.** (1998) Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence. *Nature* **393**: 537-544.
13. **Cole S.T.** (1999) Learning from the genome sequence of *Mycobacterium tuberculosis* H37Rv. *FEBS* **452**: 7-10.
14. **Cole S.T., K. Eiglmeier, J. Parkhill, K.D. James, N.R. Thomson, P.R. Wheeler, N. Honore, T. Garnier, C. Churcher, D. Harris, K. Mungall, D. Basham, D. Brown, T. Chillingworth, R. Connor, R.M. Davies, K. Devlin, S. Duthoy, T. Feltwell, A. Fraser, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, C. Lacroix, J. Maclean, S. Moule, L. Murphy, K. Oliver, M.A. Quail, M.A. Rajandream, K.M. Rutherford, S. Rutter, K. Seeger, S. Simon, M. Simmonds, J. Skelton, R. Squares, S. Squares, K. Stevens, K. Taylor, S. Whitehead, J.R. Woodward and B.G. Barrell.** (2001) Massive gene decay in the leprosy bacillus. *Nature* **409**(6823): 1007-1011.

15. **Colston M.J. and F.I. Lamb.** (1989) Molecular biology of the mycobacteria. *Lepr. Rev.* **60**: 89-93.
16. **Crawford J.T. and J.H. Bates.** (1984) Restriction endonuclease mapping and cloning of *Mycobacterium intracellulare* plasmid pLR7. *Gene* **27**: 331-333.
17. **Crawford J.T. and J.O. Falkinham III.** (1990) Plasmids of the *Mycobacterium avium* complex, p. 97-119. In J. McFadden (ed.), *Molecular Biology of the Mycobacteria*. Harcourt Brace Jovanovich Publishers, London.
18. **De Smit M.H. and J. van Duin.** (1990) Secondary structure of the ribosome binding site determines translational efficiency: A quantitative analysis. *Biochem.* **87**: 7668-7672.
19. **Del Solar G., R. Giraldo, M.J. Ruiz-Echevarria, M. Espinosa and R. Diaz-Orejas.** (1998) Replication and control of circular bacterial plasmids. *Microbiol. Mol. Biol. Rev.* **62**(2): 434-464.
20. **Del Solar G. and M. Espinosa.** (2000) Plasmid copy number control: an ever-growing story. *Molecular Microbiology* **37**(3): 492-500.
21. **De Voss J.J., K. Rutter, B.G. Schroeder and C.E. Barry III.** (1999) Iron acquisition and metabolism by mycobacteria. *Jnl. Bacteriol.* **181**(15): 4443-4451.
22. **Dye C., S. Scheele, P. Dolin, V. Pathania and M.C. Raviglione.** (1999) Global burden of tuberculosis. *Jama* **282**(7): 677-686.
23. **Erardi F.X., M.L. Failla and J.O. Falkinham III.** (1987) Plasmid-encoded copper resistance and precipitation by *Mycobacterium scrofulaceum*. *Appl. Environ. Microbiol.* **53**: 1951-1954.

24. **Falkinham III, J.O. and J.T. Crawford** (1994) Plasmids. Chapter 13 p185-198. In B.R. Bloom (ed.), *Tuberculosis, pathogenesis, protection and control*. American Society for Microbiology, Washington DC.
25. **Freedman V.H., D.E. Weinstein and G. Kaplan.** (1999) How *Mycobacterium leprea* infects peripheral nerves. *Lepr. Rev.* **70**: 136-139.
26. **French A.L., D.A. Benator and F.M. Gordin.** (1997) Nontuberculous mycobacterial infections. *Med. Clin. North Am.* **81**(2): 361-379.
27. **Fry K.L., P.S. Meissner and J. O. Falkinham III.** (1986) Epidemiology of infection by nontuberculous mycobacteria. VI. Identification and use of epidemiological markers for studies of *Mycobacterium avium*, *Mycobacterium intracellulare* and *Mycobacterium scrofulaceum*. *Am. Rev. Respir. Dis.* **134**: 39-43.
28. **Gardner G.M. and R.S. Weiser.** (1947) A bacteriophage for *Mycobacterium smegmatis*. *Proc. Soc. Exp. Biol. Med.* **66**: 250-206.
29. **Gerdes K., J. Møller-Jensen and R.B. Jensen.** (2000) Plasmid and chromosome partitioning: surprises from phylogeny. *Mol. Microbiol.* **37**(3): 405-410.
30. **Gordhan B.G., S.J. Andersen, A.R. de Meyer, and V. Mizrahi.** (1996) Construction by homologous recombination and phenotypic characterisation of a DNA polymerase domain polA mutant of *Mycobacterium smegmatis*. *Gene* **178**(1-2): 125-130.
31. **Guay D.R.P.** (1996) Nontuberculous mycobacterial infections. *Ann. Pharma.* **30**: 819-830.
32. **Hatfull G.H.** (1993) Genetic transformation of mycobacteria. *Trends Microbiol.* **1**: 310-314.

33. **Hatfull G.H. and W.R. Jacobs Jr.** (1994) Mycobacteriophages, Chapter 12 p165-183. In B.R. Bloom (ed.), *Tuberculosis, pathogenesis, protection and control*. American Society for Microbiology, Washington DC.
34. **Hayakawa T., T. Tanaka, K. Sakaguchi, N. Otake and H. Yonehara.** (1979) A linear plasmid-like DNA in *Streptomyces sp.* Producing lankacidin group antibiotics. *J. Gen. Appl. Microbiol.* **25**: 255-260.
35. **Heifets L.** (1997) Mycobacteriology laboratory. *Clinics in Chest medicine.* **18**(1): 35-53.
36. **Hinnebusch J., S. Bergstrom and A.G. Barbour.** (1990) Cloning and sequence analysis of linear plasmid telomeres of the bacterium *Borrelia burgdorferi*. *Mol. Microbiol.* **4**: 811-820.
37. **Hiraga S., T. Sugiyama and T Itoh.** (1994) Comparative analysis of the replicon regions of eleven ColE2-related plasmids. *J. Bacteriol.* **176**: 7233-7243.
38. **Hirsch C.S., J.L. Johnson and J.J. Ellner.** (1999) Pulmonary tuberculosis. *Curr. Opin. Pulm. Med.* **5**(3): 143-150.
39. **Holcık M. and V.N. Iyer.** (1997) Conditionally lethal genes associated with bacterial plasmids. *Microbiol.* **143**: 3403-3416.
40. **Holland S.M.** (2001) Nontuberculous mycobacteria. *Am. J. Med. Sci.* **321**(1): 49-55.
41. **Jacobs W.R., G.V. Kalpana, J.D. Cirilo, L. Pascopella, S.B. Snapper, R.A. Udani, W. Jones, R.G. Barlette and B.R. Bloom.** (1991) Genetic systems for mycobacteria. *Methods Enzymol.* **204**: 537-555.
42. **Jacobs W.R., M. Tuckman and B.R. Bloom.** (1987) Introduction of foreign DNA into mycobacteria using a shuttle phasmid. *Nature.* **327**: 532-535.

43. **Jacobson R.R. and J.L. Krahenbuhl.** (1999) Leprosy. *The Lancet* **353**: 655-660.
44. **Jamieson F.B. and P. Chedore.** Tuberculosis in the 1990's: From bedside to bench and back. (1999) *Can Respir J.* **6**(4): 341-344.
45. **Jenkins P.A.** (1991) Mycobacteria in the environment. *Jnl. Appl. Bacteriol. Symp. Suppl.* **70**: 137S-141S.
46. **Jensen R.B. and K. Gerdes.** (1995) Programmed cell death in bacteria: proteic plasmid stabilisation systems. *Mol Microbiol.* **17**(2): 205-10.
47. **Jucker M.T. and J.O. Falkinham III.** (1990) Epidemiology of infection by nontuberculous mycobacteria. IX. Evidence for two DNA homology groups among small plasmids in *M. avium*, *M. intracellulare*, and *M. scrofulaceum*. *Am. Rev. Respir. Dis.* **142**: 858-862.
48. **Khan S. A.** (1997) Rolling-circle replication of bacterial plasmids. *Microbiol. Mol. Biol. Rev.* **61**(4): 442-455.
49. **Kolattukudy P.E., N.D. Fernandez, A.K. Azad, A.M. Fitzmaurice and T.D. Sirakova.** (1997) Biochemistry and molecular genetics of cell-wall lipid biosynthesis in mycobacteria. *Mol. Microbiol.* **24**(2): 263-270.
50. **Konicek J., M. Konickova-Radochova and M. Slosarek.** (1991) Gene manipulation in mycobacteria. *Folia Microbiol.* **36**(5): 411-422.
51. **Labidi A., C. Dauget, K.S. Goh and H.L. David.** (1984) Plasmid profiles of *Mycobacterium fortuitum* complex isolates. *Curr. Microbiol.* **11**: 235-240.
52. **Labidi A., E. Mardis, B.A. Roe and F.J. Wallace, Jr.** (1992) Cloning and DNA sequence of the *Mycobacterium fortuitum* var *fortuitum* plasmid pAL5000. *Plasmid.* **27**: 130-140.

53. **Lim E.M., J. Rauzier, J. Timm, G. Torrea, A. Murray, B. Gicquel and D. Portnoï.** (1994) Identification of *Mycobacterium tuberculosis* DNA sequences encoding exported proteins by using *phoA* gene fusions. *Jnl. Bacteriol.* **177**(1): 59-65.
54. **Looman A. C., J. Bodlaender, M. de Gruyter, A. Vogelaar and P. H. van Knippenberg.** (1986) Secondary structure as primary determinant of the efficiency of ribosomal binding sites in *Escherichia coli*. *Nucl. Acids. Res.* **14**: 5481-5495.
55. **Lowrie D.B.** (1999) DNA vaccines against tuberculosis. *Curr. Opin. Mol. Ther.* **1**(1): 30-33.
56. **Mathews D.H., J. Sabina, M. Zuker and D.H. Turner.** (1999) Expanded Sequence Dependence of Thermodynamic Parameters Improves Prediction of RNA Secondary Structure. *J. Mol. Biol.* **288**: 911-940.
57. **Meinhardt F., R. Schaffrath and M. Larsen.** (1997) Microbial linear plasmids. *Appl. Microbiol. Biotechnol.* **47**: 329-336.
58. **Meissner P.S. and J.O.Falkinham III.** (1984) Plasmid-encoded mercuric reductase in *Mycobacterium scrofulaceum*. *J. Bacteriol.* **157**: 669-672.
59. **Garbe T.R., J. Barathi, S. Barnini, Y. Zhang, C. Abou-Seid, D. Tang, R. Mukherjee and D.B. Young.** (1994) Transformation of mycobacterial species using hygromycin resistance as selectable marker. *Microbiology* **140**: 133-138.
60. **Gordhan B.G., S.J. Andersen, A.R. Meyer and V. Mizrahi.** (1996) Construction by homologous recombination and phenotypic characterisation of a DNA polymerase domain *polA* mutant of *Mycobacterium smegmatis*. *Gene* **178**(1-2): 125-130.

61. **Murray J.F.** (1998) Tuberculosis and HIV infection: A global perspective. *Respiration* **65**:335-342.
62. **Nordström K.** (1993) Plasmid replication and maintenance. In *Plasmids: a Practical Approach*, 2nd ed., pp. 2-38. Edited by K.G. Hardy. Oxford: Oxford University Press.
63. **Opravil M.** (1997) Epidemiological and clinical aspects of mycobacterial infections. *Infection* **25**(1): 56-59.
64. **Parsons L.M., C.S. Jankowski and K.M. Derbyshire.** (1998) Conjugal transfer of chromosomal DNA in *Mycobacterium smegmatis*. *Mol. Microbiol.* **28**(3): 571-582.
65. **Picardeau M. and V. Vincent.** (1998) Mycobacterial linear plasmids have an invertron-like structure related to other linear replicons in actinomycetes. *Microbiol.* **144**: 1981-1988.
66. **Picardeau M., C. le Dantec and V. Vincent.** (2000) Analysis of the internal replication region of a mycobacterial linear plasmid. *Microbiol.* **146**: 305-313.
67. **Pozniak A. and T. Bull.** (1999) Recently recognised mycobacteria of clinical significance. *Journal of infection.* **38**: 157-161.
68. **Proctor G.N.** (1994) Mathematics of microbial plasmid instability and subsequent differential growth of plasmid-free and plasmid-containing cells, relevant to the analysis of experimental colony number data. *Plasmid* **32**: 101-130.
69. **Qin M., H. Taniguchi and Y. Mizuguchi.** (1994) Analysis of the replication region of a mycobacterial plasmid, pMSC262. *Jnl. Bacteriol.* **176**(2): 419-425.
70. **Rambukkana, I.** (2000) How does *Mycobacterium leprae* target the peripheral nervous system? [see comments]. *Trends. Microbiol.* **8**: 23-28.

71. **Ranes M.G., J. Rauzier, M. Lagranderie, M. Gheorghiu and B. Gicquel.** (1990) Functional analysis of pAL5000, a plasmid from *Mycobacterium fortuitum*: Construction of a “mini” mycobacterium-*Escherichia coli* shuttle vector. *J. Bacteriol.* **172**(5): 2793-2797.
72. **Rauzier J., J. Moniz-Pereira and B. Gicquel-Sanzey.** (1988) Complete nucleotide sequence of pAL5000, a plasmid from *Mycobacterium fortuitum*. *Gene* **71**: 315-321.
73. **Sakaguchi K.** (1990) Invertrons, a class of structurally and functionally related genetic elements that include linear DNA plasmids, transposable elements, and genomes of adeno-type viruses. *Microbiol. Rev.* **54**: 66-74.
74. **Salfinger M.** (1997) Diagnosis of tuberculosis and other diseases caused by mycobacteria. *Infection* **25**(1): 60-62.
75. **Sambrook J., E.F. Fritsch and T. Maniatis.** (1989) *Molecular cloning: a laboratory manual* (second edition). Cold Spring Harbor Laboratory Press, USA.
76. **Shinnick T.M. and R.C. Good.** (1994) Mycobacterial taxonomy. *Eur. J. Clin. Microbiol. Infect. Dis.* **13**(11): 884-901.
77. **Shinnick T.M., C. H. King and F.D. Quinn.** (1995) Molecular biology, virulence and pathogenicity of mycobacteria. *Am. Jnl. Medical Sci.* **309**(2): 92-98.
78. **Short J.M., J.M. Fernandes, J.A. Sorge and W.D. Huse.** (1989) Lambda ZAP: a bacteriophage lambda expression vector with in vivo excision properties. *Nucleic Acids Res.* **16**(15): 7583-7600.
79. **Snapper, S.B., L. Lugosi, A. Jekkel, R.E. Melton, T. Kieser, B.R. Bloom and W.R Jacobs, Jr.** (1988) Lysogeny and transformation in mycobacteria: stable expression of foreign genes. *Proc. Natl. Acad. Sci. USA* **85**: 6987-6991.

80. **Snapper S.B., B.R. Bloom and W.R. Jacobs.** (1990) Molecular genetic approaches to mycobacterial investigations. *In: McFadden JJ; ed. Molecular Biology of the Mycobacteria.* London: Academic Press; pp 119-219.
81. **Spierings E., T. de Boer, L. Zulianello and T.H. Ottenhoff.** (2000) Novel mechanisms in the immunopathogenesis of leprosy nerve damage: the role of Schwann cells, T cells and *Mycobacterium leprae*. *Immunol. Cell Biol.* **78(4)**: 349-355.
82. **Stemmer W.P.** (1991) A 20-minute ethidium bromide/high-salt extraction protocol for plasmid DNA. *Biotechniques Jun*; **10(6)**: 726.
83. **Stolt P. and N.G. Stoker.** (1996a) Functional definition of regions necessary for replication and incompatibility in the *Mycobacterium fortuitum* plasmid pAL5000. *Microbiol.* **142**: 2795-2802.
84. **Stolt P. and N.G. Stoker.** (1996b) Protein-DNA interactions in the *ori* region of the *Mycobacterium fortuitum* plasmid pAL5000. *J. Bacteriol.* **178(23)**: 6693-6700.
85. **Stolt P. and N.G. Stoker.** (1997) Mutational analysis of the regulatory region of the *Mycobacterium* plasmid pAL5000. *Nucleic Acids Res.* **25(19)**: 3840-3846.
86. **Stolt P., Q. Zhang and S. Ehlers.** (1999) Identification of promoter elements in mycobacteria: mutational analysis of a highly symmetric dual promoter directing the expression of replication genes of *Mycobacterium* plasmid pAL5000. *Nucleic Acids Res.* **24(2)**: 396-402.
87. **Takechi S. and T. Itoh.** (1995) Initiation of unidirectional ColE2 DNA replication by a unique priming mechanism. *Nucleic Acids Res.* **23(20)**: 4196-4201.

88. **Takechi S., H. Matsui and T. Itoh.** (1995) Primer RNA synthesis by plasmid-specified Rep protein for initiation of ColE2 DNA replication. *EMBO Jnl.* **14**(20): 5141-5147.
89. **Vergne A.F., A.J. Watzl and M.J. Miller.** (2000) Iron chelators from mycobacteria (1954-1999) and potential therapeutic applications. *Nat Prod Rep.* **17**(1): 99-116.
90. **Villar C.A. and J. Benitez.** (1992) Functional analysis of pAL5000 plasmid in *Mycobacterium fortuitum*. *Plasmid* **28**: 166-169.
91. **Visschedijk J., B.J. van de, H. Eggens, P. Lever, S. van Beers and P. Klatser.** (2000) *Mycobacterium leprae*-millennium resistant! Leprosy control on the threshold of a new era [see comments]. *Trop. Med. Int. Health.* **5**: 388-399.
92. **Watson M.E.E.** (1984) Compilation of published signal sequences. *Nucleic Acids Res.* **12**: 5145-5163.
93. **World Health Organization.** (1998) Progress towards leprosy elimination. *Wkly. Epid. Rec.* **73**: 153-160.
94. **Yanisch-Perron C., J. Vieira and J. Messing.** (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**(1): 103-19.
95. **Yasueda, H., T. Horii and T. Itoh.** (1989) Structural and functional organization of ColE2 and ColE3 replicons. *Mol. Gen. Genet.* **215**(2): 209-216.
96. **Young D.B. and S.T. Cole.** (1993). Leprosy, tuberculosis and the new genetics. *Jnl. Bacteriol.* **175**(1): 1-6.

97. **Zuker M., D.H. Mathews and D.H. Turner.** Algorithms and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide *In: RNA Biochemistry and Biotechnology*, 11-43, J. Barciszewski & B.F.C. Clark, eds., NATO ASI Series, KluwerAcademic Publishers, (1999).

98. **Zuker M., Mfold server:** © Copyright 1995-2000, Washington University School of Medicine.